



UNIVERSITÄT ZU LÜBECK  
INSTITUTE OF MATHEMATICS AND  
IMAGE COMPUTING

## Gemeinsames Lernen von Bildregistrierung und Veränderungskarten für Lungen-CT-Bilder

Joint Learning of Image Registration and Change Detection  
for Lung-CT-Images

### Masterarbeit

im Rahmen des Studiengangs  
**Mathematik in Medizin und Lebenswissenschaften**  
der Universität zu Lübeck

Vorgelegt von  
**Temke Kohlbrandt**

Ausgegeben und betreut von  
**Prof. Dr. Jan Lellmann**

Mit Unterstützung von  
**Dr. Alessa Hering**

Die Masterarbeit entstand im Rahmen von Arbeiten bei



Lübeck, den 10. Januar 2023



Ich versichere an Eides statt, die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt zu haben.

T. Kohlbrandt

Lübeck, den 10. Januar 2023



# Kurzfassung

Mit der stetig wachsenden Zahl an medizinischen Bilddaten wachsen die Aufgaben für Radiologen. Immer mehr Daten müssen in kürzerer Zeit analysiert werden, dabei besteht die Gefahr der „Unaufmerksamkeitsblindheit“. Um die diagnostische Auswertung zu erleichtern und die Wahrscheinlichkeit zu verringern, dass wichtige Veränderungen übersehen werden, können diese mit einer farblichen Überlagerung auf den Daten für den Radiologen hervorgehoben werden.

Die Erstellung von solchen Veränderungskarten erfolgt typischerweise mit einer sequenziellen Pipeline, bei der zunächst ein Deformationsfeld über eine Bildregistrierung bestimmt und anschließend Veränderungen über ein Differenzbild berechnet werden. Idealerweise sollten auf den Veränderungskarten nur relevante Veränderungen für die diagnostische Auswertung angezeigt werden. Durch die Differenz der Bilder werden jedoch auch andere Veränderungen und Registrierungsartefakte abgebildet. Andererseits kann es bei der Registrierung auch zu Überanpassungen an die Veränderungen kommen.

In dieser Arbeit stellen wir eine neue Methode vor, in der ein Deformationsfeld und eine Veränderungskarte gemeinsam in einem Netzwerk gelernt werden, um diese Einschränkungen zu verringern. Wir haben unsere Methode auf CT-Bildern der Lunge des National Lung Screening Trial-Datensatzes getestet, in die wir zusätzliche Läsionen des Lung Image Database Consortium eingefügt haben, um größere Veränderungen zu simulieren.

Unsere erstellte Veränderungskarte zeigt in Bezug auf die Genauigkeit und Trefferquote vergleichbare Ergebnisse wie sequenziell erstellten Veränderungskarten. In einer Anwenderstudie konnte gezeigt werden, dass die Bearbeitungszeit für einige Teilnehmer mit Unterstützung der farblichen Veränderungskarte signifikant reduziert wurde, während die Genauigkeit und Trefferquote gleich blieben.

## Abstract

Radiologists face a growing challenge with increasing amount of medical image data they have to evaluate in a limited amount of time. They may also be prone to “inattentive blindness” which can lead to missed abnormalities. To help overcome these issues, color overlays can be provided to assist in diagnostic reading.

The creation of such change maps typically consists of a sequential pipeline, in which a deformation field is first determined through image registration, followed by the calculation of a subtraction image implying the changes. However, this approach can be limited by the inclusion of irrelevant changes and registration artifacts, as well as the risk of overfitting to changes due to registration.

Our proposed method involves learning a deformation field and change map jointly in a single network in order to address these limitations. The method was tested on lung CT images from the National Lung Screening Trial dataset, with additional lesions from the Lung Image Database Consortium inserted to simulate larger changes.

In terms of precision and recall our created change maps show comparable results to change maps that were created sequentially. In a user study, we also found that for some participants processing time was significantly reduced when using the change map, while precision and recall remained the same.



# Abkürzungsverzeichnis

<b>ANN</b>	artificial neural network
<b>AUC-PR</b>	Fläche unter der Precision-Recall-Kurve
<b>CM</b>	Veränderungskarte, basierend auf der Vorhersage für Veränderungen des Netzwerks
<b>CMxDiff</b>	Veränderungskarte, kombiniert aus den Veränderungskarten der Deep-Learning basierten Registrierung und der Vorhersage der Veränderungen
<b>CNN</b>	neuronaes Faltungsnetzwerk (engl. convolutional neural network)
<b>CT</b>	Computer Tomographie
<b>Diff</b>	Veränderungskarte, basierend auf dem Differenzbild der Deep-Learning basierten Registrierung
<b>HU</b>	Hounsfield-Units
<b>klassisch</b>	Veränderungskarte, basierend auf dem Differenzbild der klassischen Registrierung
<b>LIDC</b>	Lung Image Database Consortium
<b>NGF</b>	Normalized Gradient Field
<b>NLST</b>	National Lung Screening Trial
<b>ReLU</b>	Rectified Linear Unit
<b>SSD</b>	Summe der quadratischen Differenzen (engl. sum of squared distances)





# Inhalt

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Gliederung der Arbeit . . . . .	4
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Bildregistrierung . . . . .	5
2.1.1	Einführung . . . . .	5
2.1.2	Distanzmaße . . . . .	9
2.1.3	Regularisierung . . . . .	10
2.1.4	Optimierung . . . . .	11
2.2	Maschinelles Lernen . . . . .	12
2.2.1	Künstliche Neuronale Netze . . . . .	12
2.2.2	Neuronale Faltungsnetzwerke . . . . .	14
2.2.3	U-Net . . . . .	17
2.3	Deep Learning-basierte Registrierung . . . . .	17
2.4	Statistiken . . . . .	19
<b>3</b>	<b>Verwandte Arbeiten</b>	<b>23</b>
<b>4</b>	<b>Methoden</b>	<b>29</b>
4.1	Erstellung einer Veränderungskarte mit klassischer Registrierung . . . . .	29
4.1.1	Klassische Registrierung . . . . .	29
4.1.2	Erstellung der Veränderungskarte . . . . .	31
4.2	Gemeinsames Lernen einer Registrierung und Veränderungskarte . . . . .	32
4.2.1	Aufbau des Netzwerks . . . . .	32
4.2.2	Netzwerkarchitektur . . . . .	36
4.2.3	Erstellung der Veränderungskarten . . . . .	37
4.3	Visualisierung der Veränderungskarte . . . . .	37
4.4	Vorverarbeitung der Daten fürs Training . . . . .	38
4.5	Erstellung synthetischer Daten . . . . .	39
4.5.1	Hinzufügen von Läsionen in einem Bild . . . . .	41
4.5.2	Hinzufügen von Läsionen in verschiedenen Größen in beiden Bildern	42
<b>5</b>	<b>Daten und Experimente</b>	<b>45</b>
5.1	Verwendete Daten . . . . .	45
5.1.1	Datensätze . . . . .	45
5.1.2	Synthetische Daten . . . . .	46
5.1.3	Erstellung (approximierter) Segmentierungsmasken für Läsionen	47
5.2	Experimente . . . . .	47
5.2.1	Quantitative Auswertung der Methoden . . . . .	48

---

5.2.2	Anwenderstudie . . . . .	51
<b>6</b>	<b>Ergebnisse</b>	<b>55</b>
6.1	Quantitative Auswertung . . . . .	55
6.1.1	Genauigkeit und Trefferquote . . . . .	55
6.1.2	Dice-Koeffizienten . . . . .	59
6.1.3	Durchmesserdivergenzen . . . . .	60
6.2	Anwenderstudie . . . . .	65
6.2.1	Zeit . . . . .	66
6.2.2	Genauigkeit und Trefferquote . . . . .	67
6.2.3	Feedback der Teilnehmer . . . . .	68
6.3	Qualitative Auswertung . . . . .	70
<b>7</b>	<b>Diskussion und Ausblick</b>	<b>81</b>
<b>8</b>	<b>Zusammenfassung</b>	<b>87</b>
	<b>Literaturverzeichnis</b>	<b>89</b>
<b>A</b>	<b>Anhang</b>	<b>99</b>





# 1 Einleitung

## 1.1 Motivation

In der Tumorthherapie ist es Standard, im Verlauf regelmäßig 3D-CT-Aufnahmen zu machen, sodass Daten eines Patienten von verschiedenen Zeitpunkten vorliegen. Für eine Diagnose und Anpassung der Therapie müssen die Aufnahmen von Radiologen gelesen und die Läsionen manuell lokalisiert, bemessen, klassifiziert und mit früheren Aufnahmen des Patienten verglichen werden. Neben Läsionen müssen auch andere Befunde überprüft werden. Mit der stetig wachsenden Zahl an CT-Bilddaten wachsen die Aufgaben für Radiologen. Immer mehr Daten müssen in kürzerer Zeit analysiert werden.

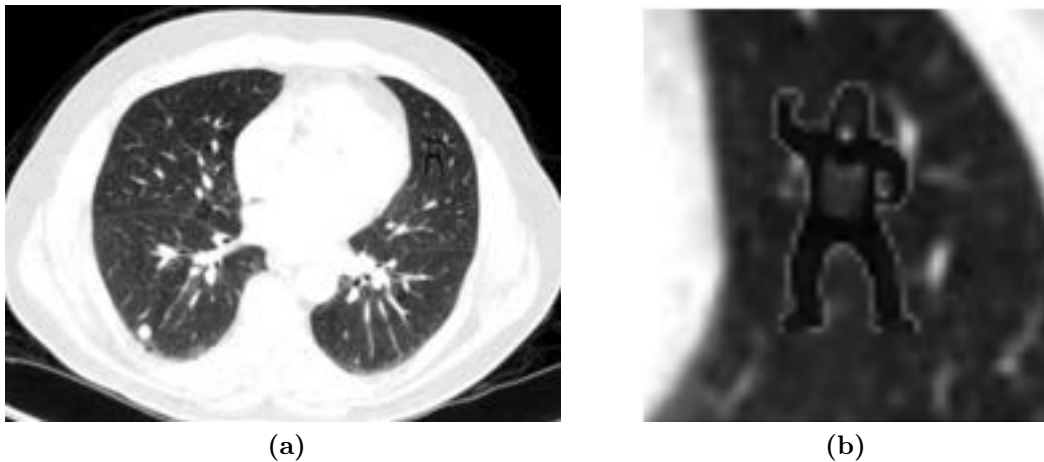
Dabei besteht die Schwierigkeit der *Unaufmerksamkeitsblindheit*, also dass Veränderungen oder Auffälligkeiten nicht entdeckt werden, da sich auf eine bestimmte andere Aufgabe fokussiert wird. In mehreren Studien wurde gezeigt, dass Menschen ihre Aufmerksamkeit nur auf eine Sache gleichzeitig fokussieren können. Eine bekannte Arbeit stammt von C. Chabris und D. Simons, die ein Gorilla durch eine Gruppe, die Ball spielten, laufen ließen und anschließend Zuschauer, die die Anzahl der Pässe zählen sollten, fragten, ob sie ein Gorilla gesehen haben [SC99].

Inspiziert von diesem Experiment wurden 2013 ähnliche Versuche in der Radiologie wiederholt und die Ergebnisse im Paper ‘The invisible gorilla strikes again: Sustained inattentive blindness in expert observers’ von Trafton Drew et al. veröffentlicht [DVW13]. Hierbei wurden Lungen-CT-Bilder mit Lungenrundherden (*lung nodules*) betrachtet, auf denen Bilder von Gorillas eingefügt wurden. Diese Gorillas sind in einem Bild über 5 Schichten mit verschiedenen Intensitäten zu erkennen und sind 48 mal größer als *lung nodules*. Ein Beispiel für eine Schicht eines Bildes, die einen eingefügten Gorilla enthält, ist in Abbildung 1.1 abgebildet.

Die 24 Radiologen und weitere 24 Probanden bekamen die Aufgabe Lungenrundherde zuerkennen. Die Aufgabe wurde gut erfüllt. In der anschließenden Befragung gaben jedoch insgesamt nur 4 Teilnehmer an, einen Gorilla in den Bildern gesehen zu haben. Der Gorilla wurde von allen erkannt, als ihnen anschließend die entsprechende Schicht gezeigt wurde.

In einem weiteren Teil wurde überprüft, ob die Gorillas in den Bildern erkannt werden, wenn danach gefragt wird. Dafür wurden anderen Probanden Videos gezeigt, in denen diese Bilder ähnlich zu einer diagnostischen Auswertung durchgescrollt werden. In über 80% der Fälle wurde eine richtige Aussage darüber gegeben, ob ein Gorilla im CT-Bild war oder nicht. Damit konnte gezeigt werden, dass sich Menschen auf die gestellte Aufgabe konzentrieren und diese gut erfüllen, aber dabei andere offensichtliche Sachen vermehrt übersehen.

Das Problem der Unaufmerksamkeitsblindheit kann in der diagnostischen Auswertung durch Radiologen ein Problem darstellen, da von Radiologen erwartet wird, auch



**Abb. 1.1:** In (a) ist eine Schicht aus einem Lungen-CT-Bild, das in der Studie zur Unaufmerksamkeitsblindheit im medizinischen Kontext von Drews et al. [DVW13] verwendet wurde, dargestellt. In dieser Schicht und den umliegenden Schichten wurde das Bild von einem Gorilla eingefügt. Dieser Teil wird vergrößert in (b) dargestellt.

Nebenbefunde zu entdecken. Um die Genauigkeit von Befundungen durch Radiologen zu erhöhen und deren Bearbeitungszeit zu reduzieren, ist die Erforschung von Ansätzen zur Unterstützung der Radiologen in der diagnostischen Auswertung von Bedeutung.

Die Bildregistrierung kann ein hilfreiches Mittel dafür sein. Zum einen kann die Bildregistrierung den Radiologen unterstützen, indem sie automatisch korrespondierende Stellen in den Bildern findet, sodass weniger Zeit zum Finden dieser Stellen benötigt wird [HPA<sup>+</sup>21]. Zum anderen kann die Bildregistrierung genutzt werden, um Veränderungen durch Differenzbilder hervorzuheben, mit sogenannten Veränderungskarten zu visualisieren [SBH<sup>+</sup>20] und damit die Aufmerksamkeit des Radiologen darauf zu lenken.

Veränderungskarten wurden schon in vielen verschiedenen Bereichen neben der medizinischen Diagnose und Behandlung genutzt, wie die Fernerkundung, Videoüberwachung und Fahrerassistenzsysteme, um Veränderungsbereiche in mehreren Bildern derselben Szene, die zu unterschiedlichen Zeitpunkten aufgenommen wurden, zu erkennen [LKB<sup>+</sup>17]. Häufig wird das Erkennen von Veränderungen mit anderen Aufgaben kombiniert oder in ein anderes Problem, wie ein Bildsegmentierungs- oder Pixelklassifizierungsproblem, umgewandelt [WPZ<sup>+</sup>21].

Aufgrund des Wachstums der verfügbaren Rechenleistung und der starken Kapazität des Repräsentationslernens werden derzeit im Bereich der Fernerkundung die besten Algorithmen zur Erkennung von Veränderungen durch neuronale Faltungsnetzwerke geliefert. Die künstliche Intelligenz (KI) hat in den vergangenen Jahren an Bedeutung gewonnen und findet in immer mehr Bereichen Anwendung. Auch im medizinischen Kontext werden meistens Faltungsnetzwerke für Segmentierungs- und Klassifikationsprobleme verwendet.

Die Bildregistrierung basiert dagegen oft noch auf konventionellen Methoden, da im Vergleich zu den anderen Problemen keine direkte Grundwahrheit zum Erlernen dieser besteht. Jedoch wurde in den vergangenen Jahren im Bereich der Deep Learning-basierte Bildregistrierung aktiv geforscht und es konnten Ansätze entwickelt werden, die gleichwertige oder sogar bessere Ergebnisse als die konventionellen Ansätze liefern

[HHM<sup>+</sup>21b, HSZ<sup>+</sup>22]. Dennoch ist die Deep Learning-basierte Bildregistrierung immer noch ein relativ neues Feld im klinischen Umfeld und die konventionelle Bildregistrierung die dominierende Technologie für die meisten Anwendungen.

Während bei der konventionellen Bildregistrierung für jedes Bildpaar ein Optimierungsproblem einzeln gelöst wird, wird bei der Deep Learning-basierte Bildregistrierung einmal ein Netz erlernt und dieses auf die Bilder angewendet, um ein Deformationsfeld zu erhalten. Die Verwendung eines Netzwerks auf Bilddaten ist in der Anwendung schneller als eine einzelne Optimierung über diese. Folglich kann durch die Verwendung eines erlernten Registrierungsnetzwerks als Unterstützung die Zeit für eine Befundung im Vergleich zur Verwendung einer konventionellen Bildregistrierung weiter reduziert werden.

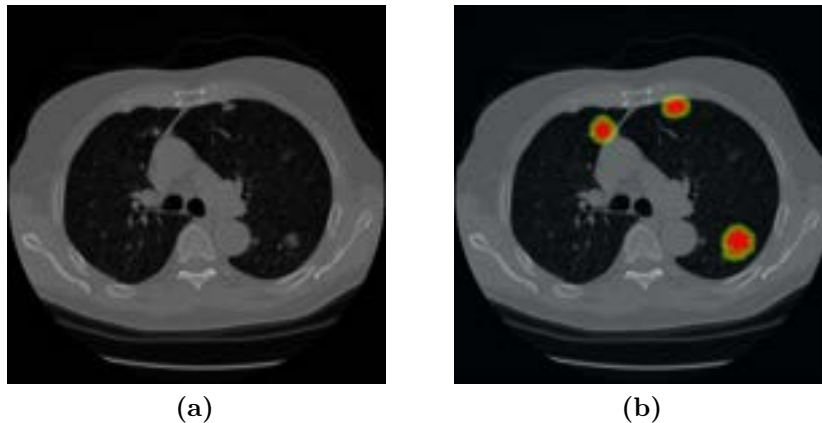
Das bisherige Vorgehen zum Erstellen von Veränderungskarten mit einer Bildregistrierung ist sequenziell. Zuerst wird eine Registrierung ausgeführt und anschließend die Bilder subtrahiert, um die Veränderungen zu erhalten. Der Nachteil dabei ist, dass es bei der Registrierung zu Überanpassungen an die Veränderungen kommen kann und entsprechend relevante Veränderungen reduziert werden. So wird etwa die Tumorgröße von einem Zeitpunkt an die Tumorgröße zum anderen Zeitpunkt angepasst, wobei genau dieser Unterschied in der diagnostischen Auswertung von Bedeutung ist.

Wählt man die Parameter der Registrierung so, dass die Tumore nicht zu stark angepasst werden, werden auch die anderen Bereiche weniger stark angepasst, sodass keine großen Vorteile in den Differenzbildern gegenüber den originalen Bilddaten vorhanden sind. Eine neue mögliche Lösung für das Problem ist das gemeinsame Lernen von Registrierung und Veränderungskarten, um relevante Veränderungen als visuelle Hilfe für den Radiologen darzustellen.

In dieser Arbeit wird ein erster Ansatz entwickelt, in dem eine Registrierung und Veränderungen gemeinsam gelernt werden. Durch die Kombination soll eine Überkompensation durch die Registrierung von relevanten Veränderungen vermieden und deren Auffälligkeit in der Veränderungskarte erhöht werden. Nach unseren Stand gibt es bisher keine Arbeiten, die das untersucht haben. Die daraus resultierenden Veränderungskarten werden mit denen eines sequenziellen Ansatzes verglichen und deren Nutzen für Radiologen in einer Anwenderstudie getestet.

Unser Ziel ist es eine Veränderungskarten für Ganzkörper-Aufnahmen mittels Deep Learning zu erstellen und damit relevante Veränderungen für den Radiologen zu markieren und ähnlich wie in Abbildung 1.2 hervorzuheben, um Radiologen eine schnellere und bessere Befundung von medizinischen Bilddaten zu ermöglichen. Dabei werden die Läsionen selbst in rot markiert.

Um die Problemstellung zunächst einzugrenzen, betrachten wir in dieser Arbeit zunächst nur Veränderungen in der Lunge auf CT-Bilder. Damit soll die Machbarkeit des Ansatzes getestet werden, welcher als Grundlage dienen soll, um später auf andere Bereiche und im besten Fall auch Ganzkörper-Aufnahmen übertragen werden zu können.



**Abb. 1.2:** (a) zeigt ein CT-Bild der Lunge, in der drei Läsionen im Vergleich zur vorherigen Aufnahme neu entstanden sind. (b) zeigt das gleiche Bild mit einer farblichen Überlagerung, um die neuen Läsionen hervorzuheben. Dabei ist die Läsion selbst in rot markiert.

## 1.2 Gliederung der Arbeit

In dieser Arbeit werden zunächst in Kapitel 2 die Grundlagen der Registrierung und des maschinellen Lernens erläutert. Dabei wird auf die Definition eines Bildes eingegangen und es werden verschiedenen Distanzmaße, Regularisierer und Optimierer für die Registrierung vorgestellt. Bei maschinellen Lernen werden künstliche neuronale Netze, Faltungnetzwerke und verschiedenen Schichten und Architekturen dieser beschrieben. Zudem wird die Deep Learning-basierte Registrierung und die Methoden zur Auswertung vorgestellt.

In Kapitel 3 werden verschiedene Arbeiten zur Visualisierung von Veränderungen vorgestellt, die als Inspirationen für unsere Methoden verwendet wurden.

Diese werden in Kapitel 4 vorgestellt und schematisch dargestellt. Zunächst werden die Methode zur Erstellung einer Veränderungskarte mit der klassischen Registrierung, des gemeinsamen Lernens einer Registrierung und einer Veränderungskarte sowie die Visualisierung der Veränderungskarten beschrieben. Zudem wird erläutert, wie die verwendeten, synthetischen Daten generiert und wie die Daten vorverarbeitet wurden, um mit dem Netzwerk trainiert werden zu können.

In Kapitel 5 werden die verwendeten Daten und Experimente zum Testen unserer Methode vorgestellt. Es werden die Datensätze, mit denen unsere synthetischen Daten erstellt wurden, und die genaue Aufteilung der verschiedenen Daten in unseren Experimenten vorgestellt. Des Weiteren wird erklärt, wie unsere kontinuierlichen Veränderungskarten bezüglich Genauigkeit und Trefferquote quantitative ausgewertet werden. Zur Auswertung eines Nutzens wird eine Anwenderstudie durchgeführt, deren Aufbau in Abschnitt 5.2 beschrieben wird.

Die Ergebnisse für die qualitative Auswertung der Veränderungskarte und für die Auswertung der Anwenderstudie werden in Kapitel 6 beschrieben. Zusätzlich werden noch verschiedenen Beispiele von unseren Veränderungskarten visualisiert und deren Unterschiede betrachtet.

Diese Ergebnisse und mögliche Lösungen für Probleme und Limitierungen werden in Kapitel 7 diskutiert.



## 2 Grundlagen

Bildregistrierung ist ein wichtiger Prozess in der digitalen Bildverarbeitung und dient dazu, zwei oder mehrere Bilder derselben Szene, oder zumindest ähnlicher Szenen, bestmöglich in Übereinstimmung miteinander zu bringen. Im ersten Abschnitt 2.1 wird eine Einführung in die Bildregistrierung gegeben, die auf die Arbeiten von J. Modersitzki [Mod03] [Mod09] basiert.

Maschinelles Lernen eröffnete in den vergangenen Jahren völlig neue Möglichkeiten in der automatischen Bildanalyse, medizinischen Diagnostik und weiteren Anwendungsgebieten. Im zweiten Abschnitt 2.2 werden die Grundlagen des maschinellen Lernens basierend auf dem Buch [GBC16] von Ian Goodfellow et al. erläutert und auf die Arten von Netze eingegangen. Im letzten Abschnitt 4.2 stellen wir die Kombination aus maschinellen Lernen und Bildregistrierung vor.

### 2.1 Bildregistrierung

Das Ziel der (medizinischen) Bildregistrierung ist es, zu einem gegebenem *Templatebild*  $\mathcal{T}$  und *Referenzbild*  $\mathcal{R}$  eine plausible *Transformation*  $u$  zu finden, sodass das transformierte Templatebild dem Referenzbild möglichst ähnlich ist.

#### 2.1.1 Einführung

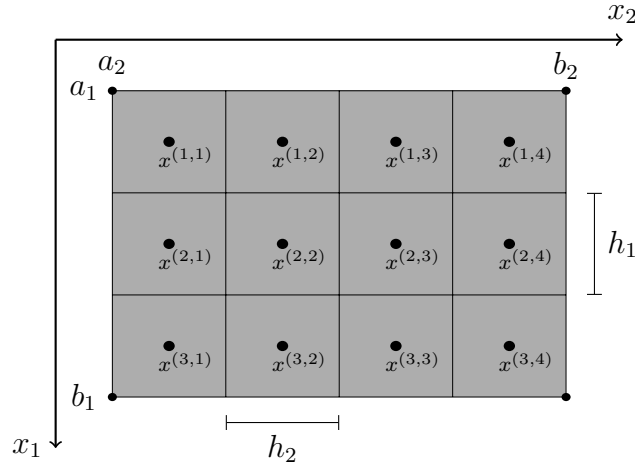
##### Bild

In der Bildregistrierung werden Bilder  $\mathcal{I}$  als stetige Abbildungen eines Bildgebiets  $\Omega \subset \mathbb{R}^d$  mit Bilddimension  $d \in \mathbb{N}$  in die (möglicherweise mehrdimensionalen) reellen Zahlen  $\mathbb{R}^k$  beschrieben:

$$\mathcal{I} : \Omega \rightarrow \mathcal{X}, \quad \Omega \subset \mathbb{R}^d, \mathcal{X} \subset \mathbb{R}^k. \quad (2.1)$$

Typischerweise wird  $d = 2$  (2D-Bilddaten) oder  $d = 3$  (3D-Intensitätsbilder oder 2D-Daten mit zusätzlicher zeitlichen Komponente) in der Bildverarbeitung genutzt. Der Parameter  $k$  beschreibt die Anzahl der Kanäle im Wertebereich  $\mathcal{X}$  des Bildes. Für ein Grauwert- oder Intensitätsbild gilt  $k = 1$  und für ein Farbbild, das jeweils ein Kanal für rote, grüne und blaue Bildanteile besitzt (RGB-Bild), gilt  $k = 3$ . Der Wertebereich  $\mathcal{X}$  kann zusätzlich eingeschränkt werden. Typische Forderungen in der Bildverarbeitung sind  $\mathcal{X} = [0,1]^k \subset \mathbb{R}^k$  oder  $\mathcal{X} = \{0,1, \dots, 255\}^k \subset \mathbb{N}_0^k$ . In dieser Arbeit beschäftigen wir uns mit dreidimensionalen Computer Tomographie (CT)-Bildern, also Grauwertbildern, daher gilt im Folgenden  $d = 3$ ,  $k = 1$ .

Der Wertebereich der CT-Bilder wird in Hounsfield-Units (HU) der Hounsfield-Skala [Buz09, Kapitel 10][Han09, Kapitel 2], die eine Abschwächung der Röntgenstrahlen im Gewebe beschreibt, angegeben. In der Praxis wird der Bereich von  $-1024$  HU



**Abb. 2.1:** Ein zellzentriertes Gitter für  $d = 2$  mit  $(m_1, m_2) = (3, 4)$  Zellen auf dem Bildgebiet  $\Omega = (a_1, b_1) \times (a_2, b_2) \subset \mathbb{R}^2$  und den Gitterabständen  $h_1$  und  $h_2$ .

bis 3071 HU genutzt, sodass die Werte als zwölfstellige Binärzahl dargestellt werden können, es sind aber auch größere Werte möglich. Es handelt sich bei den CT-Werten um standardisierte Werte basierend auf den definierten Werten von Luft ( $-1000$  HU), welche nahezu keine Röntgenstrahlung absorbiert, und Wasser ( $0$  HU). Unter Verwendung der Hounsfield-Units erhalten wir ein Wertebereich von  $\mathcal{X} = \{-1024, \dots, 3071\}^k$ .

## Diskretisierung

Für eine numerische Berechnung werden die kontinuierlichen Bilder diskretisiert. Dies geschieht durch Abtasten des Bildes auf einem Gitter. Für ein zweidimensionales Bild wird das rechteckige Bildgebiet  $\Omega = (a_1, b_1) \times (a_2, b_2) \subset \mathbb{R}^2$  in ein regelmäßiges Gitter aufgeteilt. Dabei beschreibt  $m = (m_1, m_2)$  mit  $m_i \in \mathbb{N}$  die Anzahl der Zellen (Pixel) in Richtung  $x_i$  und  $h = (h_1, h_2)$  mit  $h_i = \frac{b_i - a_i}{m_i}$  die konstant angenommenen Gitterweite (Pixelgröße) in Richtung  $x_i$  mit  $i \in \{1, 2\}$ .

In dieser Arbeit benutzen wir ein *zellzentriertes Gitter*, dessen Zellmittelpunkte durch  $x^{\bar{j}} = x^{(\bar{j}_1, \bar{j}_2)}$  mit  $\bar{j}_i \in \{1, \dots, m_i\}$  beschrieben werden. Es gilt  $x_i^{\bar{j}} = a_i + (\bar{j}_i - \frac{1}{2}) h_i$  mit  $i \in \{1, 2\}$ . Die Menge aller Zellmittelpunkte

$$\Xi = \left\{ x^{\bar{j}} : \bar{j} \in \{1, \dots, m_1\} \times \{1, \dots, m_2\} \right\} \quad (2.2)$$

ist das *zellzentriertes Gitter*. Ein Beispiel ist in Abbildung 2.1 dargestellt.

In dieser Arbeit werden 3D-Daten verwendet. Das Gitter für den dreidimensionalen Fall kann analog mit der Anzahl und Größe der Zellen (Voxel) mit  $i \in \{1, 2, 3\}$  erstellt werden.

Ein *diskretes Bild* ist nun eine Abbildung,

$$I : \Xi \rightarrow \mathcal{X}, \quad \mathcal{X} \in \mathbb{R}^k, \quad (2.3)$$

welche jedem Punkt des Gitters, genauer gesagt jeder Zelle, einen Wert in  $\mathcal{X}$  zuordnet und kann als zweidimensionale bzw. dreidimensionale Matrix betrachtet werden. Im Folgenden werden für stetige Bilder und deren Maße kalligrafische Buchstaben, wie  $\mathcal{T}$ , verwendet und für deren Diskretisierungen normale Buchstaben, wie T.

## Transformation

Eine *Transformation* oder auch *Deformation* ist eine Funktion  $y : \Omega \rightarrow \mathbb{R}^d$ , welche das Bildgebiet verändert. Die Transformation

$$y(x) = x + u(x) \quad (2.4)$$

kann in zwei Teile aufgeteilt werden: Die Identität  $x$  und die Verschiebung (engl. displacement)  $u$  der Transformation. Ein transformiertes Bild wird mit  $I[u]$  beschrieben und es gilt  $I[u](x) := I(x + u(x))$ .

Dieser Ansatz wird auch *backward mapping* oder Eulerrepräsentation genannt, da jedem Punkt im transformierten Bild eine Position und somit ein Intensitätswert des ursprünglichen Bildes zugeordnet wird. Dies hat gegenüber dem sogenannten *forward mapping*, bei welchem für jeden Punkt im Originalbild eine Position im transformierten Bild bestimmt wird, den Vorteil, dass jedem Punkt im transformierten Bild auch sicher ein Grauwert zugeordnet wird [Kö, Kapitel 2].

Bei der Transformation kann zwischen parametrischen und nicht-parametrischen Modellen unterschieden werden. Beispiele für parametrische Transformationen sind affinen Transformationen der Form  $u(x) = Ax + b$ . Bei Registrierungen, bei denen diese Art von Transformation vorkommt, spricht man von *parametrischer Registrierung*. Im Speziellen spricht man von *rigider Registrierung*, falls  $A$  orthonormal ist, die Transformation also längen- und winkelerhaltende Eigenschaften hat. Nicht-parametrische Transformationen bzw. Deformationen können im Allgemeinen nicht über eine lineare Gleichung angegeben werden. Die Veränderungen werden dabei mit einem Parameter oder einem Parametervektor für jede Zelle angegeben. Man spricht bei Verwendung von nicht-parametrischen Transformationen von *nicht-parametrischer Registrierung* oder *deformierbarer Registrierung*.

## Interpolation

Zum Auswerten von Zwischenstellen, die nicht auf den Punkten des Gitters liegen, wird eine Interpolationsethode benötigt, da nach Anwendung einer Transformation die Bildpunkte des diskreten Templatebildes nicht zwingend wieder auf Gitterpunkten liegen müssen. In dieser Arbeit werden die trilineare Interpolation und die Nearest-Neighbor-Interpolation verwendet.

Bei der Nearest-Neighbor-Interpolation wird einem Zwischenpunkt die Bildintensität jenes Gitterpunktes zugewiesen, der ihm am nächsten liegt. Die bilineare Interpolation für zweidimensionale Bilder nutzt statt eines Nachbarpunktes die vier Nachbarpunkte eines Zwischenpunktes und berechnet die Bildintensität des Zwischenpunktes als gewichtete Summe der Bildintensitäten dieser vier Nachbarn. Dabei verhalten sich die Gewichte antiproportional zum Abstand zwischen neuem Zwischenpunkt und jeweiligen Gitterpunkt. Ein Zwischenpunkt  $\tilde{x} = (\tilde{x}_1, \tilde{x}_2) \in [x_1^{(1,1)}, x_1^{(m_1,1)}] \times [x_2^{(1,1)}, x_2^{(1,m_2)}]$  des zellzentrierten Gitters  $\Omega$  wird zunächst in zwei Anteile für die Auswertung des Templatebildes an dieser Stelle zerlegt. Der erste Anteil  $\tilde{x}^G = (\tilde{x}_1^G, \tilde{x}_2^G)$  mit

$$\tilde{x}_i^G = \max \left\{ x_i^{\bar{i}} : x_i^{\bar{i}} \leq \tilde{x}_i, x_i^{\bar{i}} \in \Omega \right\}, \quad i = 1, 2, \quad (2.5)$$

beschreibt den (auf beide Dimensionen bezogenen) nächstkleineren Gitterpunkt. Der zweite Anteil  $\xi = (\xi_1, \xi_2)$  mit

$$\xi_i = \frac{\tilde{x}_i - \tilde{x}_i^G}{h_i}, \quad 0 \leq \xi_i < 1, \quad i = 1, 2, \quad (2.6)$$

bemisst den normierten Abstand zu den nächstkleineren Gitterpunkten. Der interpolierte Funktionswert lautet dann

$$\begin{aligned} T^{\text{bilinear}}(\tilde{x}) &= (1 - \xi_1)(1 - \xi_2)T(\tilde{x}_1^G, \tilde{x}_2^G) + \xi_1(1 - \xi_2)T(\tilde{x}_1^G + h_1, \tilde{x}_2^G) \\ &+ (1 - \xi_1)\xi_2T(\tilde{x}_1^G, \tilde{x}_2^G + h_2) + \xi_1\xi_2T(\tilde{x}_1^G + h_1, \tilde{x}_2^G + h_2). \end{aligned} \quad (2.7)$$

Die Nearest-Neighbor-Interpolation

$$T^{NN}(\tilde{x}) = \begin{cases} T(\tilde{x}_1^G, \tilde{x}_2^G) & \text{für } \xi_1 \leq \frac{h_1}{2} \text{ und } \xi_2 \leq \frac{h_2}{2}, \\ T(\tilde{x}_1^G + h_1, \tilde{x}_2^G) & \text{für } \xi_1 > \frac{h_1}{2} \text{ und } \xi_2 \leq \frac{h_2}{2}, \\ T(\tilde{x}_1^G, \tilde{x}_2^G + h_2) & \text{für } \xi_1 \leq \frac{h_1}{2} \text{ und } \xi_2 > \frac{h_2}{2}, \\ T(\tilde{x}_1^G + h_1, \tilde{x}_2^G + h_2) & \text{für } \xi_1 > \frac{h_1}{2} \text{ und } \xi_2 > \frac{h_2}{2} \end{cases} \quad (2.8)$$

führt zu einem unstetigen, stückweise konstanten Modell, in welchem die Funktion in einem Punkt entweder die Ableitung 0 besitzt oder dort nicht differenzierbar ist. Da die Ableitung jedoch wichtig im Prozess der Bildregistrierung ist, benutzen wir in dieser Arbeit die trilineare Interpolation zur Berechnung der Transformation der Bilder. Diese ist stückweise differenzierbar und stetig. Segmentierungsmasken enthalten nur ganzzahlige Bildintensitäten, die nicht ordinal skaliert sein müssen. In diesem Fall ist die Verwendung der Nearest-Neighbor-Interpolation der Verwendung von bilinearer Interpolation und anschließendem Runden zu bevorzugen.

Ebenso wie die Gitter kann die bilineare Interpolation und die Nearest-Neighbor-Interpolation für den dreidimensionalen Fall erweitert werden. Für dreidimensionale Bilder wird die Bildintensität des Zwischenpunktes dann mit 8 Nachbarpunkten ( $2^d$ ) und bei ersterem als trilinearen Interpolation berechnet.

Es existieren weitere Interpolationsmethoden, von denen eine bekannte die *Splines* sind. Diese setzen sich aus einer Linearkombination sogenannter *B-Splines*  $n$ -ten Grades zusammen, welche an den Gitterpunkten  $(n - 1)$ -mal stetig differenzierbar sind und für glatte Übergänge sorgen [Mod09].

## Ziel der Bildregistrierung

Da bei der Bildregistrierung die Distanz zwischen Referenzbild und transformiertem Templatebild minimiert werden soll, werden Registrierungsproblem meist in Form eines Optimierungsproblems betrachtet. Die Lösung des Optimierungsproblems

$$E(\phi) = \mathcal{D}[\mathcal{T} \circ \phi, \mathcal{R}] + \alpha \mathcal{S}[\phi], \quad (2.9)$$

$$\phi^* = \arg \min_{\phi: \Omega \rightarrow \mathbb{R}^d} E(\phi) \quad (2.10)$$

ergibt die am Anfang beschriebene gesuchte Transformation  $\phi^* : \Omega \rightarrow \mathbb{R}^d$ . Dabei ist  $E$  das *Energiefunktional*, welches aus einem *Distanzmaß*  $\mathcal{D}$  und einem mit Parameter  $\alpha$  gewichteten *Regularisierer*  $\mathcal{S}$  besteht. Je größer  $\alpha$  gewählt wird, desto mehr Gewicht wird auf die Plausibilität der Transformation und weniger Gewicht auf die Ähnlichkeit der Bilder gelegt.

Im Folgenden werden typische Beispiele für Distanzmaße und Regularisierer vorgestellt.

### 2.1.2 Distanzmaße

Es existieren viele verschiedene Distanzmaße  $\mathcal{D}$  mit denen sich die Ähnlichkeit zweier Bilder bestimmen lässt. Im Allgemeinen gilt: je ähnlicher sich zwei Bilder sind, desto kleiner wird das Distanzmaß. Im Folgenden werden zwei Distanzmaße vorgestellt, die in dieser Arbeit Anwendung finden: *Summe der quadrierten Differenzen* und *Normalized Gradient Field*.

#### Summe der quadrierten Differenzen (SSD)

Ein intuitives Distanzmaß ist die Summe der quadratischen Differenzen (engl. sum of squared distances) (SSD), die die Ähnlichkeit zweier Bilder anhand der punktwisen Differenzen der Bildintensität bestimmt. Für zwei Bilder  $\mathcal{T}, \mathcal{R}$  und eine Verschiebung  $u$  ist die SSD definiert als

$$\mathcal{D}^{\text{SSD}}[\mathcal{T}, \mathcal{R}, u] = \frac{1}{2} \int_{\Omega} (\mathcal{T}(x + u(x)) - \mathcal{R}(x))^2 dx. \quad (2.11)$$

Da wir im diskretisierten Bild die Funktionswerte der Mittelpunkte der Zellen gegeben haben, lässt sich die SSD für den zweidimensionalen Fall über die Approximationen der Integrale mit

$$D^{\text{SSD}}[T, R, u] = \frac{1}{2} h_1 h_2 \|T[u] - R\|_2^2 \quad (2.12)$$

berechnen. Ein Nachteil der SSD ist, seine Annahme, dass korrespondierende Zellen der Bilder die gleichen Bildintensitäten besitzen, weswegen es für Vergleich zwischen multimodalen Bildern ungeeignet ist. Ein mögliches Distanzmaß für diesen Fall ist das Normalized Gradient Field.

#### Normalized Gradient Field (NGF)

Das folgende Distanzmaß betrachtet zum Vergleich zweier Bilder nicht deren Bildintensitäten in den Zellen, sondern deren Ortsableitungen. Da Gradienten an Kanten besonders hoch und auf ebenen Flächen klein sind, bestimmt das Distanzmaß, ob in korrespondierenden Bildbereichen Kanten/Strukturen an den gleichen Stellen auftreten. Beim Normalized Gradient Field (NGF) [HM07] werden Gradienten mit dem inneren Produkt verglichen, sodass Kanten zweier Bilder als korrespondierend angenommen werden, wenn diese in gleiche oder entgegengesetzte Richtung verlaufen. Zusätzlich werden die Gradienten normiert, da korrespondierende Kanten nicht den gleichen Intensitätsunterschied aufweisen müssen. Das NGF-Distanzmaß ist definiert als

$$\mathcal{D}^{\text{NGF}}[\mathcal{T}, \mathcal{R}, u] = \int_{\Omega} 1 - \left( \frac{\langle \nabla \mathcal{T}(x + u(x)), \nabla \mathcal{R}(x) \rangle + \tau \varrho}{\|\nabla \mathcal{T}(x + u(x))\|_{\tau} \|\nabla \mathcal{R}(x)\|_{\varrho}} \right)^2 dx, \quad (2.13)$$

wobei  $\|\cdot\|_\varepsilon = \sqrt{\langle \cdot, \cdot \rangle + \varepsilon^2}$  mit dem Parameter  $\varepsilon \in \mathbb{R}_{\geq 0}$ , welcher als Kantendetektor Bereiche mit weniger Struktur geringer gewichtet, ist. Mit dieser Definition von  $\mathcal{D}^{\text{NGF}}$  liefert der Integrand an jeder Position  $x$  Werte in  $[0, 1]$ . Werte größer als Null bezeichnen Gradienten mit unterschiedlicher Ausrichtung, wobei größere Werte für größere Unterschiede stehen. NGF kann als Maß bezüglich des Winkels zwischen den zwei Gradienten verstanden werden. Ein Wert von 0 bedeutet, dass die Bildgradienten parallel oder antiparallel ausgerichtet sind und ein Wert von 1 bezeichnet einen Gradienten, welche einen Winkel von  $90^\circ$  bilden.

Im diskreten Fall werden die Gradienten über eine Kombination aus Vorwärts- und Rückwärtsdifferenzen als finite Differenzen  $\tilde{\nabla}I$  angenähert. Das approximierte diskrete NGF-Distanzmaß wird im zweidimensionalen Fall durch

$$D^{\text{NGF}}[T, R, u] = h_1 h_2 \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \left( 1 - \left( \frac{\frac{1}{2} \langle \tilde{\nabla}T^{(i_1, i_2)}[u], \tilde{\nabla}R^{(i_1, i_2)} \rangle + \tau \varrho}{\|\tilde{\nabla}T^{(i_1, i_2)}[u]\|_\tau^* \|\tilde{\nabla}R^{(i_1, i_2)}\|_\varrho^*} \right)^2 \right) \quad (2.14)$$

mit  $\|\cdot\|_\varepsilon^* = \sqrt{\frac{1}{2} \langle \cdot, \cdot \rangle + \varepsilon^2}$  und  $\varepsilon > 0$  berechnet. Hier dient der Parameter  $\varepsilon \in \{\tau, \varrho\}$  neben dem Gewichten von größeren Strukturen zusätzlich zur numerischen Stabilität. Der Faktor  $\frac{1}{2}$  entspringt aus der Berechnung der finiten Differenzen und kann in [Kö, Kapitel 2] genauer nachvollzogen werden.

### 2.1.3 Regularisierung

Die Bildregistrierung ist ein Optimierungsproblem. Ein Problem ist nach Hadamard *gut gestellt*, wenn es eine Lösung hat, diese Lösung eindeutig bestimmt ist und stetig von den Eingabedaten abhängt [Had02]. Ein Distanzmaß allein ist in der Regel ein schlecht gestelltes Problem [Had02], da die Lösung (wenn existent) nicht immer eindeutig ist. Deshalb wird im *Energiefunktional*  $E$  zusätzlich zum Distanzmaß ein Regularisierer eingeführt, über den Anforderungen an die Plausibilität der Transformation gestellt werden können. Die Plausibilität der Transformation wird vom Anwender unter Formulierung apriorischen Wissens bestimmt und nimmt für gewünschte Transformationen kleine Werte an.

Ein Beispiel für einen Regularisierer ist der *diffusive Regularisierer*

$$\mathcal{S}^{\text{diff}}[\phi] = \frac{1}{2} \cdot \int_{\Omega} \sum_{i=1}^d \|\nabla u_i\|^2 dx,$$

der durch das Bestrafen großer erster Ableitungen des *Verschiebungsfeldes*  $u(x) = \phi(x) - x$  die Glattheit der Transformation fordert.

Der Krümmungsregularisierer minimiert die Krümmungsenergie der Verschiebung, um eine möglichst glatte Transformation zu erhalten. Er ist über die Ableitungen zweiter Ordnung des Verschiebungsfeldes  $u$  wie folgt definiert:

$$\mathcal{S}^{\text{curv}}[\phi] = \frac{1}{2} \cdot \int_{\Omega} \sum_{i=1}^d (\Delta u_i)^2 dx. \quad (2.15)$$

Neben den zwei vorgestellten expliziten Regularisierungstermen wird eine Transformation auch über *implizite Regularisierung* beeinflusst. Ein Beispiel für eine implizite Regularisierung ist die Einschränkung möglicher Transformationen, beispielsweise auf rigide Transformationen. Aber auch die Berechnung auf größeren Gittern als die ursprüngliche Bildauflösung und das anschließende Interpolieren ist eine implizite Regularisierung, bei der das Deformationsfeld (implizit) geglättet wird.

### 2.1.4 Optimierung

Wie zuvor beschrieben, wird das Registrierungsproblem (2.10) als ein Optimierungsproblem betrachtet, das das Energiefunktional (2.9) minimiert. Im Folgenden werden die Grundlagen zur Berechnung eines Minimierers

$$x^* = \arg \min_{x \in \mathbb{R}^d} f(x) \quad (2.16)$$

einer beliebigen, stetig differenzierbaren Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  beschrieben. Üblicherweise werden zum Lösen des Registrierungsproblems iterative Verfahren verwendet, welche  $x$  schrittweise verändern und den Funktionswert  $f(x)$  dabei jeweils verringern, bis gewisse Abbruchbedingungen erfüllt sind.

#### Gradientenabstieg

Ein sehr bekanntes Verfahren ist der *Gradientenabstieg* [Ber97, Nes03], bei dem der Minimierer über den steilsten Abstieg des Gradienten gefunden werden soll. Beim Gradientenabstiegsverfahren wird für eine stetig differenzierbare Funktion  $f$  zunächst ein Startwert  $x^0 \in \mathbb{R}^d$  gewählt. Dieser wird iterativ über das Update

$$x^k = x^{k-1} - \alpha^k \cdot \nabla f(x^{k-1}) \quad (2.17)$$

aktualisiert, bis eine zuvor festgelegte Abbruchbedingung erfüllt wird. Dabei beschreibt  $k \in \mathbb{N}_0$  die aktuelle Iteration und  $-\nabla f(x^{k-1})$  die Richtung des steilsten Abstieges, in der das Gradientenabstiegsverfahren den Minimierer vermutet.

Das  $\alpha^k \in \mathbb{R}_{>0}$  beschreibt die Schrittweite, welche fest gewählt oder in jeder Iteration beispielsweise mit der Armijo-Regel [Arm66] unterschiedlich gewählt werden kann. Die Größe der Schrittweite ist von Bedeutung, da eine zu große Schrittweite zur Folge haben kann, dass ein Schritt (2.17) an dem Minimierer vorbeigeht, der nur mittels einer geringen Schrittweite erreichbar ist. Eine zu kleine Schrittweite kann hingegen dazu führen, dass der Algorithmus in einem lokalen Minimum endet oder durch eine höhere Anzahl an Iterationen vermeidbarer Rechenaufwand benötigt wird.

Als Abbruchbedingungen kann unter anderem eine maximale Anzahl an Iterationen  $k_{max}$  gewählt werden, aber auch Werte abhängig vom Ergebnis des Iterationsschrittes, z.B. ein Schwellwert für die Genauigkeit oder Veränderungen. Eine Kombination aus diesen Möglichkeiten können von Vorteil sein und wird beispielsweise bei den Gill-Murray-Wright-Abbruchbedingungen verwendet [GMW81].

Weitere Optimierungsverfahren sind das stochastische Gradientenabstiegsverfahren, das (Gauß-)Newton-Verfahren [Bec14] oder das (L-)BFGS-Verfahren (Limited-Memory Broyden-Fletcher-Goldfarb-Shanno) [LN89] [WN<sup>+</sup>99]. Das Newton-Verfahren löst ein

nicht lineares Minimierungsproblem auf Basis der zweiten Ableitung durch lokale Linearisierungen und der schrittweisen Annäherung an die Nullstelle. Dabei wird für die Suchrichtung die Hesse-Matrix benötigt. Verwandt mit diesem ist das Gauß-Newton-Verfahren, welches häufig für nicht lineare inverse Probleme, wie in der Bildverarbeitung, verwendet wird. Dieses Verfahren eignet sich für große Probleme, da die Berechnung der Hesse-Matrix entfällt. In dieser Arbeit findet das Gauß-Newton-Verfahren Verwendung in der parametrischen Registrierung. Das L-BFGS Verfahren verwendet die Schätzung der inverse Hesse-Matrix und speichert statt der gesamten Matrix nur einige Vektoren. Durch den daraus resultierenden geringen Speicherbedarf ist diese Methode besonders gut für Optimierungsprobleme mit vielen Variablen geeignet und findet daher Anwendung bei der deformierbaren Registrierung in dieser Arbeit [WN<sup>+</sup>99].

### Multi-Level-Strategie

In konventionellen Registrierungsansätzen haben sich mehrstufige Fortsetzungs- und Skalenraumtechniken, sogenannte *Multilevel-Strategien*, als sehr effizient erwiesen, um lokale Minima während des Optimierungsprozesses der Kostenfunktion zu vermeiden, topologische Veränderungen oder Faltungen zu reduzieren und die Laufzeiten zu beschleunigen [BK89, HM04].

Dabei wird von grob zu fein registriert. Das bedeutet, dass zuerst ein Deformationsfeld auf einer geringeren Auflösungsstufe berechnet wird, um erst einmal grobe Deformationen zu erkennen. Dies benötigt weniger Rechenaufwand als auf der höheren Auflösung. Anschließend wird die Auflösung des Deformationsfeld mit Interpolation verfeinert und die Bilder werden erneut mit Verwendung dieser Startdeformation auf einer höheren Auflösung registriert. Dies kann so lange wiederholt werden, bis die gewünschte Auflösung der Bilder erreicht ist. Die verschiedenen Auflösungen können passend zum Registrierungsproblem gewählt werden. Durch die Verwendung der initialen Schätzungen des Deformationsfeldes aus der vorherigen Stufe sinkt der Rechenaufwand und lokale Minima können vermieden werden [HM04].

## 2.2 Maschinelles Lernen

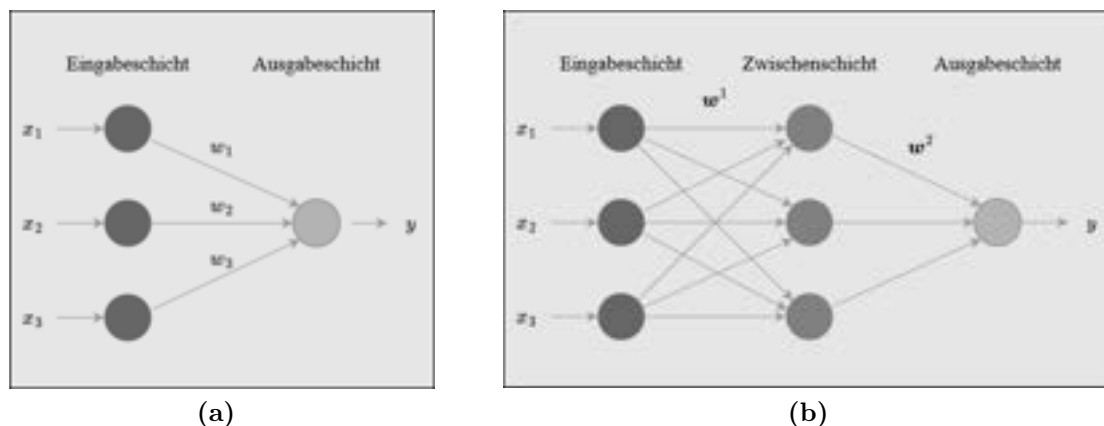
Im Folgenden werden Grundlagen zum maschinellen Lernen basierend [GBC16] und deren Einsatz in der Bildverarbeitung erläutert.

### 2.2.1 Künstliche Neuronale Netze

Die kleinste Einheit eines *künstlichen neuronalen Netzes* (engl. artificial neural network (ANN)) ist ein *Perzeptron* oder auch *Neuron*, welches eine Funktion des menschlichen Gehirns mathematisch nachahmt und mit den Grundstein fürs Maschinelle Lernen (ML) legt.

Das Gehirn besteht aus etwa 100 Milliarden Neuronen, welche über Synapsen miteinander verbunden sein können. Die Nervenzellen nehmen über Dendriten, die Fortsätze an dem einen Zellende, Impulse von anderen, über die Synapsen verbundenen, Zellen auf und lösen ein Aktionspotenzial aus, wenn die Summe der eingehenden Impulse einen gewissen Schwellwert überschreitet. Dies sorgt dafür, dass ein weiterer Impuls über das





**Abb. 2.2:** (a) schematische Darstellung eines Perzeptrons mit  $n = 3$  Eingabeneuronen  $x_i$ , Ausgabeneuron  $y$  und Gewichten  $w_i$  (b) Beispiel eines vollständig verbundenen neuronalen Netzes mit einer Zwischenschicht mit  $n = 3$  Eingabeneuronen  $x_i$ , die über die Gewichte  $w^1$  mit allen Neuronen der Zwischenschicht verbunden sind. Diese sind mit den Gewichten  $w^2$  mit dem Ausgabeneuron  $y$  verbunden

Axon zum anderen Ende der Zelle weitergeleitet und verbundene Neuronen angeregt werden. Diese Grundform lässt sich mathematisch als gewichteten Summe über die eingehenden Neuronen  $x_i$ ,  $i \in \{1, \dots, n\}$  als

$$y = \sigma \left( \sum_{i=1}^n w_i \cdot x_i + b_i \right) \quad (2.18)$$

modelliert. Dabei verbinden die Gewichte  $w_i$  die Eingabeneuronen  $x_i$  mit dem Ausgabeneuron  $y$  und bestimmen deren Einfluss auf das Ausgabeneuron: Je größer das Gewicht eines Neurons ist, desto stärker ist sein Einfluss auf die Ausgabe. Die *Aktivierungsfunktion*  $\sigma$  wird meist als nichtlineare Funktion gewählt, z.B. als Sigmoidfunktion, um auch nicht lineare Zusammenhänge modellieren zu können. Der negative Schwellwert der Aktivierungsfunktion (Bias) wird mit  $b_i$  bezeichnet. Diese Grundform eines neuronalen Netzes ist in Abbildung 2.2a abgebildet.

Allgemein lässt sich dieser Zusammenhang als Funktion  $f : \Theta \times X \rightarrow Y$  beschreiben, die für gegebene Gewichte  $w \in \Theta$  und Eingaben  $x \in X$  die Ausgabe  $y \in Y$  bestimmt. Um bei Bedarf mehrere Werte aus einem Netz zurückzugeben, kann  $y$  auch mehrdimensional konstruiert werden. Zusätzlich zur Ein- und Ausgabeschicht kann ein Netz noch weitere Zwischenschichten enthalten. Wenn ein Netz viele ( $\gg 10$ ) solcher Zwischenschichten enthält, wird es als *tiefes neuronales Netz* (engl. deep neural network) bezeichnet. Falls jedes Neuron einer Schicht mit allen Neuronen der folgenden Schicht verbunden ist, spricht man von einer *vollständig verbundenen* (engl. fully connected) Schicht. In Abbildung 2.2b wird ein vollständig verbundenes neuronales Netz mit einer Zwischenschicht dargestellt.

Damit ein Netzwerk lernen kann, wie es auf bestimmte Eingaben reagieren soll, wird die Ausgabe  $y$  für Eingabe  $x$  mit einem bekannten *Referenzwert* (ground truth)  $y^*$  verglichen und die Gewichte  $w$  entsprechend angepasst. Eine Funktion  $L : Y \times Y \rightarrow \mathbb{R}_{>0}$ , welche die Ausgabe mit der Referenz vergleicht, wird *Kosten-* oder auch *Verlustfunktion* (engl. *loss function*) genannt. Je ähnlicher sich der Ausgabe- und Referenzwert sind,

desto kleiner wird der *Verlust* (engl. *loss*). Daher ist es das Ziel, zu gegebenen Eingaben  $x$  die Gewichte  $w$  zu bestimmen, die den Verlust minimieren:

$$L(\mathbf{y}, \mathbf{y}^*) = L(f(\mathbf{w}, \mathbf{x}), \mathbf{y}^*) \rightarrow \min_{\mathbf{w} \in \Theta}. \quad (2.19)$$

Für die Minimierung der Verlustfunktion (2.19) wird ein Optimierungsverfahren verwendet. Ein Beispiel hierfür ist der in Abschnitt 2.1.4 vorgestellte Gradientenabstieg. Hierfür wird der Gradient der Funktion  $L$  bezüglich der Gewichte  $\mathbf{w}$  benötigt:

$$\nabla_{\mathbf{w}} L = \left( \frac{\partial L}{\partial w_1}, \dots, \frac{\partial L}{\partial w_n} \right)^\top.$$

Je komplexer das Netz und damit die Funktion  $f$  ist, desto komplizierter wird die Berechnung des Gradienten. Dementsprechend kann es für tiefe Netze sehr zeit- und rechenaufwändig sein, den Gradienten als gesamten Ausdruck zu berechnen.

Der Algorithmus der *Backpropagation* bietet eine Methode, die Gradientenberechnung des Netzes äußerst effizient zu gestalten. Dafür wird zunächst im Vorwärtsdurchlauf von der Eingabe bis zum Verlust jeder berechnete Zwischenwert gespeichert. Im anschließenden Rückwärtsdurchlauf werden mithilfe der Kettenregel die Ableitungswerte von Zwischenschritt zu Zwischenschritt zurück propagiert. Diese Methode profitiert davon, dass nur lokale Gradienten berechnet werden müssen, und kann somit die Gradientenberechnung für jeden Parameter des neuronalen Netzes auch bei vielen Schichten effizient gestalten dreyfus1990artificial, kinnebrock1994neuronale.

## 2.2.2 Neuronale Faltungsnetzwerke

Neuronale Faltungsnetzwerke (engl. neuronales Faltungsnetzwerk (engl. convolutional neural network) (CNN)) sind eine spezielle Art von künstlichen neuronalen Netzen, welches sich besonders für Daten auf regulären Gittern, insbesondere Bilder, eignen. Deshalb haben sie sich in den vergangenen Jahren in der Bildverarbeitung etabliert und werden zum Beispiel für Mustererkennung und Bildklassifikation verwendet.

In solchen Anwendungsbereichen ist es wichtig, gleiche Merkmale an unterschiedlichen Positionen detektieren zu können. Aus diesem Grund wird statt einem vollständig verbundenen neuronalen Netzes ein Netz mit flexiblen Faltungsschichten verwendet, deren lernbare Gewichte die Faltungskerne verschiedener Faltungen sind. Dabei hat das Faltungsnetzwerk weniger Parameter, die gelernt werden müssen, im Vergleich zum vollständig verbundenen neuronalen Netz.

Für ein diskretes zweidimensionales Bild  $I$  der Größe  $m_1 \times m_2$  auf dem Gitter  $\{1, \dots, m_1\} \times \{1, \dots, m_2\}$  und ein Faltungskern  $K$  der Größe  $k_1 \times k_2$  auf dem Gitter  $\{-h_1, \dots, h_1\} \times \{-h_2, \dots, h_2\}$  ist die *diskrete Faltung* als

$$(I * K)(i, j) = \sum_{s=-h_1}^{h_1} \sum_{t=-h_2}^{h_2} I(i-s, j-t) \cdot K(s, t) \quad (2.20)$$

definiert. Für ungerade  $k_1$  und  $k_2$  gilt  $h_1 = \frac{k_1-1}{2}$  und  $h_2 = \frac{k_2-1}{2}$ .

Damit sich die Größe des Bildes vor und nach der Faltung nicht unterscheidet, werden Randwerte benötigt. Für die Wahl dieser Randwerte gibt es mehrere Möglichkeiten,

wovon zwei bekannte die Neumann-Randbedingung, welche das Bild mit den äußeren Bildwerten fortsetzt, und die Dirichlet-Randbedingung, welche das Bild mit Nullen fortsetzt, sind.

Dieses Erweitern der Bilder mit Randwerten wird *Padding* genannt. In den Faltungsschichten wird mit jeder Faltung eine Merkmalskarte (channel) erstellt. Pro Schicht können verschiedene Merkmalskarten über das Falten mit unterschiedlichen Faltungskernen erzeugt werden.

Jede Merkmalskarte stellt einen Kanal der Ausgabe der Faltungsschicht dar. Durch die verschiedenen Faltungskerne können verschiedenen Gewichte gelernt werden, durch die unterschiedliche Merkmale detektiert werden können. Durch die geringere Größe der Filterkerne im Vergleich zur Eingabe, können lokale Merkmale, wie Kanten, detektiert werden. Mit der Kombination mehrerer Schichten und der dadurch zunehmenden Tiefe eines Netzes ist es möglich, auch komplexere Strukturen zu extrahieren. Durch die dünnbesetzte Vernetzungsstruktur und das Wiederverwenden der Gewichte in Form eines Faltungskerns ist das neuronale Faltungsnetzwerk effizienter und weniger speicherintensiv im Vergleich zu neuronalen Netzen.

Die Verschachtelung über mehrere Schichten vergrößert das rezeptive Feld der folgenden Faltungsschicht, sodass die letzten Schichten die gesamte Semantik des Bildes und die gesamten globalen Merkmale erfassen können.

Ein Faltungsnetzwerk besteht typischerweise aus einer Abfolge von Schichten, die verschiedene Funktionen ausüben. Im Folgenden werden zusätzlich zur Faltungsschicht noch weitere vorgestellt.

## Pooling

Auf die Faltungsoperation folgt häufig eine *Pooling*-Operation, die benachbarte Informationen in einer Merkmalskarte mittels einer Aggregationsfunktion zusammenfasst. Diese Pooling-Schicht sorgt für Verminderung des Rechenaufwandes und des Speicherbedarfs, denn die Funktion dieser ist es, semantisch ähnliche Merkmale zu einem zu vereinen und so die Größe der Merkmalskarte zu verringern. Beim Average-Pooling wird beispielsweise der Bereich der Größe  $k_1 \times k_2$  (Patch) um ein Pixel betrachtet und der durchschnittliche Wert dieses Bereichs in dem Pixel gespeichert. Analog kann auch das Maximum oder Minimum dieses Bereichs gespeichert werden.

Die Verringerung der Größe der Merkmalskarte kann auch durch eine Schrittweite  $s > 1$  zwischen den Iterationen des Kerns (Stride) implementiert werden. Die Pooling-Schicht ermöglicht dem Netz zusätzlich in tieferen Schichten die Informationen aus Pixeln zu verknüpfen, deren Distanz in den Eingabebildern größer als die Größe des Filterkerns ist. Damit sorgt sie dafür, dass lokale Merkmale global werden. Die Ausdehnung des Bereichs in der Eingabeschicht, mit welchem ein Neuron verbunden ist, wird *rezeptives Feld* genannt.

## Aktivierungsfunktionen

Aufgrund der Linearität von Faltungen lassen sich durch Kombination dieser nur lineare Funktionen darstellen. Um auch nichtlineare Zusammenhänge lernen zu können, werden nichtlineare Schichten als Aktivierungsfunktion  $\sigma$  eingeführt.

Ein Beispiel ist die *Sigmoid*-Funktion, genauer die *Logistische Funktion* [HM95]

$$\text{logistic}(x) = \frac{1}{1 + e^{-x}}. \quad (2.21)$$

Die Funktion nimmt einen Eingangswert  $x$  entgegen und hat eine Ausgabe im Intervall  $(0, 1)$  zurück. Da die Ableitung für Ausgabewerte nahe 0 und 1 geringe Werte annehmen, werden die Gradienten der Verlustfunktion ebenfalls klein, was die Aktualisierung der Gewichte und damit den gesamten Lernprozess verhindert, weshalb sie in den vergangenen Jahren weniger Anwendung findet.

Die *Rectified Linear Unit (ReLU)* [Bro19, HZRS15] ist dagegen in den vergangenen Jahren sehr populär geworden. Sie ist definiert als

$$\text{ReLU}(x) = \begin{cases} x, & \text{für } x > 0, \\ ax, & \text{sonst} \end{cases} \quad (2.22)$$

mit  $a = 0$  und ist weniger rechenaufwendig im Vergleich zur Logistischen Funktion. Durch  $a = 0$  entstehen jedoch Nullgradienten für Eingaben kleiner 0. Das bedeutet, dass der Gradient, der durch diese ReLU-Neuronen fließt, von diesem Zeitpunkt an ebenfalls Null ist. Eine Abwandlung der ReLU ist die *leakyReLU* [MHN<sup>+</sup>13], die mit  $a = 0.01$  definiert ist und entsprechend kleine Aktivierung bei negativen Eingaben zulässt. Heutzutage wird meist die leakyReLU verwendet, da sie Nullgradienten vermeidet und somit zu einem stabileren Training des Netzes führt.

Des Weiteren gibt es noch die *Softmax*-Aktivierungsfunktion [BN06]. Sie wird nur in der letzten Schicht angewandt und nur dann, wenn das neuronale Netz bei Klassifizierungsaufgaben Wahrscheinlichkeitswerte vorhersagen soll. Sie ist definiert als

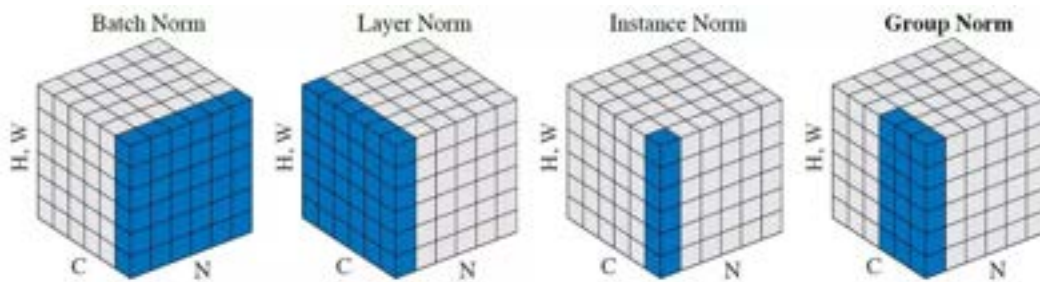
$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \quad (2.23)$$

Jede Ausgabe eines bestimmten Neurons  $i$  nicht nur von dem Wert  $x_i$  ab, den das Neuron erhält, sondern von allen Werten im Vektor  $x$ .

Einfach gesagt, zwingt die Softmax-Aktivierungsfunktion die Werte der Ausgangsneuronen dazu, Werte zwischen 0 und 1 anzunehmen, sodass diese Wahrscheinlichkeitswerte der verschiedenen Klassen im Intervall  $[0, 1]$  darstellen können, welche sich zu 1 summieren.

## Normalisierungen

Mit der Änderung von Parametern innerhalb der Faltungsschicht verändert sich auch die Verteilung der weitergegebenen Ausgaben während des Trainings [IS]. Dies verlangsamt das Training, da niedrigere Lernraten und eine sorgfältige Parameterinitialisierung erforderlich sind. Aus diesem Grund werden die Eingaben einer Schicht normalisiert, sodass der Gradient schneller konvergieren kann. Eine Möglichkeit ist die *Batchnormalisierung*, die einer Normalisierung über jeden Mini-Batch, eine disjunkte Teilmenge des Trainingsdatensatzes, berechnet. [IS]. Die Batchnormalisierung liefert gute Ergebnisse für große Batchgrößen ( $> 8$ ). Durch die Beschränkung der Hardware beim Lernen, insbesondere in der Deep Learning-basierte Registrierung, ist oft nur eine Batchgröße von 1 möglich. Für diese ist die Batchnormalisierung nicht geeignet.



**Abb. 2.3:** Grafische Darstellung der Normalisierungen: Jedes Bild zeigt einen Merkmalskarten-Tensor. Dabei werden bei den verschiedenen Normalisierungen die in blau markieren Pixel mit dem gleichen Mittelwert und gleicher Varianz normalisiert. Für die Group Norm wurde eine Gruppengröße von 2 gewählt. H und W stehen für die Größe eines Bildes, C für die Anzahl der Kanäle und N beschreibt die Anzahl der Batches. [WH18, Figure 2]

Die *Instanznormalisierung* berechnet eine ähnliche Normalisierung für jede Merkmalskarte und ist daher für eine Batchgröße von 1 geeignet. In Abbildung 2.3, welche aus dem Paper von Yuxin Wu zur Gruppennormalisierung [WH18] entstammt, werden die verschiedenen Normalisierungen schematisch dargestellt. Dabei wurde für die Gruppennormalisierung eine Gruppengröße von 2 gewählt.

### 2.2.3 U-Net

Im Jahr 2015 wurde von Ronneberger et al. eine neue Art von Faltungsnetzwerk vorgestellt: das *U-Net* [RFB15]. Neben Faltungs- und Poolingschichten enthält das Netz Verbindungen, die einige Schichten im Netz überspringen, sogenannte *skip connections*. Der Name des Netzes ist durch die U-förmige Architektur motiviert (Abb 2.4).

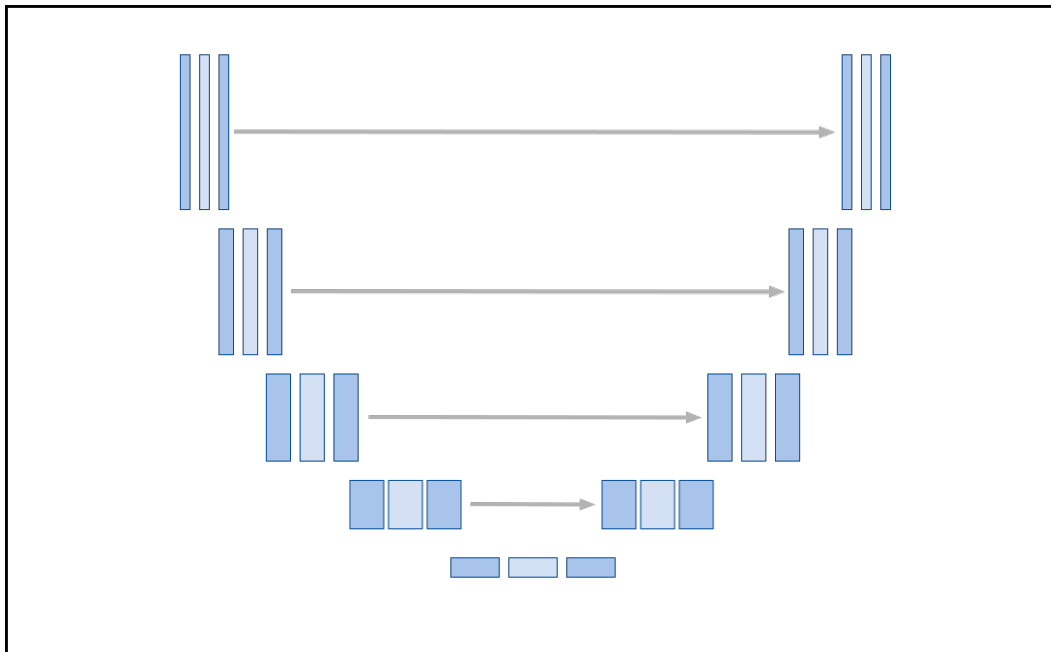
Das Netz lernt Merkmale durch Faltungen und Reduzierung der Größe durch Pooling im kontrahierenden Pfad. Anschließend wird die Größe wieder durch Faltungen erhöht und über die *skip connections* Informationen aus vorherigen Schichten hinzugefügt, welche auf die passende Größe angepasst werden. Durch diese Verbindungen und damit das Überspringen der Faltungsschichten mit geringerer Auflösung können in dem expandierenden Pfad räumliche Informationen der Merkmale gelernt werden. Entwickelt wurde das U-Net erstmals für eine Segmentierungsaufgabe und konnte im Vergleich zu damaligen Methoden bessere Segmentierungsergebnisse zeigen [RFB15].

Mittlerweile gibt es viele Abwandlungen und Erweiterungen des U-Nets, wie hinzugefügte Normalisierungsschichten oder Erweiterungen auf 3D-Bilder, und es wird häufig in der Bildverarbeitung verwendet [IJK<sup>+</sup>21, BZS<sup>+</sup>18b, TSC<sup>+</sup>21].

## 2.3 Deep Learning-basierte Registrierung

Deep Learning findet bereits Anwendung in vielen Bereichen, auch im Kontext von medizinischen Bildern, insbesondere in Segmentierungs- und Klassifizierungsproblemen [ZBLY16, RFB15, ZTLS18, KBLG<sup>+</sup>21]. In der Radiologie müssen immer mehr Bilddaten analysiert werden.

Die Verwendung von Deep Learning-basierte Ansätzen könnte die Befundung beschleunigen, da kein Optimierungsproblem für jedes Bildpaar einzeln gelöst werden



**Abb. 2.4:** Schematische Darstellung der U-Net Architektur von Ronneberger et al. [RFB15]. Die blauen Boxen stellen Merkmalskarten dar. Auf jeder Ebene haben die Merkmalskarten unterschiedliche Auflösungen. In der untersten Ebene ist die Auflösung am geringsten und die Anzahl der Kanäle am größten. Durch die skip connections (graue Pfeile) werden Informationen von Merkmalskarten derselben Auflösung direkt übertragen.

muss. Stattdessen muss nur einmal ein Netz trainiert werden, welches anschließend auf einem Bildpaar angewendet werden kann, um eine Deformation zu erhalten.

In der klinischen Routine werden jedoch eher die klassischen Methoden für die Registrierung verwendet, da es in der Registrierung keine annotierten Grundwahrheiten gibt, wie bei Segmentierungsaufgaben, die direkt gelernt werden können (supervised learning). Die Erstellung solcher Grundwahrheiten ist sehr aufwendig, da für jeden Voxel der Bildpaare ein dreidimensionaler Verschiebungsvektor erstellt werden muss, anstatt ganzen Bereichen in einem Bild einen Wert zuzuordnen, wie bei der Erstellung einer Segmentierung.

Neben dem hohen Aufwand ist es in den realen Daten in der Regel nicht möglich, jedem Voxel genau einen korrespondierenden Voxel zuzuordnen, da im gleichen Gewebe starke Ähnlichkeiten unter den Voxelintensitäten bestehen. Zudem können sich zwischen zwei Bildern die Voxelintensitäten gleicher Gewebe durch Dichte- oder Zusammensetzungsveränderungen, wie ein höherer Flüssigkeitsanteil oder ausgedehntere Lunge in der Einatmung, unterscheiden. Eine manuell bestimmte Grundwahrheit stellt demnach nur eine von vielen plausiblen Deformationen dar.

Trotzdem kann ein Registrierungsnetzwerk mit einer Grundwahrheit trainiert werden, diese kann z.B. aus einer klassischen Registrierungsmethode stammen [EP18, SVB<sup>+</sup>17]. Jedoch lernt das Netz dann nur diese Methode und ist durch deren Genauigkeit beschränkt. Eine andere Möglichkeit des Lernens einer Registrierung wäre das Definieren der Verlustfunktion des Netzes ähnlich zum Energiefunktional der klassischen Registrierung. Hier wird keine Grundwahrheit benötigt (unsupervised learning) [BZS<sup>+</sup>18a, DV BV<sup>+</sup>19, LF18, HH19, FOGM18].

Zusätzlich zu diesem Energiefunktional kann weiteres Vorwissen in die Verlustfunktion eingebaut und zum Training genutzt werden (weakly-supervised learning). Zum Beispiel kann eine Metrik über die Segmentierungsmasken vom Referenzbild und transformierten Templatebild in der Verlustfunktion verwendet werden. Diese Masken werden nur während des Trainings genutzt und nicht für die spätere Anwendung des Netzes auf anderen Daten gebraucht [HGH19, BZS<sup>+</sup>19, HMG<sup>+</sup>18a, HMG<sup>+</sup>18b].

In den vergangenen Jahren wurde von A. Hering et al. [HHM<sup>+</sup>21b] die *Learn2Reg* Challenge organisiert, die den Vergleich von Registrierungsmethoden möglich machen sollte. Registrierungsmethoden werden aufgrund der Datenbeschränkung häufig nur auf einzelnen Datensätzen entwickelt und können so schlecht untereinander verglichen werden. Dies erschwert eine Übertragung der Methoden in die Anwendung in der Klinik, da die Methoden nicht auf einer breiten Bandbreite an Daten getestet werden konnten.

Die Learn2Reg Challenge deckt ein breites Spektrum an Anatomien (z.B. Thorax, Abdomen, Gehirn), Modalitäten (CT, MR, Ultraschall), Verfügbarkeit von Annotationen sowie Bewertung der Registrierung innerhalb und zwischen Patienten ab. Somit liefert sie eine Möglichkeit des Vergleichs von verschiedenen Registrierungsmethoden auf mehreren Problemstellungen. Bei der Challenge wurden sowohl klassische als auch lernbasierte Methoden eingereicht und verglichen. In der Auswertung der Challenge von 2021 [HHM<sup>+</sup>21b] konnte gezeigt werden, dass einige Ansätze gute Ergebnisse für alle Aufgaben lieferten und direkt auf weitere Datensätze angewendet werden können, ohne stark an Genauigkeit zu verlieren und somit Vorteile für die Anwendung in der Praxis bringen.

Auch wenn bisher hauptsächlich die klassische Registrierung in der Klinik Anwendung findet, wurde in der lernbasierten Registrierung in den vergangenen Jahren viel geforscht. So wurde der Anteil der lernbasierten und kombinierten Methoden, die auf der WBIR (Workshop on Biomedical Image Registration) 2022 vorgestellt wurden, größer als der Anteil der klassischen Methoden [HSZ<sup>+</sup>22]. In einigen Bereichen können diese lernbasierten Methoden mittlerweile gleich gute, wenn nicht sogar bessere Ergebnisse, als die klassische Registrierung liefern [HHM<sup>+</sup>21b].

## 2.4 Statistiken

### Gütemaße für binäre Klassifikatoren

Für die Bewertung eines binären Klassifikators werden vier Fälle unterschieden, wobei  $y$  der Grundwahrheit entspricht und  $f(x)$  das Ergebnis des Klassifikators für eine Beobachtung  $x$  ist:

- richtig positiv (true positive (TP)):  $y = 1, f(x) = 1$
- falsch negativ (false negativ (FN)):  $y = 1, f(x) = 0$
- falsch positiv (false positive (FP)):  $y = 0, f(x) = 1$
- richtig negativ (true negative (TN)):  $y = 0, f(x) = 0$

Über die Werte der (absoluten) Häufigkeiten können noch weitere Kennzahlen bestimmt werden:

$$\begin{aligned}
\text{Sensitivität} &= \frac{TP}{TP + FN} \\
\text{Spezifität} &= \frac{TN}{TN + FP} \\
\text{Genauigkeit} &= \frac{TP}{TP + FP} \\
\text{Korrektklassifikationsrate} &= \frac{TP + TN}{TP + TN + FP + FN}
\end{aligned}$$

Die *Sensitivität/Trefferquote* (engl. sensitivity, recall, true positive rate (TPR)) gibt an, wie gut ein Klassifikator ein positives Ergebnis erkennt und die *Spezifität* (engl. specificity, true negative rate (TNR)), wie gut ein Klassifikator ein negatives Ergebnis erkennt. Die *Genauigkeit* (engl. precision, positive prediction value (PPV)) beschreibt den Anteil der richtig positiven Ergebnisse im Verhältnis zu allen gefundenen Ergebnissen, also dem Anteil der positiven Vorhersagen, weshalb er auch positiver Vorhersagewert genannt wird.

Das Ziel eines guten Klassifikators ist es, diese genannten Kennzahlen zu maximieren. Jedoch beeinflussen diese Maße sich gegenseitig, so sinkt die Trefferquote bei steigender Genauigkeit und umgekehrt. Um trotzdem einen Eindruck über den Klassifikator zu bekommen, werden diese Werte gemeinsam in einem *Precision-Recall-Kurve* dargestellt und die Fläche unter der Kurve berechnet [SHJ19]. Dafür werden auf die Ausgabe des Klassifikator verschiedene Schwellwerten angewendet und für die resultierende binäre Entscheidung jeweils die Genauigkeit und Trefferquote berechnet. Die Genauigkeiten werden dann im Diagramm gegen die Trefferquoten aufgetragen. Ein weiteres Beispiel ist die *ROC-Kurve* [PGJ04], die die Sensitivität gegen die Falsch-Positiv-Rate aufträgt. Zudem kann ein Verhältnis von Genauigkeit und Trefferquote berechnet werden. Dieses ist als *F<sub>β</sub>-Maß* [TH15, S<sup>+</sup>07] bekannt:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{\text{Genauigkeit} \cdot \text{Trefferquote}}{\beta^2 \cdot \text{Genauigkeit} + \text{Trefferquote}}. \quad (2.24)$$

Für  $\beta = 1$  ist das *F<sub>β</sub>-Maß* das harmonische Mittel von Genauigkeit und Trefferquote und ist äquivalent zum Dice-Koeffizienten [S<sup>+</sup>07]. Je größer  $\beta$  gewählt wird, desto stärker wird die Trefferquote gewichtet. Für  $\beta < 1$  wird die Genauigkeit stärker gewichtet, sodass für  $F_0 = \text{Genauigkeit}$  gilt.

### Dice-Koeffizient

Eine Segmentierungsmaske ist ein diskretes Bild, welches eine (anatomische)  $j$ -ten Struktur darstellt, häufig als Binärbild. Die Maske hat den Wert  $j \in \mathbb{N}$  an Positionen, an denen sie mit der entsprechenden Struktur korrespondiert. Die nicht zugeordneten Zellen beschreiben den Hintergrund und haben den Wert 0. Da wir an der Übereinstimmung zweier Segmentierungsmasken interessiert sind, liefert uns eine zellenweise Berechnung von Intensitätsdifferenzen oder Gradienten nicht genügend Informationen. In solchen Fällen kann ein Ähnlichkeitsmaß berechnet werden. Ein bekanntes Ähnlichkeitsmaß ist der *Dice-Koeffizient* (engl. dice similarity coefficient (DSC))[Dic45, Sor48], welcher



den Schnitt zweier Mengen im Verhältnis zur Summe deren Mächtigkeit betrachtet. Für zwei Segmentierungsmasken  $R^S$  und  $T^S(u)$  einer Registrierung, deren Ähnlichkeit bestimmt werden soll, betrachten wir die Zellen innerhalb einer Maske mit identischen Werten als eine Menge. Der Dice-Koeffizient ist definiert als

$$D^{\text{Dice}}(T^S, R^S, u) = 2 \cdot \frac{|T^S(u) \cap R^S|}{|T^S(u)| + |R^S|}. \quad (2.25)$$

Der Dice-Koeffizient nimmt Werte zwischen 0 und 1 an, wobei 1 für eine komplette Übereinstimmung und 0 für keine Übereinstimmung steht. Bei Segmentierungsmasken, die aus mehreren Labeln bestehen, kann der Dice-Koeffizient für jeden Labelwert einzeln bestimmt werden. Der Hintergrund mit Zellenwert 0 wird dabei nicht verglichen [Han09, Kapitel 5].

### Jaccard-Koeffizient

Der *Jaccard-Koeffizient* [Jac12] betrachtet das Verhältnis der Fläche des Schnitts zweier Masken zu deren Vereinigung. Er wird daher auch als *Intersection over Union (IoU)* bezeichnet und ist für Segmentierungsmasken definiert als

$$D^{\text{Jaccard}}(T^S, R^S, u) = 2 \cdot \frac{|T^S(u) \cap R^S|}{|T^S(u) \cup R^S|}. \quad (2.26)$$

Der Jaccard-Koeffizient nimmt ebenfalls Werte zwischen 0 und 1 an die analog zu den Werten des Dice-Koeffizientens interpretiert werden können.

Der Dice-Koeffizient und Jaccard-Koeffizient sind verwandt und können ineinander umgerechnet werden:

$$D^{\text{Jaccard}} = \frac{D^{\text{Dice}}}{2 - D^{\text{Dice}}} D^{\text{Dice}} = \frac{2D^{\text{Jaccard}}}{D^{1+\text{Jaccard}}}$$

### Wilcoxon-Rangsummen-Test

Der Wilcoxon-Rangsummen-Test [Wil92] wird in der Literatur auch als Mann-Whitney-U-Test oder Wilcoxon-Mann-Whitney-Test bezeichnet. Für den Test müssen die untersuchten Variablen müssen dafür mindestens ordinal sein. Es wird keine Normalverteilung der Stichprobenwerte vorausgesetzt, sondern sie können eine beliebige Verteilung annehmen. Daher wird der Wilcoxon-Rangsummen-Test als Alternative zum t-Test [Hen11] verwendet, wenn dessen Bedingungen der Normalverteilung oder intervallskalierter Variablen verletzt sind.

Betrachtet werden zwei unabhängige Stichproben  $X$  und  $Y$  der Größen  $n_1$  und  $n_2$  aus zwei Populationen und diese werden bezüglich der Hypothesen

$$H_0 : P(X < Y) = P(X > Y) \quad H_1 : P(X < Y) \neq P(X > Y) \quad (2.27)$$

getestet.

Allgemein besagt die Nullhypothese, dass es gleich wahrscheinlich ist, dass ein zufällig aus der einen Population ausgewählter Wert größer oder kleiner ist als ein zufällig

ausgewählter Wert aus der anderen Population. Die Gegenhypothese besagt dagegen, dass dies nicht gleich wahrscheinlich ist.

Wenn zusätzlich zu der allgemeinen Formulierung angenommen wird, dass die Werte kontinuierlich sind und die Verteilungen bis auf einen Shift-Parameter  $\mu$  identisch sind, lauten die Hypothesen

$$H_0 : F_0(x) = F_1(x) \quad H_1 : F_0(x) = F_1(x + \mu). \quad (2.28)$$

Eine Ablehnung der Nullhypothese  $H_0$  bedeutet eine Ungleichheit der Mediane der zwei Stichproben besteht. Eine Formänderung der Verteilungen lässt sich mit diesem Test nur ungenügend erfassen.

## 3 Verwandte Arbeiten

Zur Unterstützung der Radiologen in Tumorthherapie wollen wir Veränderungen mittels einer farbliche Überlagerung auf den CT-Bildern entwickeln. Die typische Vorgehensweise, um solche Veränderungskarten zu erstellen, ist eine sequenzielle, bei der zuerst eine Registrierung und anschließend eine Veränderungserkennung, beispielsweise durch Differenzbildern, geschätzt wird. Problem hierbei ist häufig, dass relevante Veränderungen, wie Tumorwachstum, zu stark registriert werden, sodass wertvolle Informationen verloren gehen.

Daher wollen wir eine Veränderungskarte zusammen mit einem Deformationsfeld lernen, sodass relevante Veränderungen stärker hervorgehoben und Veränderungen weniger stark kompensiert werden. Nach unserer Kenntnis gibt zum jetzigen Zeitpunkt keine Arbeit, die gemeinsam eine Veränderungskarte (engl change map) mit einer Registrierung lernt.

Im Folgenden werden verschiedene Arbeiten zu Veränderungskarten, Erkennung von Veränderungen und kombinierten Lernen mit einer Registrierung vorgestellt, die zur Inspiration unserer Arbeit dienen und diese bezüglich unseres Kontextes eingeordnet.

### **Datenterme als Gewicht für Veränderungen**

In dem Paper ‘Joint Registration and Change Detection in Longitudinal Brain MRI’ stellen E. Dufresne et al. [DFK<sup>+</sup>20] einen Ansatz vor, welcher eine Registrierung und Veränderungen gemeinsam schätzt, um Überkompensation der Registrierung bei relevanten Veränderungen zu vermeiden.

Dafür wurden Magnetresonanz-Gehirn-Bilder von zwei verschiedenen Zeitpunkten betrachtet und eine binäre Veränderungskarte (change map) für diese als Überlagerung erstellt. Dabei sollen relevanten Veränderungen, wie Läsionsveränderungen, in der Registrierung nicht überkompensiert werden, während nicht relevante Veränderungen, wie natürlicher Gewebeschwund (Atrophie), nicht zu wenig registriert werden soll. In dieser Studien wurde zusätzlich versucht, die Veränderungen durch Atrophie gering zu halten, indem ein geringer Zeitabstand zwischen den Bildaufnahmen gewählt wurde.

Um die Überkompensation zu verringern, wurde in dem Ansatz das Deformationsfeld und die Veränderungen gemeinsam geschätzt. Dafür wurde eine binäre Veränderungskarte basierend auf der Größe des Datenterms an einem Punkt erstellt. Mit dieser Veränderungskarte wurde der Datenterm im Minimierungsproblem gewichtet, sodass in Bereichen von großen Veränderungen der Datenterm nicht ins Minimierungsproblem einbezogen wurde.

In dem Paper wurde gezeigt, dass diese Methode bessere Ergebnisse gegenüber dem sequenziellen Ansatz sowohl bei synthetischen Daten als auch realen Daten liefert. Das Tumorwachstum wird weniger überkompensiert und somit wird die Veränderung des Tumors für Radiologen wahrheitsgetreuer dargestellt.

Einen ähnlichen Ansatz verfolgte D. Drobny in seiner Masterarbeit 'Automatic Weighting of Non-corresponding Regions in Image Registration'[Dro15], welcher sich mit nicht korrespondierenden Regionen in medizinischen Bildern vom Abdomen beschäftigt. Der Ansatz versucht Fehler in der Registrierung zu reduzieren, die durch Resektion, oder Hinzufügen von Materialien (Schrauben etc.) und der Anforderung der Glattheit einer Registrierung entstehen.

Dafür wurde nicht wie im oben vorgestellten Paper von E. Dufresne das Gewicht binär definiert, sondern als Wahrscheinlichkeit der Ähnlichkeit bestimmt, sodass nicht korrespondierende Regionen schwächer gewichtet werden.

Dieser Ansatz wurde für die zwei im Abschnitt 2.1.2 Distanzmaße SSD und NGF getestet. Gegenüber der klassischen, ungewichteten Registrierung konnte beim neuen Ansatz unter Verwendung von SSD eine Verbesserung festgestellt werden, unter der Verwendung von NGF konnte weder eine signifikante Verbesserung noch Verschlechterung festgestellt werden.

Beide Ansätze verwenden den Datenterm bzw. Distanzmaße zur Bestimmung eines Gewichts, was dafür sorgt, dass gleich starke Veränderungen gleich behandelt werden. In unserem Ansatz soll jedoch zwischen relevanten und nicht relevanten Veränderungen unabhängig von der Ähnlichkeit unterschieden werden, da kleinere Veränderungen der Lungenrundherde und Veränderungen durch den Herzschlag einen gleichen Unterschied des Datenterms haben können, aber nur einer der beiden Fälle für die Diagnostik entscheidend ist. Diese müssen erlernt und können nicht über einfache Optimierungsmethoden unterschieden werden.

### Deep Learning-basierte Ansätze

Das Erkennen von Veränderungen (engl. change detection) zwischen zwei Bildaufnahmen derselben Szenen oder desselben Objekts zu unterschiedlichen Zeitpunkten ist eine grundlegende Aufgabe im Bereich der Computer Vision.

Dafür wurden einige Arbeiten im nicht-medizinischen Bereich der Fernerkundung vorgestellt, die zur Erkennung von Veränderungen ein Siamese Network, welches die Eingaben zuerst getrennt lernt und später konkateniert, verwenden [CKZ22, CLY<sup>+</sup>22, APD<sup>+</sup>22, JEH<sup>+</sup>21]

Ein Beispiel hierfür ist das Paper von Zhixue Wang et al. [WPZ<sup>+</sup>21], in welchem ein Algorithmus zur Erkennung von Veränderungen vorgeschlagen wird, der auf vollständig gefalteten siamesischen Netzen für optische Luftbilder basiert und mithilfe eines verbesserten kontrastiven Verlusts (engl. contrastive loss), dem fokalen kontrastiven Verlust (FCL), trainiert wird.

Merkmale werden mit dem vorgeschlagenen kontrastivem Verlust direkt aus Bildpaaren extrahiert und Veränderungen mithilfe einer Abstandsmetrik gemessen, welche als Messung des Grades der Übereinstimmung zwischen Pixelpaaren angesehen werden kann.

Damit verringert die Methode die Intra-Klassen-Varianz und vergrößert die Inter-Klassen-Differenz, sodass die binarisierte Veränderungskarte durch einen einfachen Schwellenwert erhalten werden kann. Mit dem neu vorgeschlagenen fokalen kontrastiver Verlust wird das Problem des Ungleichgewichts zwischen den Daten (mehr Daten ohne Veränderung, als Daten mit Veränderung) begegnet und die Auswirkung von der

Unausgewogenheit wird verringert. Das Modell wird dazu gebracht, sich auf schwierige Daten zu konzentrieren und konnte im Vergleich zu anderen Basismethoden bessere Ergebnisse und eine bessere Leistung in Bezug auf das F1-Maß (siehe (2.24)) erzielen.

Auch in unserem Fall wollen wir eine große Inter-Klassen-Differenz bei kleiner Intra-Klassen-Varianz. Ebenso gibt es einen großen Unterschied in der Pixelmenge, die eine Veränderungen aufweisen, und in der Pixelmenge, die keine Veränderungen aufweisen. Jedoch wird das Problem der Unausgeglichenheit über eine binäre Veränderungskarte gelöst, sodass für unseren Ansatz mit einer kontinuierlichen Veränderungskarte, das Problem bezüglich Unausgeglichenheit der Daten nicht gelöst wird. Zudem handelt es sich hier auch um einen sequenziellen Ansatz, da die Eingaben in das Siamese Network vorab registriert werden, sodass allein die Veränderungen im Netz gelernt werden.

Während die Registrierung zwischen den Luftbildern in der Fernerkundung eher auf parametrischen Transformationen beruht, sind es in unserem Fall nicht parametrische Deformationen, was eine passende Vorregistrierung, die relevante Veränderungen nicht zu stark kompensiert, erschwert und erneut gegen einen sequenziellen Ansatz spricht.

Im medizinischen Bereich gibt es bereits Methoden, die eine Registrierung gemeinsam mit einer Segmentierung lernen [XN19, MGSC18, QR21]. Ein Beispiel dafür ist das Paper ‘A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes’ von B. Li et al. aus dem Jahr 2019 [LNK<sup>+</sup>19], in dem eine Lösung bezüglich konsistenter Segmentierungen für Veränderungen im Gehirn betrachtet wurde.

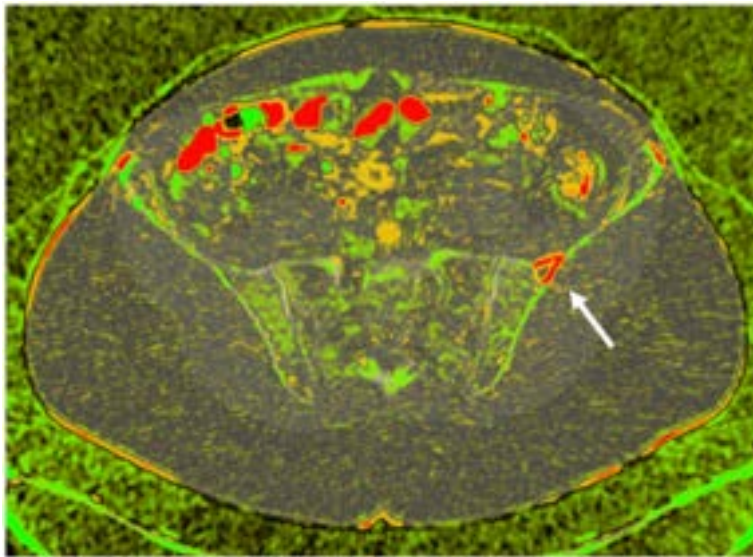
Für diese Lösung wurde ein hybrides Faltungsnetzwerk und eine Verlustfunktion bestehend aus vier Komponenten (Ähnlichkeit der registrierten Bilder, Genauigkeit der Segmentierung, Glattheit des Deformationsfeld und Konsistenz der Segmentierung) verwendet. In dem Paper konnte eine Verbesserung der Segmentierungsgenauigkeit und Konsistenz gegenüber einem sequenziellen Ansatz, Registrierung mit anschließender Segmentierung, gezeigt werden.

In dieser Arbeit soll nicht nur die Segmentierung, sondern sämtliche Veränderungen in Form von Veränderungskarten betrachtet werden, wobei der Übergang zwischen Segmentierung und Veränderungskarten nicht eindeutig definiert ist. Aufgrund der Komplexität der Veränderungskarten und dem begrenzten Zeitraum dieser Masterarbeit, werden nur vereinfachte Veränderungskarten im Bereich der Tumordiagnostik betrachtet, die große Ähnlichkeiten zu Segmentierungsmasken von Läsionen haben können.

## Darstellungen von Veränderungen

In den zuvor vorgestellten Arbeiten wurde Veränderungen binäre betrachtet. Eine Methode, Veränderungen kontinuierlich darzustellen, sind sogenannte Heatmaps. In dem Paper ‘Development and Validation of a Convolutional Neural Network for Automated Detection of Scaphoid Fractures on Conventional Radiographs’ [HSV<sup>+</sup>21] von Nils Hendrix et al. wurde ein CNN trainiert, welches Knochenbrüche in der Hand mittels einer Heatmap auf Röntgenbildern vorhersagt. Die aus dem Netz resultierenden Ergebnisse erreichten eine Leistung auf dem Niveau von Radiologen.

Fürs Training wurden Röntgenbilder verwendet, die einen oder keinen Bruch zeigen. Im Training wurden zunächst die Handknochen segmentiert, die Bildausschnitte entsprechend vergrößert und diese mit einem DenseNet [HLVDMW17] trainiert, ob ein



**Abb. 3.1:** Ein Beispiel für eine farbkodierte Veränderungskarte aus dem Paper [SBH<sup>+</sup>20]. Dabei sind Osteolysen in grün und Sklerosen in rot dargestellt. Mit dem weißen Pfeil wird auf eine Sklerose einer früheren Osteolyse hingedeutet.

Bruch vorhanden war oder nicht. Der Bereich, der zur Entscheidung geführt hat, wird anschließend visuell als Heatmap mit der Smooth Grad-CAM++ Methode [OSCW19] auf dem Röntgenbild dargestellt, um potenzielle Frakturen besser zu lokalisieren.

Diese visuelle Überlagerung sollte möglichst die Fraktur abdecken, was nach einer manuellen Betrachtung laut dem Paper geschieht. Somit kann die visuelle Darstellung die Aufmerksamkeit des Radiologen auf die Bereiche lenken, die das Netz zur Entscheidungsfindung als relevant ansah.

Da pro Bild entweder eine oder keine Fraktur vorhanden ist, lernt das Netzwerk nur eine binäre Entscheidung (Bruch oder kein Bruch) und eine Position. In unserer Problemstellung haben wir jedoch dreidimensionale Bilddaten, mehrere Veränderungen von Läsionen in einem Bild und zudem keine feste Struktur, die erkannt werden soll, sondern Veränderungen zwischen zwei Bildern, die teilweise nicht eindeutig segmentiert werden können. Daher eignet sich die Methode für unser Problem nicht direkt, jedoch zeigt sie eine Möglichkeit, eine kontinuierliche Veränderungskarte oder Entscheidungskarte darzustellen.

Eine weitere kontinuierliche Veränderungskarte wird im Paper 'Rapid study assessment in follow-up whole-body computed tomography in patients with multiple myeloma using a dedicated bone subtraction software' von Sieren et al. [SBH<sup>+</sup>20] vorgestellt. In dieser wurden einfache Differenzbilder aus niedrig dosierten Ganzkörper-CT-Aufnahmen als Veränderungskarte zur Unterstützung der Erkennung von Multiplen Myelomen ("Knochenmarkkrebs") erstellt und bezüglich der diagnostischen Genauigkeit und dem Nutzen bewertet. In diesen Differenzkarten werden Osteolysen ("Knochenauflösung") in grün und Sklerosen ("Knochenverhärtungen") in rot dargestellt. Ein Beispiel aus dem Paper ist in Abbildung 3.1 abgebildet.

Zur Beurteilung wird die Bearbeitungszeit sowie die richtig markierten Läsionen und richtigen Klassifikationen von zwei Radiologen bei Verwendung dieser Unterstützung mit den Ergebnissen unter Verwendung des Standard-Softwareprogramms ohne

Differenzkarten verglichen. Die Krankheiten wurden mit beiden Methoden bei allen Patienten korrekt klassifiziert. Für beide Radiologen verkürzt sich die Bearbeitungszeit und die Sensitivität erhöht sich, sodass diese Methode für den Einsatz in der klinischen Routine geeignet scheint.

Durch die Verwendung eines einfachen Differenzbildes enthalten die Differenzkarten Rauschen sowie Markierungen von nicht relevanten Veränderungen zum Beispiel im Bereich des Verdauungstraktes (siehe Abb. 3.1 obere rote Markierungen), sodass es vermehrt zu falsch positiv gewählten Markierungen von den Radiologen kommt.

Basierend auf diesem Paper haben wir uns für eine farbkodierte Veränderungskarte entschieden, bei der jedoch nur absolute Veränderungen betrachtet werden, sodass nur eine Signalfarbe für Veränderungen verwendet wird. Zudem sollen Rauschen und Markierungen von nicht relevanten Veränderungen durch gemeinsames Lernen einer Registrierung und Veränderungskarte reduziert werden.

Zur Auswertung unserer Methoden soll ebenfalls eine Anwenderstudie durchgeführt werden, um den Nutzen in der klinischen Anwendung abzuschätzen.





## 4 Methoden

Im folgenden Kapitel werden die Methoden zum Erstellen einer Veränderungskarte  $V$  zum einen mit der klassischen Registrierung, zum anderen mit einem gelernten Netzwerk vorgestellt. Diese sind schematisch in Abbildung 4.1 dargestellt. In dem Netzwerk soll ein Deformationsfeld und eine Veränderungskarte gemeinsam gelernt werden, um Überanpassungen von Veränderungen durch die Registrierung zu verringern und um relevante Veränderungen stärker hervorzuheben. In der Beschreibung wird insbesondere auf die Netzwerkarchitektur (Abb. 4.3) in Abschnitt 4.2 eingegangen. Des Weiteren wird in den jeweiligen Abschnitten beschrieben, wie diese Veränderungskarten für den Radiologen visualisiert werden. Zum Testen dieser Methoden verwenden wir synthetische Daten. Das Vorgehen zum Erstellen dieser wird schematisch in Abbildung 4.5 dargestellt und in Abschnitt 4.5 erläutert.

Die Registrierung ist ein zentraler Teil der vorgestellten Methoden. Für diese verwenden wir (in allen Fällen) das Normalized Gradient Field als Distanzmaß und den Krümmungsregularisierer zur Regularisierung, welche wir bereits in den Grundlagen in Abschnitt 2.1 vorgestellt haben. Wir betrachten Bilder von zwei verschiedenen Zeitpunkten, in denen verschiedene natürliche Variationen vorkommen. Sowohl eine Veränderung in der Dichte als auch in der Zusammensetzung von Gewebe verändern einen Intensitätswert. Dementsprechend können ein höherer Wasseranteil, eine verbesserte Durchblutung oder etwa eine unterschiedliche Ausdehnung der Lunge beim Atmen zu einer Variation an Intensitätswerten innerhalb einer Gewebeart führen. Unsere Wahl fiel insbesondere auf das NGF, da wir ein Distanzmaß unabhängig von Modalitäten und nicht direkt auf den Intensitätswerten für unser Problem verwenden wollten.

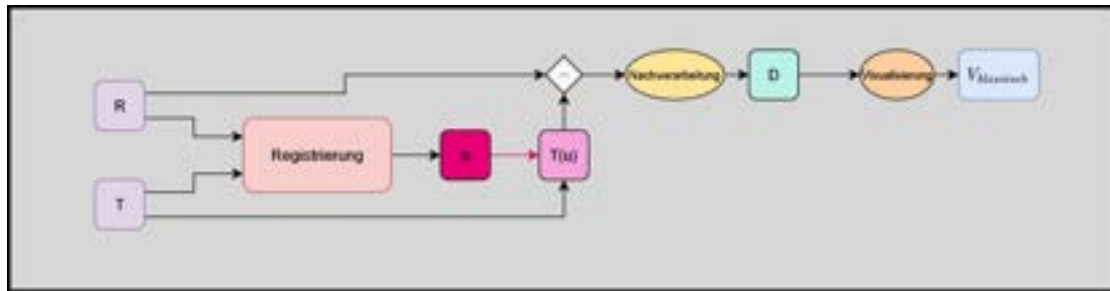
Der Krümmungsregularisierer bestraft nach Abschnitt 2.1.3 besonders starke Krümmungen und erzwingt eine Glattheit. Dies ist nötig, um Überanpassungen in einigen Bereichen zu vermeiden und eine gleichmäßige Deformation zu bevorzugen.

Weitere verwendete Maße und Regularisierer für die Registrierung sind bei den einzelnen Methoden vorgestellt. Auf entscheidende Parameter wird in der jeweiligen Beschreibung der Methode eingegangen.

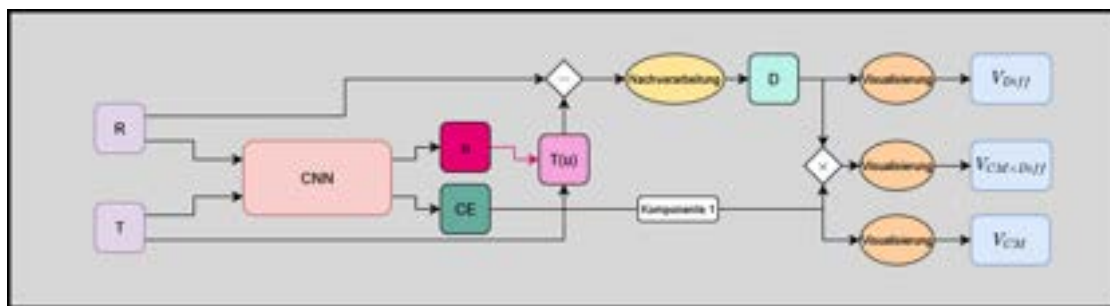
### 4.1 Erstellung einer Veränderungskarte mit klassischer Registrierung

#### 4.1.1 Klassische Registrierung

Die klassische Registrierung löst für jedes Bildpaar einzeln ein Optimierungsproblem, um ein Deformationsfeld zu finden. Wir verwenden dafür Module aus dem internen MevisLab-Paket [mev]. Zuerst lesen wir die beiden Bilder (Referenz- und Templatebild) ein und passen deren Bildbereiche an, sodass bei beiden Bildern der Grauwertbereich  $\mathcal{X} = [0, \dots, 4096]^k$  ist. Dies ist für unsere Registrierung selbst nicht notwendig, wird aber



(a) Erstellung einer Veränderungskarte mit klassische Registrierung



(b) Gemeinsames Lernen einer Registrierung und Veränderungskarte

**Abb. 4.1:** Schematische Darstellung der verschiedenen Veränderungskarten. (a) Es wird die Erstellung einer Veränderungskarte mit einer klassischen Registrierung dargestellt. Dafür wird das Bildpaar (R, T) registriert und das resultierende Deformationsfeld  $u$  auf das Templatebild angewendet. Über die Differenz  $D$  des Referenzbildes und des deformierten Templatebildes und Nachverarbeitungs-Schritte, wird die Veränderungskarte  $V_{klassisch}$ , welche farblich visualisiert wird. (b) Es wird die Erstellung von Veränderungskarten durch das gemeinsame Lernen einer Registrierung und Veränderungskarte dargestellt. Dabei ist die Erstellung von der Veränderungskarte  $V_{Diff}$  analog zu der Erstellung von  $V_{klassisch}$  mit dem Unterschied, dass das Deformationsfeld  $u$  durch ein Faltungsnetzwerk (CNN) gelernt wird. Zusätzlich lernt dieses Netzwerk einer Vorhersage von Veränderungen (CE), von dem die Komponente 1 als Veränderungskarte  $V_{CM}$  verwendet wird. Des Weiteren wird die Veränderungskarte  $V_{CM \times Diff}$  durch die voxelweise Multiplikation von  $V_{Diff}$  und  $V_{CM}$  bestimmt. Alle drei Veränderungskarten werden ebenfalls farblich visualisiert.

später bei der Auswertung benötigt und daher jetzt schon ausgeführt. Im ersten Schritt der Registrierung bestimmt das *FASTA*-Modul eine Translation für das Templatebild, wobei dieses das Normalized Gradient Field verwendet, welches auf den überlappenden Bildbereichen ausgewertet wird (NGFOverlap).

Als Nächstes folgt eine rigide Registrierung mittels dem *RegLibParametric*-Modul mit einer aus der Translation generierten Anfangsschätzung. Für die Registrierung wird die Multi-Level-Strategie aus Abschnitt 2.1.4 verwendet. Dafür bestimmen wir die Bildauflösung für die verschiedenen Bild-Level über die Glättung mit einem Gaußkern und Downsampling [YADG15] (Multi Level Type: Gaussian downsampling). Wir optimieren über drei Level, wobei wir für die feinste Bildauflösung Level 1 wählen, sodass sie die halbe Größe im Vergleich zum originalen Bild pro Dimension hat, da mit der rigiden Registrierung nur eine Vorregistrierung für die deformierbare Registrierung erstellt werden soll. Bei der rigiden Registrierung wird NGF als Distanzmaß und das Gauß-Newton-Verfahren als Optimierungsverfahren verwendet.

Als Letztes folgt eine deformierbare Registrierung mit dem *RegLibNonParametric*-Modul. Hier wird ebenfalls die Multi-Level-Strategie verwendet. Das Modul optimiert mit dem L-BFGS-Verfahren über drei Level, wobei wir als feinste Auflösung die originale Auflösung (Level 0) wählen. Als Distanzmaß wählen wir ebenfalls NGF mit einem Kantenparameter (Edge Parameter NGF) von 10 und als Regularisierer den Krümmungsregularisierer (curvature) mit einem Gewicht von  $\alpha = 100$ .

Es werden noch zwei zusätzliche Distanzmaße ins Energiefunktional ((2.9)) eingefügt: Zum einen die Summe der quadrierten Differenzen (SSD) der zwei zu den Bildern gehörenden Segmentierungsmasken der Lunge, zum anderen die Distanz zwischen Keypoints. Die Segmentierungsmasken für die Lunge werden mit dem *LungSegmentation*-Modul erstellt, welches basierend auf dem Paper [LKS<sup>+</sup>11] implementiert wurde. Die beiden so segmentierten Lungenflügel werden in einem Binärbild ( $\mathcal{X} = \{0,1\}$ ) zusammengefügt. Die Keypoints werden mit dem *DetectLungKeypoints*-Modul bestimmt, welches auf dem Paper von Rühak [RPH<sup>+</sup>17, II.C] basiert und das Prinzip von Heinrich [HHS15] verwendet. Das Gewicht im Energiefunktional für das Distanzmaß der Segmentierungsmasken beträgt 1 und das der Keypoints 100.

Nach der Optimierung erhalten wir ein Deformationsfeld und ein deformiertes Templatebild.

### 4.1.2 Erstellung der Veränderungskarte

Unsere Veränderungskarte basiert auf einem Differenzbild, welches durch die Subtraktion des deformierten Templatebildes vom Referenzbild berechnet wird. Wenn die Bildbereiche nicht vor der Registrierung angepasst wurden, müssen sie vor dem Subtrahieren der Bilder angepasst werden, damit identische Gewebe im Differenzbild Werte um 0 haben.

Um nur die Veränderungen in dem zu untersuchenden Bereich zu betrachten, wird das Differenzbild mit der Segmentierungsmaske des Organs vom Referenzbild multipliziert. In unserem Fall multiplizieren wir unser Differenzbild mit der Lungensegmentierungsmaske des Referenzbildes. In der erstellten Veränderungskarte sollen nur absolute Veränderungen betrachtet werden, dementsprechend werden die Veränderungskarten zu absoluten Veränderungskarten umgewandelt. Um die Verzerrung durch Ausreißer, z.B. Artefakte, zu unterbinden, definieren wir 1000 als obere Schranke. Anschließend

skalieren wir den Wertebereich der Veränderungskarte von  $[0, 1000]$  auf den Bereich  $[0, 1]$ , um unsere finale Veränderungskarte, basierend auf dem Differenzbild der klassischen Registrierung,  $V_{klassisch}$  zu erhalten.

## 4.2 Gemeinsames Lernen einer Registrierung und Veränderungskarte

Das Erstellen einer Veränderungskarte mittels gemeinsamen Lernen einer Registrierung und einer Veränderungskarte ist in Abbildung 4.1 in der unteren Hälfte schematisch dargestellt. In der folgenden Erläuterung wird insbesondere auf die Kombination der Verlustfunktionen (Abschnitt 4.2.1) sowie den Aufbau unseres Faltungsnetzwerkes (Abschnitt 4.2.2) eingegangen.

### 4.2.1 Aufbau des Netzwerkes

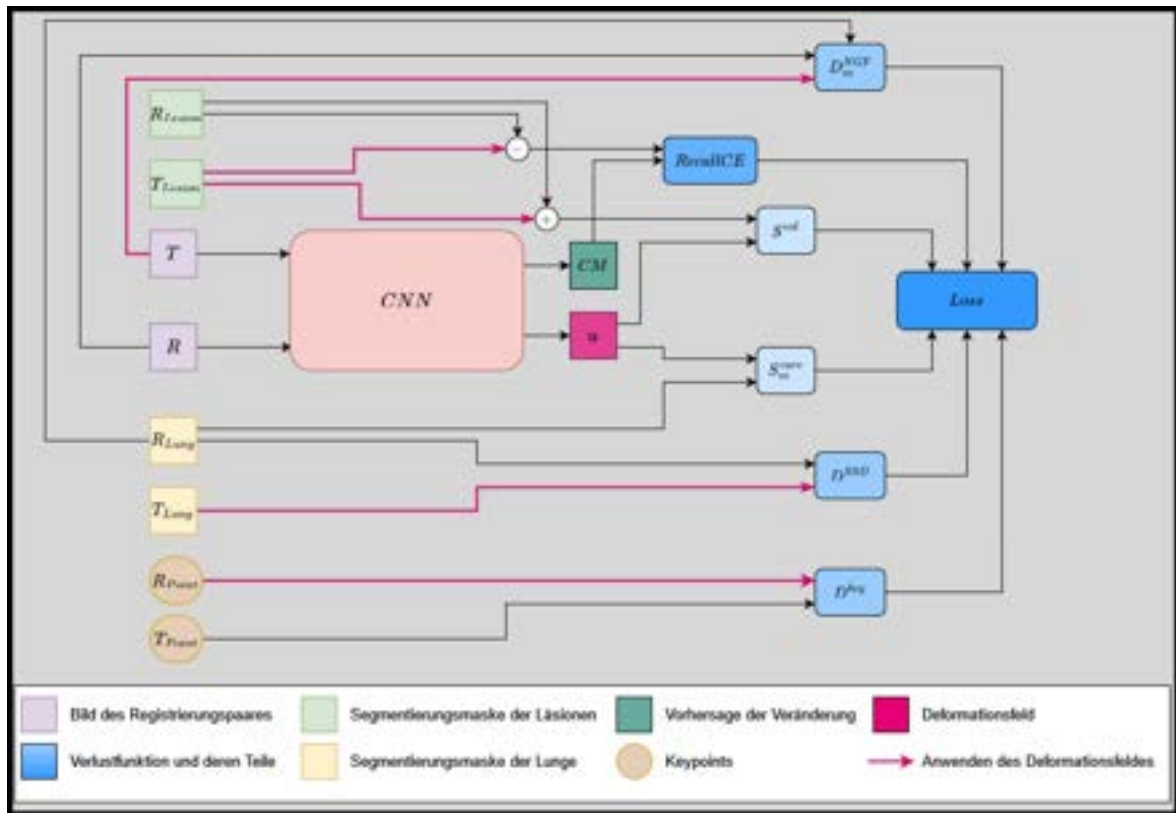
Der Ablauf des Trainings in unserer Methode mit allen Eingaben und allen Verlustfunktionen ist in Abbildung 4.2 schematisch dargestellt. Das Faltungsnetz erhält ein Paar aus Referenzbild  $R$  und Templatebild  $T$ , welche wir im ersten Schritt des Netzwerkes in der Kanaldimension konkatenieren. Der genaue Aufbau des Faltungsnetzwerkes wird in Abschnitt 4.2.2 ausführlich beschrieben. Als Ausgabe erhalten wir zum einen ein dreikanaliges Deformationsfeld  $u$ , wobei jeder Kanal eine Dimension des Deformationsfeldes darstellt. Zum anderen erhalten wir eine zweikanalige Ausgabe für unsere Veränderungskarte, basierend auf der Vorhersage für Veränderungen des Netzwerkes (CM). Dabei beschreibt ein Kanal den Hintergrund und ein Kanal die Veränderungen.

Das Deformationsfeld  $u$  wenden wir sowohl auf das Templatebild als auch auf die zugehörigen Segmentierungen der Lunge und der Läsionen an. Wir erhalten ein transformiertes Templatebild  $T(u)$ , eine transformierte Lungensegmentierung  $T^S(u)$  und transformierte Segmentierungen der Läsionen  $T^L(u)$ . Zudem wenden wir die Deformation auf die Keypoints des Referenzbildes an. Die Ähnlichkeit des transformierten Templatebildes mit dem Referenzbild wird mit dem NGF-Maß aus Abschnitt 2.1.2 berechnet. Davon betrachten wir nur die Werte aus dem Bereich der Segmentierungsmaske der Lunge in unserer Verlustfunktion. Wir verwenden also ein maskiertes Normalized Gradient Field ( $D^{NGF_m}$ ) als Distanzmaß.

Für den Vergleich der binären Segmentierungsmasken der Lunge des transformierten Templatebildes und des Referenzbildes wollen wir absolute Differenzen der Maskenwerte betrachten, um die Anzahl der Voxel, die sich unterscheiden, zu erhalten. Dafür verwenden wir die ebenfalls in Abschnitt 2.1.2 vorgestellte SSD) als Distanzmaß ( $D^{SSD}$ ). Zusätzlich bestimmen wir den Mittelwert über die Distanzen der korrespondierenden Keypoints des Templatebildes  $K^T$  zu den transformierten Keypoints des Referenzbildes  $K_u^F$ :

$$D^{key} = \frac{1}{N} \sum_{i=1}^N (\sqrt{((K_u^F)_i - (K^T)_i)^T ((K_u^F)_i - (K^T)_i)}) \quad (4.1)$$

mit Anzahl  $N$  der korrespondierenden Keypoints pro Bildpaar.



**Abb. 4.2:** Übersicht des Trainingsablaufs. Beim Training werden die Bilder des Registrierungs-paares  $R, T$  dem Netz ( $CNN$ ) übergeben. Dieses hat als Ausgabe ein Deformationsfeld  $u$  und eine Vorhersage der Veränderungen  $CE$ . Ein pinkfarbener Pfeil besagt, dass das Deformationsfeld auf diesen Daten angewendet wird. Die Teilfunktionen sowie die Verlustfunktion sind in Blautönen dargestellt. Es gibt insgesamt sechs Teilfunktionen, die Distanzmaß  $D$ , Regularisier  $R$  und weitere aufgeteilt werden. Bei Funktionen, die mit einem  $m$  markierten sind, werden die Werte durch eine Lungen-Segmentierungsmaske maskiert. Das erste Distanzmaß ist das Normalized Gradient Field für die (deformierten) Bilder des Registrierungs-paares  $D_m^{NGF}$ , welches deren Ähnlichkeit bestimmt. Dieses wird mit der Lungen-Segmentierungsmaske des Referenzbildes  $R_{Lung}$  gewichtet. Die Summe der quadrierten Differenzen der Lungen-Segmentierungsmasken vom Referenzbild  $R_{Lung}$  und vom Templatebild  $T_{Lung}$ , welche deformiert wird, ist das zweite Distanzmaß  $D^{SSD}$ , welches beschreibt, wie genau die Lungenbereiche aufeinander registriert wurden. Das letzte Distanzmaß  $D^{key}$  bestimmt die Abstände von Keypoints von bestimmten Strukturen der beiden Bilder. Dafür werden die Keypoints vom Referenzbild deformiert. Es werden zwei Regularisierer bestimmt. Der maskierte Krümmungsregularisierer  $S_m^{curv}$  bestimmt die Glattheit des Deformationsfeld  $u$  im Bereich der Lungen-Segmentierungsmaske des Referenzbildes  $R_{Lung}$ . Beim zweiten Regularisierer  $S^{vol}$  werden starke Volumenveränderungen durch das Deformationsfeld  $u$  in den Bereichen, in denen Läsionen vorkommen, bestraft. Dafür werden die Bereiche der (deformierten) Läsions-Segmentierungsmasken  $R_{Lesion}, T_{Lesion}$  in eine Segmentierungsmaske zusammengefügt (+). Die letzte Teilfunktion  $RecallCE$  bewertet die Vorhersage der Veränderungen  $CE$ . Dafür werden die Bereiche, die nur in einer der beiden Läsions-Segmentierungsmasken markiert sind, bestimmt (-) und der Funktion als Grundwahrheit übergeben. Diese sechs Teilfunktionen gehen mit unterschiedlichen Gewichten in die Verlustfunktion ( $Loss$ ) ein.

Die Plausibilität der Deformation  $u$  wird durch den Krümmungsregularisierer aus Abschnitt 2.1.3 bemessen. Da wir in unserer Veränderungskarte nur den Bereich der Lunge betrachten, werden die Werte des Krümmungsregularisierers auch mit der binären Segmentierungsmaske der Lunge gewichtet, sodass nur Werte innerhalb der Lunge in die Verlustfunktion eingehen ( $S_m^{curv}$ ). Zusätzlich sollen starke Deformationen in den Bereichen der Läsionen vermieden werden. Dafür verwenden wir ein *Volume-Change*-Regularisierer, der von Hering et al. [HHM<sup>+</sup>21a, §3.2.3] vorgestellt wurde, und gewichten diesen mit der kombinierten binären Segmentierungsmaske der Läsionen, sodass nur Veränderungen im Bereich der Läsionen in unsere Verlustfunktion ( $S^{vol}$ ) eingehen. Dabei ist die aus den Segmentierungsmasken vom Referenz-  $R^L$  und transformierten Templatebild  $T^L$  kombinierte Segmentierungsmaske  $RT^L$  definiert als

$$RT_v^L = \text{sgn}(R_v^L + T_v^L) = \begin{cases} 1 & \text{für } R_v^L + T_v^L > 0 \\ 0 & \text{für } R_v^L + T_v^L = 0, \end{cases} \quad (4.2)$$

mit  $R_v^L, T_v^L \in \{0,1\}$  für alle Voxel  $v$ .

Neben der Registrierung lernen wir eine Veränderungskarte. Zur Bestimmung dieser wird der *Recall Loss* [TMS<sup>+</sup>20], eine gewichtete Kreuzentropie, verwendet (*RecallCE*). Eine feste Gewichtung der Kreuzentropie bei stark unterschiedlich gewichteten Klassen mit einem festen Wert pro Epoche, z.B. der Frequenz, führt laut Tian et al. [TMS<sup>+</sup>20] meist zur Überanpassung, sodass man eine gute Genauigkeit (engl. precision) erhält, aber eine schlechte Trefferquote (engl. recall). Daher entwickelten sie eine Methode, die das Gewicht in jeder Epoche neu bestimmt und für ein ausgeglichenes Genauigkeit-Trefferquote-Verhältnis sorgt, den Recall Loss.

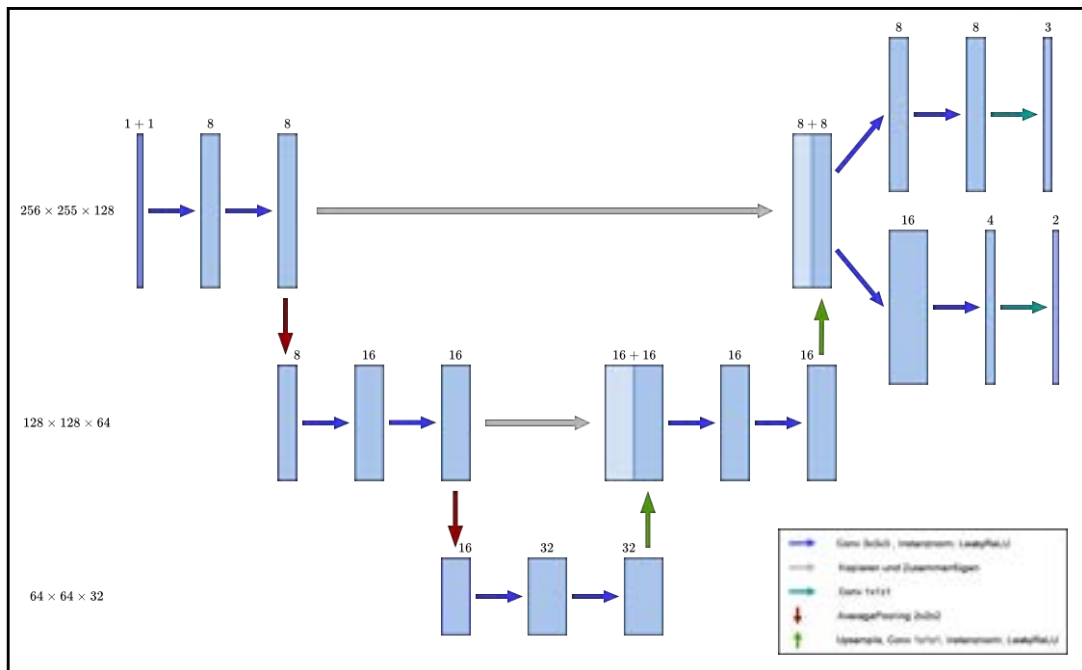
Wir betrachten als eine Klasse die Veränderungen, die über die Differenz der Segmentierungsmasken der Läsionen definiert sind, und als andere Klasse den Hintergrund. Dabei machen die Veränderungen (Maskenwert 1) nur einen sehr geringen Anteil (0.01%) des Bildes im Vergleich zum Hintergrund (Maskenwert 0) aus. Daher verwenden wir den Recall Loss.

Die einzelnen Verlustfunktionen gewichten und summieren wir, sodass wir als finalen Verlust  $L$  erhalten:

$$L = \lambda \cdot D_m^{NGF}(T, R, u) + \alpha \cdot S_m^{curv}(u) + \beta \cdot D^{key}(K^R, K^T, u) + \gamma \cdot D^{SSD}(T^S, R^S, u) + \delta \cdot S^{vol}(T^L, R^L, u) + \eta \cdot \text{RecallCE}(CM, T^L, R^L) \quad (4.3)$$

Die Gewichte unseres Netzes werden im Anschluss mittels ADAM [KB14] angepasst. Dieser Ablauf wiederholt sich für eine gewählte Anzahl an Epochen, bis das Training abgeschlossen ist.

Unser Netz ist so konstruiert, dass es lediglich das zu registrierende Bildpaar als Eingabe erwartet. Zum Testen des Netzwerks auf ungesesehenen Bildpaaren werden keine Segmentierungen oder Keypoints benötigt. Das trainierte Netz liefert ein dreikanaliges Deformationsfeld, wobei ein Kanal eine Dimension des Verschiebungsvektors darstellt. Zudem liefert das Netz eine zweikanalige Ausgabe für unsere Veränderungskarte, bei der die Kanäle einen Wert für den Hintergrund und einen für Veränderungen beschreibt.



**Abb. 4.3:** Die erstellte Netzwerkarchitektur, die zum gemeinsamen Lernen einer Registrierung und einer Veränderungskarte verwendet wird. Die blauen Rechtecke stehen für Merkmalskarten und die schwarzen Zahlen darüber beschreiben deren Anzahl an Kanälen. An der linken Seite sind die Größen der Merkmalskarten für die verschiedenen Ebenen gegeben. Die Architektur basiert auf dem U-Net [RFB15] und 3 Schichten. Die Eingabe sind zwei Merkmalskarten, die konkateniert sind, und es gibt zwei Ausgaben mit drei bzw. zwei Kanälen. Ein blauer Pfeil beschreibt eine Faltung mit einem Kernel der Größe  $3 \times 3 \times 3$ , eine Instanznormalisierung und anschließend eine Anwendung der leakyReLU. Die Kanaldimension passen sich entsprechend der gegebenen schwarzen Zahlen über den Merkmalskarten an. Ein roter Pfeil beschreibt ein Average-Pooling mit einem Kernel der Größe  $2 \times 2 \times 2$ . Durch diesen wird die Größe der Merkmalskarten halbiert. Mit einem grünen Pfeil verdoppelt sich diese wieder. Ein grüner Pfeil beschreibt ein Hochtasten (Upsampling), sowie anschließend eine Faltung mit einem Kernel der Größe  $1 \times 1 \times 1$ , eine Instanznormalisierung und eine Anwendung der leakyReLU. Ein türkiser Pfeil beschreibt nur eine Faltung mit einem Kernel der Größe  $1 \times 1 \times 1$ . Die Architektur ist u-förmig aufgebaut und mit einem grauen Pfeil werden Merkmalskarten kopiert und mit Merkmalskarten derselben Größe konkateniert. Nach dem letzten Hochtasten werden unterschiedliche Faltungen auf die Merkmalskarte ausgeführt. So entsteht zum einen eine Merkmalskarte mit 8 Kanälen, die nach weiteren Faltungen 3 Kanäle hat. Jeder Kanal dieser Merkmalskarte stellt eine Dimension des Deformationsfeld dar. Zum anderen entsteht eine Merkmalskarte, die zunächst 16 und nach weiteren Faltungen 2 Kanäle hat. Dies ist unsere Vorhersage der Veränderungen, wobei ein Kanal für den Hintergrund und ein Kanal für die Veränderungen steht.

### 4.2.2 Netzwerkarchitektur

Die Architektur unseres Faltungsnetzwerkes, welches eine Deformationsfeld und eine Veränderungskarte gemeinsam lernen soll, basiert auf der des U-Nets [RFB15] und ist in Abbildung 4.3 zusammengefasst. Die Architektur besteht aus einem kontrahierenden und einem expandierenden Pfad, der sich in der letzten Schicht in zwei Pfade (Heads) aufteilt, und hat eine Tiefe von 2.

Das einkanalige Referenz- und Templatebild stellen die Eingabedaten des Netzwerkes dar und werden im ersten Schritt entlang der Kanaldimension konkateniert, wenn diese getrennt dem Netzwerk übergeben wurden, sodass die Eingabe aus zwei Kanälen besteht.

Ein grundlegender Teil des Netzwerkes stellt die sogenannte *Doppelfaltung* dar, welche aus einer zweifachen Ausführung einer Faltung mit einem Faltungskern der Größe  $3 \times 3 \times 3$ , Instanznormalisierung und leakyReLU-Aktivierung besteht:

$$\begin{aligned} & \text{Faltung} \rightarrow \text{Instanznormalisierung} \rightarrow \text{leakyReLU-Aktivierung} \\ & \rightarrow \text{Faltung} \rightarrow \text{Instanznormalisierung} \rightarrow \text{leakyReLU-Aktivierung}. \end{aligned} \quad (4.4)$$

Dabei wird die Größe der Kanaldimension, wenn nicht anders beschrieben, innerhalb der ersten Faltung um Faktor 2 erhöht und innerhalb der zweiten Faltung beibehalten.

Eine weitere Abfolge, die sich wiederholt, ist die des Hochtastens, um die räumliche Dimension wieder herzustellen. Dafür führen wir zunächst eine trilineare Interpolation aus, die die Auflösung der Merkmalskarte verdoppelt. Anschließend halbieren wir mittels einer Faltung mit einem Faltungskern der Größe  $1 \times 1 \times 1$  die Anzahl der Kanäle. Zudem führen wir eine Instanznormalisierung und eine leakyReLU-Aktivierung aus:

$$\text{Interpolation} \rightarrow \text{Faltung} \rightarrow \text{Instanznormalisierung} \rightarrow \text{leakyReLU-Aktivierung} \quad (4.5)$$

Auf die konkatenierten Merkmalskarten führen wir eine solche Doppelfaltung aus, wobei in der ersten Faltung die Anzahl der Kanäle auf 8 erhöht wird. Anschließend wenden wir ein Average-Pooling auf nicht überlappenden Ausschnitten der Größe  $2 \times 2 \times 2$  an, um die Größe der Merkmalskarte zu verringern. Auf die resultierende Merkmalskarte mit halbiertes Auflösung wenden wir die Operationen (Doppelfaltung und Average-Pooling) erneut an, sodass wir eine Merkmalskarte mit einem Viertel der Auflösung der Eingabedaten und 16 Kanälen erhalten. Auf diese Merkmalskarte wenden wir erneut die Doppelfaltung an, die die Anzahl der Kanäle erneut verdoppelt, aber es folgt keine Average-Pooling. Damit haben wir unseren kontrahierenden Pfad definiert, über den lokale Informationen abstrahiert, zusammengefasst und als semantische Informationen in die Kanäle der Merkmalskarte projiziert werden.

Im darauffolgenden expandierenden Pfad wird zunächst die Auflösung aufs Doppelte hochskaliert. Anschließend konkatenieren wir die Merkmalskarte mit der vorherigen Merkmalskarte mit der entsprechenden Auflösungsstufe und mit der identischen Anzahl an Kanälen, damit globale Informationen übergeben werden.

Diese drei Schritte (Doppelfaltung, Hochskalieren, Konkatenieren) wenden wir erneut auf die konkatenierte Merkmalskarte an, sodass die resultierende Merkmalskarte die gleiche Auflösung wie die Eingabedaten und 16 Kanäle hat.

Auf diese Merkmalskarte wenden wir getrennt zwei Doppelfaltungen an, um zwei verschiedene Merkmalskarten zu erhalten. Auf die erste Doppelfaltung folgt eine einfache



Faltung mit einem Faltungskern der Größe  $1 \times 1 \times 1$ , welche die Kanalanzahl auf drei reduziert. Diese Merkmalskarte geben wir als Deformationsfeld zurück.

Bei der zweiten Doppelfaltung wird im Gegensatz zu den anderen Doppelfaltungen die Anzahl der Kanäle erst in der zweiten Faltung verändert. In der ersten Faltung dieser Doppelfaltung werden zunächst die 16 Kanäle beibehalten und erst anschließend auf 4 Kanäle reduziert, um mehr Parameter zum Lernen zu lassen. Auf diese zweite Doppelfaltung folgt eine einfache Faltung mit einem Faltungskern der Größe  $1 \times 1 \times 1$ , welche die Kanalanzahl auf 2 reduziert. Diese Merkmalskarte dient als Vorhersage für Veränderungen und wird zur Bestimmung der Veränderungskarte verwendet.

Das Deformationsfeld und Veränderungskarte teilen sich möglichst lange die Gewichte, da die Veränderungen mit der Deformation, genauer mit der Differenz von Referenz- und deformierten Templatebild, korrespondieren sollen. Jedoch soll die Veränderungskarte kein einfaches Differenzbild sein, sondern optimalerweise zwischen den verschiedenen Veränderungen unterscheiden, und Einfluss auf die Registrierung haben, indem sie starke Deformationen in den Bereichen von relevanten Veränderungen vermeidet. Daher wird sie parallel und nicht anschließend in einem weiteren Netz gelernt.

### 4.2.3 Erstellung der Veränderungskarten

Mit den Ausgaben des Netzwerks erstellen wir drei verschiedenen Veränderungskarten  $V \in [0, 1]$ . Die Veränderungskarte, basierend auf dem Differenzbild der Deep-Learning basierten Registrierung (Diff)  $V_{Diff}$  wird analog  $V_{klassisch}$  aus der Differenz des deformierten Templatebildes und dem Referenzbild wie in Abschnitt 4.1.2 erstellt.

Für die CM,  $V_{CM}$  wird der zweite Kanal der Vorhersage der Veränderungen verwendet. Zuvor wird eine Softmax ((2.23)) auf die Kanaldimension der Vorhersage der Veränderungen angewendet, um Werte zwischen 0 und 1 zu erhalten, sowie das Verhältnis von Hintergrund (Label 0) und Veränderung (Label 1) beizubehalten.

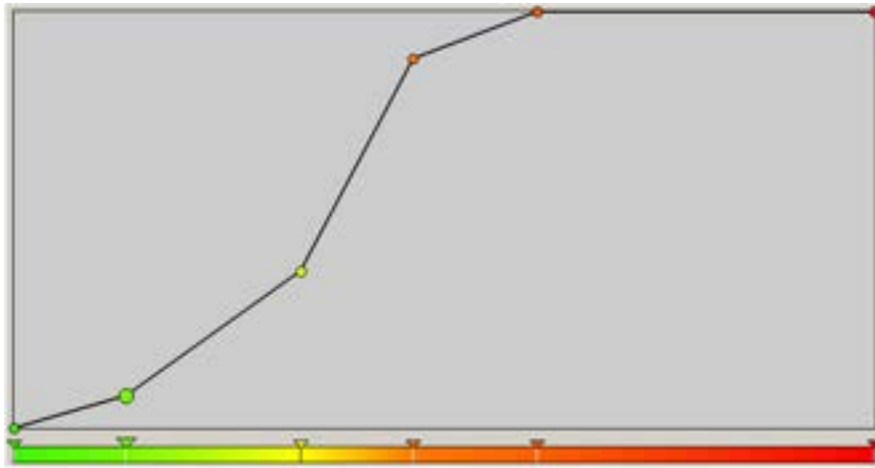
Die Veränderungskarte, kombiniert aus den Veränderungskarten der Deep-Learning basierten Registrierung und der Vorhersage der Veränderungen (CMxDiff),  $V_{CM \times Diff}$  wurde durch voxelweise Multiplikation und anschließender Skalierung auf das Intervall  $[0, 1]$  erstellt:

$$V_{CM \times Diff} = \frac{V_{CM} \odot V_{Diff}}{\max(V_{CM} \odot V_{Diff})}$$

## 4.3 Visualisierung der Veränderungskarte

Die Veränderungskarte wird als farbliche Überlagerung auf dem Referenzbild angezeigt. Dafür verwenden wir die MevisLab-Module *View2D*, *GVROrthoOverlay* und *SoLUTEditor* [mev]. Zur Visualisierung der Werte der Veränderungskarte werden verschiedene Farben und Intensitäten verwendet, die wir im *SoLUTEditor*-Modul setzen. Diese Zuordnung ist in Abbildung 4.4 abgebildet.

Kleinere Veränderungen werden grün und mit einer geringen Intensität angezeigt. Die Werte der Veränderungskarte im Bereich um werden gelb mit einer Intensität um  $\frac{1}{3}$  angezeigt. Werte größer als 0.5 haben eine Intensität über 0.9 und nehmen einen orangen bis roten Farbton. Werte größer als 0.6 haben die komplette Intensität.



**Abb. 4.4:** Die Abbildung zeigt die Zuordnung von Farbwerten und -intensitäten auf dem Bereich  $[0,1]$  im *SoLUTE*Editor-Modul für die Visualisierung der Veränderungskarte. Die Werte im Bereich um  $\frac{1}{3}$  werden gelb mit einer Intensität um  $\frac{1}{3}$  angezeigt. Werte größer als 0.5 haben eine Intensität über 0.9 und nehmen einen orangen bis roten Farbton. Werte größer als 0.6 haben zudem die komplette Intensität.

Umso größer der Wert der Veränderungskarte, desto stärker geht der Farbton ins Rote und desto stärker ist die Intensität. Dabei soll rot als Signalfarbe besonders aufmerksam auf Veränderungen machen, während grün die Aufmerksamkeit nicht so stark auf sich lenkt, aber dennoch kleinere Veränderungen markiert.

## 4.4 Vorverarbeitung der Daten fürs Training

Da wir fürs Training eine einheitliche und kleinere Auflösung der Bilder benötigen, werden die Bilder mit MevisLab vorarbeitet. Zudem wenden wir auch schon eine Translation und rigide Registrierung auf den Bildern in der Vorverarbeitung an, sodass im Netz nur die deformierbare Registrierung gelernt wird. Dies ist ähnlich zur klassischen Registrierung, bei der wir der deformierbaren Registrierung eine initiale Schätzung der Transformation aus der vorhergehenden Translation und rigiden Registrierung übergeben.

Für die Vorverarbeitung passen wir zuerst die Bildbereiche beider Bilder aneinander an. Als Nächstes bestimmen wir die Schwerpunkte der Bilder und verschieben das Templatebild so, dass die Schwerpunkte beider Bilder übereinander liegen, und gleichen die Weltkoordinaten an. Zudem bestimmen wir die Translation als ersten Registrierungsschritt, wobei wir das Normalized Gradient Field auf den überlappenden Bildbereichen (NGFOverlap) als Distanzmaß verwenden. Die kombinierte Transformationsmatrix aus den beiden vorherigen Schritten verwenden wir als initiale Schätzung für die folgende rigide Registrierung.

Bei der rigiden Registrierung verwenden wir die Multi-Level-Strategie, wobei wir über drei Level optimieren und für die feinste Auflösung Level 2 wählen, da wir nur eine grobe Registrierung in der Vorverarbeitung haben wollen. Die Bildauflösungen der verschiedenen Level bestimmen wir über Gaussian-Downsampling [YADG15]. Das Referenz- und das transformierte Templatebild mit geringerer Bildauflösung erstellen

wir mit dem *Resample3d*-Modul und dem *Lanczos*-Filter, welcher für dreidimensionale Bilder bessere Ergebnisse bezüglich der strukturellen Ähnlichkeit im Vergleich zu trilinearere und trikubische Interpolation liefert [MADSP20], und speichern diese.

Die Transformationsmatrix aus der rigiden Registrierung wenden wir neben dem Templatebild auch auf die Segmentierungsmasken an, die zum Templatebild gehören. In unserem Fall wenden wir diese sowohl auf der Segmentierungsmaske der Lunge als auch auf der Segmentierungsmaske der Läsionen an. Die Bildauflösung der Segmentierungsmasken, sowohl der transformierten Segmentierungsmasken des Templatebildes als auch die Segmentierungsmasken des Referenzbildes, verringern wir mit *Nearest-Neighbor*-Interpolation und speichern die neuen Masken.

In diesem Fall verwenden wir die Nearest-Neighbor-Interpolation, da wir nur ganzzahlige Werte und keine Zwischenergebnisse erhalten wollen. Dies ist insbesondere wichtig, wenn man keine binären Segmentierungsmasken, sondern zum Beispiel Segmentierungsmasken der Lungenlappen verwendet, bei denen jeder Lungenlappen einen eigenen Wert hat, diese aber in keiner Beziehung zueinander stehen.

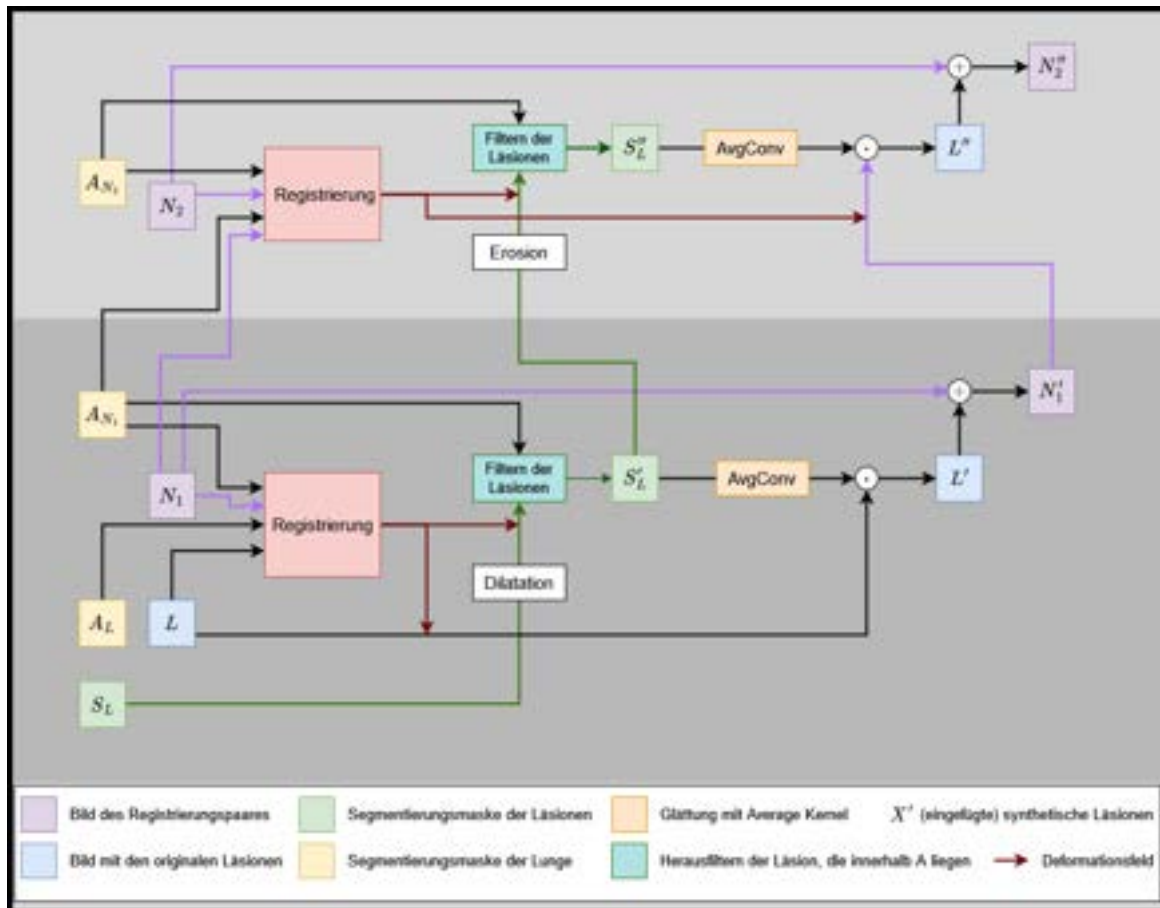
Zudem bestimmen wir mit dem *DetectLungKeypoints*-Modul, welches auf dem Paper von Rühaak [RPH<sup>+</sup>17] und dem Prinzip von Heinrich [HHS15] basiert, markante korrespondierende Punkte in der Lunge. Dafür verwenden wir das Referenzbild und das transformierte Templatebild sowie deren Lungensegmentierungsmasken, in der originalen Auflösung. Die Punkte speichern wir sowohl als Weltkoordinaten, als auch konvertiert als Voxelkoordinaten passend zu den Bildern mit geringerer Auflösung.

## 4.5 Erstellung synthetischer Daten

Für unsere Experimente haben wir synthetische Daten erstellt, damit wir genaue Segmentierungen von bekannten Läsionen haben. Dafür werden Läsionen, deren Segmentierung bekannt ist, in Bilder eingefügt. Da wir Veränderungen, partielle Remission und Progression, von Läsionen betrachten wollen, werden zwei verschiedene Arten von synthetischen Daten erstellt.

Für diese Methode werden Bilder eines Datensatzes benötigt, in dem jeweils korrespondierende Bilder ( $N_1, N_2$ ) existieren, z.B. Bilder von zwei verschiedenen Zeitpunkten von einem Patienten. Dieses Bilderpaar soll in den Experimenten miteinander registriert werden. Zudem werden Bilder  $L$  aus einem Datensatz benötigt, der die gleiche Körperregion, wie des ersten Datensatzes abbildet, und bei dem Segmentierungsmasken für auftretende Läsionen vorhanden sind.

In dem einen Fall der erstellten synthetischen Daten wird nur in einem Bild von dem Registrierungspaar eine Läsion eingefügt (*'add lesion'*), im zweiten Fall wird in dem einen Bild die komplette Läsion eingefügt und in dem anderen eine verkleinerte Version von dieser (*'growth lesion'*). Diese Verfahren sind in Abbildung 4.5 schematisch dargestellt und werden im Folgenden genauer erläutert. Dabei umfasst der hellere Kasten die Methode zum Einfügen von Läsionen in einem Bild und der dunkleren Kasten die Erweiterung zum Einfügen einer verkleinerten Läsion im zweiten Bild, welche fürs Hinzufügen von Läsionen in verschiedenen Größen in beiden Bildern benötigt wird.



**Abb. 4.5:** In der Abbildung ist die Methode zum Erstellen der synthetischen Daten schematisch dargestellt. Dabei beschreibt der Teil im hellen Kasten die Methode zum Einfügen von Läsionen im ersten Bild  $N_1$  und der Teil im dunkleren Kasten die Erweiterung, wenn zusätzlich im zweiten Bild  $N_2$  eine kleinere Läsion eingefügt werden soll. Die Einfärbung einiger Pfeile (lila und grün) dient zur besseren Übersicht. In beiden Fällen wird zunächst eine Deformationsfeld mit einer Registrierung bestimmt. Dieses ist als roter Pfeil dargestellt und wird auf die entsprechenden Bilder angewendet. Das Templatebild beinhaltet in beiden Fällen die Läsionen, die im Referenzbild eingefügt werden sollen. Die Segmentierungsmasken der Läsionen  $S_L$  werden entsprechend deformiert und anschließend gefiltert, sodass nur Läsionen, die auch innerhalb der Segmentierungsmaske des Organs  $A$  liegen, eingefügt werden. Beim Einfügen in das zweite Bild wird die Segmentierungsmaske der Läsionen vom ersten zuvor erodiert, um kleinere Läsionen zu erhalten. Die Masken werden geglättet (AvgConv) und mit den Templatebildern multipliziert, um die Läsionen auszuschneiden. Diese werden in die Referenzbilder mit einer Addition eingefügt.

### 4.5.1 Hinzufügen von Läsionen in einem Bild

Um Läsionen aus einem Bild  $L$  in ein Bild des Registrierungs-paares  $N_1$  einzufügen, haben wir zunächst den Grauwertbereich der beiden Bilder angepasst und so verschoben, dass für Luft ein Wert von 0 angenommen wird und der Grauwertbereich  $\mathcal{X} = [0, \dots, 4096]^k$  annimmt. Die Grauwerte lassen sich dann als eine um 1024 verschobene Hounsfield-Skala interpretieren.

Danach haben wir den Schwerpunkt und das Weltkoordinatensystem von Bild  $L$  an  $N_1$  angepasst und anschließend eine Translation als ersten Schritt der Registrierung berechnet. Dafür wurde das NGF, welches auf den sich überlappenden Bereichen der Bilder berechnet wird (NGFOverlap), als Distanzmaß verwendet.

Da die Läsionen an möglichst passender Stelle eingefügt werden sollen, haben wir eine Multi-Level-Registrierung der beiden Bilder ausgeführt, sowohl eine rigide als auch eine deformierbare Registrierung, für die wir die *MevisLab*-Module *RegLibParametric* und *RegLibNonParametric* genutzt haben. Das Bild  $R_1$  des Registrierungs-paares ist hierbei das Referenzbild und das Bild  $L$  das Templatebild, bei dem die Läsionen bekannte sind. Dementsprechend wird das Bild  $L$  auf das Bild  $N_1$  registriert, da die Läsionen in diesem eingefügt werden sollen.

Für die Registrierung wird die Bildauflösung für die verschiedenen Bild-Level über die Glättung mit einem Gaußkern und dem Downsampling berechnet [YADG15] (Multi Level Type: Gaussian downsampling). Sowohl bei der rigiden als auch deformierbaren Registrierung verwenden wir das NGF als Distanzmaß und optimieren über drei Level. Dabei setzen wir für die rigide Registrierung die feinste Auflösung auf Level 1 und optimieren mit dem Gauß-Newton Verfahren [WN<sup>+</sup>99], um eine gute rigide Registrierung zu erhalten.

Bei der deformierbaren Registrierung setzen wir die feinste Auflösung auf Level 2, um eine gröbere deformierbare Registrierung zu erhalten, da wir nur die groben Positionen zum Einfügen bestimmen wollen. Auch bei der deformierbaren Registrierung haben wir das NGF als Distanzmaß gewählt und gewichten dieses mit 1. Den Kantenparameter  $\varepsilon$  haben wir für die beide Bilder auf 10 gesetzt.

Als Regularisierer haben wir den Krümmungsregularisierer gewählt und diesen mit Faktor 100 gewichtet. Dabei verwenden wir das L-BFGS-Verfahren [WN<sup>+</sup>99] zum Optimieren, da dieses die inverse Hessematrix schätzt und reduziert abspeichert, sodass es weniger Speicherbedarf bei der Vielzahl an Variablen benötigt im Vergleich zum Gauß-Newton-Verfahren [WN<sup>+</sup>99].

Wenn vorhanden, können weitere Informationen in Form von Masken (oder Punkten) für die deformierbare Registrierung verwendet werden. Zum Beispiel können Segmentierungsmasken des Organs, in dem die Läsionen eingefügt werden sollen, verwendet werden, um besonders diese Bereiche bzw. die Kanten der Organe gut aufeinander zu registrieren. In unserem Fall möchten wir Läsionen in der Lunge einfügen und haben dafür die Segmentierung der Lungenflügel nach [LKS<sup>+</sup>11] bestimmt und zu einer Segmentierungsmaske der Lunge zusammengesetzt. Die Summe der quadrierten Differenzen (2.1.2) der beiden Lungen-Segmentierungsmasken gehen gewichtet mit 0.01 als weiteres Distanzmaß in die Optimierung ein.

Für die Registrierung berechnen wir nacheinander eine Translation, eine rigide Transformation und eine (nicht-parametrische) Deformation. Bei der rigiden Registrierung

wird die kombinierte Transformationsmatrix aus dem vorherigen Anpassen der Schwerpunkte und der vorherigen Translation als erste Anfangsschätzung verwendet und bei der deformierbaren Registrierung die Transformationsmatrix der rigiden Registrierung.

Anschließend haben wir die Deformation der deformierbaren Registrierung auf die Läsionssegmentierungsmasken angewendet und mit einem Kernel der Größe  $(3 \times 3 \times 3)$  dilatiert [Zam89], um einen kleinen Rand um die Läsion zu erhalten, da sich meist das Gewebe um Läsionen ebenfalls verändert bzw. die Ränder keine eindeutige Kante sind.

Um das Einfügen von Läsionen außerhalb der Lunge zu vermeiden, haben wir anschließend die Segmentierungsmaske einer einzelnen Läsion mit der Segmentierungsmaske der Lunge multipliziert und anschließend den Jaccard-Koeffizient der Multiplikation und der Läsionssegmentierungsmaske berechnet. Der Jaccard-Koeffizient beschreibt die Ähnlichkeit zweier Mengen und kann Werte zwischen 0 und 1 annehmen, dabei ist die Übereinstimmung umso größer, je näher der Jaccard-Koeffizient an der 1 liegt.

Wenn der Koeffizient kleiner oder gleich 0.95 ist, soll die Läsion nicht eingefügt werden, da sie nicht ausreichend in der Lunge liegt. Nur Läsionen, die einen Koeffizienten über 0.95 haben, haben wir in der neuen Läsionssegmentierungsmaske aufgenommen. Aufgrund der vorherigen Dilatation wurde ein Schwellwert kleiner 1 gewählt.

Diese neue Maske haben wir mit einer Faltung mit einem *Average Kernel* der Größe  $3 \times 3$  geglättet, um Übergänge an den Rändern zu erhalten, aber die Läsion selbst möglichst nicht zu glätten. Anschließend haben wir diese mit dem deformierten Bild  $L$  multipliziert. Die so erhaltenen Läsionen haben wir auf das originale Bild  $N_1$  addiert und somit die Läsionen eingefügt.

Da die Bildwerte bei 0 starten, kann diese einfache Addition ausgeführt werden und Strukturen des originalen Bildes bleiben erhalten, anstatt überschrieben zu werden. Das Erstellen dieser synthetischen Daten ist in Abbildung 4.5 im hellen Kasten schematisch dargestellt.

### 4.5.2 Hinzufügen von Läsionen in verschiedenen Größen in beiden Bildern

Wenn in beiden Bildern eine Läsion eingefügt werden soll, deren Größe sich aber unterscheidet, wird für das Einfügen einer Läsion in das erste Bild  $N_1$  des Registrierungs-paares identisch zum vorherigen Abschnitt 4.5.1.

Zum Einfügen der Läsion im zweiten Bild des Registrierungs-paares  $N_2$  gehen wir analog vor. Das erste Bild des Registrierungs-paares  $N_1$ , bei dem die Läsion in voller Größe eingefügt wird, ist hier das Templatebild und das zweite Bild des Registrierungs-paares  $N_2$ , bei dem die verkleinerte Läsion eingefügt wird, ist das Referenzbild. Zunächst passen wir den Grauwertbereich des zweiten Registrierungsbildes  $N_2$  genau wie den des ersten Registrierungsbildes  $N_1$  an, indem wir ihn so verschieben, dass für Luft ein Wert von 0 angenommen wird und es sich um eine verschobene Hounsfieldskala handelt.

Danach haben wir den Schwerpunkt und das Weltkoordinatensystem von Bild  $N_1$  an  $N_2$  angepasst und anschließend eine Translation als ersten Schritt der Registrierung berechnet. Das FASTA-Modul nutzt dazu das Normalized Gradient Field auf den überlappenden Bereichen als Distanzmaß (NGFOverlap). Diese Transformationsmatrix, die die beiden vorherigen Schritte zusammenfasst, dient wieder als erste Anfangsschätzung für die rigide Registrierung.

Auch hier wird sowohl eine rigide, als auch eine deformierbare Registrierung angewendet, bei denen die Bildauflösungen über *Gaussian downsampling* bestimmt wurden, und über 3 Level optimiert.

Diese rigide Registrierung nutzt ebenfalls Level 1 als feinste Auflösung und optimiert mit dem Gauß-Newton-Verfahren. Als Distanzmaß wird analog das NGF verwendet. Bei der deformierbaren Registrierung haben wir diesmal Level 1 statt Level 2 als feinste Auflösung gewählt, da wir beim Registrieren von den Bildern des Registrierungs-paares  $(N_1, N_2)$  eine höhere Genauigkeit benötigen, damit die Läsionen an korrespondierenden Stellen einzufügen werden. Aus diesem Grund wird bei dieser deformierbaren Registrierung der analog gewählte Krümmungsregularisierer mit 10 statt 100 gewichtet, um hier eine geringere Gewichtung auf die Glattheit und damit eine verhältnismäßig höhere Gewichtung der Distanz für die korrespondierenden Stellen zu erhalten, an denen die Läsionen eingefügt werden sollen.

Auch bei dieser deformierbaren Registrierung haben wir das NGF als Distanzmaß verwendet und mit Faktor 1 gewichtet. Um die Registrierung der Lungenbereiche zu verstärken, nutzt diese deformierbare Registrierung ebenfalls die Segmentierungsmasken der Lunge, welche über [LKS<sup>+</sup>11] erstellt wurden, als weiteres Distanzmaß. Dazu geht die SSD der beiden Lungensegmentierungsmasken der Bilder  $N_1$  und  $N_2$  mit 0.01 gewichtet in die Optimierung ein. Auch hier verwenden wir das L-BFGS-Verfahren [WN<sup>+</sup>99] für die Optimierung.

Die Deformation aus dieser Registrierung wenden wir auf das bereits deformierte Bild  $L$ , welches die Läsionen enthält, und auf die deformierte Segmentierungsmaske der Läsionen an. Die Segmentierungsmaske einer Läsion wird mittels Erosion [Zam89] verkleinert. Das benutzte Modul (FastMorphology) erwartet eine ungerade Größe des Kernels pro Dimension, dafür wird die Größe des Kernels pro Dimension mit  $2n + 1$  berechnet.

Damit Variationen in der Erosion vorkommen, definieren wir die Eingaben für die Größe des Kernels  $n$  zufällig. Dafür bestimmen wir für jede Läsion der Segmentierungsmaske den Begrenzungsrahmen (engl. Bounding Box), um damit den maximalen Durchmesser  $d_i$ ,  $i \in \{x, y, z\}$ , bezüglich der Dimensionsrichtungen zu erhalten.

Den Vektor  $v_i$  mit den möglichen Werten, die zur Bestimmung des Kernel pro Dimension verwendet werden sollen, bestimmen wir über  $r_i = \lfloor \frac{d_i - 1}{2} \rfloor$ ,  $v_i = [0, 1, \dots, r_i]$  mit  $i \in \{x, y, z\}$ . Da Werte im mittleren Bereich bevorzugt werden sollen, haben wir mit der Gaußverteilung ein Verteilungsvektor  $f$  erstellt, mit dem die Größe des Kernels pro Dimension zufällig gewählt wird.

$$v_i = \left[ \frac{-r_i}{2}, \dots, \frac{r_i}{2} \right], \quad g_i = \frac{1}{\sqrt{2\pi}} \exp \frac{-(v_i + 1)^2}{2}, \quad f = \frac{g_i}{\sum g_i} l = [0, \dots, r_i] \quad (4.6)$$

Mit diesen Eingaben wird ein kubischer Kernel erstellt, mit dem wir die Läsion erodieren und anschließend überprüfen, ob diese innerhalb des Lungenbereiches liegt. Dazu multiplizieren wir die Segmentierungsmaske der Läsion mit der Segmentierungsmaske der Lunge und berechnen wie zuvor der Jaccard-Koeffizient. Wenn der Koeffizient größer als 0.95 ist, bleibt die Läsion in der Segmentierungsmaske erhalten.

Für glatte Übergänge wird die Segmentierungsmaske aller Läsionen mit einem *Average Kernel* der Größe  $3 \times 3$  gefaltet. Anschließend extrahieren wir die Läsionen aus dem Bild, in dem wir die Segmentierungsmaske der Läsionen mit dem zweifach deformierten

Bild  $L$  multiplizieren, und addieren anschließend diese auf das zweite Bild des Registrierungspaars  $N_2$ . Auch hier kann die Addition durchgeführt werden, da zuvor der Bildbereich so verschoben wurde, dass der minimale Wert 0 ist.

Wir erhalten so zwei Bilder, in denen Läsionen mit unterschiedlicher Größe eingefügt wurden, und somit ein Wachstum bzw. eine Rezession der Läsionen nachgeahmt wird. Dabei kann in der späteren Registrierungen das modifizierte Bild  $N_1$  als Referenz- oder Templatebild und das modifizierte Bild  $N_2$  entsprechend als Template- oder Referenzbild verwendet werden, abhängig davon, ob eine partielle Remission oder eine Progression imitiert werden soll.



# 5 Daten und Experimente

## 5.1 Verwendete Daten

In unserer Arbeit verwenden wir Bildpaare aus dem *National Lung Screening Trial (NLST)* Datensatz [Tea13, Tea11]. Zur Erstellung von synthetischen Daten werden annotierte Läsionen aus dem *Lung Image Database Consortium (LIDC)* Datensatz [AI15, AIMB<sup>+</sup>11], wie in Abschnitt 4.5 beschrieben, in die NLST Bildpaare eingefügt.

### 5.1.1 Datensätze

#### National Lung Screening Trial (NLST)

Die National Lung Screening Trial (NLST) [Tea13, Tea11] ist eine vom National Cancer Institute (NCI) geförderte und vom American College of Radiology Imaging Network (ACRIN) und der Lung Screening Study Group durchgeführte Studie zur Lungenkrebsvorsorge. Zur Früherkennung von Lungenkrebs werden sowohl Röntgenaufnahmen des Brustkorbs als auch Spiral-Computertomografien (CT) eingesetzt. Die Auswirkungen dieser Screening-Techniken auf die Sterblichkeitsrate bei Lungenkrebs waren damals noch nicht weitreichend geklärt. Daher startete 2002 die NLST-Studie, bei der die zwei Methoden zur Erkennung von Lungenkrebs hinsichtlich ihrer Auswirkungen auf die Sterberate an Lungenkrebs in einer Hochrisikopopulation miteinander verglichen wurden. Über einen Zeitraum von 21 Monaten nahmen 53454 derzeitige oder ehemalige starke Raucher im Alter von 55 bis 74 Jahren an der NLST-Studie teil. In Rahmen dieser Arbeit verwenden wir die Follow-Up CT-Aufnahmen, welche im Abstand von einem Jahr aufgenommen wurden.

Einige dieser CT-Daten wurden für die *Learn2Reg*-Challenge 2022 verwendet [DHHH22]. Dabei wurden Daten ausgewählt, bei denen Bilder von einem Patienten von mindestens zwei Zeitpunkten vorhanden sind und die mindestens eine Läsion oder Lungenrundherd (engl. nodule) in jedem Bild enthalten ist. Die Aufnahmen von den verschiedenen Zeitpunkten eines Patienten bilden ein Bildpaar für die Bildregistrierung. Zudem wurden Informationen über die Läsionen in den Bildern, wie Schwerpunkt (engl. center of gravity) und axialer Durchmesser, extrahiert.

Eine Teilmenge von 250 der Learn2Reg NLST-Daten wird in dieser Arbeit zur Entwicklung und Auswertung unserer Methode verwendet. Zur Vereinfachungen sprechen wir nur von Läsionen und soll die Lungenrundherde einschließen.

#### Lung Image Database Consortium (LIDC)

Das *Lung Image Database Consortium (LIDC)* und die *Image Database Resource Initiative (IDRI)* haben eine Datenbank erstellt [AI15, AIMB<sup>+</sup>11], die aus thorakalen

Computertomografien (CT) zur Diagnose und Früherkennung von Lungenkrebs mit annotierten Läsionen besteht.

Sieben akademische Zentren und acht Unternehmen der medizinischen Bildgebung arbeiteten zusammen, um diesen Datensatz zu erstellen, der 1018 Fälle umfasst. Jeder Fall enthält Bilder einer klinischen Thorax-CT-Untersuchung und eine zugehörige XML-Datei, in der die Ergebnisse eines zweistufigen Bildannotationsprozesses aufgezeichnet sind, der von vier erfahrenen Thoraxradiologen durchgeführt wurde.

Die LIDC-IDRI-Datenbank soll eine wichtige Ressource für die Forschung im Bereich der medizinischen Bildgebung darstellen, um die Entwicklung, Validierung und Verbreitung von CAD in der klinischen Praxis voranzutreiben.

### 5.1.2 Synthetische Daten

Wir haben für unsere Experimente 200 synthetische Datenpaare, wie in Abschnitt 4.5 beschrieben, mit dem NLST- und LIDC-Datensatz erstellt und zusätzlich 50 Datenpaare unverändert aus dem NLST-Datensatz genommen. Davon wurden 200 Datenpaare beim Trainieren des Netzwerks 4.2.2 und 50 Datenpaare zur Evaluierung (Test) verwendet. Für die gleichen 50 Datenpaare wurde auch eine Deformation über die klassische Registrierungsmethode (Abschnitt 4.1) bestimmt, um die Methoden vergleichen zu können.

Keines der NLST-Datenpaare wurde doppelt verwendet. Für die Erstellung der synthetischen Daten stehen uns 200 LIDC-Daten mit zugehörigen Segmentierungsmasken der Läsionen zur Verfügung. Die LIDC-Daten wurden zufällig den NLST-Datenpaaren zugeordnet. Dafür wurde bei den Testbildern nur die ersten 50 LIDC-Daten und bei den Trainingsdaten nur die restlichen 150 LIDC-Daten verwendet, um disjunkte Trainings- und Testmengen zu garantieren. Innerhalb der Mengen werden durch die geringe Anzahl der LIDC-Daten diese teilweise doppelt verwendet, jedoch werden diese dann für die zwei unterschiedlichen Typen von synthetischen Daten genutzt. Die Läsionen werden also einmal nur in einem Bild ('add lesion') eingefügt und einmal mit verschiedenen Größen in den beiden Bildern ('growth lesion') eingefügt.

Das Bild vom NLST-Datenpaar vom früheren Zeitpunkt wird als Referenzbild und das Bild vom späteren Zeitpunkt als Templatebild verwendet. In welchem Bild die (größere) Läsion eingefügt wird, wird bei der Erstellung der synthetischen Daten zufällig bestimmt.

Die 200 Datenpaare, die beim Trainieren verwendet wurden, bestehen aus 80 synthetischen Datenpaaren, bei denen in ein Bild Läsionen eingefügt wurden ('add lesion'), 80 synthetischen Datenpaaren, bei denen in beiden Bild Läsionen eingefügt wurden ('growth lesion'), und 40 Datenpaaren, die nicht verändert wurden.

Fürs Training wurden die Datenpaare zufällig in Trainings- und Validierungsdatsatz im Verhältnis 4 : 1 aufgeteilt, sodass 160 Datenpaare zum Trainieren und 40 zum Validieren verwendet wurden. In allen drei Datensätzen (Training, Validierung und Test) ist das Verhältnis von den Bildpaaren, in denen in einem, beiden bzw. keinem Bilder Läsionen eingefügt wurden, identisch und beträgt 2 : 2 : 1. Da die Läsionen der NLST-Datenpaare auch ein Teil der synthetischen Datenpaare sind, haben wir uns dazu entschieden, einen kleineren Anteil an unveränderten Datenpaaren zu verwenden. Die genauen Anzahlen und Aufteilung sind in Tabelle 5.1 aufgelistet.

**Tab. 5.1:** Anzahl und Aufteilung der 250 Datenpaare für Training, Validierung und Testen, sowie die Aufteilung bezüglich der Anzahl der Bilder, in die synthetisch Läsionen eingefügt werden.

	gesamt	Einfügen von künstliche Läsionen in		
		einem Bild	beiden Bildern	keinem Bild
Training	160	64	64	32
Validierung	40	16	16	8
Test	50	20	20	10

Die Bilder haben eine Auflösung von 512 Voxeln in  $x$ - und  $y$ -Dimension und eine unterschiedliche Anzahl an Schichten in die  $z$ -Richtung, die zwischen 96 und 455 bei den verwendeten 50 Testbildern liegt (durchschnittlich 167). Dabei beträgt die Voxelgröße für die  $x$ - und  $y$ -Dimension zwischen 48 mm und 0.97 mm (durchschnittlich 0.66 mm) und für die  $z$ -Dimension zwischen 0.625 mm und 2.5 mm (durchschnittlich 2 mm). Für das Netzwerk der Deep Learning-basierten Registrierung wird eine kleinere und einheitliche Bildgröße benötigt, weshalb mit der in Abschnitt 4.4 beschriebenen Vorverarbeitung die Auflösung für alle Bilder auf  $256 \times 256 \times 128$  angepasst wird.

### 5.1.3 Erstellung (approximierter) Segmentierungsmasken für Läsionen

Für den LIDC-Datensatz stehen uns Segmentierungsmasken der Läsionen zur Verfügung, für den NLST-Datensatz jedoch nur Schwerpunkte und axiale Durchmesser  $d$  der korrespondierenden Läsionen.

Wir approximieren daher die Segmentierungsmasken als Kugeln und verwenden dafür die MevisLab-Module *SetVoxelValue* und *LabelDilation* [mev]. Um die komplette Läsion und auch Gewebe um die Läsion zu erfassen, verwenden wir für die Dilatation einen größeren Radius  $r$  als den halben Durchmesser  $d$ . Dieser wird je nach Größe des Durchmessers unterschiedlich definiert:

$$r = \begin{cases} \frac{d}{2} & \text{für } d > 20 \text{ mm} \\ 1.5 \cdot \frac{d}{2} & \text{für } d \leq 10 \text{ mm} \\ (2 - 0.05d) \cdot \frac{d}{2} & \text{sonst.} \end{cases} \quad (5.1)$$

Für die Segmentierungsmaske der synthetischen Daten werden die Segmentierungsmasken der LIDC-Daten und die approximierten Segmentierungsmasken der NLST-Daten addiert und überprüft, ob sich Läsionsbereiche überschneiden und gegebenenfalls dieser Fall aussortiert.

## 5.2 Experimente

In dieser Arbeit werden verschiedene Ansätze zur Erstellung von Veränderungskarten miteinander verglichen:

Für die Erstellung einer Veränderungskarte mit klassischer Registrierung, wie in Abschnitt 4.1 der Methoden beschrieben, verwenden wir das interne MevisLab vom

Fraunhofer MEVIS [mev], ein modulares Framework für die Forschung und Entwicklung im Bereich der Bildverarbeitung, welches auf Python [van95] und C++ [ISO20] basiert.

Für die Daten mit einer Auflösung von 512 Voxeln in x- und y-Dimension und unterschiedlicher Auflösung in z-Dimension haben wir die Gitterabstände für alle deformierbaren Registrierungen in dem *RegLibNonParametric*-Modul auf 55 für x- und y-Richtung und 33 in z-Richtung gesetzt. Die anderen gewählten Parameter sind bereits in der Methode 4.1 festgesetzt worden.

Das gemeinsame Lernen einer Registrierung und einer Veränderungskarte, wie in Abschnitt 4.2 der Methoden beschrieben, ist in Python mithilfe von PyTorch [NEU] umgesetzt. Dies ist eine Open-Source-Bibliothek, welche die effiziente Implementierung von neuronalen Netzen und die Beschleunigung von Berechnungen durch Verwendung des Grafikprozessors ermöglicht [van95]. Durch die Limitierung der Hardware (Intel(R) Xeon(R) Silver 4208 CPU @ 2.10GHz , NVIDIA GeForce RTX 2080 Ti) werden die Bilder, wie in Abschnitt 4.4 beschrieben, vorverarbeitet, sodass sie eine Auflösung von  $256 \times 256 \times 128$  haben. Die Parameter für die Verlustfunktion (4.3)

$$L = \lambda \cdot D^{NGF_m}(T, R, u) + \alpha \cdot S_m^{curv}(u) + \beta \cdot D^{key}(K^R, K^T, u) + \gamma \cdot D^{SSD}(T^S, R^S, u) + \delta \cdot S^{vol}(T^L, R^L, u) + \eta \cdot RecallCE(CM, T^L, R^L)$$

werden empirisch auf einem Teildatensatz der Trainingsbilder bestimmt. Dafür wird  $\lambda$  auf 1 gesetzt und die restlichen Parameter basierend auf visuellen Ergebnissen, sowie dem Verhältnis der einzelnen Verlustfunktionen angepasst. Die folgenden Parameter werden für die Experimente verwendet:

<i>Parameter</i>	<i>Parameterwert</i>
$\lambda$	1
$\alpha$	$10^5$
$\beta$	$10^8$
$\gamma$	$10^2$
$\delta$	$10^5$
$\eta$	$10^7$

Zur Minimierung unserer Verlustfunktion verwenden wir die Adaptive Moment Estimation (ADAM) Methode [KB14], welche eine Weiterentwicklung des Gradientenabstiegs ist. Sie verwendet unterschiedliche adaptive Lernraten für verschiedene Parameter, welche anhand der ersten Momente der Gradienten geschätzt werden. ADAM ist robust und gut geeignet für eine Vielzahl an nicht-konvexen Problemen im Bereich des maschinellen Lernens [KB14]. Insgesamt werden die Parameter über 50 Epochen optimiert.

### 5.2.1 Quantitative Auswertung der Methoden

Das eigentliche Ziel der Veränderungskarte ist die visuelle Hervorhebung der Veränderungen. Trotzdem soll in diesem ersten Experiment eine quantitative Auswertung erfolgen.

**Tab. 5.2:** Verwendete Schwellwert  $\tau$  und zugehörige Schwellwertnummer  $\vartheta$  mit logarithmischem Zusammenhang auf vier Dezimalstellen gerundet.

$\vartheta$	1	2	3	4	5	6	7	8	9	10
$\tau$	0	0.1383	0.243	0.3274	0.3979	0.4586	0.5119	0.5593	0.6021	0.641

$\vartheta$	11	12	13	14	15	16	17	18	19	20
$\tau$	0.6767	0.7097	0.7404	0.769	0.7959	0.8212	0.8451	0.8678	0.8893	0.9098

$\vartheta$	21	22	23	24	25
$\tau$	0.9294	0.9482	0.9661	0.9834	1

Es soll im ersten Schritt betrachtet werden, ob in den Veränderungskarten die bekannten Veränderungen der Läsionen markiert werden und im zweiten Schritt, ob die Größe der Veränderung für diese Markierungen entscheidend ist.

Zur Auswertung erstellen wir mittels verschiedener Schwellwerte  $\tau$  aus unserer kontinuierlichen Veränderungskarte  $V \in [0,1]$  eine binäre Veränderungskarte  $V^\tau \in \{0,1\}$ . Dabei wurden 25 Schwellwerte mit einem logarithmischen Zusammenhang bestimmt (5.2), sodass im unteren Bereich größere Schritte zwischen den Schwellwerten und im oberen kleinere Schritte zwischen den Schwellwerten sind, da die Unterschiede im oberen Bereich eine größere Bedeutung für unsere Auswertung haben.

Für die Erstellung der binären Veränderungskarte werden die Werte der Voxel größer dem Schwellwert auf 1 gesetzt und alle Werte der Voxel kleiner gleich dem Schwellwert auf 0 gesetzt. Anschließend werden für diese die einzelnen zusammenhängenden Komponenten mit Funktionen aus dem *scikit-image* Bibliothek [vdWSN<sup>+</sup>14] bestimmt. Ebenso bestimmen wir für unsere Grundwahrheit die zusammenhängenden Komponenten. Für diese werden die Segmentierungsmaske der Läsionen des Referenzbildes und die deformierten Segmentierungsmasken des Templatebildes in einer zusammengefügt.

Über beide Mengen der Komponenten wird iteriert und in einer Liste gespeichert, ob in den Bereichen der Komponente in der Grundwahrheit und in der Veränderungskarte mindestens ein Voxel mit einem positiven Wert, als einer Markierung, vorhanden ist.

Zuerst wird über die Komponenten der Grundwahrheit iteriert und eine Überlagerung von den Markierungen in der Grundwahrheit mit denen in der Veränderungskarte als 'true positive' und ohne Überlagerung als 'false negative', in die Liste angefügt. Zudem werden in die Liste bekannte Informationen zu den Läsionen, wie die Differenz der axialen Durchmesser der Läsionen von den beiden Zeitpunkten, hinzugefügt.

Als Zweites wird über die Komponenten der binären Veränderungskarte iteriert und überprüft, ob eine Überlagerung dieser Markierung mit der Grundwahrheit vorhanden ist, es sich also um ein 'true positive' handelt oder um ein 'false positive'. Hierbei werden jedoch nur die 'false positive' Fälle der Liste hinzugefügt, weil es sonst zur doppelten Betrachtung käme. Da diese Komponenten keine Läsionen sind, kann nur ein axialer Durchmesser der Komponente und keine Differenz der ursprünglichen Läsionen abgespeichert werden.

Dieses Vorgehen bedeutet, dass eine Komponente der Veränderungskarte, die über zwei Komponenten der Grundwahrheit liegt, als zwei Fälle eingeht und damit bei-

de Komponenten der Grundwahrheit in die Auswertung eingehen. Wenn aber eine Komponente der Grundwahrheit über zwei Komponenten der Veränderungskarte liegt, geht nur ein Fall, die der Komponente der Grundwahrheit, in die Auswertung mit ein. Die Komponenten der binären Veränderungskarte, die keine Markierung in den entsprechenden Bereichen der Grundwahrheit haben, gehen alle einzeln ein.

Durch das Iterieren über die Komponenten der beiden Masken gibt es keine Bereiche, die richtiger weiser nicht positiv erkannt wurden (true negative (TN)). Durch die verändernde Größe der Komponenten und einer Auswirkung dieser Größe, sind in einem Bereich zwei kleine oder eine große Komponente, auf die Auswertung, werden keine Komponenten dieser Art erzeugt und dementsprechend gehen sie nicht in die Bewertung ein. Die so erstellten Listen können für folgenden Auswertungen verwendet werden.

### Genauigkeit und Trefferquote

Für die verschiedenen Schwellwerte betrachten wir für diese jeweils die Genauigkeit, auch positiver Vorhersagewert (PPV, engl. precision), und die Trefferquote, auch Sensitivität (engl. recall), bezüglich der binären Veränderungskarte und unserer Grundwahrheit über die erstellten Listen und stellen diese in einer Precision-Recall-Kurve (PR-Kurve) zueinander dar.

Mit der Fläche unter der Precision-Recall-Kurve (AUC-PR) (engl. area under the precision-recall-curve) kann dieses in einem Wert zusammengefasst werden, dem AUC-PR-Wert. Im Allgemeinen gilt: Je höher der AUC-PR-Wert, desto besser schneidet ein Klassifikator für die jeweilige Aufgabe ab. Da sich in unserem Fall die Grundmenge für jeden Schwellwert verändert, kann dieser Wert für die einzelnen Methoden verzerrt sein. Jedoch können die Kurven und der AUC-PR-Wert für die verschiedenen Methoden verglichen werden, da jeweils ein höherer Wert auch für höhere Genauigkeiten und Trefferquoten spricht.

### Dice-Koeffizienten

Um auszuschließen, dass der gesamte Bereich als Veränderung markiert wird, wird für alle Schwellwerte  $\tau$  der Dice-Koeffizient von der binären Veränderungskarte  $V^\tau$  und der Lungensegmentierungsmaske des Referenzbildes berechnet. Hierbei gibt es keinen idealen Wert, da das Volumen der Läsionen im Vergleich zum Lungenvolumen verhältnismäßig gering ist, sollte sich dies auch im Dice-Koeffizienten widerspiegeln. Zusätzlich wird auch der Dice-Koeffizient von den binären Veränderungskarten  $V^\tau$  und der Grundwahrheit betrachtet, um abschätzen zu können, wie viel von den Läsionen markiert wurde. Im Gegensatz zu typischen Segmentierungsproblemen streben wir keinen Dice-Koeffizienten von 1 an, da die Veränderungskarte auch eine größere Fläche als die binäre Veränderung markieren können, wodurch die Veränderung deutlicher hervorgehoben wird.

### Durchmesserdiffenzen

Um zu überprüfen, welche Größe von Veränderungen erkannt werden, werden für jede Methode die F2-Maße (2.24) der verschiedenen Schwellwerte  $\tau$  bestimmt. Es werden jeweils die zwei Schwellwerte  $\tau \neq 0$  mit den besten F2-Werten ausgewählt und zusätzlich ein

**Tab. 5.3:** Übersicht der Quantile (Grenzen der Intervalle) und Anzahl der Läsionen in entsprechenden Intervallen. Intervallgrenzen in mm.

Intervall	Anzahl
[0, 0.2)	21
[0.2, 0.5)	32
[0.5, 1.1)	30
[1.1, 2.9)	28
[2.9, 4.2)	28
[4.2, 6.0)	27
[6.0, 7.4)	28
[7.4, 9.2)	29
[9.2, 13.3)	28
[13.3, 33.8)	28

fester Schwellwert (0.68) für alle Methoden im mittleren Bereich, der in der farblichen Visualisierung, wie in 4.3 beschrieben, einen orange-roten Farbton annimmt.

Für die Differenzen des axialen Durchmessers zwischen Referenz- und Templatebild der 281 Läsionen werden 10 Quantile in Schritten von 0.1 berechnet und diese als Grenzen für die Eingruppierung der Durchmesserdiffenzen verwendet. Die Grenzen für diese Gruppen und die Anzahl der Läsionen innerhalb einer Gruppe sind in Tabelle 5.3 aufgelistet.

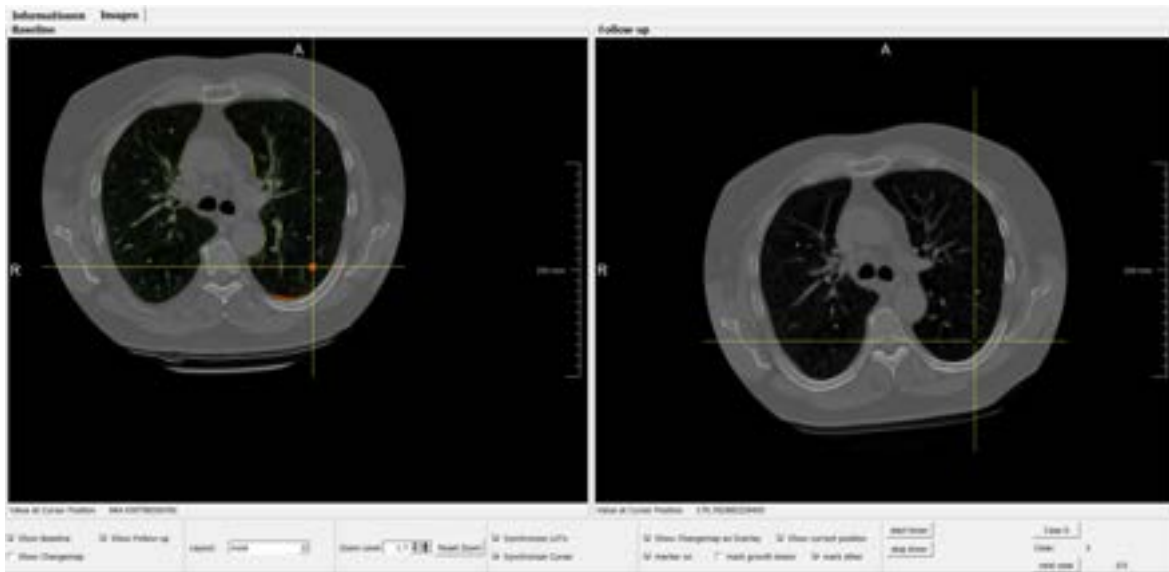
Es wurde sich für die Aufteilung mittels Quantile entschieden, damit die Gruppen möglichst gleich viele Läsionen beinhalten. Für jeder dieser Gruppen wird die relative Häufigkeit von richtig markierten (true positive (TP)) und nicht erkannten (false negative (FN)) Läsionen für jede Methode und den gewählten Schwellwerten berechnet.

### 5.2.2 Anwenderstudie

Das eigentliche Ziel der Veränderungskarten ist die Hervorhebung von Veränderungen für die Radiologen, sodass der Befundungsprozess beschleunigt wird und die Aufmerksamkeit auch auf Veränderungen, die leicht übersehen werden können, gelenkt wird. Aus diesem Grund haben wir eine Anwenderstudie durchgeführt, um den Nutzen und den Unterschied der farblichen Überlagerung zu testen. Dafür haben wir eine Applikation in MevisLab gebaut, in dem ein Referenzbild (Baseline) und ein Templatebild (Follow-up) angezeigt werden und Punkte in diesem markiert werden können. Das Interface dieser ist in Abbildung 5.1 dargestellt.

Das Templatebild wird mit einer rigiden Registrierung transformiert, sodass Template- und Referenzbild ähnlicher werden, aber keine Längen sowie Winkel und damit Größen verändert werden.

Es gibt drei Studienteile, in denen sechs Bildpaare betrachtet werden und Läsionen markiert werden sollten. Dabei geht immer das erste Bildpaar nicht in die Wertung mit ein, damit es zum Gewöhnen an die Software verwendet werden kann. Zudem wird für dieses immer das gleiche Bildpaar verwendet. Die anderen fünf Bildpaare werden in jedem Studienteil in unterschiedlicher Reihenfolge angezeigt. Zudem gab es eine *Cool*



**Abb. 5.1:** Ein Beispiel aus dem zweiten Teil der Anwenderstudie mit einer grünen Markierung für eine Veränderung von einer Läsion und einer rosa Markierung für weitere Auffälligkeiten. Auf dem Referenzbild (Baseline) ist eine Veränderungskarte als farbliche Überlagerung angezeigt. Das gelbe Kreuz stellt die aktuelle Position des Cursors im Referenzbild dar, sowie die korrespondierende Position im Templatebild (Follow-up) und hat in der Mitte eine Läsion, die im Referenzbild in die Veränderungskarte markiert wurde

*Down Phase* von sieben Tagen, um zu verhindern, dass die Studienteilnehmer sich an die Fälle erinnern.

Die Bildpaare werden händisch ausgesucht. Dabei sollen in den Bildern möglichst viele annotierte Läsionen und alle Arten von Bildtypen vorhanden sein. Es gibt also ein Bildpaar, bei dem nur im Referenzbild Läsionen eingefügt sind, ein Bildpaar, bei dem nur im Templatebild Läsionen eingefügt sind, sowie jeweils ein Bildpaar, bei dem im Referenz- bzw. Templatebild die größeren Läsionen und im anderen Bild die kleineren Läsionen eingefügt sind. Als Letztes wurde noch ein Bildpaar gewählt, bei dem keine zusätzlichen Läsionen eingefügt wurden.

Allgemein werden die Bilder ausgewählt, die keine markante Auffälligkeiten haben, sodass ein Lerneffekt vermieden wird. Zudem soll zwischen den einzelnen Studienteilen ungefähr eine Woche Zeit liegen.

Beim ersten Studienteil sollen die Teilnehmer ohne farbliche Überlagerung die Läsionen suchen und markieren. Im zweiten Teil werden zusätzlich die Veränderungskarte, die aus klassischer Registrierung 4.1 resultiert, als farbliche Überlagerung auf dem Referenzbild gegeben. Im dritten Teil werden ebenfalls eine Veränderungskarte, die diesmal aus einer Kombination der Ausgaben der Deep Learning-basierten Registrierung 4.2 resultiert, als farbliche Überlagerung auf dem Referenzbild gegeben.

Beim Referenzbild kann nach Belieben zwischen der Ansicht mit oder ohne dieser farblichen Überlagerung gewechselt werden. Läsionen sollen in der Studie in grün markiert werden und weitere Auffälligkeiten sollen mit rosa markiert werden. Ein Beispiel des Moduls mit zufällig gesetzten Markierungen in grün und rosa sowie einer markierten Läsion ist in Abbildung 5.1 zu sehen. Die genaue Aufgabenstellung für die Studienteilnehmer ist in der Box Aufgabenstellung zu finden.



**AUFGABENSTELLUNG**

Markiere Läsionen (Nodules), deren Größe sich signifikant verändert, in grün. (Richtwert: Laut RECIST Kriterien handelt es sich um eine Rezession, wenn die Summe der Durchmesser um mindestens 30% gesunken und als Progression, wenn die Summe der Durchmesser um mindestens 20% gewachsen ist)

Die Stelle muss nur auf einem der beiden Bilder markiert werden. Markiere möglichst auf dem Bild, auf dem die Läsion größer ist und möglichst mittig von dieser.

Markiere andere Veränderungen, Auffälligkeiten und Läsionen, deren Größe sich nicht verändert, in rosa.

Am Anfang jedes Falles drücke auf *start timer*, am Ende zuerst *stop timer* und wenn du weiter machen möchtest *next case*.

Du kannst auch zwischen Fällen pausieren, dann denke aber besonders daran *start timer* und *stop timer* zu klicken.

Innerhalb eines Falles mache keine Pause, da die Zeit gestoppt wird.

Der erste Fall (Case 0) geht nicht in die Wertung mit ein und ist zum Austesten der Funktionen gedacht. Die 5 folgenden Fälle (Case 1-5) gehen in die Wertung ein.

Eine Anleitung zum Modul findest du unter *Anleitung*.

Vielen Dank, dass du mitmachst!

Zudem bekommen die Studienteilnehmer eine Anleitung zu dem Modul. Diese ist für den zweiten Studienteil im Anhang unter Anleitung zu finden. In der Anleitung zum ersten Studienteil werden die Anmerkungen bezüglich der farblichen Überlagerung, die erst ab dem zweiten Studienteil vorkommt, ausgelassen.

Sowohl die Verteilung der Zeiten als auch die Verteilungen der Genauigkeit und Trefferquote werden für die verschiedenen Teilnehmer auf signifikante Veränderungen mittels eines zweiseitigen Wilcoxon-Rangsummen-Tests [Wil92] getestet. Die Hypothese der Gleichheit zweier Verteilungen werden für einen  $p$ -Wert  $p < 0.05$  abgelehnt und damit signifikante Unterschiede in den Verteilungen angenommen.

Um Ergebnisse besser einordnen zu können und Rückmeldung für zukünftige Anwenderstudien zu erhalten, werden den Teilnehmern nach der Studie Fragen gestellt, zu denen Beispiele gegeben werden, die während der Studie oder in der Entwicklung aufkamen:

#### FRAGEN

1. Was könnte deine Leistung in der Studie verzerrt haben?  
(Beispiel: Verschiebung beim Zoomen, sodass die Bilder nicht mehr parallel lagen)
2. Was kann in einer zukünftigen Anwenderstudie noch verbessert werden, was sollte zukünftig eingebaut oder beachtet werden?  
(Beispiel: Messtool, um die Größe der Läsionen messen zu können)
3. Hast du weiteres Feedback, was nicht zu den vorherigen Fragen passt?

## 6 Ergebnisse

Die verschiedenen Methoden zum Erstellen einer Veränderungskarte werden sowohl quantitativ als auch qualitativ ausgewertet. Dabei kann keine eindeutige Überlegenheit einer Methode bezüglich der quantitativen Merkmale (Trefferquote, Genauigkeit, Dice-Koeffizient und Erkennung der verschiedenen Durchmesserdiffenzen) erkannt werden.

Zudem wird der Nutzen zweier Methoden mit einer Anwenderstudie getestet, bei der Verbesserungen bezüglich der Bearbeitungszeit, jedoch keine signifikanten Verbesserungen, zu beobachten sind. Bezüglich der annotierten Läsionen zeigt sich kein eindeutiger Nutzen durch Verwendung von Veränderungskarten. Im Folgenden werden die Ergebnisse der in Kapitel 5 vorgestellten Experimente genauer beschrieben.

### 6.1 Quantitative Auswertung

#### 6.1.1 Genauigkeit und Trefferquote

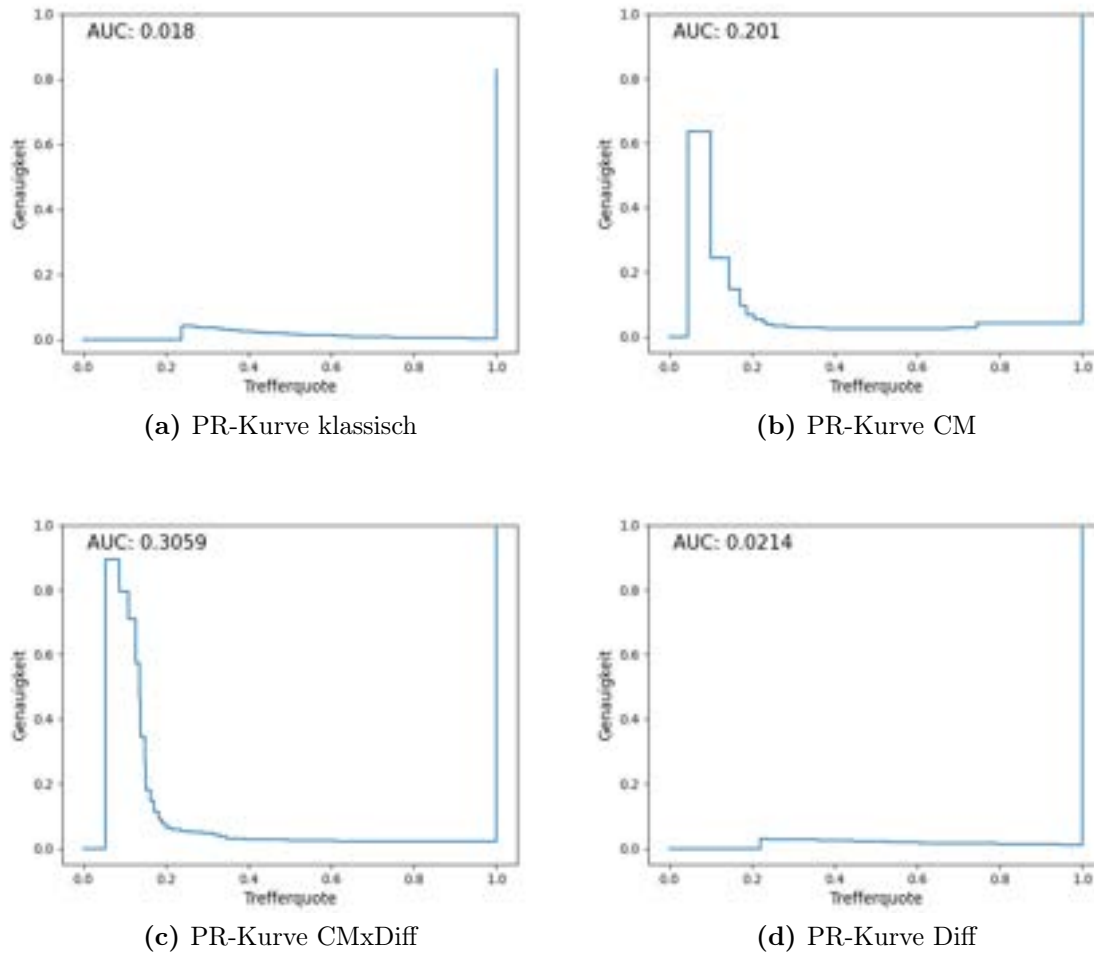
Für die verschiedenen Methoden wurden Precision-Recall-Kurven über alle Bilder erstellt, die das Verhältnis der Genauigkeiten (Precision) und Trefferquoten (Recall) der verschiedenen binären Veränderungskarten, welche mit verschiedenen Schwellwerte  $\tau$  erstellt wurden, abbilden.

In Abbildung 6.1 wurden diese für die Veränderungskarte, basierend auf dem Differenzbild der klassischen Registrierung (klassisch), und für die drei verschiedenen Veränderungskarten des Deep Learning-basierten Ansatzes (CM, CMxDiff, Diff) dargestellt. Zusätzlich sind die Werte, auf die die PR-Kurven beruhen, in Abbildung 6.2 veranschaulicht und im Anhang als Tabelle A.1 genauer aufgeführt. Bei der Veranschaulichung wurden die Nummern der Schwellwerte  $\vartheta$  auf der  $x$ -Achse und die Trefferquoten (a) bzw. die Genauigkeiten (b) auf der  $y$ -Achse aufgetragen.

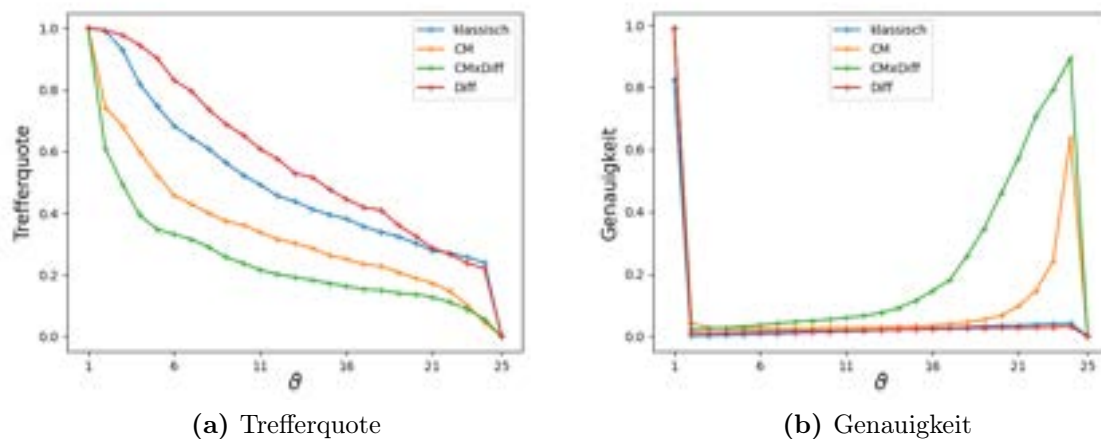
Bei dem Deep Learning-basierten Ansatz wird ein Netzwerk mit Bildpaaren trainiert, welchen als Ausgabe ein Deformationsfeld und eine Vorhersage für Veränderungen hat. Mit diesen Ausgaben werden, wie in Abschnitt 4.2.3) beschrieben, einzeln (CM, Diff) und auch kombiniert (CMxDiff) Veränderungskarten erstellt, sodass drei verschiedene Versionen einer Deep Learning-basierten Veränderungskarten miteinander verglichen werden.

Bei den beiden Methoden, die nur auf der Differenz des Referenz- und des deformierten Templatebildes beruhen (klassisch, Diff), erkennt man in Abbildung 6.1, dass die Veränderungskarten für sämtliche Schwellwerte eine niedrige Genauigkeit haben. Diese Veränderungskarten haben viele Komponenten, die nicht zu einer Läsion gehören.

Im Vergleich dazu haben die Veränderungskarten, die mit der Vorhersage der Veränderungen erstellt wurden (CM, CMxDiff), bei Schwellwerten mit niedriger Trefferquote eine höhere Genauigkeit. Das bedeutet, dass in diesen Fällen die Komponenten, die in der Veränderungskarte markiert wurden, häufiger Läsionen richtig markieren. Gleich-



**Abb. 6.1:** Die Precision-Recall-Kurve für die binären Veränderungskarten der verschiedenen Methoden über alle Bilder. Auf der  $x$ -Achse beschreibt die Trefferquoten für die verschiedenen Schwellwerte, die zum Erstellen der binären Veränderungskarte verwendet werden. Die  $y$ -Achse beschreibt die entsprechenden Genauigkeiten. Der Wert für die Fläche unter der Kurve ist links mit 'AUC:' angegeben. Bei CM und CMxDiff werden hohe Genauigkeiten bei kleinen Trefferquoten angenommen und haben damit eine größere Fläche unter der Kurve. Diese Trefferquoten werden bei den anderen beiden Methoden jedoch für keinen Schwellwert angenommen.



**Abb. 6.2:** Die Trefferquoten und Genauigkeiten der verschiedenen Methoden über alle Bilder gemeinsam in Abhängigkeit der Schwellwerte in einem Plot. Auf der  $x$ -Achse sind die Schwellwertnummern  $\vartheta$  aufgeführt. Die zugehörigen Schwellwerte  $\tau$  können der Tabelle 5.2 entnommen werden. Auf der  $y$ -Achse sind die Trefferquoten (a) bzw. Genauigkeiten (b) zu den Schwellwerten aufgeführt. Die genauen Werte sind im Anhang A.1 aufgelistet. Bei der Trefferquote ist ein ähnlicher Verlauf der Kurven zu beobachten, während bei den Genauigkeiten CM und CMxDiff höhere Werte annehmen. Keine der Methoden ist in beiden Fällen den anderen Methoden überlegen.

zeitig bedeutet die geringe Trefferquote, dass von den vorhandenen Läsionen wenige durch Komponenten der Veränderungskarte markiert wurde, also wenige richtig erkannt wurden.

Es ist zu vermuten, dass es bei diesen Schwellwerten wenige Komponenten in der Veränderungskarte gibt. Abbildung 6.2 zeigt, dass die niedrigen Trefferquoten zu hohen Schwellwerten gehören.

Die Fläche unter der Precision-Recall-Kurve (AUC-PR) beschreibt, wie gut die Trefferquote und die Genauigkeit im Verhältnis zueinander sind. Beide Werte sollen möglichst hohe Werte annehmen. Der durchschnittliche AUC-PR-Wert der verschiedenen Methoden über die 50 verschiedenen Veränderungskarten ist in den jeweiligen PR-Kurven in Abbildung 6.1 mit aufgeführt. Dabei nimmt die AUC-PR größere Werte bei größeren Genauigkeiten zu kleineren Trefferquoten an, sodass in unseren Ansätzen die mit der Vorhersage der Veränderungen einen größeren AUC-PR-Wert haben (0.201 und 0.3059). Der AUC-PR-Wert der anderen beiden Methoden beträgt ungefähr ein Zehntel dieser. Es werden in allen vier Methoden kleine Werte für die Flächen unter den Kurven angenommen. Diese kleinen AUC-PR basieren auf den Trefferquoten und insbesondere auf den kleinen Genauigkeiten, die jeweils in Abbildung 6.2 für die einzelnen Schwellwerte dargestellt sind.

Die Trefferquoten sinken kontinuierlich mit steigendem Schwellwert, wohingegen die Genauigkeiten keine komplette, kontinuierliche Steigung aufweisen, da der erste Schwellwert oder die ersten Schwellwerte größer sind. Für klassisch und Diff sind fast alle Genauigkeiten kleiner als 0.05. Ausnahme ist in beiden Fällen die Genauigkeit für Schwellwert  $\tau = 0$ .

In allen vier Fällen nimmt die Genauigkeit für diesen Schwellwert den höchsten Wert an und die Trefferquote ein Wert von 1, da bei einem niedrigen Schwellwert am meisten markiert ist. Bei einem kleinen Schwellwert sind die Komponenten größer und daher zum Teil deren Anzahl kleiner. Sobald eine Komponente in einem Voxel die Grundwahrheit der Läsionen schneidet, wird diese als richtige Komponente in die Bewertung aufgenommen. Durch die geringere Anzahl kann somit eine hohe Genauigkeit erzielt werden. In der klassischen und kombinierten Methode (CMxDiff) nimmt die Genauigkeit des zweiten Schwellwert den kleinsten Wert an und steigen ab diesem kontinuierlich, bei den anderen beiden Methoden bei dem dritten Schwellwert.

Die Werte für  $\tau = 1$  nehmen komplett den Wert 0 an, da es keine Komponenten mit Werten größer 1 bei einem maximalen Wert von 1 gibt und somit keine richtig markierten Läsionen (TP), die für die Berechnung der Genauigkeit und Trefferquote verwendet werden können.

Die Methoden, die die Vorhersage einer Wahrscheinlichkeit aus dem Netz verwenden, nehmen, wie bereits in den PR-Kurven (Abb. 6.1) erkannt, deutlich höhere Genauigkeiten an. So liegen die maximalen Genauigkeiten für  $\tau \neq 0$  bei diesen Methoden bei 0.6363 und 0.8947, während sie bei den anderen Methoden Werte kleiner als 0.05 (0.042, 0.0317) annehmen (Tab. A.1.)

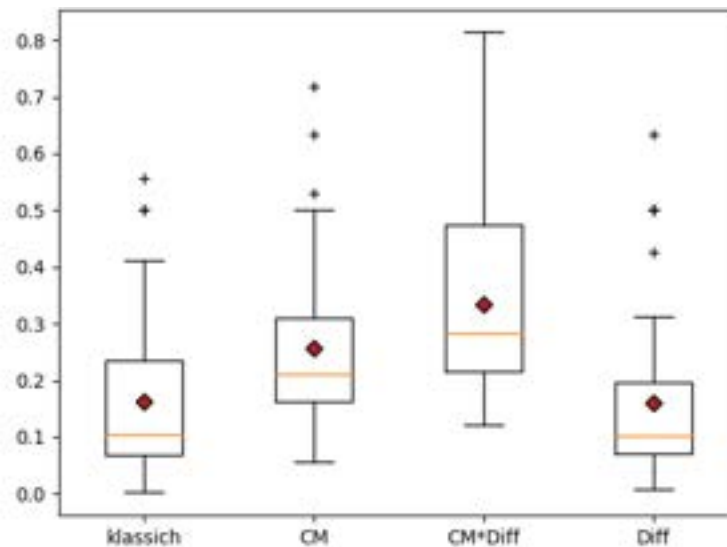
Gleichzeit nehmen die Methoden mit einer höheren Genauigkeit kleinere Trefferquoten bei diesem Schwellwert an, welcher bei allen Methoden  $\tau = 0.98$  ist. So betragen bei CM und CMxDiff für den Schwellwert  $\tau = 0.98$  die Trefferquoten 0.0455 und 0.0552, während sie bei klassisch und Diff 0.2384 und 0.2208 betragen.

Die Methoden haben im größten Schwellwert  $\tau \neq 1$  somit entweder eine höhere Trefferquote oder eine höhere Genauigkeit, jedoch hat keine Methode in beiden einen höheren Wert im Vergleich zu den anderen, sodass keine Methode den anderen in diesen Schwellwert überlegen ist.

Dies ist im Graphen von Abbildung 6.2 auch für die meisten anderen Schwellwerte zu beobachten. Die Methode, die bei der Trefferquote die höchsten Werte annimmt, nimmt bei der Genauigkeit die schlechtesten an, ebenso bei den zweit-, dritt- und viertniedrigsten Werte. Es wird im Verhältnis zu den anderen Methoden also nur eine bessere Trefferquote oder eine bessere Genauigkeit angenommen, aber nicht beides.

In Abbildung 6.3 wurde für alle 50 Fälle die AUC-PR für jedes Bild einzeln berechnet und für die verschiedenen Methoden in einem Boxplot mit dem Mittelwert aufgetragen, um die Stabilität der Methoden zu betrachten. Bei den Mittelwerten der AUC-PR über die einzelnen Fälle ist die gleiche Tendenz wie bei den AUC-PR-Werten über alle Fälle (Tab. A.1) zu erkennen, jedoch verringern sich die Unterschiede zwischen den Methoden. Die kombinierte Methode CMxDiff hat den größten Mittelwert, aber auch die größte Spannweite, sodass die Trefferquoten und insbesondere die Genauigkeiten innerhalb der einzelnen Bilder stärker variieren im Vergleich zu den anderen Methoden.

Die Verteilungen der AUC-PR-Werte für CM und Diff sind dagegen weniger gestreut. In allen Fällen ist der Mittelwert größer als der Median.



**Abb. 6.3:** Boxplot für zur Visualisierung der Verteilung der Flächen unter den Precision-Recall-Kurven der einzelnen 50 Bilder bei den vier verschiedenen Methoden. Der Median dieser Werte ist mit einer orangefarbenen Linie markiert. Die Boxen beschreiben den Bereich vom 0.25-Quantil bis zum 0.75-Quantil und Ausreißer vom Interquantilsabstand werden mit Kreuzen dargestellt. Zusätzlich wurde der Mittelwert mit einer Raute visualisiert. Die Verteilung der Werte bei CMxDiff ist deutlich größer im Vergleich zu den anderen Methoden. Das Verhältnis der Mittelwerte ist ähnlich zu den Werten für die Fläche unter der Kurve über alle Bilder, die in Abbildung 6.1 gegebene sind

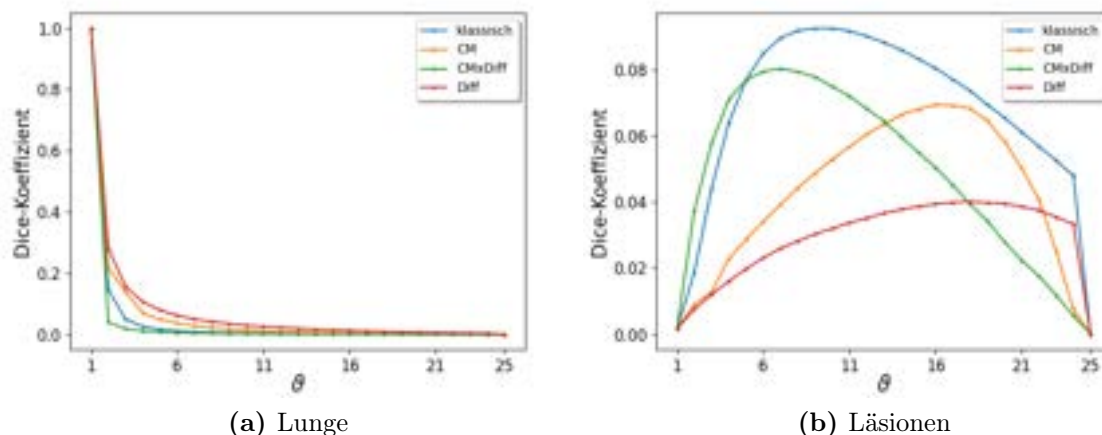
### 6.1.2 Dice-Koeffizienten

Neben der Genauigkeit und Trefferquote wird auch noch betrachtet, wie groß der Dice-Koeffizient von der Veränderungskarte und Lungensegmentierungsmaske sowie der Läsionssegmentierungsmaske ist.

Bei ersterem bedeutet ein Dice-Koeffizient von 1, dass die Veränderungskarte alle Bereiche der Lunge markiert und dementsprechend wenig Informationen über Veränderungen selbst liefert. Er sollte demnach keinen großen Wert annehmen. In unserem Fall gilt nicht, dass ein kleinerer Dice-Koeffizient immer besser ist. Je nach Größe und Vorkommen der Veränderungen ändert sich der optimale Wert. Bei dem Dice-Koeffizient von Veränderungskarte und Segmentierungsmaske der Läsionen hängt der optimale Wert von der Größe der Veränderungen im Verhältnis zu den Läsionen ab. Bei einem Wert von 1 würden beide Masken übereinstimmen. Da die Veränderungskarte jedoch nur Veränderungen und nicht die ganzen Läsionen markieren soll, wäre ein Wert von 1 bei sich verändernden Läsionen nicht optimal.

In Abbildung 6.4 wurde für die verschiedenen Methoden beide Dice-Koeffizienten für die Veränderungskarten der verschiedenen Schwellwerte aufgetragen. Für alle vier Fällen ist der Dice-Koeffizient mit der Lunge für den Schwellwert 0 nahezu 1 und nimmt mit den folgenden Schwellwerten stark ab. Dabei nimmt der Koeffizient bei den Methoden unterschiedlich schnell mit wachsendem Schwellwert ab, am stärksten bei der kombinierten Veränderungskarte (CMxDiff) und am schwächsten bei Diff.

Bei dem Dice-Koeffizient mit den Läsionen erreichen die Methoden für unterschiedliche Schwellwerte den maximalen Wert. Bei klassisch und bei CMxDiff wird dieser bei



**Abb. 6.4:** Durchschnittliche Werte des Dice-Koeffizienten über die 50 verschiedenen Bilder für die vorhergesagte Maske der Veränderungskarte und die Maske der Lunge (a), sowie die Maske der Grundwahrheit der Läsionen (b) in Bezug auf die 25 verschiedenen Schwellwerte 5.2.

kleineren Schwellwerten angenommen. Bei klassisch ist es der zehnte Schwellwert (0.641) mit Wert 0.0925, bei CMxDiff der siebte Schwellwert (0.5119) mit einem Wert von 0.0803.

Die anderen beiden Methoden nehmen den maximalen Dice-Koeffizienten bei einem Schwellwert um die 0.85 an. Bei CM ist es der 16. Schwellwert (0.8212) mit einem Wert von 0.0694 und Diff der 18. Schwellwert (0.8678) mit einem Wert von 0.0401.

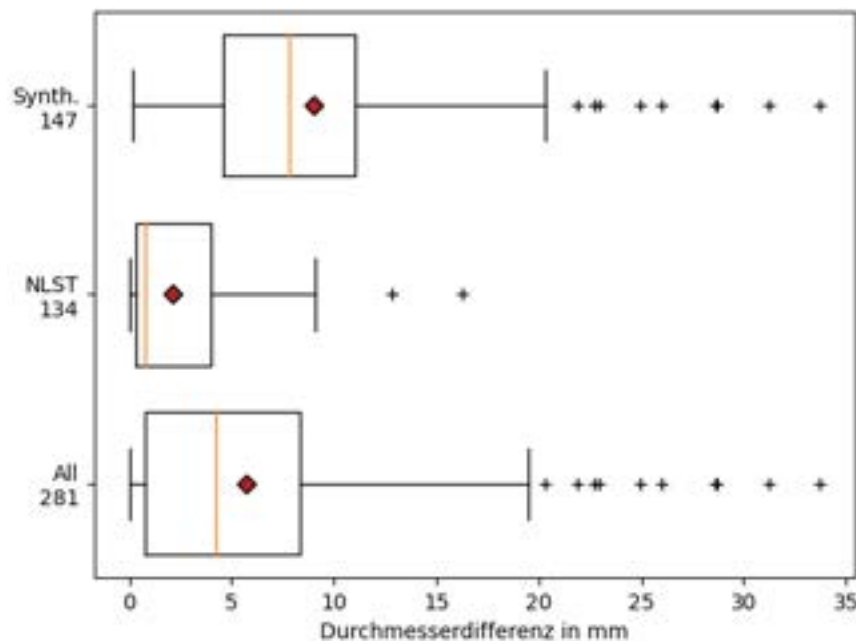
### 6.1.3 Durchmesserdifferenzen

Zur Übersicht über die axialen Differenzen der Durchmesser der verschiedenen Läsionen wurden diese in Abbildung 6.5 für alle 281 Läsionen, und aufgeteilt in synthetische und originalen NLST-Läsionen als Boxplot dargestellt.

Zudem wurde die Differenz der Durchmesser mit dem maximal angenommen Durchmesser jeder Läsion in Abbildung 6.6 visualisiert. Dabei wurde farblich zwischen den beiden Läsionstypen unterschieden. Die Anzahl von NLST-Läsionen und künstlich eingefügten Läsionen sind annähernd gleich (134 und 147). Die Durchmesserdifferenzen der NLST-Läsionen sind jedoch deutlich kleiner im Vergleich zu den Differenzen der synthetisch eingefügten Läsionen. Sie haben zudem weniger Ausreißer in Richtung der größeren Differenzen.

Der im Vergleich zum Mittelwert (2.11 mm) kleinere Medianwert (0.75 mm) deutet auf eine Häufung kleinerer Werte hin. Dies ist auch in Abbildung 6.6 (rote Punkte) zu erkennen. Die Durchmesserdifferenzen der NLST-Läsionen nehmen für verschiedenen Durchmesser kleinere Werte an oder den kompletten Wert des Durchmessers, wenn eine Läsion vollständig verschwunden oder neu entstanden ist. Die zweitgenannten Punkte





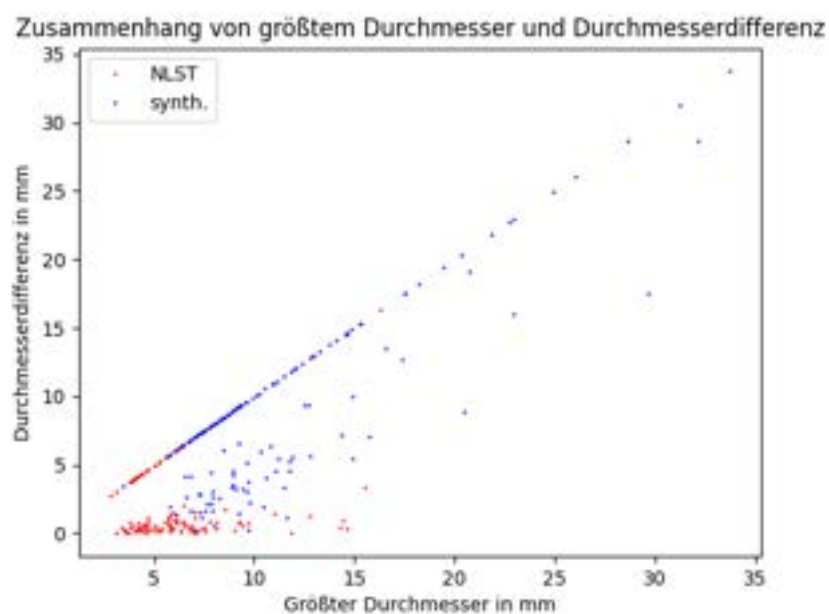
**Abb. 6.5:** Boxplot für zur Visualisierung der Verteilung der Durchmesserdifferenzen insgesamt und bei den unterschiedlichen Läsionstypen. Der Median dieser Differenzen ist mit einer orangefarbenen Linie markiert. Die Boxen beschreiben den Bereich vom 0.25-Quantil bis zum 0.75-Quantil und Ausreißer vom Interquantilsabstand werden mit Kreuzen dargestellt. Zusätzlich ist noch die Anzahl der jeweiligen Läsionen unter deren Bezeichnung an der  $y$ -Achse angegeben und der Mittelwert mit einer roten Raute markiert.

sind auf der Diagonalen des Diagramms abgebildet und haben eine geringere Anzahl (42) als erstgenannte (92).

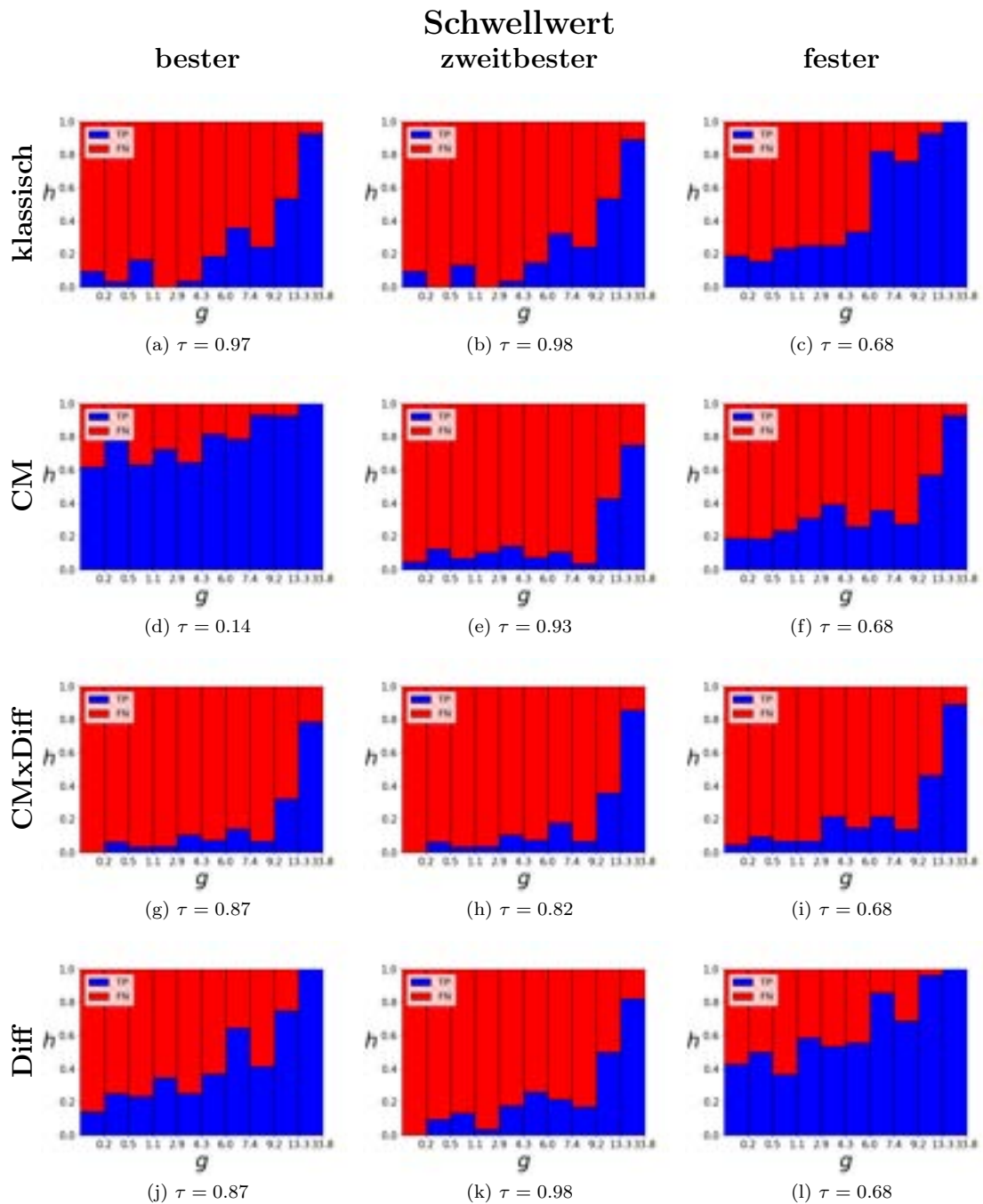
Die Durchmesserdifferenzen der synthetischen Läsionen nehmen größere Werte an (Mittelwert: 8.99 mm, Median: 7.78 mm), sind mehr gestreut und haben mehr Ausreißer in Richtung höherer Werte im Vergleich zu den NLST-Läsionen. Zudem ist die Anzahl der Läsionen, die komplett verschwinden oder sich neu entwickeln, höher (88) als die der Läsionen, die sich nur im Durchmesser verändern (59). Daher sind in Abbildung 6.6 vermehrt die blauen Punkte, die zu den synthetisch eingefügten Läsionen gehören, auf der Diagonalen zu finden. Die anderen blauen Punkten sind tendenziell auf einer Nebendiagonalen zu finden. Damit steigt bei den synthetisch eingefügten Läsionen mit der maximalen Größe, die die Läsion in einem der beiden Bilder annimmt, die Durchmesserdiffferenz, während bei den NLST-Läsionen die Durchmesserdifferezenzen klein bleiben. Insgesamt nehmen die Läsionen einen Mittelwert von 5.71 mm und ein Median von 4.24 mm an.

In Abbildung 6.7 wurden für alle vier Methoden jeweils für die verschiedenen Schwellwerte  $\tau$  die F2-Scores berechnet und der beste und zweitbeste Schwellwert  $\tau \neq 0$  bezüglich diesem zur Auswertung der Erkennung der verschiedenen Durchmesserdifferezenzen gewählt. Zusätzlich wurde ein fester Schwellwert für alle vier Methoden gewählt.

Es wurden die relativen Häufigkeiten  $h$  der richtig erkannten (TP) und nicht erkannten (FN) Läsionen pro Gruppe der Durchmesserdifferezenzen  $g$  betrachtet. Die meisten durch den F2-Score gewählten Schwellwerte gehören zur oberen Hälfte dieser und nehmen Werte über 0.8 an.



**Abb. 6.6:** Verhältnis größter axialer Durchmesser pro Läsionen und Durchmesserdiffferenz. Die  $x$ -Achse beschreibt den größten axialen Durchmesser in mm, welcher von der Läsion entweder im Referenz- oder Templatebild angenommen wird. Die  $y$ -Achse beschreibt die Differenz der axialen Durchmesser vom Referenz- und Templatebild in mm. Wenn eine Läsion nur in einem der beiden Bilder vorkommt, ist diese im Diagramm auf der Diagonalen markiert. Dabei stehen rote Punkte für Läsionen aus dem NLST-Datensatz und blaue Punkte für synthetisch eingefügte Läsionen. Die NLST-Läsionen haben im Vergleich zu den synthetisch eingefügten Läsionen einen geringeren axialen Durchmesser und eine verhältnismäßig kleiner Durchmesserdiffferenz.



**Abb. 6.7:** Die Diagramme stellen die relative Häufigkeit  $h$  der richtig erkannten (TP) und nicht erkannten (FN) Läsionen für verschiedene Durchmesserdifferenzen, die in Gruppen  $g$  unterteilt sind, für die vier verschiedenen Methoden jeweils für den erst- und zweitbesten Schwellwert  $\tau \neq 0$  bezüglich des F2-Scores, sowie ein fest gewählter Schwellwert dar. Die relative Häufigkeit  $h$  der richtig erkannten (TP) ist in Blau dargestellt, die relative Häufigkeit von den nicht erkannten Läsionen (FN) in Rot. Die relativen Häufigkeiten ergeben addiert 1 und wurden in der Darstellung übereinandergesetzt. Auf der  $x$ -Achse sind die Grenzen der Gruppen für Durchmesserdifferenzen angegeben. Dabei ist der Wert links neben dem Balken die untere Schranke, die auch mit angenommen werden darf, und rechts die obere Schranke der Gruppe, die nicht mit angenommen wird. Die Grenzen der Gruppen sind mit Quantilen bestimmt worden, sodass die Gruppen unterschiedliche Breiten haben. Sie sind jedoch im Plot zur besseren Lesbarkeit gleich breit dargestellt. In Tabelle 5.3 sind die genauen Werte und die Anzahl der Läsionen in den entsprechenden Gruppen, die zur Berechnung der relativen Häufigkeiten verwendet wurden, angegeben

**Tab. 6.1:** Mittelwerte für richtig erkannte (TP) und nicht erkannte (FN) Veränderungen von Läsionen über die verschiedenen Gruppen der Durchmesserdiffenzen. Diese werden für die gewählten Schwellwerte der verschiedenen Methoden aus in Abbildung 6.7 aufgeführt. Der Schwellwert steht in der Zeile 'Schwellwert'. Der Mittelwert von TP und der Mittelwert von FN addieren sich immer zu 1. Für alle Methoden sinkt der Mittelwert TP bei größeren Schwellwerten und der Mittelwert FN steigt entsprechend.

	klassisch			CM			CMxDiff			Diff		
	1	2	fest	1	2	fest	1	2	fest	1	2	fest
Schwellwert	0.97	0.98	0.68	0.14	0.93	0.68	0.87	0.82	0.68	0.87	0.98	0.68
Mittelwert TP	0.26	0.24	0.49	0.79	0.19	0.37	0.16	0.18	0.23	0.44	0.24	0.65
Mittelwert FN	0.74	0.76	0.51	0.21	0.81	0.63	0.84	0.82	0.77	0.56	0.76	0.35

Eine Ausnahme ist der beste Schwellwert  $\tau$  bezüglich des F2-Scores für CM. Dieser beträgt 0.14 und ist damit der kleinste Schwellwert  $\tau \neq 0$ , der angenommen werden kann. Durch die deutlich erhöhte Genauigkeit bei  $\tau = 0.14$  im Vergleich zu den anderen Methoden, hat CM den hohen F2-Score (Tab A.1). Bei dem niedrigen Schwellwert sind die relative Häufigkeit für die verschiedenen Gruppen insgesamt höher im Vergleich zu den anderen Schwellwerten und Methoden.

Insgesamt ist bei allen Abbildungen zu erkennen, dass größere Durchmesserdiffenzen besser erkannt werden. Verändert sich der Schwellwert geringfügig, so verändern sich die relativen Häufigkeiten geringfügig. Bei CM verändern sich dagegen die Schwellwerte und die relativen Häufigkeiten stark.

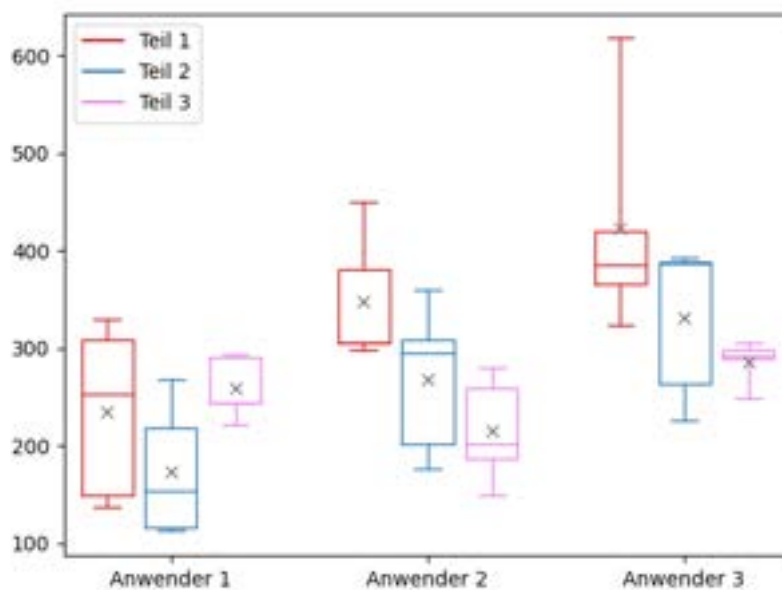
In Tabelle 6.1 sind für die vier Methoden die Mittelwerte der relativen Häufigkeiten über den Gruppen für richtig erkannte (TP) und nicht erkannte (FN) zu den gewählten Schwellwerten aufgeführt. Tendenziell gilt: Je kleiner der Schwellwert, desto größer die relative Häufigkeit für richtig erkannte. Es gibt jedoch Unterschiede zwischen den verschiedenen Methoden. So haben CMxDiff und Diff den gleichen besten Schwellwert (0.87), jedoch hat Diff höhere relative Häufigkeiten für richtig erkannt, durchschnittlich 0.44, als CMxDiff (durchschnittlich 0.16).

Dies erkennt man insbesondere auch beim fest gewählten Schwellwert für alle Methoden. So haben die Methoden, die nur auf der Differenz der Bilder beruhen (klassisch, Diff), eine durchschnittliche relative Häufigkeit über 0.48 für richtig erkannte (TP). Die beiden Methoden, die die Vorhersage der Veränderung des Netzes verwenden (CM, CMxDiff), haben eine relative Häufigkeit unter 0.38.

In fast allen Fällen sinkt die relative Häufigkeit für richtig erkannte vom besten zum zweitbesten Schwellwert. Die Ausnahme ist CMxDiff, bei dem die durchschnittliche relative Häufigkeit (TP) leicht steigt für den zweitbesten Schwellwert. Die durchschnittlichen relativen Häufigkeiten sind bis auf zwei Werte kleiner als 0.5.

Während bei allen vier Methoden die höchste Durchmesserdiffenz eine relative Häufigkeit für richtig erkannte von über 0.5 hat, so sind in der zweithöchsten Durchmesserdiffenz es nur noch zwei Drittel und in der drittgrößten Durchmesserdiffenz nur noch die Hälfte.

In den meisten Gruppen werden somit nur weniger als die Hälfte der Läsionen bei den gewählten Schwellwerten markiert.



**Abb. 6.8:** Die 5 Bearbeitungszeiten in s der verschiedenen Anwender werden farblich unterschiedlich für die Studienteile als Boxplot dargestellt. Der Boxplot ist über den Median und 0.25- und 0.75-Quantil bestimmt. Die Antennen (Whisker) wurden so definiert, dass sie das Minimum und Maximum umfassen, damit alle 5 Werte abgelesen werden können. Zusätzlich sind die Mittelwerte der Bearbeitungszeit für jeden Studienteil als graues x visualisiert. Während die Bearbeitungszeit von Anwender 2 und Anwender 3 mit jedem Studienteil sinkt, variiert diese für Anwender 1.

## 6.2 Anwenderstudie

An der Anwenderstudie haben insgesamt 3 Personen teilgenommen und es wurden pro Studienteil 5 Fälle in die Bewertung einbezogen.

Der erste Anwender (Anwender 1) hat 2,5 Jahre Vollzeit als Radiologe gearbeitet und ist seit 6 Jahren nebenberuflich in dem Bereich tätig. Anwender 2 und 3 sind in der Forschung und Entwicklung für die medizinische Bildverarbeitung tätig. Sie beschäftigen sich in aktuellen Projekten ebenfalls mit Erkrankungen und Läsionen in der Lunge, sodass diese Erfahrungen mit solchen Bildern haben, aber keine ausgebildeten Radiologen sind.

Im ersten Teil der Studie (Teil 1) hatten die Teilnehmer keine Hilfestellung, in Teil 2 hatten sie die Veränderungskarte, basierend auf dem Differenzbild der klassischen Registrierung (klassisch), und in Teil 3 die Veränderungskarte, kombiniert aus den Veränderungskarten der Deep-Learning basierten Registrierung und der Vorhersage der Veränderungen (CMxDiff), als visuelle Überlagerung zur Unterstützung.

Für die verschiedenen Studienteile wurde die Bearbeitungszeit sowie die Trefferquote und Genauigkeit der einzelnen Anwender ausgewertet.

**Tab. 6.2:** Bearbeitungszeiten in Sekunden  $s$  der Anwender pro Studienteil. Alle Studienteile sind als Mittelwert  $\pm$  Standardabweichung der Bearbeitungszeit und das zweiseitige 95%-Konfidenzintervalle angegeben, vorausgesetzt die Annahme einer Normalverteilung wurde mit dem Shapiro-Wilk-Test [SW65] nicht abgelehnt.

	Anwender 1		Anwender 2		Anwender 3	
Teil 1	235.6 $\pm$ 79.5 s	[157.7; 313.5]	348.0 $\pm$ 59.4 s	[289.8; 406.2]	422.5 $\pm$ 102.4 s	[322.2; 522.8]
Teil 2	173.6 $\pm$ 60.5 s	[114.3; 233.0]	268.3 $\pm$ 69.1 s	[200.6; 336.0]	331.8 $\pm$ 71.9 s	
Teil 3	258.6 $\pm$ 28.7 s	[230.4; 286.7]	215.3 $\pm$ 47.6 s	[168.6; 262.0]	286.7 $\pm$ 20.1 s	[267.0; 306.4]

**Tab. 6.3:** Auflistung der  $p$ -Werte zwischen den jeweiligen Studienteilen für die verschiedenen Anwender.  $p_{i,j}$  beschreibt den  $p$ -Wert des zweiseitigen Wilcoxon-Rangsummen-Tests zwischen den Studienteilen  $i$  und  $j$  bezüglich der Bearbeitungszeit. Es werden für  $p < 0.05$  signifikante Veränderungen angenommen.

	Anwender 1	Anwender 2	Anwender 3
$p_{1,2}$	0.3095	0.2222	0.5476
$p_{1,3}$	1.0	0.0079	0.0079
$p_{2,3}$	0.0556	0.3095	0.5476

### 6.2.1 Zeit

In Abbildung 6.8 sind die Zeiten für die 5 verschiedenen Bildpaare als Boxplot, sowie der Mittelwert als graues  $\times$  für die Studien dargestellt. Die Antennen (Whisker) sind so definiert, dass sie bis zum Minimum und Maximum reichen. Daher können alle 5 Werte aus dem Boxplot abgelesen werden.

Die Mittelwerte sowie die Standardabweichung als auch das 95%-Konfidenzintervall für die Zeiten sind für die verschiedenen Studien in Tabelle 6.2 aufgelistet. Die Zeiten jeder Studie sind mittels des Shapiro-Wilk-Tests [SW65] und einem Signifikanzniveau von  $\alpha = 0.05$  auf eine Normalverteilung getestet. Wurde die Hypothese der Normalverteilung nicht abgelehnt, wurde das Konfidenzintervall bestimmt.

Bei Teil 2 von Anwender 3 wird die Hypothese der Normalverteilung der Zeiten mit einem  $p$ -Wert von 0.0424 abgelehnt und dementsprechend kein Konfidenzintervall bestimmt. Die restlichen  $p$ -Werte betragen 0.0974 oder höher. Bei den Anwendern, die keine Radiologen sind (2, 3), ist zu erkennen, dass die Zeiten und damit auch die durchschnittliche Zeit pro Bild mit jedem Teil sinkt. Der Unterschied zwischen den beiden Teilen mit Unterstützung (Teil 2 und 3) ist ähnlich groß, wie der Unterschied von Teil 1 ohne Unterstützung zum ersten Teil mit Unterstützung (Teil 2). Somit haben Anwender 2 und 3 die beste mittlere Zeit in Teil 3 der Studie, welcher die Veränderungskarte, die auf der Kombination der Veränderungskarte der Deep Learning-basierten Registrierung und der Vorhersage der Veränderungen beruht, als Unterstützung hat.

Beim Radiologen verringern sich ebenfalls die Zeiten vom Teil ohne Unterstützung (Teil 1) zum Teil mit der Veränderungskarte basierend auf dem Differenzbild der klassischen Registrierung als Unterstützung. Bei Teil 3 hat Anwender 1 den größten Mittelwert (258.6) im Vergleich zu den anderen Studienteilen. Bei allen drei Anwendern hat Teil 3 die geringste Spannweite und Standardabweichung. Dadurch ist bei Anwender 1 trotz des hohen Mittelwerts die maximale Zeit von Teil 3 geringer als die zwei größten Zeiten von Teil 1. Für Anwender 1 und 3 sinken die Standardabweichungen mit den

**Tab. 6.4:** Auswertung der Trefferquote und Genauigkeit der Anwender in den verschiedenen Studienteilen. Dabei wird zwischen den Markierungen für Veränderungen von Läsionen (Läsion) und den Markierungen für alle Auffälligkeiten unterschieden. Es werden die Mittelwerte  $\pm$  Standardabweichungen der Trefferquoten und Genauigkeiten über die fünf Bilder im jeweiligen Studienteil angegeben.

		Anwender 1		Anwender 2		Anwender 3	
		Trefferquote	Genauigkeit	Trefferquote	Genauigkeit	Trefferquote	Genauigkeit
Teil 1	Läsion	$0.37 \pm 0.23$	$0.65 \pm 0.37$	$0.32 \pm 0.15$	$0.88 \pm 0.24$	$0.08 \pm 0.1$	$0.28 \pm 0.21$
	alle	$0.57 \pm 0.29$	$0.67 \pm 0.23$	$0.66 \pm 0.19$	$0.5 \pm 0.23$	$0.12 \pm 0.1$	$0.16 \pm 0.14$
Teil 2	Läsion	$0.22 \pm 0.2$	$0.67 \pm 0.41$	$0.34 \pm 0.17$	$0.67 \pm 0.35$	$0.18 \pm 0.19$	$0.62 \pm 0.41$
	alle	$0.49 \pm 0.17$	$0.62 \pm 0.18$	$0.66 \pm 0.15$	$0.61 \pm 0.28$	$0.32 \pm 0.22$	$0.26 \pm 0.13$
Teil 3	Läsion	$0.29 \pm 0.27$	$0.57 \pm 0.44$	$0.26 \pm 0.14$	$0.83 \pm 0.29$	$0.05 \pm 0.06$	$0.27 \pm 0.39$
	alle	$0.56 \pm 0.27$	$0.69 \pm 0.28$	$0.61 \pm 0.08$	$0.67 \pm 0.21$	$0.16 \pm 0.12$	$0.13 \pm 0.1$

Teilen kontinuierlich. Anwender 2 hat im Gegensatz dazu die größte Standardabweichung bei Teil 2. Es ist zu beobachten, dass Anwender 1, der Radiologe, insgesamt am wenigsten Zeit pro Bild gebraucht und Anwender 3 sich am meisten Zeit pro Bild genommen hat.

Im Weiteren wurde die Zeiten der Studien auf signifikante Veränderungen überprüft. In der Tabelle 6.3 sind die  $p$ -Werte des zweiseitigen Wilcoxon-Rangsummen-Tests aufgeführt. Signifikante Änderungen werden für  $p < 0.05$  angenommen. Dabei beschreibt  $p_{i,j}$  mit  $i, j \in \{1, 2, 3\}$  den  $p$ -Wert für den Wilcoxon-Rangsummen-Test für Teil  $i$  und  $j$ .

Nur zwischen Teil 1 und Teil 3 bei Anwender 2 und 3 werden signifikante Veränderungen der Zeiten angenommen. Bei beiden sind alle Zeiten aus Teil 3 kleiner als die Zeiten aus Teil 1, sodass es sich um eine signifikante Verbesserung in Teil 3 handelt. Auffällig ist zudem der  $p_{2,3}$ -Wert von Anwender 1. Dieser ist nur knapp über dem Signifikanzniveau, sodass die Zeiten von Teil 3 nahezu eine signifikante Verschlechterung zu Teil 2 für Anwender 1 aufweisen.

Ansonsten sind zwischen den  $p_{1,2}$ - und  $p_{2,3}$ -Werten keine größeren Unterschiede zu erkennen, sodass die Unterschiede der Zeiten von Teil 2 zu Teil 1 und Teil 3 ähnlich sind und es somit keine eindeutige Abgrenzung von den Zeiten ohne Unterstützung durch eine visuelle Veränderungskarte zu denen mit Unterstützung gibt.

### 6.2.2 Genauigkeit und Trefferquote

Für alle Anwender werden die verschiedenen Markierungen bezüglich Genauigkeit und Trefferquote in dem jeweiligen Studienteil untersucht. Die Mittelwerte und Standardabweichungen über die 5 Bilder sind in Tabelle 6.4 aufgeführt, wobei zu einem nur die Markierungen für Läsionen mit Veränderungen und zum anderen alle Markierungen betrachtet werden. Zusätzlich wurde die Genauigkeiten und Trefferquoten über alle Bilder pro Studienteil betrachtet, welche im Anhang A.2 zu finden sind.

In allen Fällen steigen die Mittelwerte der Trefferquoten von nur Markierungen von Läsionen zu allen Markierungen, d.h. es wurde häufig eine Läsion mit rosa markiert statt grün.

Die Trefferquoten der verschiedenen Anwender unterschieden sich stark. So liegen die Trefferquoten von Anwender 2 für die Läsionen um 0.3 und für alle Markierungen um 0.65. Bei Anwender 1 liegen sie zwischen 0.22 und 0.37 sowie 0.49 und 0.57 und haben damit mit eine größere Spannweite. Dabei erzielt der Radiologe die höchste Trefferquote

**Tab. 6.5:**  $p_{i,j}$  beschreibt den  $p$ -Wert des zweiseitigen Wilcoxon-Rangsummen-Tests zwischen den Studienteilen  $i$  und  $j$  und wurde für die Trefferquoten und die Genauigkeit bestimmt. Es werden für  $p < 0.05$  signifikante Veränderungen dieser Werte zwischen den Studienteilen angenommen.

		Anwender 1		Anwender 2		Anwender 3	
		Trefferquote	Genauigkeit	Trefferquote	Genauigkeit	Trefferquote	Genauigkeit
P <sub>1,2</sub>	Läsion	0.3917	0.8973	0.8315	0.2877	0.5008	0.3636
	alle	0.6742	0.9166	1.0	0.6905	0.1376	0.2948
P <sub>1,3</sub>	Läsion	0.832	1.0	0.9155	0.8668	0.9049	0.7491
	alle	1.0	1.0	1.0	0.4633	0.8304	0.8325
P <sub>2,3</sub>	Läsion	0.7457	1.0	0.7518	0.5023	0.3661	0.2972
	alle	0.9166	0.833	0.6723	0.6004	0.1706	0.2222

für den Studienteil ohne Unterstützung. Anwender 3 hat die kleinsten durchschnittlichen Trefferquoten, die zwischen 0.05 und 0.18 für die Läsionen und zwischen 0.12 und 0.32 für alle Markierungen liegen. Insgesamt wurden nie mehr als zwei Drittel der Läsionen gefunden.

Anders sieht es bei den Genauigkeiten aus. Diese sinken in allen Fällen für Anwender 2 und 3 von den Markierungen von nur veränderten Läsionen zu allen Läsionen, d. h. dass mehr Markierungen in rosa gesetzt wurden, die keine Läsionen sind, anstatt Markierungen, die Läsionen sind. Bei Anwender 1, dem Radiologen, sinkt die Genauigkeit nur in einem Fall und steigt in zwei Fällen. Dabei erreichen Anwender 1 und 2 in allen Fällen eine Genauigkeit über 0.5, in den meisten Fällen über 0.65. Anwender 3 erreicht dagegen nur in einem Fall eine Genauigkeit über 0.5 und in den anderen unter 0.3.

Es ist für alle Anwender in keinem Studienteil eine Überlegenheit bezüglich deren anderen Studienteile zu erkennen. Dies wird auch in Tabelle 6.5 deutlich, welche die  $p$ -Werte für die Wilcoxon-Rangsummen-Tests zwischen den Studienteilen aufführt. Für  $p_{i,j} < 0.05$  werden signifikante Unterschiede für die Verteilungen der Werte innerhalb eines Studienteils angenommen. Dabei beschreibt  $p_{i,j}$  mit  $i, j \in \{1,2,3\}$  den  $p$ -Wert für den Wilcoxon-Rangsummen-Test für Teil  $i$  und  $j$ . Alle  $p$ -Werte sind größer als 0.1, sodass in keinem Fall signifikante Unterschiede angenommen werden. Dabei werden insgesamt häufig hohe Werte über 0.65 angenommen, was gegen Veränderungen der Verteilungen spricht.

### 6.2.3 Feedback der Teilnehmer

Nach der Studie wurden den Teilnehmern noch drei Fragen zur Studie gestellt. Nach Teil 1 gab es bereits den Veränderungsvorschlag eine Anzeige in die Oberfläche der Anwenderstudie einzubauen, damit die Anwender sehen können, ob der *start*- bzw. *stop*-Button gedrückt wurde. Dies wurde bereits für Teil 2 und 3 eingebaut, da dies die eigentlichen Ergebnisse der Anwender nicht beeinflusst, sondern nur die richtige Anwendung erleichtert.

#### Was könnte deine Leistung in der Studie verzerrt haben?

Alle drei Anwender nennen das Problem, dass beim Zoomen das Baseline und das Follow-Up Bild unterschiedlich verschieben. Wird gezoomt, so verschiebt sich das Follow-



Up Bild etwas in die obere linke Ecke, während das Baseline Bild zentriert bleibt. So sind die Bilder nicht mehr auf eine Höhe und ein direkter Vergleich wird erschwert.

Von den Anwendern, die in der Forschung und Entwicklung tätig sind (Anwender 2 und 3), sind noch weitere Punkte angemerkt worden. Zum einen wird angemerkt, dass der Wechsel zwischen 'mark growth lesion' und 'mark other' nur geht, indem man 'mark other' klickt. Daher kam es dazu, dass eine Läsion in grün markiert wurde und anschließend zu rosa korrigiert werden musste. Dies hat zeitliche Auswirkungen oder führt zu falsch klassifizierten Strukturen, wenn dies nicht bemerkt wird.

Zum anderen wird angemerkt, dass die Schichten oft eher suboptimal gelinkt sind, sodass teilweise etwas hoch- und runtergescrollt werden musste, um die passenden Schichten im Follow-Up Bild zu sehen. Es sieht dann erst so aus, als wäre im Follow-Up Bild keine Läsion (korrespondierende Struktur) und man findet diese dann doch paar Schichten weiter. Dies kann auch zu falschen Entscheidungen führen, wenn es übersehen wird.

Zuletzt wird von einem Anwender noch angemerkt, dass er allein durch die Übung schneller geworden sein könnte, da er zuvor noch keine Lungen Läsionen annotiert hat.

### **Was kann in einer zukünftigen Anwenderstudie noch verbessert werden, was sollte zukünftig eingebaut oder beachtet werden?**

Anwender 1 würde fest hinterlegte Fenster-Einstellungen, wie es sie in der Praxis gibt, einbauen, z.B. eine Lungen-Fenster-Einstellung auf der F1-Taste. Zudem wäre ein Reset-Button für die Fensterung hilfreich. Ein weiterer Anwender würde eine automatische Fensterung auf die Lunge, während die Bilder laden, hilfreich finden.

Die Curser-Synchronisation könnte laut einem Anwender verbessert werden, wenn sie durch Registrierung anstatt Voxelkoordinaten verbunden wäre, sodass das Finden und Vergleichen von auffälligen Strukturen vereinfacht und deutlich beschleunigt werden kann.

Anwender 2 und 3 würde weitere oder andere Darstellungen der Veränderungskarte hilfreich finden. Dafür gibt es viele verschiedene Vorschläge: Zum einen, dass die Veränderungskarte auf einem beliebigen Bild angezeigt werden kann und nicht nur auf dem Baseline Bild, zum anderen, dass das Baseline Bild mit Veränderungskarte und ohne Veränderungskarte parallel angezeigt werden kann und nicht innerhalb eines Bildes gewechselt werden muss. Wenn es nur in einem Bild angezeigt wird, musste dieses nach dem Empfinden eines Anwenders zweimal durchgegangen werden, einmal mit und einmal ohne Veränderungskarte, um alles zu sehen. Dies führt zu einer verlängerten Bearbeitungszeit.

Auch eine veränderte Position der Veränderungskarte (ohne CT-Bild) Richtung Baseline Bild könnte den Vergleich erleichtern und weniger Wechseln erfordern.

### **Weiteres Feedback**

Des Weiteren hätte ein Anwender ohne radiologische Ausbildung eine Zeitanzeige, um seine Zeit beobachten zu können, hilfreich gefunden, um nicht zu viel Zeit pro Fall zu verlieren.

## 6.3 Qualitative Auswertung

Neben der quantitativen Auswertung wurden die Visualisierungen der Veränderungskarte der verschiedenen Läsionen betrachtet.

Zur Einschätzung der Markierungen in den binären Veränderungskarten wurde, für die in Abbildung 6.7 gewählten Schwellwerte, die binäre Veränderungskarte der verschiedenen Methoden in einer Schicht eines Bildes in Abbildung 6.9 visualisiert. Zur Orientierung sind Ränder der Lungensegmentierungsmaske in gelb, sowie die Segmentierungsmaske der Läsionen des Referenzbildes in pink und die des Templatebildes in blau, gegeben. In dieser Schicht sind mehrere Läsionen zu erkennen, wobei es sich bei den Läsionen im rechten Teil des Bildes um vollständig rezessive und bei der Läsion im linken Teil des Bildes um eine wachsende Läsion handelt.

Es ist zu erkennen, dass für hohe Schwellwerte kaum Markierungen vorhanden sind. Bei dem kleinsten Schwellwert  $\tau = 0.14$  bei CM sind am meisten Markierungen vorhanden. Alle Veränderungen von Läsionen werden markiert, aber es werden auch Veränderungen am Lungenrand und in Bereichen des Bronchialbaums markiert. Zudem ist der Bereich der wachsende Läsion vollständig markiert und nicht nur die Veränderung selbst.

Auch bei dem fest gesetzten Schwellwert  $\tau = 0.68$  sind nur wenige Veränderungen zu erkennen. Dabei sind für diesen Schwellwert neben Veränderungen von Läsionen bei klassisch und Diff Veränderungen am linken Lungenrand und bei CM Veränderungen im Bereich des Bronchialbaums markiert. Bei CMxDiff sind keine Markierungen zu erkennen.

Da die binären Karten nur zur Auswertung dienen, sind in Abbildung 6.10 die farblichen kontinuierlichen Veränderungskarten der verschiedenen Methoden für die gleiche Schicht sowie die farbliche Legende abgebildet. Dabei sind wieder die Ränder der Lungensegmentierungsmaske in weiß, sowie die Segmentierungsmaske der Läsionen des Referenzbildes in pink und die des Templatebildes in blau zur Orientierung gegeben.

Ähnlich zu den binären Veränderungskarten sind bei CM neben den Läsionen, Veränderung im Bereich des Bronchialbaums und den Lungenrändern markiert. Ebenso sind bei klassisch und Diff weitere Veränderungen markiert und bei CMxDiff ist kaum etwas markiert. Die Veränderungskarten, die auf einer Differenz beruhen (klassisch, Diff), weisen deutlich mehr Rauschen in den Bildern auf. CM hat dagegen glattere farbliche Übergänge.

In Abbildung 6.11 sind Veränderungskarten für eine andere Schicht eines anderen Bildes als farbliche Überlagerung auf dem Referenzbild gegeben. Zusätzlich werden das Referenz- und Templatebild abgebildet. Auch hier ist die Glattheit von CM im Vergleich zu den anderen Methoden zu erkennen. In Diff und CM, die auf dem gleichen Deformationsfeld beruhen, sind viele Veränderungen im Bereich des Bronchialbaums zu erkennen. Auf dieser Schicht ist keine Läsion abgebildet, jedoch ist eine Veränderung des Lungengewebes im rechten Lungenflügel (links im Bild). Dieser wird von allen Methoden erkannt, jedoch von klassisch und Diff deutlich stärker hervorgehoben, während sie von CM nur etwas grünlich dargestellt wird.

In den Abbildungen 6.12 bis 6.16 werden Ausschnitte der Veränderungskarten in Bereichen von Läsionen für die verschiedenen Methoden abgebildet. Zudem sind jeweils das Referenz- und Templatebild gegeben sowie die Läsionssegmentierungsmasken der

verschiedenen Methoden visualisiert. Die vollständigen sowie weitere Bilder sind im Anhang (ab A.1) zu finden.

In der oberen Reihe sind jeweils das Template- und Referenzbild und das Referenzbild mit einer Visualisierung der Segmentierungsmasken der Läsionen abgebildet. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in Lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in Blau dargestellt. In allen Fällen wird die Segmentierungsmaske des deformierten Templatebildes mit der Deformation aus der klassischen Registrierung gezeigt  $M_K$ . Wenn in der Deep Learning-basierten Deformation deutliche Abweichungen vorhanden sind, werden die Segmentierungsmasken zusätzlich abgebildet  $M_{DL}$ . In der unteren Reihen sind die Veränderungskarten der verschiedenen Methoden als farbliche Überlagerung sowie die zu den Farben gehörende Legende abgebildet.

In den ersten beiden Abbildungen (6.12 6.13) sind Bereiche von NLST-Läsionen und deren Veränderungskarte abgebildet. Bei Abbildung 6.12 sind zwei NLST-Läsionen in der Schicht vorhanden. Durch die verschiedenen Veränderungen der Segmentierungsmasken der größeren Läsion wird bei den Deep Learning-basierten Veränderungskarten eine Veränderung angezeigt. Die kleinere Läsion ist bei Diff nur durch eine Markierung von anderen Veränderungen und in den anderen Methoden nicht markiert. Bei klassisch sind in beiden Fällen die Masken nahezu identisch, sodass keine Veränderung für die kleiner Läsion und nur ein kleiner Rand für die größere Läsion angezeigt wird.

In der Schicht von Abbildung 6.13 ist eine NLST-Läsion abgebildet. Für klassisch sind die Segmentierungsmasken  $M_K$  nahezu identisch und es wird nur eine leichte Veränderung bei der Läsion angezeigt. Die Masken  $M_{DL}$  unterscheiden sich jedoch. In der angezeigten Schicht ist nur die Segmentierungsmaske der Läsion im Referenzbild abgebildet, während die des deformierten Templatebildes erst in späteren Schichten auftritt, da diese Positionen nicht optimal aufeinander registriert wurden. Demnach ist in den Veränderungskarten von CM, Diff und CMxDiff der Bereich der Läsion markiert. Zudem sind Bereiche des Bronchialbaums markiert.

In den Abbildungen 6.14 und 6.15 ist jeweils ein Ausschnitt einer Schicht aus unterschiedlichen Bildpaaren, die eine synthetisch eingefügte Läsion des LIDC-Datensatzes beinhaltet, mit den Veränderungskarten visualisiert. In beiden Fällen reduziert sich die Größe der Läsion von Referenz- zum Templatebild. Für die sequenziellen Ansätze, klassisch und Diff, lässt sich in den Veränderungskarten der Unterschied genauer erkennen, während in CM die Veränderungskarte einen glatteren Farbverlauf hat und die vollständige Läsion des Referenzbildes markiert ist.

In der letzten Abbildungen ist erneut eine NLST-Läsion abgebildet. Zusätzlich ist eine weitere Veränderung in Lungengewebe zu erkennen. Durch die verschiedenen Veränderungen der Segmentierungsmasken durch die Registrierung wird bei CM und Diff eine Veränderung angezeigt. Dabei hat sich die Maske  $M_{DL}$  nicht von der Größe verändert, sondern deren Position ist um einige Schichten verschoben. Bei klassisch sind die Masken nahezu identisch, sodass keine Veränderung angezeigt wird.

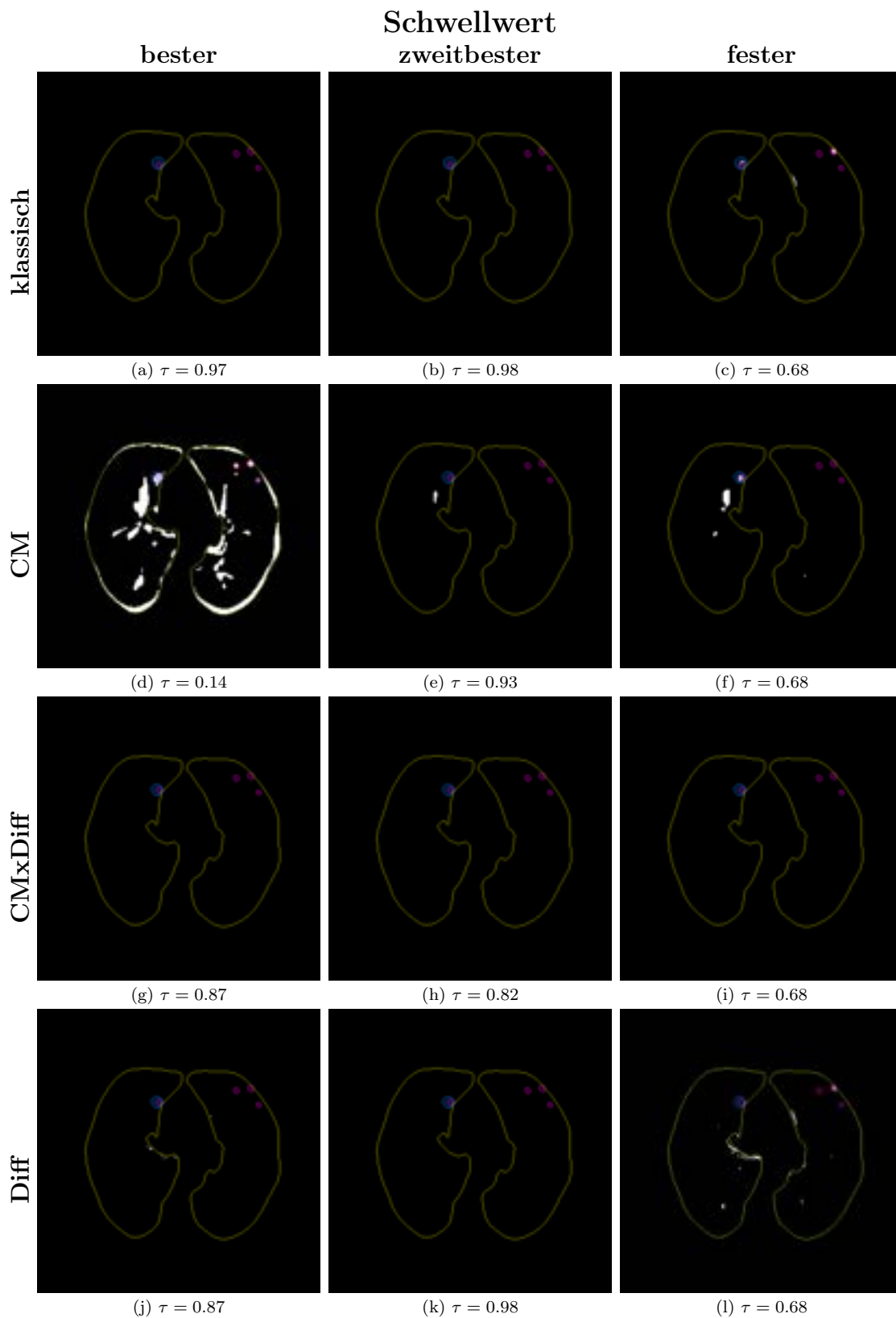
In allen Methoden wird jedoch die Veränderung des Lungengewebes in der Nähe der Läsion im rechten Lungenflügel (links im Bild) angezeigt. Am stärksten ist diese bei Diff visualisiert. In klassisch und CM ist die ähnlich stark hervorgehoben. Die Veränderungskarte von CM hat einen glatteren Farbverlauf und visualisiert die Veränderung der Läsion, die im gleichen Bereich liegt durch einen stärkeren Farbton, sodass beide

Veränderung wahrzunehmen sind. Bei den anderen drei Methoden geht die Veränderung der Läsion in der Markierung von der Veränderung im Lungengewebe unter.

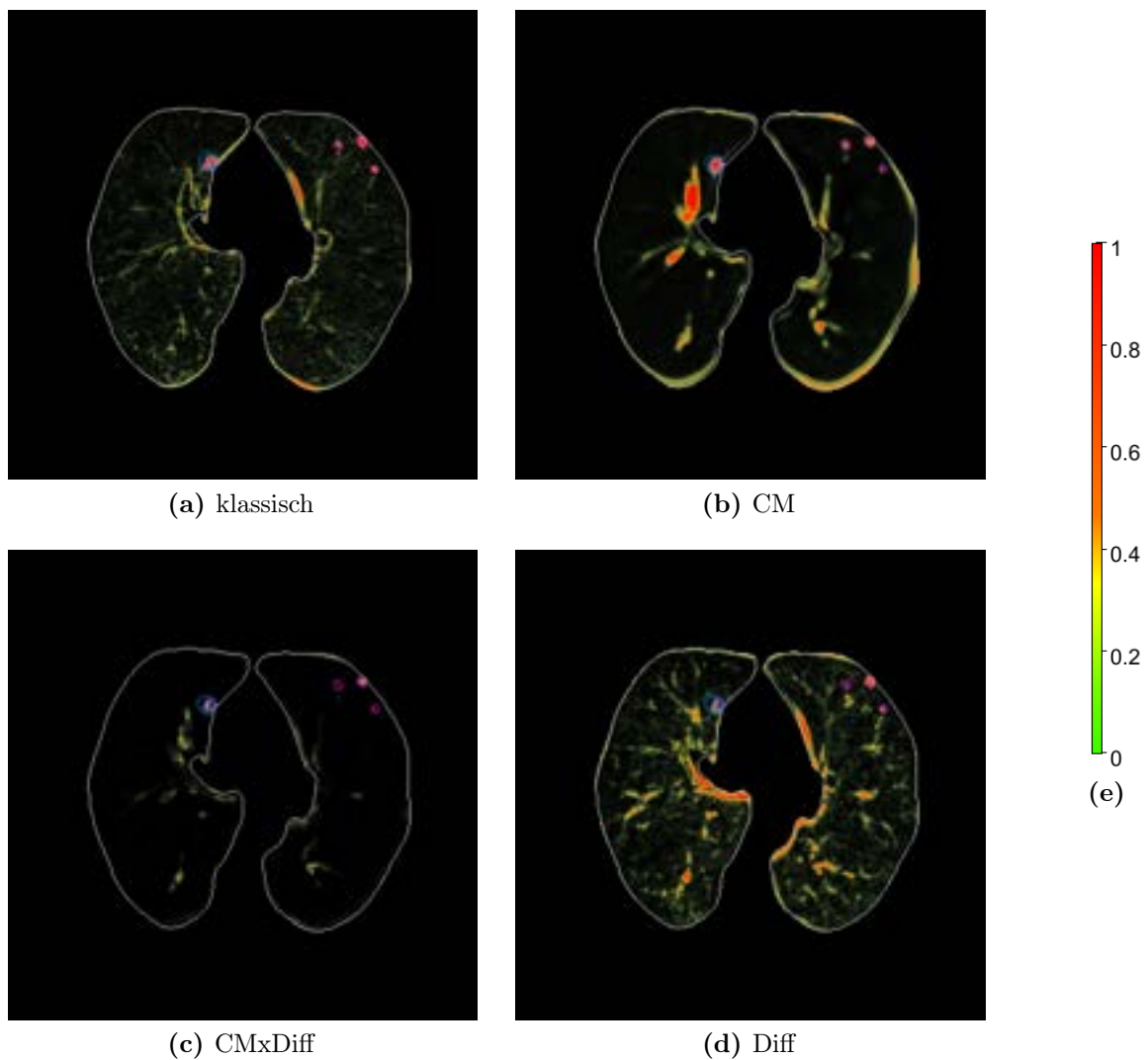
Diese zweidimensionalen Abbildungen zeigen nur einen Teil der dreidimensionalen Veränderungskarten.

Allgemein ist zu beobachten, dass NLST-Läsionen weniger markiert werden, da diese kleiner sind und nur geringere Veränderungen zwischen den Segmentierungsmasken aufweisen. Zudem ist in einigen Bildern zu erkennen, dass die Ränder der Lunge und die Äste des Bronchialbaums nicht optimal aufeinander registriert sind. Dies trifft bei der Deep Learning-basierten Registrierung häufiger im Vergleich zur klassischen Registrierung auf.

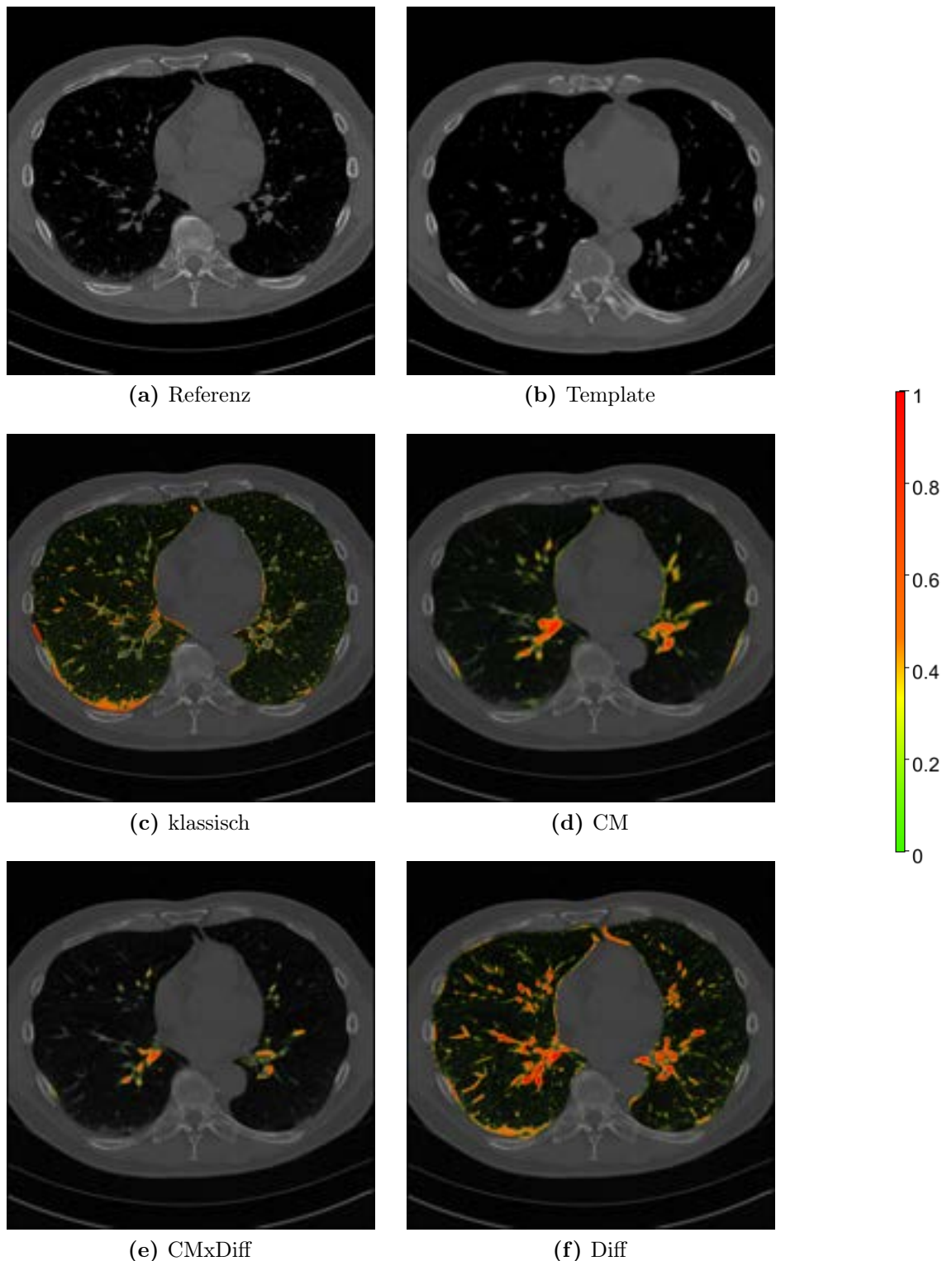
Allgemein hängen die Markierungen von Veränderungen von der Registrierengenauigkeit ab. So ist eine größere Übereinstimmung der Positionen der Läsionen in der klassischen Registrierung zu erkennen, wohingegen Läsionen in der Deep Learning-basierten Registrierung nahe beieinander lokalisiert sind und durch den Versub häufiger als eine Veränderung markiert sind.



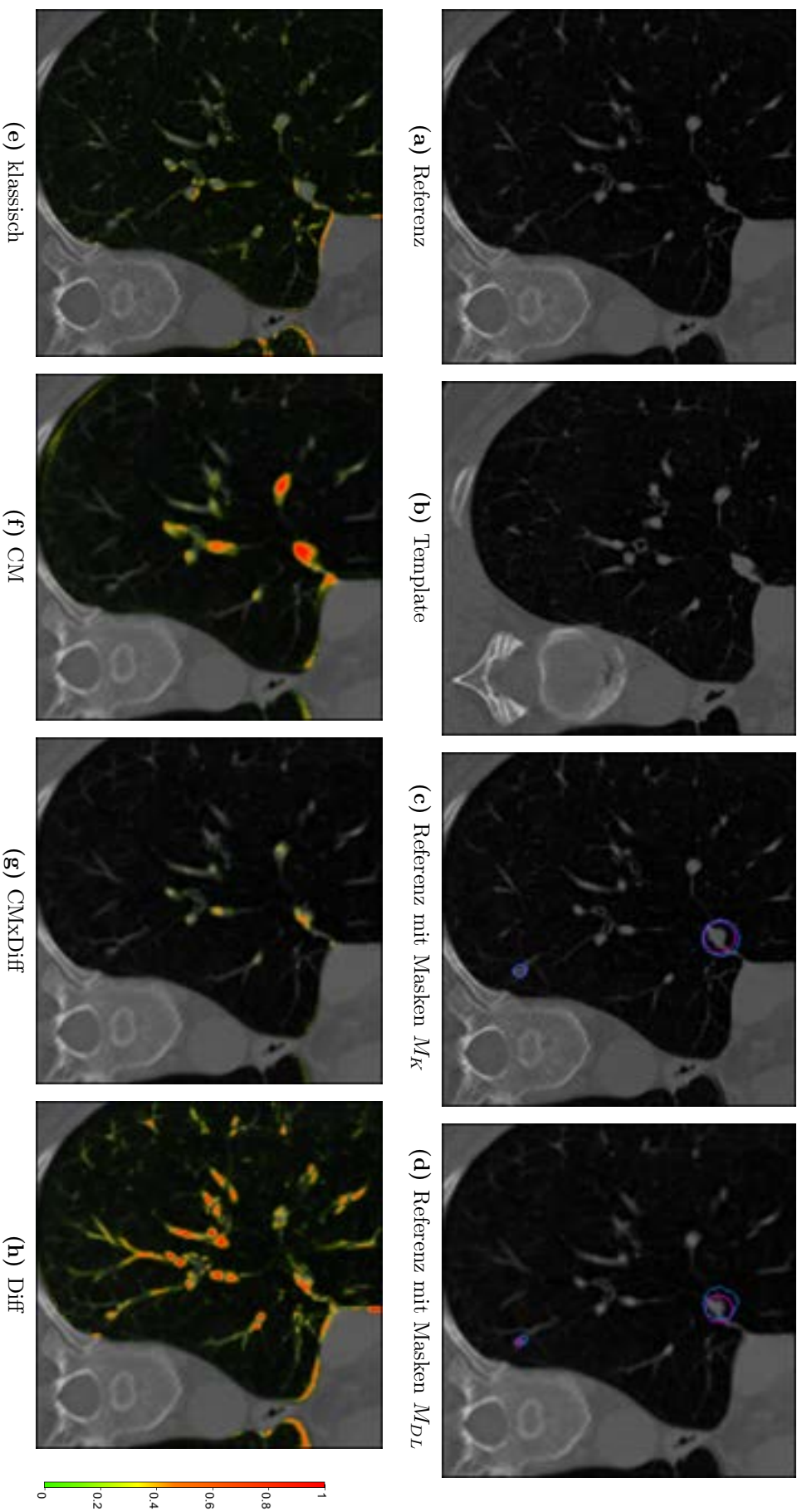
**Abb. 6.9:** Visualisierung der verschiedenen binären Veränderungskarten. Es wird die gleiche Schicht, wie bei den binären Veränderungskarten in Abbildung 6.7 dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila, die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau und die Segmentierungsmaske der Lunge in weiß dargestellt. Die Veränderungskarten wurden mit denselben Schwellwerten wie Abbildung 6.7 verwendet erstellt.



**Abb. 6.10:** Farbliche Visualisierung der verschiedenen kontinuierlichen Veränderungskarten. Es wird die gleiche Schicht, wie bei den binären Veränderungskarten in Abbildung 6.9 dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila, die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau und die Segmentierungsmaske der Lunge in weiß dargestellt.

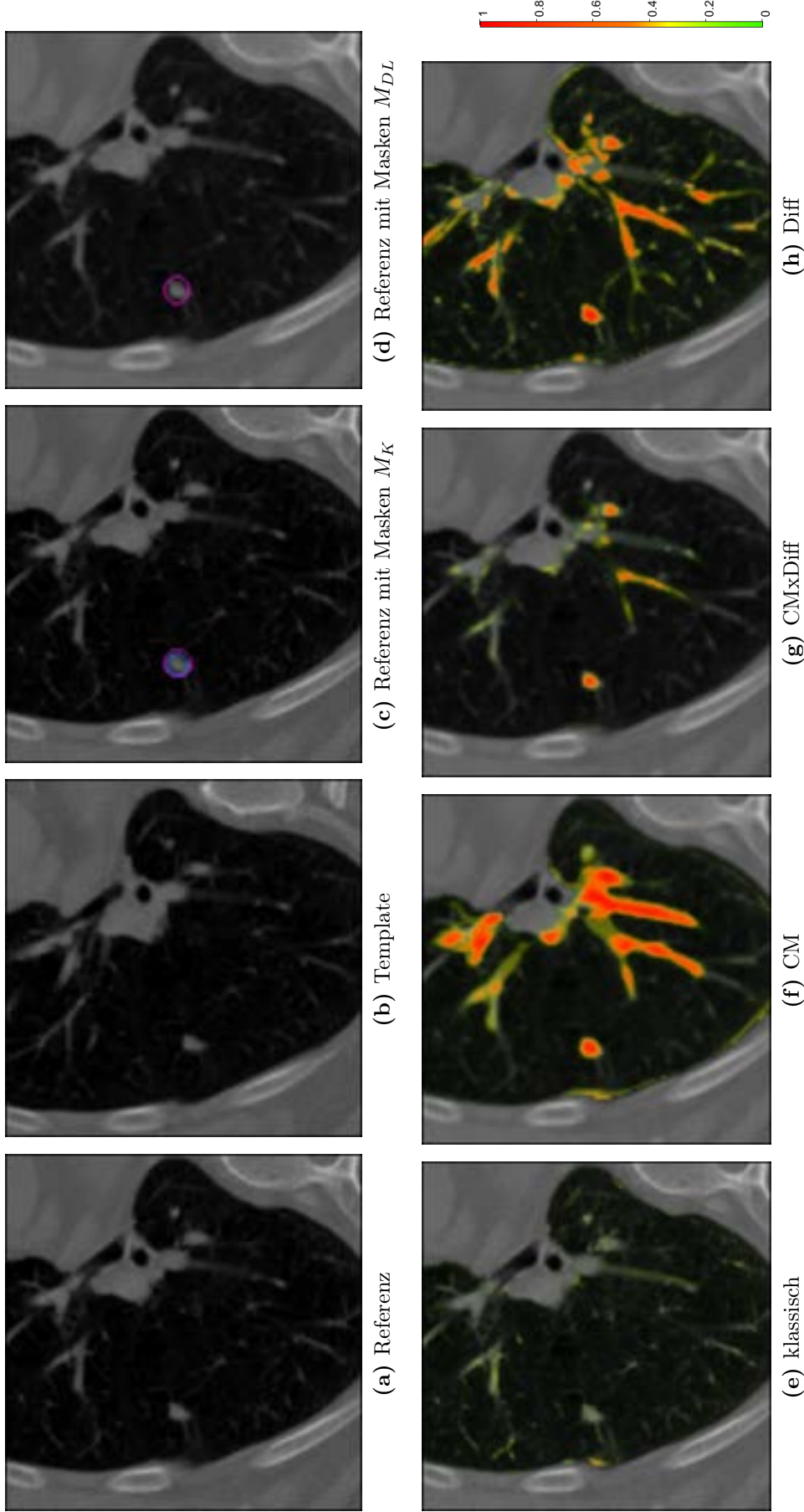


**Abb. 6.11:** Darstellung der verschiedenen Veränderungskarten auf einer Schicht in einem Bild, die keine Läsion enthält. In der oberen Reihe sind das Referenz- und Templatebild dieser Schicht dargestellt. In den unteren beiden Reihen sind die Veränderungskarten als farbliche Überlagerung auf dem Referenzbild angezeigt. Seitlich ist die Farbskala für die Veränderungskarten abgebildet. Es gibt keine bekannten Läsionen in dieser Schicht, jedoch ist eine andere Veränderung im Referenzbild zu erkennen. Die Veränderung wurde in den Veränderungskarten von klassisch und Diff farblich stark hervorgehoben. Jedoch sind auch Veränderungen am Bronchialbaum markiert, sodass die Markierung der Veränderung weniger auffällt. Bei CM ist Bereich der genannten Veränderung eine leichte Markierung zu erkennen, jedoch nicht so deutlich, wie bei den anderen beiden Methoden. Auch bei CMxDiff ist diese leicht markiert.

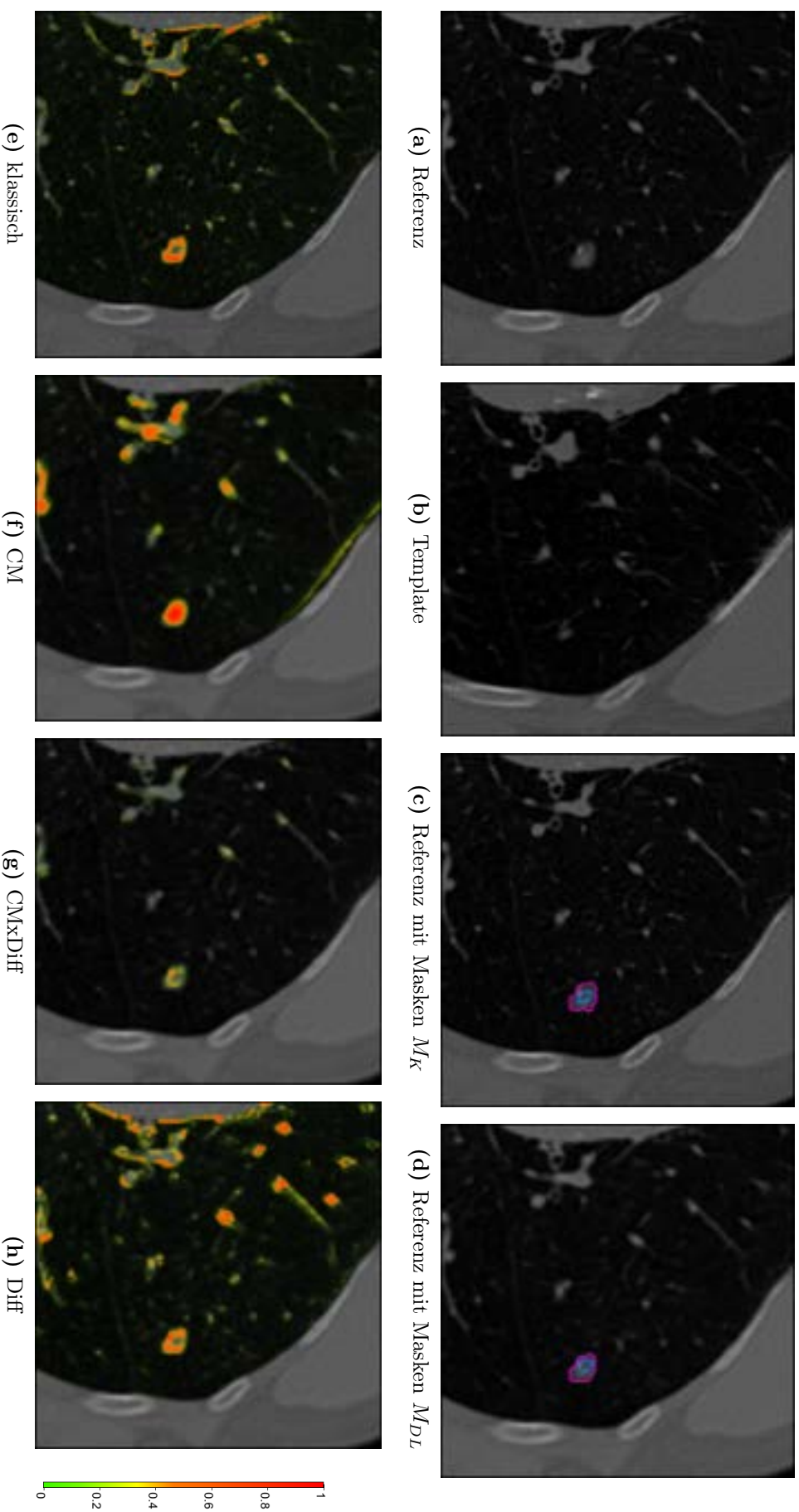


**Abb. 6.12:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die zwei NIST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Durch die verschiedenen Veränderungen der Segmentierungsmasken der größeren Läsion wird bei den Deep Learning-basierten Veränderungskarten eine Veränderung angezeigt. Die kleinere Läsion ist jedoch nicht bzw. bei Diff nur durch eine Markierung von anderen Veränderungen markiert. Bei klassisch sind in beiden Fällen die Masken nahezu identisch, sodass keine Veränderung für die kleiner Läsion und nur ein kleiner Rand für die größere Läsion angezeigt wird. Die vollständigen Veränderungskarten sind unter A.2 zu finden.

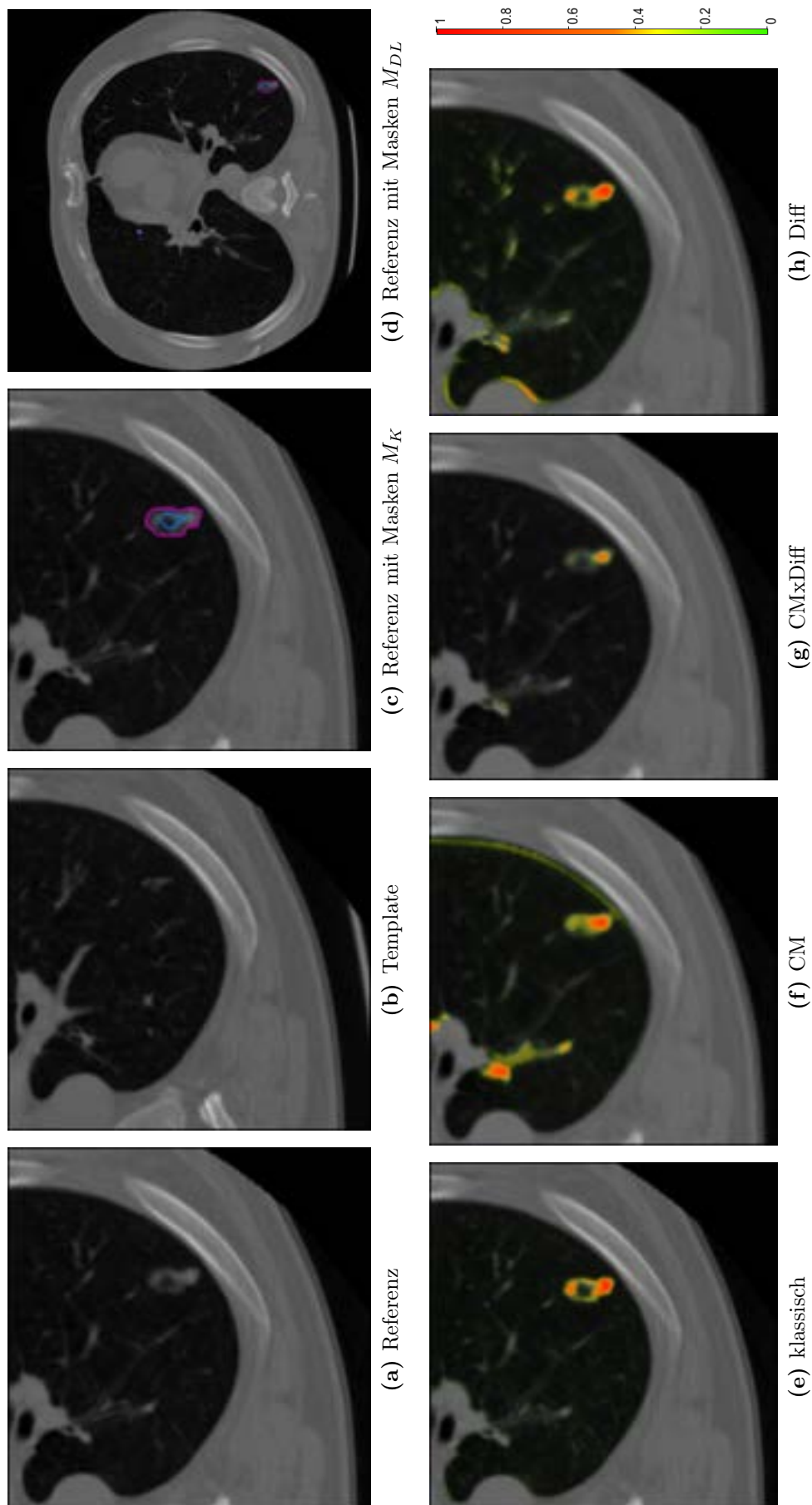




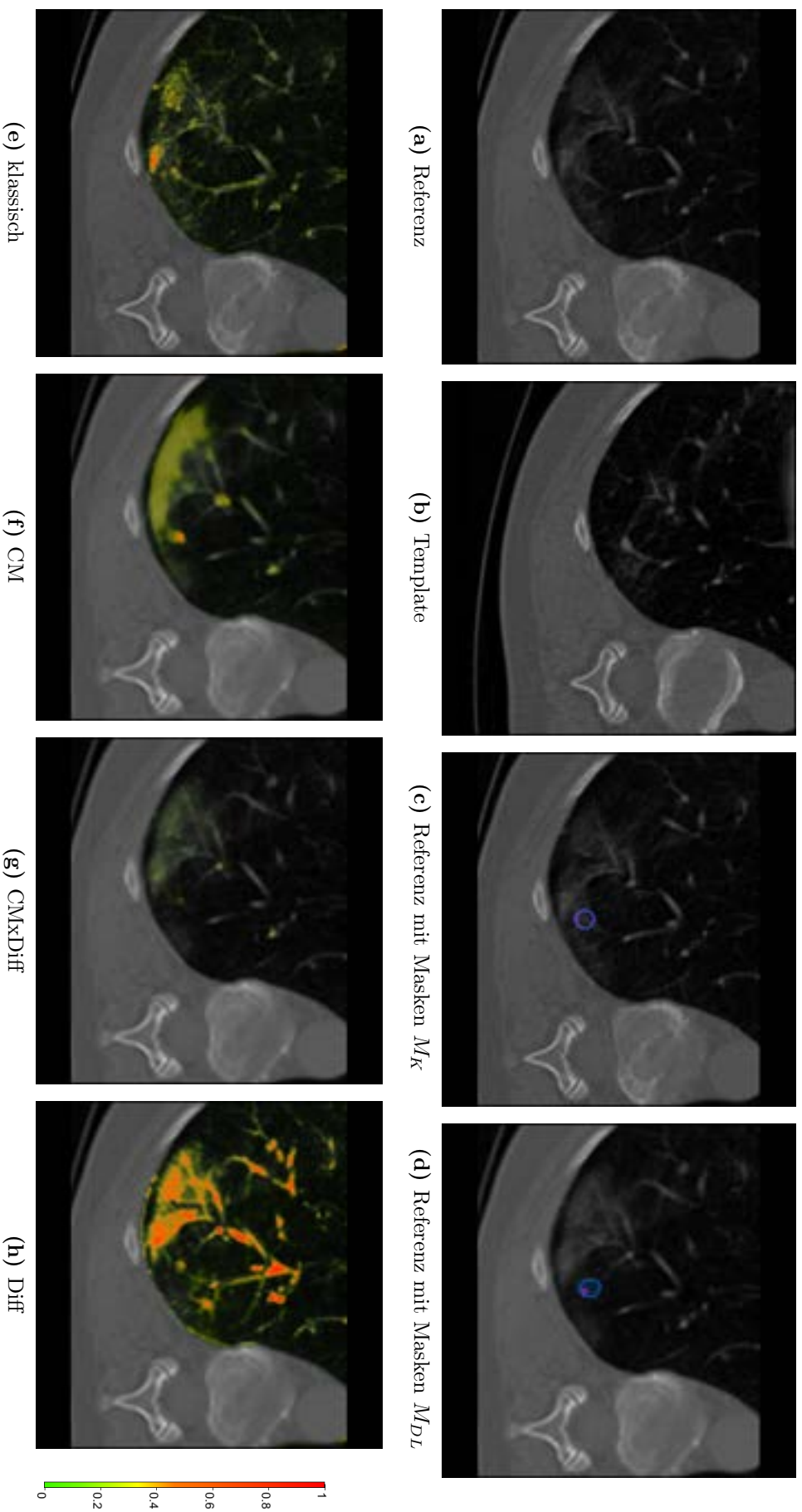
**Abb. 6.13:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die eine NLST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}(d)$  in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Fürklassisch sind die Segmentierungsmasken  $M_K$  nahezu identisch und es wird nur eine leichte Veränderung bei der Läsion angezeigt. Die Masken  $M_{DL}$  unterscheiden sich jedoch. In der angezeigten Schicht ist nur die Segmentierungsmaske der Läsion in Referenzbild abgebildet, während die des deformierten Templatebildes erst in späteren Schichten auftritt. In den Veränderungskarten von CM, Diff und CMxDiff ist der Bereich der Läsion markiert. Zudem sind Bereiche des Bronchialbaums markiert. Die vollständigen Veränderungskarten sind unter A.8 zu finden.



**Abb. 6.14:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die eine synthetisch eingefügte Läsion aus dem LIDC-Datensatz beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die synthetische Läsion hat im Referenz- und Templatebild eine unterschiedliche Größe. Für die sequenziellen Ansätze klassisch und Diff lässt sich dieser Unterschied erkennen, während in CM die vollständige Läsion des Referenzbildes markiert ist. Die vollständigen Veränderungskarten sind unter A.4 zu finden.



**Abb. 6.15:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die eine synthetisch eingefügte Läsion des LIDC-Datensatzes beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihen werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die synthetische Läsion hat im Referenz- und Templatebild eine unterschiedliche Größe. Für die sequenziellen Ansätze klassisch und Diff lässt sich dieser Unterschied erkennen, während in CM die Veränderungskarte einen glatteren Farbverlauf hat und die vollständige Läsion des Referenzbildes markiert ist. Die vollständigen Veränderungskarten sind unter A.4 zu finden. Die vollständigen Veränderungskarten sind unter A.12 zu finden.



**Abb. 6.16:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die eine NLST-Läsion und Veränderungen des Lungengewebes beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Darstellbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Durch die verschiedenen Veränderungen der Segmentierungsmasken durch die Registrierung wird bei CM und Diff eine Veränderung angezeigt. Dabei hat sich die Maske nicht von der Größe verändert, sondern deren Position ist um einige Schichten verschoben. Bei klassisch sind die Masken nahezu identisch, sodass keine Veränderung angezeigt wird. In allen Methoden wird die Veränderung des Lungengewebes in der Nähe der Läsion im rechten Lungenflügel (links im Bild) dargestellt.

## 7 Diskussion und Ausblick

Das Ziel dieser Arbeit war es verbesserte Veränderungskarten zu erstellen, die relevante gegenüber irrelevanten Veränderungen stärker hervorhebt. Dafür wurde ein Netzwerk trainiert, welches ein Deformationsfeld und Veränderungen gemeinsam lernt. Dadurch soll zudem die Überkompensation der Veränderungen durch die Registrierung verringert werden. Aus den Ausgaben des Netzwerkes wurden Veränderungskarten erstellt.

Zum Vergleich haben wir Veränderungskarten über einen sequenziellen Ansatz mit einer klassischen Registrierung erstellt. Wir haben diese Methoden einerseits quantitative als auch qualitative ausgewertet und den Nutzen der Veränderungskarten in der diagnostischen Auswertung mit einer Anwenderstudie überprüft.

Als Erstes wurden für die verschiedenen Methoden zur Erstellung einer Veränderungskarte die Precision-Recall-Kurven, welche auf den Genauigkeiten und Trefferquoten beruhen, sowie die Fläche unter der Precision-Recall-Kurve bestimmt. Dabei haben die beiden Methoden, die die Vorhersage der Veränderung verwenden (CM, CMxDiff), einen höheren AUC-PR-Wert im Vergleich zu den sequenziellen Methoden, (klassisch, Diff), da sie bei einer kleinen Trefferquote hohe Genauigkeiten annehmen. Diese kleinen Trefferquoten treten bei hohen Schwellwerten auf. Es gibt in diesen Methoden bei hohen Schwellwerten meist wenige Komponenten, aber diese markieren zum großen Teil Veränderungen von Läsionen.

Vom Netzwerk wurden vermehrt Läsionen mit Veränderungen mit einem hohen Wert der Veränderungskarte versehen, während nicht relevante Veränderungen einen kleineren Wert erhalten. Die geringe Trefferquote bei den hohen Schwellwerten besagt jedoch, dass nur ein Bruchteil der vorhandenen Veränderungen der Läsionen mit einem hohen Wert versehen sind.

Bei klassisch und Diff werden hingegen für keinen Schwellwert so kleine Trefferquoten angenommen und es werden demnach bei hohen Schwellwerten ein größerer Anteil an Läsionen gefunden. Da wir mit unseren Methoden möglichst alle Läsionen markieren wollen, sind hohe Trefferquoten zu bevorzugen. Es sollte jedoch auch nicht zu viele Markierungen in der Veränderungskarte geben, die keine Veränderungen der Läsionen sind, weshalb eine hohe Genauigkeit ebenso erwünscht ist. Diese nimmt jedoch für klassisch und Diff nur Werte kleiner als 0.05 an.

Da wir die Veränderungskarten binär über Komponenten, dessen Anzahl sich mit den verschiedenen Schwellwerten verändert, auswerten, können die Genauigkeiten für die verschiedenen Schwellwerte verzerrt sein. Wäre etwa die Lunge als eine große Komponente markiert, so hätten wir eine Genauigkeit von 1. Gäbe es dahingegen 100 kleinere Komponenten, die nahezu die ganze Lunge markieren und zur Hälfte mit Veränderungen von Läsionen überlappen, so hätte wir eine Genauigkeit von 0.5, bei fast gleich markierten Bereich. Daher ist eine geringere Genauigkeit für viele kleine Komponenten in der Veränderungskarte wahrscheinlicher im Vergleich zu wenigen großen Komponenten.

Mit den Dice-Koeffizienten über die Veränderungskarte und Lungensegmentierungsmaske (Abb. 6.4a) kann erkannt werden, dass mit steigendem Schwellwert, die überlagernden Voxel geringer werden. Dadurch kann ausgeschlossen werden, dass die gesamte Lunge mit hohen Werten markiert ist. Jedoch sagt dies nichts über die Aufteilung der einzelnen Komponenten der Veränderungskarte aus.

Die hohen Genauigkeiten bei Schwellwert  $\tau = 0$  bei allen Methoden sowie die hohen Genauigkeiten bei klassisch und CMxDiff bei den kleineren Schwellwert sind durch die große Größe und geringe Anzahl der Komponenten der Veränderungen zu vermuten.

Kleine Größenänderungen von Läsionen sind oft im Differenzbild nach der Registrierung nicht zu erkennen. Dies kann dadurch erklärt werden, dass der Distanzterm während der Optimierung die Anpassung der Läsionskanten aufeinander bewirkt. Da dies nur mit einem glatte Deformationsfeld in Einklang zu bringen ist und der Regularisierungsterm für kleine Veränderungen ein geringes Gewicht im Energiefunktional hat, kann es eher zu Überanpassungen der Veränderungen kommen. In der klassischen Registrierung weisen die Voxel in  $x$ - und  $y$ -Richtung durchschnittlich eine Größe von 0.66 mm auf, weshalb die Markierungen von Veränderungen bereits vor der Registrierung nur wenige Voxel entsprechen. Für die Deep Learning-basierten Methoden wurde zusätzlich die Auflösung in  $x$ - und  $y$ -Richtung zuvor halbiert, sodass kleine Veränderungen noch weniger Voxel umfassen.

Da die NLST-Läsionen einen Medianwert bezüglich der axialen Durchmesserdifferenzen von 0.75 mm haben, sind diese schwer zu in der Veränderungskarte zu visualisieren. Zudem sind die wahren Veränderungen schwieriger zu erlernen, da die Bereiche und somit die Differenzen der Läsionen durch Kugeln approximiert sind. Der Medianwert der synthetischen Läsionen beträgt dagegen 7.78 mm, sodass diese mehr Voxel umfassen und besser erkannt werden können.

Trotz der geringen Voxelanzahl der Veränderungen werden in allen Methoden auch kleine Veränderungen erkannt und markiert (Abb. 6.7). Diese haben vermehrt einen helleren Farbton, also einen geringeren Wert in der Veränderungskarte. Größere Veränderungen der Läsionen haben dagegen häufiger einen dunklen Farbton, also einen höheren Wert in der Veränderungskarte. Dies kann auch auf strukturelle Unterschiede zurückgeführt werden. Zum einen machen Anpassungen von wenigen Millimetern bzw. Voxeln durch die Registrierung prozentual einen geringeren Anteil der Unterschiede aus, zum anderen korrespondieren Veränderungen der axialen Durchmesser mit dem maximalen axialen Durchmesser des Referenz- und Templatebildes (Abb. 6.6). Da eine Läsion als richtig gefunden klassifiziert wird, sobald ein Voxel im Bereich dieser Läsionen auch in der Veränderungskarte markiert ist, ermöglicht die größere Voxelanzahl statistisch eine höhere Trefferquote in der binären Auswertung. Um diesem Ungleichgewicht entgegenzuwirken, könnte statt der absoluten Überlagerung von einem Voxel eine prozentuale Überlagerung als Entscheidung für richtig markierte Veränderungen verwendet werden.

Trotz der unterschiedlichen Auflösungen und den damit verbunden Schwierigkeiten, sind keine eindeutigen Unterschiede zwischen den Ergebnissen der Methoden und somit kein Vorteil für klassisch durch eine feinere Auflösung zu erkennen. Es bleibt jedoch offen, ob klassisch durch eine geringere oder die Deep Learning-basierten Methoden durch eine höhere Auflösung schlechtere bzw. bessere Ergebnisse erzielen würden.

Insgesamt gibt diese Auswertung einen Überblick über die Trefferquoten und Genauigkeiten für verschiedene Schwellwerte. Jedoch ist die Auswertung auf Basis dieser Werte nicht absolut zu betrachten.

Zum einen kommt es durch unterschiedliche Verteilungen der Werte der Veränderungskarten zu verschiedene Darstellungen der Werte. So nimmt bei der Vorhersage der Veränderungen nur ein kleinerer Teil Werte nahe 1 an, während beim Erstellen der Veränderungskarten mittels Differenzbildern durch den gewählten Schwellwert, der Artefakte ausschließt, vermehrt der Wert 1 angenommen wird. Dadurch nehmen die Trefferquoten für Veränderungskarten, die auf Differenzen beruhen, bei hohen Schwellwerten größere Werte an.

Zum anderen werden in der Auswertung sechs Schwellwerte zwischen 0.9 und 1 betrachtet, welche damit über ein Fünftel der Werte ausmachen. Jedoch sind in der farblichen Überlagerung (siehe Abb. 4.4 & 6.10e) kaum Unterschiede in der farblichen Wahrnehmung zwischen diesen Werten zu erkennen, da alle Werte über 0.5 mit einer orange-roten Nuance dargestellt werden. So machen diese Schwellwerte einen Unterschied in der Auswertung der Methoden, jedoch kaum einen Unterschied in der gewählten Visualisierung der Veränderungen.

Zudem betrachtet der Dice-Koeffizient über die Veränderungskarte und der Grundwahrheit bezüglich der Bereiche der Läsionen, wie gut diese übereinstimmen. Dieser verschlechtert sich jedoch, wenn für die Läsion ein größerer Bereich als die Veränderung selbst markiert ist und wenn vermehrt der Bereich neben einer Läsion markiert ist. In beiden Fällen wird jedoch die Aufmerksamkeit des Radiologen durch diese Markierungen auf den Bereich der Läsion gelenkt, die bei einer kleineren Markierung eventuell übersehen worden wäre. Daher eignet sich dieser Dice-Koeffizient, ebenso der Dice-Koeffizient mit der Lungen-Segmentierungsmaske, nur zur Abschätzung von Tendenzen und nicht für absolute Aussagen.

Wir konnten zeigen, dass unsere Veränderungskarte, basierend auf der Vorhersage für Veränderungen des Netzwerks (CM), neben Läsionen auch andere Veränderungen erkennt (Abb. 6.11,6.16), diesen jedoch teilweise einen geringeren Wert zuweist, sodass sie bei CM im Vergleich zu den sequenziell erstellten Veränderungskarten weniger auffällig ist. Zudem wurden auch Veränderungen an nicht optimal registrierten Gefäßen markiert. Diese sind bei allen vier Methoden in den Veränderungskarten zu beobachten. Zudem können in der Betrachtung der Veränderungskarten keine Unterschiede bezüglich Überkompensation und Filterung bezüglich relevanter und nicht relevanter Veränderungen bei großen Markierungen erkannt werden. Jedoch sind die Übergänge der Werte der Veränderungskarten von CM glatter, sodass weniger Rauschen in den Markierungen vorhanden ist.

In der Anwenderstudie von [SBH<sup>+</sup>20] konnte gezeigt werden, dass auch mit vielen weiteren markierten Komponenten die Anwender eine verbesserte Bearbeitungszeit sowie Sensitivität erzielen konnte. Eine hohe Genauigkeit ist demnach nicht entscheidend über den Nutzen einer Veränderungskarte, da die Radiologen nicht relevante, markierte Veränderungen, wie nicht optimal registrierte Gefäße, automatisch herausfiltern und nicht genauer betrachten.

In unserer Machbarkeitsstudie konnten wir ebenfalls Verbesserungen bezüglich der Bearbeitungszeit erkennen. In der Studie sollten die Teilnehmer Läsionen, die sich verändern, in 5 Bildpaaren markieren. Insgesamt gab es drei Studienteile, wobei die

Teilnehmer im zweiten Teil die Veränderungskarte, basierend auf dem Differenzbild der klassischen Registrierung (klassisch), und im dritten Teil die Veränderungskarte, kombiniert aus den Veränderungskarten der Deep-Learning basierten Registrierung und der Vorhersage der Veränderungen (CMxDiff), zur Unterstützung erhielten.

Trotz der Aussage einzelner Teilnehmer, dass sie bei den Teilen 2 und 3 die Bilder teils zweifach, einmal mit und einmal ohne Veränderungskarte, durchgegangen sind, konnten für die Teile einer Verbesserung teils eine signifikante Verbesserung der Bearbeitungszeit beobachtet werden. Die signifikanten Verbesserungen wurden für die Teilnehmer ohne radiologische Ausbildung festgestellt. Dabei ist eine zeitliche Verbesserung sowohl durch einen sichereren Umgang in der Befundung als auch die Gewöhnung an die Applikation im Verlauf der Zeit einzubeziehen, da sich die Zeiten mit jedem Studienteil verbesserten.

Neben der Bearbeitungszeit haben wir die Trefferquote oder Genauigkeiten bezüglich markierter Läsionen der Teilnehmer betrachtet. Bei diesen sind keine starken Veränderungen zu erkennen. Durch den geringen Umfang der Fälle und nicht repräsentativen Teilnehmer können wir im Rahmen dieser Masterarbeit nur wenige signifikanten Veränderungen wahrnehmen.

Da jedoch leichte Verbesserungen erzielt wurden, ist davon auszugehen, dass mit einem größeren Umfang an Fällen und mehr repräsentativen Teilnehmern sowie mit einfachen Anpassungen der Veränderungskarten signifikante Verbesserungen durch Verwendung einer Veränderungskarte, wie in [SBH<sup>+</sup>20], gezeigt werden können.

Eine einfache Anpassung wäre eine veränderte Kombination der Veränderungskarten der Deep Learning-basierten Methode, die statt einer voxelweisen Multiplikation eine voxelweise Addition verwenden, um möglichst viele Veränderungen beizubehalten oder eine Skalierung der Vorhersage der Veränderungen ähnlich zum Schwellwert bei der Erstellungen der Veränderungskarten mittels Differenzen der Bilder. Da die Veränderungskarte von CMxDiff in der visuellen Betrachtung die schwächsten und geringsten Markierungen in den Veränderungskarten hat, könnten bereits durch die Wahl einer anderen Veränderungskarte weitere Verbesserungen erzielt werden. Um unsere neue Methode zu testen, sollte bevorzugt die Veränderungskarte von CM der sequenziellen von Diff für eine weitere Anwenderstudie gewählt werden.

Durch die geringen Trefferquoten wäre eine Verwendung von Veränderungskarten im klinischen Alltag bisher nicht von signifikanten Nutzen, wäre jedoch auch nicht nachteilig und kann vereinzelt die Aufmerksamkeit des Radiologen auf Veränderungen lenken.

Durch die geringe Größe der Veränderungen sowie die Bevorzugung von größeren Markierungen, um die Aufmerksamkeit auf die entsprechenden Bereiche zu lenken, ist die Visualisierung von Veränderungen zu überdenken.

Zum einen könnten Veränderungen von wenigen Millimetern in den Bildern sowie Mengen an anderen Markierungen untergehen, wenn nur die Veränderungen selbst angezeigt werden. Gerade solche Veränderungen können in der Befundung für Radiologen schwerer zu erkennen sein und sollen durch eine Veränderungskarte besser wahrgenommen werden. Es bleibt jedoch die Frage offen, ab welcher Größe von einer relevanten Veränderung gesprochen wird, die dokumentiert werden sollte, da sie zu einer Veränderung der Therapie führen würde.

Um die Bereiche, in denen kleine Veränderungen erkannt wurden, hervorzuheben, könnten diese, ähnlich zur Arbeit von Hendrix et al. [HSV<sup>+</sup>21], mit einer Heatmap



eindeutig markiert werden. Diese zeigen selbst nicht die genaue Veränderung an und können auch die Größe der Veränderungen nur indirekt angeben. Damit bleibt das Problem zum Erkennen von den genauen (Größen-)Unterschieden für die Radiologen mit dieser Visualisierung allein bestehen.

Um die Vorteile beider Möglichkeiten zum Anzeigen von Veränderungen in einer Veränderungskarte zu kombinieren, könnten die genauen Veränderungen erlernt werden und wie zuvor visualisiert werden. Zusätzlich können mit der Smooth Grad-CAM++ Methode [OSCW19] Heatmaps als eine *Aufmerksamkeitskarte* für die zuvor bestimmten Veränderungen mit anderen Farbtönen visualisiert werden.

Damit ein Netzwerk nicht nur bestimmte Arten von Veränderungen lernt, sollte Methoden des halb überwachten Lernens sowie Transfer-Lernens einbezogen werden.

Unsere Methode mit einem einfachen Ansatz zur Vorhersage einer Veränderung zeigt für große Veränderungen bereits gute Ergebnisse. Die Veränderungskarten weisen jedoch glatten Übergänge auf und markieren dadurch das Öfteren den gesamten Bereich der Läsion, sodass die genauen Veränderungen schlechter im Vergleich zu den sequenziellen Methoden ablesbar sind. Zudem ist in der Betrachtung der Veränderungskarten zu erkennen, dass die Positionen der Läsionen mit der Deformation der Deep Learning-basierten Methode weniger genau übereinstimmen und somit Veränderungen der Läsionen selbst ungenauer im Vergleich zur klassischen Registrierung dargestellt werden. Da die Vorhersage der Veränderungen mit den deformierten Segmentierungsmasken der Läsionen gelernt wird, hat dieses Deformationsfeld auch Einfluss auf die Genauigkeit der Veränderungskarte von CM und damit auch von CMxDiff. Die Genauigkeit der Veränderungskarte hängt dementsprechend von der Genauigkeit der zugrundeliegenden Registrierung ab.

Um die Veränderungskarte für kleinere Veränderungen zu bessern und die Veränderungen genauer anzuzeigen, könnte die Verlustfunktion optimiert werden und mit der Multi-Level-Strategie gelernt werden. Statt dem recall loss könnte eine Erweiterung des contrastive loss für kontinuierliche Veränderungskarten entwickelt und getestet werden.

Mit dem von Wang et al. vorgestellten contrastive loss [WPZ<sup>+</sup>21] konnten bereits Verbesserungen für die Veränderungserkennung sowie gleichwertige Ergebnisse bezüglich des F1-Scores gezeigt werden. Auch in Bereichen des selbst überwachten, halb überwachten Lernens sowie Transfer-Lernens konnte in den vergangenen Jahren Verbesserungen durch ein contrastive loss gezeigt werden [CKNH20, CEKK21].

Um möglichst genau Veränderungen, insbesondere von kleinen Veränderungen, zu lernen, wären des Weiteren optimierte Datenannotationen von Vorteil. So wäre eine genauere Approximation als eine Kugel von den Bereichen der NLST-Läsionen sowie Annotation von anderen Veränderungen von Vorteil. Da vermehrt auch Veränderungen im Bereich des Bronchialbaums markiert werden, könnten durch weitere Informationen, z.B. ein Atlas der Atemwegen [FHHH22], die Veränderungen besser bezüglich relevant und nicht relevant gefiltert werden. Diese Informationen könnten auch für eine Verbesserung des Deformationsfeldes genutzt werden, ebenso wie die Verwendung von Segmentierungsmasken der Lungenlappen [XJCVG20] anstatt der Segmentierungsmasken der gesamten Lunge.

Für eine Filterung könnte auch ein sogenanntes *attention gate* [OSF<sup>+</sup>18] für die Vorhersage der Veränderungen in unserem Netzwerk verwendet werden. Diese attention gates konzentrierten sich auf die Zielregionen und unterdrückt die Aktivierung von Merk-

malen in irrelevanten Regionen. Dies findet bereits in der medizinischen Bildverarbeitung und auch in der Veränderungserkennung Anwendung [SOS<sup>+</sup>19, CKZ22].

Da das Deformationsfeld und die Vorhersage der Veränderungen in unserer Methode zum größten Teil sich die Gewichte teilen, müsste das Netzwerk so umgebaut und das attention gate so eingefügt werden, dass das Deformationsfeld durch dieses nicht negativ beeinflusst wird.

Es bleibt zu überprüfen, ob bei der Vorhersage der Veränderungen des Netzwerks die synthetischen Daten selbst oder Veränderungen gelernt werden, da große Veränderungen von Läsionen mit den synthetischen Daten korrelieren. Diesbezüglich sind keine Veränderungen in den Ergebnissen zur Erkennung der verschiedenen Durchmesserunterschieden zwischen der Deep Learning-basierten und der klassischen Registrierung zu erkennen. Ein Training nur auf den NLST-Daten ohne synthetisch eingefügten Läsionen würde Klarheit bringen.

Insgesamt konnte mit unserer Methoden gezeigt werden, dass ein gemeinsames Lernen einer Veränderungskarte und eines Deformationsfeldes möglich ist und in der Anwendung vergleichbare Ergebnisse zur sequenziellen klassischen Registrierung liefern. Insbesondere werden auch andere Veränderungen als die gegebenen Läsionsveränderungen gelernt. Diese Arbeit beschränkt sich auf Lungen-CT-Bilder und Veränderungen im Lungenbereich. Das Ziel ist es, eine Veränderungskarte für Ganzkörper-Bildaufnahmen zu entwickeln, die sämtliche Veränderungen visualisiert. Zur Überprüfung der Allgemeingültigkeit muss unsere bisherige Methode auf weiteren Datensätzen sowie andere Regionen im Körper getestet werden.

## 8 Zusammenfassung

Mit der stetig wachsenden Zahl an medizinischen Bilddaten wachsen die Aufgaben für Radiologen. Immer mehr Daten müssen in kürzerer Zeit analysiert und die Therapien entsprechend angepasst werden.

So ist es in der diagnostischen Auswertung schwierig, die genauen Veränderungen in medizinischen Bildern von zwei verschiedenen Zeitpunkten zu erkennen. Zudem besteht die Schwierigkeit der Unaufmerksamkeitsblindheit, also dass Veränderungen oder Auffälligkeiten nicht entdeckt werden, da sich auf eine bestimmte Aufgabe in der Befundung fokussiert wird, z.B. das Erkennen von Läsionen.

Veränderungskarten bieten eine Möglichkeit die Aufmerksamkeit der Radiologen auf Veränderungen zu lenken. Zur Erstellung dieser werden bisher unter anderem Segmentierungen der Strukturen oder sequenzielle Ansätze, bestehend aus einer Registrierung und Erstellung eines Differenzbildes, verwendet. Erstere zeigen jedoch keine Veränderungen von Geweben an und bei zweitem können die für die Diagnostik wichtigen Veränderungen durch die Registrierung reduziert werden.

In unserer Arbeit stellen wir neben einem sequenziellen Ansatz zum Erstellen einer Veränderungskarte, welcher aus einer Registrierung einzelner Bildpaare und anschließendem Erstellen und Nachverarbeiten eines Differenzbildes besteht, einen Ansatz vor, der gemeinsam ein Deformationsfeld und eine Veränderungskarte in einem Netzwerk lernt. Dabei wird mit dem Deformationsfeld des zweiten ebenfalls eine Veränderungskarte über einen sequenziellen Ansatz mit einem Differenzbild der Bildpaare erzeugt. Zusätzlich erhalten wir eine Vorhersage für Veränderungen aus dem Netzwerk, die wir als eigene Veränderungskarte verwenden. Zudem erstellen wir durch die voxelweise Multiplikation der beiden Veränderungskarten eine kombinierte Veränderungskarte.

Wir haben die Methoden zum Erstellen von Veränderungskarten im Bereich der Lunge auf CT-Bildpaaren des NLST-Datensatzes getestet, wobei in einigen Bildern zusätzlich Läsionen aus dem LIDC-Datensatz, dessen genauen Segmentierungen bekannt sind, eingefügt werden.

Diese Methoden wurden bezüglich Genauigkeit und Trefferquote für unterschiedliche Schwellwerte betrachtet und des Nutzens der Veränderungskarten für Radiologen in einer Anwenderstudie untersucht.

Insgesamt konnte mit unserer Methoden gezeigt werden, dass ein gemeinsames Lernen einer Veränderungskarte und eines Deformationsfeldes möglich ist und in der Anwendung vergleichbare Ergebnisse zur sequenziellen klassischen Registrierung liefern. Insbesondere werden auch andere Veränderungen als die gegebenen Läsionsveränderungen gelernt und visualisiert. In den Trefferquoten und Genauigkeiten gab es keine eindeutigen Unterschiede zwischen den verschiedenen Methoden. In der Anwenderstudie konnte eine Verbesserung der Bearbeitungszeit in der Befundung mit Veränderungskarte im Vergleich zur Befundung ohne Veränderungskarte gezeigt werden, die zum Teil signifikant sind. Dabei hat sich die Genauigkeit und Trefferquote der gefundenen Läsionen nicht bei den Teilnehmern nicht verändert.

Mit unserem ersten Ansatz konnte bereits ein Nutzen für die Befundung von Veränderungen in Lungen-CT-Bildern erkannt werden. Zukünftig sollen dieser noch weiter optimiert werden, sodass zum einen die Veränderungen selbst genauer angezeigt werden und zum anderen mit Heatmaps auf die Regionen der Veränderungen aufmerksam gemacht wird, damit auch kleine Markierungen von veränderten Läsionen nicht übersehen werden. Damit soll die Befundung für Radiologen noch weiter erleichtert und verbessert werden.

# Literaturverzeichnis

- [AI15] McLennan G. Bidaut L. McNitt-Gray M. F. Meyer C. R. Reeves A. P. Zhao B. Aberle D. R. Henschke C. I. Hoffman E. A. Kazerooni E. A. MacMahon H. Van Beek E. J. R. Yankelevitz D. Biancardi A. M. Bland P. H. Brown M. S. Engelmann R. M. Laderach G. E. Max D. Pais R. C. Qing D. P. Y. Roberts R. Y. Smith A. R. Starkey A. Batra P. Caligiuri P. Farooqi A. Gladish G. W. Jude C. M. Munden R. F. Petkovska I. Quint L. E. Schwartz L. H. Sundaram B. Dodd L. E. Fenimore C. Gur D. Petrick N. Freymann J. Kirby J. Hughes B. Castele A. V. Gupte S. Sallam M. Heath M. D. Kuhn M. H. Dharaiya E. Burns R. Fryd D. S. Salganicoff M. Anand V. Shreter U. Vastagh S. Croft B. Y. Clarke L. P. Armato III, S. G. Data From LIDC-IDRI [Data set], 2015.
- [AIMB<sup>+</sup>11] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, 38(2):915–931, 2011.
- [APD<sup>+</sup>22] Sakshi Ahuja, Bijaya Ketan Panigrahi, Nilanjan Dey, Arpit Taneja, and Tapan Kumar Gandhi. McS-Net: Multi-class Siamese network for severity of COVID-19 infection classification from lung CT scan slices. *Applied Soft Computing*, 131:109683, 2022.
- [Arm66] Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3, 1966.
- [Bec14] Amir Beck. *Introduction to nonlinear optimization: Theory, algorithms, and applications with MATLAB*. SIAM, 2014.
- [Ber97] Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.
- [BK89] Ruzena Bajcsy and Stane Kovačič. Multiresolution elastic matching. *Computer vision, graphics, and image processing*, 46(1):1–21, 1989.
- [BN06] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.
- [Bro19] Jason Brownlee. A gentle introduction to the rectified linear unit (relu). *Machine Learning Mastery*, 6, 2019.

- [Buz09] Thorsten M Buzug. Computed tomography: from photon statistics to modern cone-beam ct, 2009.
- [BZS<sup>+</sup>18a] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9252–9260, 2018.
- [BZS<sup>+</sup>18b] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. September 2018.
- [BZS<sup>+</sup>19] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019.
- [CEKK21] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. *arXiv preprint arXiv:2112.09645*, 2021.
- [CKNH20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.
- [CKZ22] Sol Cummings, Lukas Kondmann, and Xiao Xiang Zhu. Siamese Attention U-Net for Multi-Class Change Detection. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 211–214. IEEE, 2022.
- [CLY<sup>+</sup>22] Tao Chen, Zhiyuan Lu, Yue Yang, Yuxiang Zhang, Bo Du, and Antonio Plaza. A Siamese Network Based U-Net for Change Detection in High Resolution Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:2357–2369, 2022.
- [DFK<sup>+</sup>20] Eleonore Dufresne, Denis Fortun, Babloo Kumar, Stephane Kremer, Vincent Noblet, Eléonore Dufresne, and Stéphane Kremer. Joint Registration and Change Detection in Longitudinal Brain MRI. pages 104–108, 2020.
- [DHHH22] Adrian Dalca, Alessa Hering, Lasse Hansen, and Mattias Heinrich. Learn2Reg-Challenge 2022. <https://learn2reg.grand-challenge.org/learn2reg-2022>, letzter Zugriff 1.Januar 2023, 2022.
- [Dic45] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

- [Dro15] David Drobny. Automatic Weighting of Non-corresponding Regions in Image Registration, 2015.
- [DVBV<sup>+</sup>19] Bob D De Vos, Floris F Berendsen, Max A Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143, 2019.
- [DVW13] Trafton Drew, Melissa L-H Võ, and Jeremy M Wolfe. The invisible gorilla strikes again: Sustained inattentional blindness in expert observers. *Psychological science*, 24(9):1848–1853, 2013.
- [EP18] Koen AJ Eppenhof and Josien PW Pluim. Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE transactions on medical imaging*, 38(5):1097–1105, 2018.
- [FHHH22] Fenja Falta, Lasse Hansen, Marian Himstedt, and Mattias P Heinrich. Learning an Airway Atlas from Lung CT Using Semantic Inter-patient Deformable Registration. In *Bildverarbeitung für die Medizin 2022*, pages 75–80. Springer, 2022.
- [FOGM18] Enzo Ferrante, Ozan Oktay, Ben Glocker, and Diego H Milone. On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains. In *International Workshop on Machine Learning in Medical Imaging*, pages 294–302. Springer, 2018.
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [GMW81] Philip E Gill, Walter Murray, and Margaret H. Wright. Practical optimization. *Academic Press Inc. [Harcourt Brace Jovanovich Publishers]*, 1981.
- [Had02] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.
- [Han09] Heinz Handels. *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie*. Springer-Verlag, 2009.
- [Hen11] Norbert Henze. *Stochastik für Einsteiger: eine Einführung in die faszinierende Welt des Zufalls*. Springer-Verlag, 2011.
- [HGH19] Alessa Hering, Bram van Ginneken, and Stefan Heldmann. mlvirnet: Multilevel variational image registration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 257–265. Springer, 2019.
- [HH19] Alessa Hering and Stefan Heldmann. Unsupervised learning for large motion thoracic CT follow-up registration. In *Medical Imaging 2019: Image Processing*, volume 10949, pages 331–337. SPIE, 2019.

- [HHM<sup>+</sup>21a] Alessa Hering, Stephanie Häger, Jan Moltz, Nikolas Lessmann, Stefan Heldmann, and Bram van Ginneken. CNN-based lung CT registration with multiple anatomical constraints. *Medical Image Analysis*, 72:102139, 2021.
- [HHM<sup>+</sup>21b] Alessa Hering, Lasse Hansen, Tony C. W. Mok, Albert C. S. Chung, Hanna Siebert, Stephanie Häger, Annkristin Lange, Sven Kuckertz, Stefan Heldmann, Wei Shao, Sulaiman Vesal, Mirabela Rusu, Geoffrey Sonn, Théo Estienne, Maria Vakalopoulou, Luyi Han, Yunzhi Huang, Mikael Brudfors, Yaël Balbastre, SamuelJ outard, Marc Modat, Gal Lifshitz, Dan Raviv, Jinxin Lv, Qiang Li, Vincent Jaouen, Dimitris Visvikis, Constance Fourcade, Mathieu Rubeaux, Wentao Pan, Zhe Xu, Bailiang Jian, Francesca De Benetti, Marek Wodzinski, Niklas Gunnarsson, Jens Sjölund, Huaqi Qiu, Zeju Li, Christoph Großbröhmer, Andrew Hoopes, Ingerid Reinertsen, Yiming Xiao, Bennett Landman, Yuankai Huo, Keelin Murphy, Nikolas Lessmann, Bram van Ginneken, Adrian V. Dalca, and Mattias P. Heinrich. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. pages 1–15, 2021. arXiv: 2112.04489.
- [HHS15] Mattias P Heinrich, Heinz Handels, and Ivor JA Simpson. Estimating large lung motion in copd patients by symmetric regularised correspondence fields. In *International conference on medical image computing and computer-assisted intervention*, pages 338–345. Springer, 2015.
- [HLVDMW17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [HM95] Jun Han and Claudio Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International workshop on artificial neural networks*, pages 195–201. Springer, 1995.
- [HM04] Eldad Haber and Jan Modersitzki. Cofir: coarse and fine image registration. In *Real-Time PDE-Constrained Optimization*. Citeseer, 2004.
- [HM07] Eldad Haber and Jan Modersitzki. Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine*, 46(03):292–299, 2007.
- [HMG<sup>+</sup>18a] Yipeng Hu, Marc Modat, Eli Gibson, Nooshin Ghavami, Ester Bonmati, Caroline M Moore, Mark Emberton, J Alison Noble, Dean C Barratt, and Tom Vercauteren. Label-driven weakly-supervised learning for multimodal deformable image registration. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1070–1074. IEEE, 2018.



- [HMG<sup>+</sup>18b] Yipeng Hu, Marc Modat, Eli Gibson, Wenqi Li, Nooshin Ghavami, Ester Bonmati, Guotai Wang, Steven Bandula, Caroline M Moore, Mark Emberton, et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis*, 49:1–13, 2018.
- [HPA<sup>+</sup>21] Alessa Hering, Felix Peisen, Teresa Amaral, Sergios Gatidis, Thomas Eigentler, Ahmed Othman, and Jan Hendrik Moltz. Whole-Body Soft-Tissue Lesion Tracking and Segmentation in Longitudinal CT Imaging Studies. In *Medical Imaging with Deep Learning*, pages 312–326. PMLR, 2021.
- [HSV<sup>+</sup>21] Nils Hendrix, Ernst Scholten, Bastiaan Vernhout, Stefan Bruijnen, Bas Maresch, Mathijn de Jong, Suzanne Diepstraten, Stijn Bollen, Steven Schalekamp, Maarten de Rooij, et al. Development and validation of a convolutional neural network for automated detection of scaphoid fractures on conventional radiographs. *Radiology: Artificial Intelligence*, 3(4):e200260, 2021.
- [HSZ<sup>+</sup>22] Alessa Hering, Julia Schnabel, Miaomiao Zhang, Enzo Ferrante, Mattias Heinrich, and Daniel Rueckert, editors. *Biomedical Image Registration: 10th International Workshop, WBIR 2022, Munich, Germany, July 10–12, 2022, Proceedings*, volume 13386 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham, 2022.
- [HZRS15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [IJK<sup>+</sup>21] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- [IS] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*.
- [ISO20] ISO. *ISO/IEC 14882:2020: Programming languages — C++*. pub-ISO, 2020.
- [Jac12] Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912.
- [JEH<sup>+</sup>21] Anirudh Joshi, Sabri Eyuboglu, Shih-Cheng Huang, Jared Dunnmon, Arjun Soin, Guido Davidzon, Akshay Chaudhari, and Matthew P Lungren. OncoNet: Weakly Supervised Siamese Network to automate cancer treatment response assessment between longitudinal FDG PET/CT examinations. *arXiv preprint arXiv:2108.02016*, 2021.

- [KB14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
- [KBLG<sup>+</sup>21] Hoel Kervadec, Houda Bahig, Laurent Letourneau-Guillon, Jose Dolz, and Ismail Ben Ayed. Beyond pixel-wise supervision: semantic segmentation with higher-order shape descriptors. 2021.
- [Kö] Lars König. *Matrix-free approaches for deformable image registration with large-scale and real-time applications in medical imaging*. PhD thesis.
- [LF18] Hongming Li and Yong Fan. Non-rigid image registration using self-supervised fully convolutional networks without training data. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1075–1078. IEEE, 2018.
- [LKB<sup>+</sup>17] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [LKS<sup>+</sup>11] Bianca Lassen, Jan-Martin Kuhnigk, Michael Schmidt, Stefan Krass, and Heinz-Otto Peitgen. Lung and lung lobe segmentation methods at fraunhofer mevis. In *Fourth international workshop on pulmonary image analysis*, volume 18, pages 185–99, 2011.
- [LN89] Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.
- [LNK<sup>+</sup>19] Bo Li, Wiro J Niessen, Stefan Klein, Marius de Groot, M Arfan Ikram, Meike W Vernooij, and Esther E Bron. A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 645–653. Springer, 2019.
- [MADSP20] Thiago Moraes, Paulo Amorim, Jorge Vicente Da Silva, and Helio Pedrini. Medical image interpolation based on 3D lanczos filtering. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 8(3):294–300, 2020.
- [mev] MevisLab. <https://www.mevislab.de/>, letzter Zugriff 1.Januar 2023.
- [MGSC18] Dwarikanath Mahapatra, Zongyuan Ge, Suman Sedai, and Rajib Chakravorty. Joint registration and segmentation of xray images using generative adversarial networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 73–80. Springer, 2018.

- [MHN<sup>+</sup>13] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Atlanta, Georgia, USA, 2013.
- [Mod03] Jan Modersitzki. *Numerical methods for image registration*. OUP Oxford, 2003.
- [Mod09] Jan Modersitzki. *FAIR: flexible algorithms for image registration*. SIAM, 2009.
- [Nes03] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.
- [NEU] pytorch: An imperative style, high-performance deep learning library.
- [OSCW19] Daniel Omeiza, Skyler Speakman, Celia Cintas, and Komminist Weldermariam. Smooth Grad-CAM++: An enhanced inference level visualization technique for deep convolutional neural network models. *arXiv preprint arXiv:1908.01224*, 2019.
- [OSF<sup>+</sup>18] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [PGJ04] Seong Ho Park, Jin Mo Goo, and Chan-Hee Jo. Receiver operating characteristic (roc) curve: practical review for radiologists. *Korean journal of radiology*, 5(1):11–18, 2004.
- [QR21] Liang Qiu and Hongliang Ren. U-RSNet: An unsupervised probabilistic model for joint registration and segmentation. *Neurocomputing*, 450:264–274, 2021.
- [RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [RPH<sup>+</sup>17] Jan Rühaak, Thomas Polzin, Stefan Heldmann, Ivor JA Simpson, Heinz Handels, Jan Modersitzki, and Mattias P Heinrich. Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. *IEEE transactions on medical imaging*, 36(8):1746–1757, 2017.
- [S<sup>+</sup>07] Yutaka Sasaki et al. The truth of the F-measure. *Teach tutor mater*, 1(5):1–5, 2007.
- [SBH<sup>+</sup>20] MM Sieren, F Brenne, A Hering, H Kienapfel, N Gebauer, TH Oechtering, A Fürschke, F Wegner, E Stahlberg, S Heldmann, et al. Rapid study assessment in follow-up whole-body computed tomography in

- patients with multiple myeloma using a dedicated bone subtraction software. *European Radiology*, 30(6):3198–3209, 2020.
- [SC99] Daniel J Simons and Christopher F Chabris. Gorillas in our midst: Sustained inattentive blindness for dynamic events. *perception*, 28(9):1059–1074, 1999.
- [SHJ19] Helen R Sofaer, Jennifer A Hoeting, and Catherine S Jarnevich. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577, 2019.
- [Sor48] Thorvald A Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biol. Skar.*, 5:1–34, 1948.
- [SOS<sup>+</sup>19] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53:197–207, 2019.
- [SVB<sup>+</sup>17] Hessam Sokooti, Bob de Vos, Floris Berendsen, Boudewijn PF Lelieveldt, Ivana Išgum, and Marius Staring. Nonrigid image registration using multi-scale 3D convolutional neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 232–239. Springer, 2017.
- [SW65] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
- [Tea11] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.
- [Tea13] National Lung Screening Trial Research Team. Data from the national lung screening trial (NLST) [data set], 2013.
- [TH15] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.
- [TMS<sup>+</sup>20] Junjiao Tian, Niluthpol Chowdhury Mithun, Zachary Seymour, Hanpang Chiu, and Zsolt Kira. Recall loss for imbalanced image classification and semantic segmentation. 2020.
- [TSC<sup>+</sup>21] Minh-Son To, Ian G Sarno, Chee Chong, Mark Jenkinson, and Gustavo Carneiro. Self-supervised Lesion Change Detection and Localisation in Longitudinal Multiple Sclerosis Brain Imaging. June 2021.

- [van95] Guido vanRossum. Python reference manual. *Department of Computer Science [CS]*, (R 9525), 1995.
- [vdWSN<sup>+</sup>14] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.
- [WH18] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
- [Wil92] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer, 1992.
- [WN<sup>+</sup>99] Stephen Wright, Jorge Nocedal, et al. Numerical optimization. *Springer Science*, 35(67-68):7, 1999.
- [WPZ<sup>+</sup>21] Zhixue Wang, Chaoyong Peng, Yu Zhang, Nan Wang, and Lin Luo. Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss. *Neurocomputing*, 457:155–167, 2021.
- [XJCVG20] Weiyi Xie, Colin Jacobs, Jean-Paul Charbonnier, and Bram Van Ginneken. Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans. *IEEE transactions on medical imaging*, 39(8):2664–2675, 2020.
- [XN19] Zhenlin Xu and Marc Niethammer. Deepatlas: Joint semi-supervised learning of image registration and segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 420–429. Springer, 2019.
- [YADG15] Arvind R Yadav, Radhey Shyam Anand, ML Dewal, and Sangeeta Gupta. Gaussian image pyramid based texture features for classification of microscopic images of hardwood species. *Optik*, 126(24):5570–5578, 2015.
- [Zam89] Piero Zamperoni. Morphologische operatoren. In *Methoden der digitalen Bildsignalverarbeitung*, pages 216–244. Springer, 1989.
- [ZBLY16] Tao Zhou, Harish Bhaskar, Fanghui Liu, and Jie Yang. Graph regularized and locality-constrained coding for robust visual tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(10):2153–2164, 2016.
- [ZTLS18] Tao Zhou, Kim-Han Thung, Mingxia Liu, and Dinggang Shen. Brain-wide genome-wide association study for alzheimer’s disease via joint projection learning and sparse regression model. *IEEE Transactions on Biomedical Engineering*, 66(1):165–175, 2018.



# Anhang A

## Anhang

### ANLEITUNG ANWENDERSTUDIE

Erst einmal Danke, dass du an meiner Anwenderstudie für die Masterarbeit teilnimmst! Hier einmal paar Erklärungen bzw. eine Anleitung.

Die Studie besteht aus 3 Teilen. Du bist jetzt im zweiten Teil.  
Deine Aufgabe ist es Läsionen, die sich signifikant verändert haben zu markieren.

Dafür öffne zuerst in der MevisLab-Datei das Modul mit dem Namen *ReaderStudy2*.

In dem ersten Tab gib bitte deinen Namen ein und den Pfad, indem du den Ordner für den entsprechenden Studienteil gespeichert hast.

Zusätzlich gebe die Nummer des Studienteils an. Hier *Study Number: 2*

Als Nächstes klicke bitte auf *check path*. Dies überprüft, ob das erste Bild gefunden wird, also ob der Pfad stimmt, und lädt das erste Bild.

Zudem wird ein Ordner erstellt, in dem deine Werte gespeichert werden. Der Name von diesem Ordner wird neben *SaveFolder* angezeigt, wenn dieser erstellt wurde und besteht aus dem Pfad, Studienteilordner und deinem Namen.

Achte hier besonders einmal drauf, ob *ReaderStudy2* im Pfad steht.

Wenn der Pfad nicht korrekt ist, also das erste Bild nicht gefunden werden kann, steht in der *Info* Zeile **folder path is not correct!**

Kontrolliere dann nochmal den Pfad und klicke erneut auf *check path*. Wenn es alles geklappt hat, wird neben in der *Info* Zeile **folder checked. You can start with the images.**

Du kannst dann zum Tab *Images* wechseln.

Im Tab *Images* siehst du 2 Bilder, das *Baseline* und das *Follow-up* Bild, wobei auf dem *Baseline*-Bild eine farbliche Überlagerung bei der Lunge zu sehen sein sollte, welches mit dem Drücken der Taste *O* entfernt oder hinzugefügt werden kann.

Der erste Fall ist zum Testen der Funktionen und geht **NICHT** in die Bewertung mit ein.

Unten gibt es eine Zeile mit mehreren Optionen.

Links kann man händisch auswählen, welches der Bilder angezeigt werden soll. Es gibt jetzt zusätzlich die Möglichkeit sich die *Changemap* und das registrierte *Follow-up* Bild sich anzeigen zu lassen.

Mit dem Feld *Layout* kann gewählt werden, in welcher Form das Bild gezeigt

werden soll (Axial, Coronal,...).

Bei den Bildern sind sowohl die LUT (variierbar durch Klicken der rechten Maustaste und ziehen) als auch der Cursor miteinander synchronisiert. Beides kann mit den entsprechenden Feldern händisch an und aus geschaltet werden. Ebenso gibt es das Feld *Show current position*, über das der Position Cursor (gelbe Hilfslinien) an und ausgestellt werden können.

Die restlichen Felder sind für die Anwenderstudie.

Deine Aufgabe in der Studie ist es, Läsionen zu markieren, deren Größe sich signifikant verändert hat.

Da es kein Messtool in dieser Anwendung gibt, schätze dieses grob ab, bzw. entscheide für jede Läsion einzeln.

Bitte klicke beim Starten von einem Fall auf den Button *start timer* und wenn du meinst, dass du fertig bist auf den Button *stop timer*. Mit diesen beiden Feldern wird deine Zeit gemessen, die anschließend mit ausgewertet werden soll.

Mit dem Drücken der Taste *O* (für Overlay) auf der Tastatur oder der Checkbox *Show Changemap als Overlay* kann das Overlay hinzugefügt bzw. entfernt werden.

Markiert werden sollen die Läsionen mit einem '+' markiert werden, welches über einen Klick eingefügt wird. Durch einen erneuten Klick kann das '+' wieder entfernt werden. Zusätzlich kann der Markierer über das Feld *marker on* oder durchs Drücken der Taste *X* auf der Tastatur an- und ausgestellt werden.

In beiden Bildern kann markiert werden.

Da es nicht nur Veränderungen bei den Läsionen gibt bzw. es manchmal schwer ist zu entscheiden, ob eine Läsion gewachsen ist, gibt es zwei Möglichkeiten zum Markieren.

In GRÜN sollen die Stellen markiert werden, an denen sich eine Läsion verändert hat.

In ROSA können sämtlich andere Sachen markiert werden (andere Veränderungen, Auffälligkeiten, Läsionen, deren Größe sich nicht verändert hat).

Bitte markiert Läsionen, die ihr gefunden habt, aber der Meinung seid, dass die Größe sich nicht verändert hat, in ROSA.

Zwischen den beiden Farben kann durch Klicken auf *mark other* oder durch Drücken der Taste *M* auf der Tastatur gewechselt werden.

Wenn du mit einem Fall fertig bist und *stop timer* gedrückt hast, kannst auf *next case* klicken.

Deine Werte werden gespeichert und das nächste Bild geladen. Bitte drücke wieder auf *start timer*, wenn du anfängst Läsionen zu suchen.

Falls du aus irgendeinem Grund wieder an den Anfang möchtest, kannst du *Case*



0 klicken und das erste Bild (Case 0) wird wieder geladen.

Insgesamt gibt es 5 Fälle (Case 1-5), die in die Bewertung eingehen.

Die genaue Aufgabenstellung ist noch einmal in Aufgabenstellung beschrieben.

Danke!

Schwellwert $\vartheta$	$\tau$	klassisch		CM		CMxDiff		Diff	
		Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
1	0.0	1.0	0.8265	1.0	0.9904	1.0	0.9904	1.0	0.9904
2	0.1383	0.9929	0.0011	0.7435	0.0418	0.6104	0.0221	0.9935	0.0108
3	0.243	0.9288	0.0021	0.6818	0.029	0.4935	0.025	0.9773	0.0105
4	0.3274	0.8185	0.0037	0.5974	0.0249	0.3929	0.0275	0.9448	0.0114
5	0.3979	0.7473	0.0052	0.5227	0.0247	0.3474	0.0306	0.9026	0.0127
6	0.4586	0.6833	0.0066	0.4578	0.0242	0.3312	0.0358	0.8312	0.0136
7	0.5119	0.6441	0.0084	0.4286	0.0249	0.3149	0.0411	0.7955	0.0147
8	0.5593	0.6085	0.0104	0.3994	0.0253	0.289	0.0465	0.737	0.0153
9	0.6021	0.5623	0.0121	0.3734	0.0255	0.2565	0.0495	0.6883	0.0162
10	0.641	0.5231	0.0139	0.3604	0.0268	0.237	0.054	0.6526	0.017
11	0.6767	0.4911	0.0156	0.3377	0.0272	0.2143	0.0592	0.6071	0.0176
12	0.7097	0.4555	0.0172	0.3149	0.0275	0.2013	0.0653	0.5747	0.0185
13	0.7404	0.4377	0.0192	0.3019	0.0288	0.1916	0.0754	0.5292	0.0193
14	0.769	0.4128	0.0208	0.2857	0.0307	0.1818	0.0906	0.5162	0.0211
15	0.7959	0.395	0.023	0.263	0.0322	0.1721	0.1142	0.4773	0.0215
16	0.8212	0.3808	0.0251	0.25	0.0347	0.1623	0.1458	0.4448	0.0223
17	0.8451	0.3559	0.0267	0.2338	0.0385	0.1526	0.1808	0.4188	0.0237
18	0.8678	0.3381	0.0292	0.2273	0.0455	0.1494	0.2584	0.4091	0.0259
19	0.8893	0.3238	0.032	0.2078	0.0533	0.1396	0.3468	0.3604	0.0258
20	0.9098	0.3025	0.0343	0.1883	0.0671	0.1364	0.4615	0.3247	0.0266
21	0.9294	0.2776	0.0347	0.1721	0.0971	0.1266	0.5735	0.2857	0.0266
22	0.9482	0.2705	0.0377	0.1461	0.1466	0.1104	0.7083	0.2662	0.0279
23	0.9661	0.2562	0.0401	0.1006	0.2422	0.0877	0.7941	0.237	0.0289
24	0.9834	0.2384	0.042	0.0455	0.6364	0.0552	0.8947	0.2208	0.0317
25	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AUC-PR		0.018028		0.200991		0.305855		0.021424	

**Tab. A.1:** Übersicht aller Trefferquoten (recall) und Genauigkeiten (precision) für die verschiedenen Schwellwerte ( $\tau$ ) über alle 50 Bilder, mit denen die PR-Kurven 6.1 für die vier Methoden erstellt wurden. Die Werte sind in Abbildung 6.2 dargestellt. Der Schwellwert  $\tau$  ist die untere Grenze zum Erstellen eines binären Bildes, die nicht eingeschlossen ist. In der letzten Zeile wird die Fläche unter der Precision-Recall-Kurve (AUC-PR) für jede Methode angegeben.

Im Folgenden sind Beispiele für die verschiedenen Veränderungskarten aufgeführt. Dafür werden einzelne Schichten dargestellt. In der oberen Reihe sind das Template- und Referenzbild und das Referenzbild mit einer Visualisierung der Segmentierungsmasken der Läsionen abgebildet. Dabei ist die Segmentierungsmaske der Läsionen

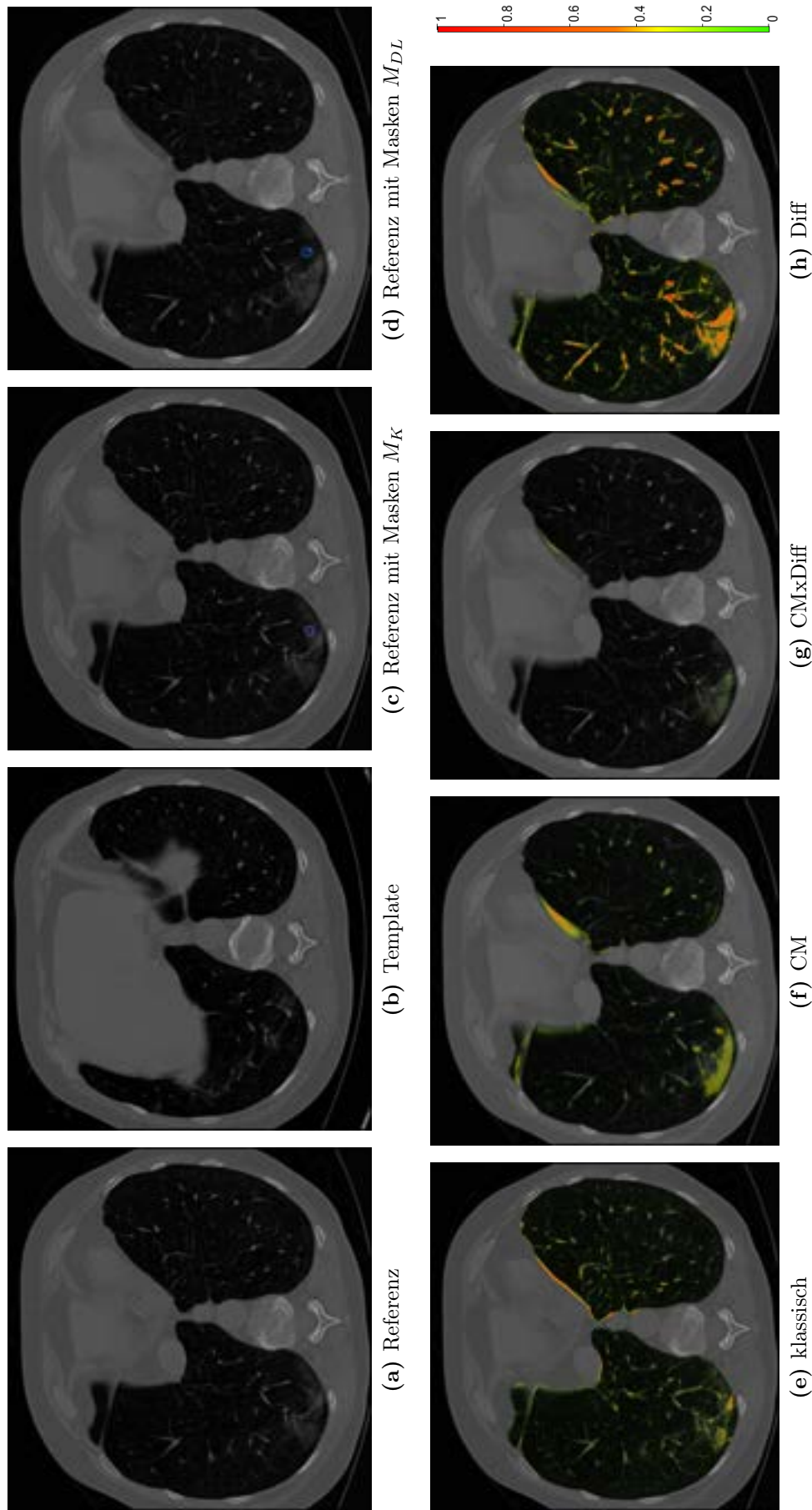
**Tab. A.2:** Trefferquoten und Genauigkeiten der Anwender über alle Läsionen berechnet. Es zeigt sich ein ähnliches Muster, wie bei den mittleren Trefferquoten und Genauigkeiten in Tabelle 6.4

		Anwender 1		Anwender 2		Anwender 3	
		Trefferquote	Genauigkeit	Trefferquote	Genauigkeit	Trefferquote	Genauigkeit
Teil 1	Läsion	0.4615	0.75	0.3333	0.8125	0.0769	0.375
	alle	0.6667	0.6667	0.6923	0.4737	0.1538	0.1429
Teil 2	Läsion	0.2821	0.8466	0.3333	0.52	0.1538	0.5455
	alle	0.5385	0.6563	0.6667	0.4815	0.3077	0.2609
Teil 3	Läsion	0.3846	0.7143	0.2821	0.7333	0.0513	0.2
	alle	0.641	0.7143	0.6154	0.6316	0.2051	0.1379

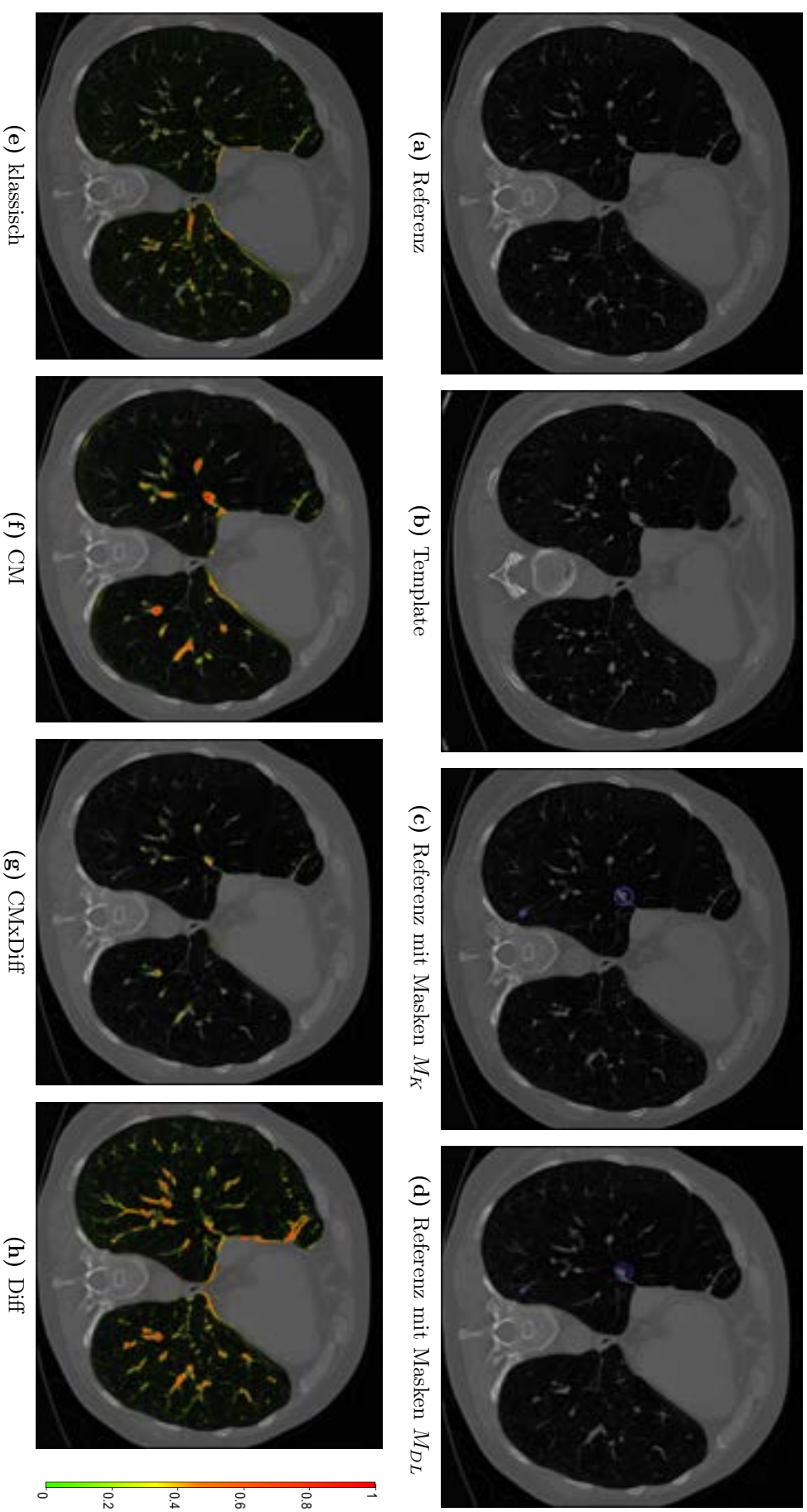
im Referenzbild in Lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in Blau dargestellt. In allen Fällen wird die Segmentierungsmaske des deformierten Templatebildes mit der Deformation aus der klassischen Registrierung gezeigt  $M_K$ . Wenn in der Deep Learning-basierten Deformation deutliche Abweichungen vorhanden sind, werden die Segmentierungsmasken zusätzlich abgebildet  $M_{DL}$ . In der unteren Reihen sind die Veränderungskarten der verschiedenen Methoden also farbliche Überlagerung, sowie die zu den Farben gehörende Legende abgebildet.

Allgemein lassen sich die dreidimensionalen Veränderungen nur teilweise in zweidimensionalen Bildern darstellen. Diese Bilder dienen damit einer groben Übersicht.

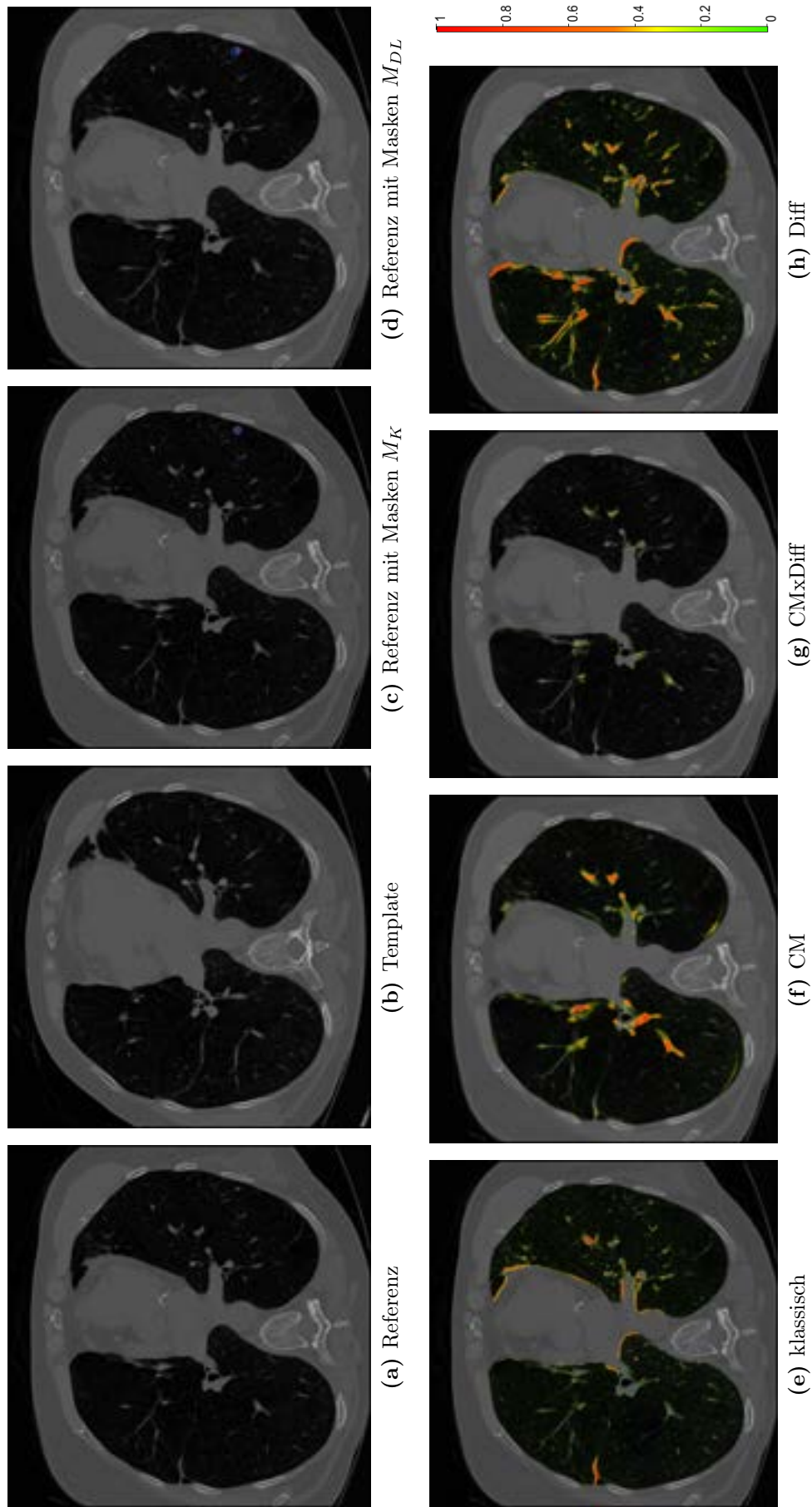
Die Abbildungen A.1A.2A.4A.3 stellen Schichten desselben Bildes dar, ebenso die Abbildungen A.5A.6A.7A.8. Die Abbildungen A.9A.10A.11A.12 stammen aus unterschiedlichen Bildern.



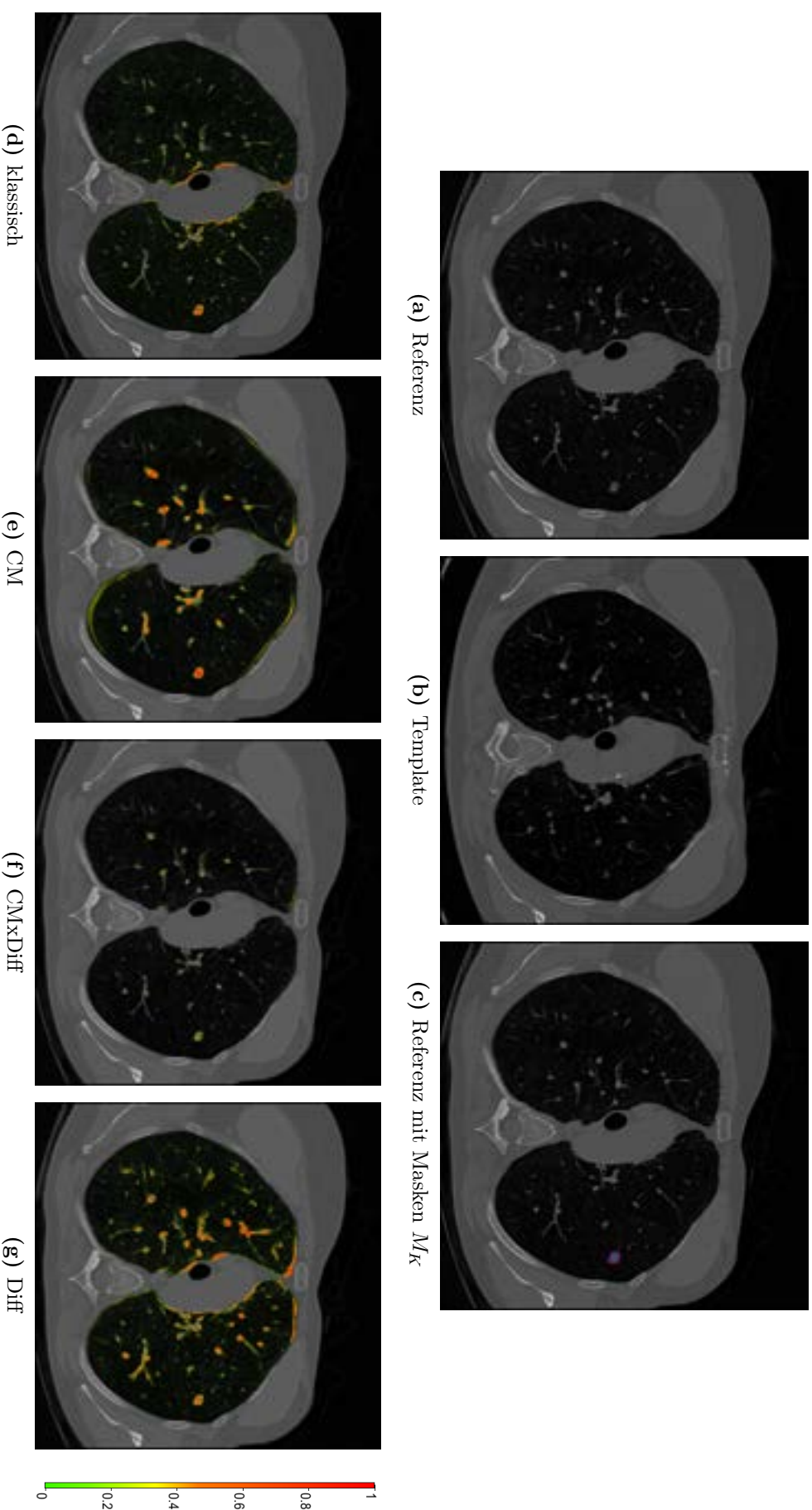
**Abb. A.1:** Visualisierung einer Schicht eines Bildpaares, die eine NLSST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Durch die verschiedenen Veränderungen der Segmentierungsmasken durch die Registrierung wird bei CM und Diff eine Veränderung angezeigt. Bei klassisch sind die Masken nahezu identisch, sodass keine Veränderung angezeigt wird. In allen Methoden wird die Veränderung des Lungengewebes in der Nähe der Läsion im rechten Lungenflügel (links im Bild) dargestellt.



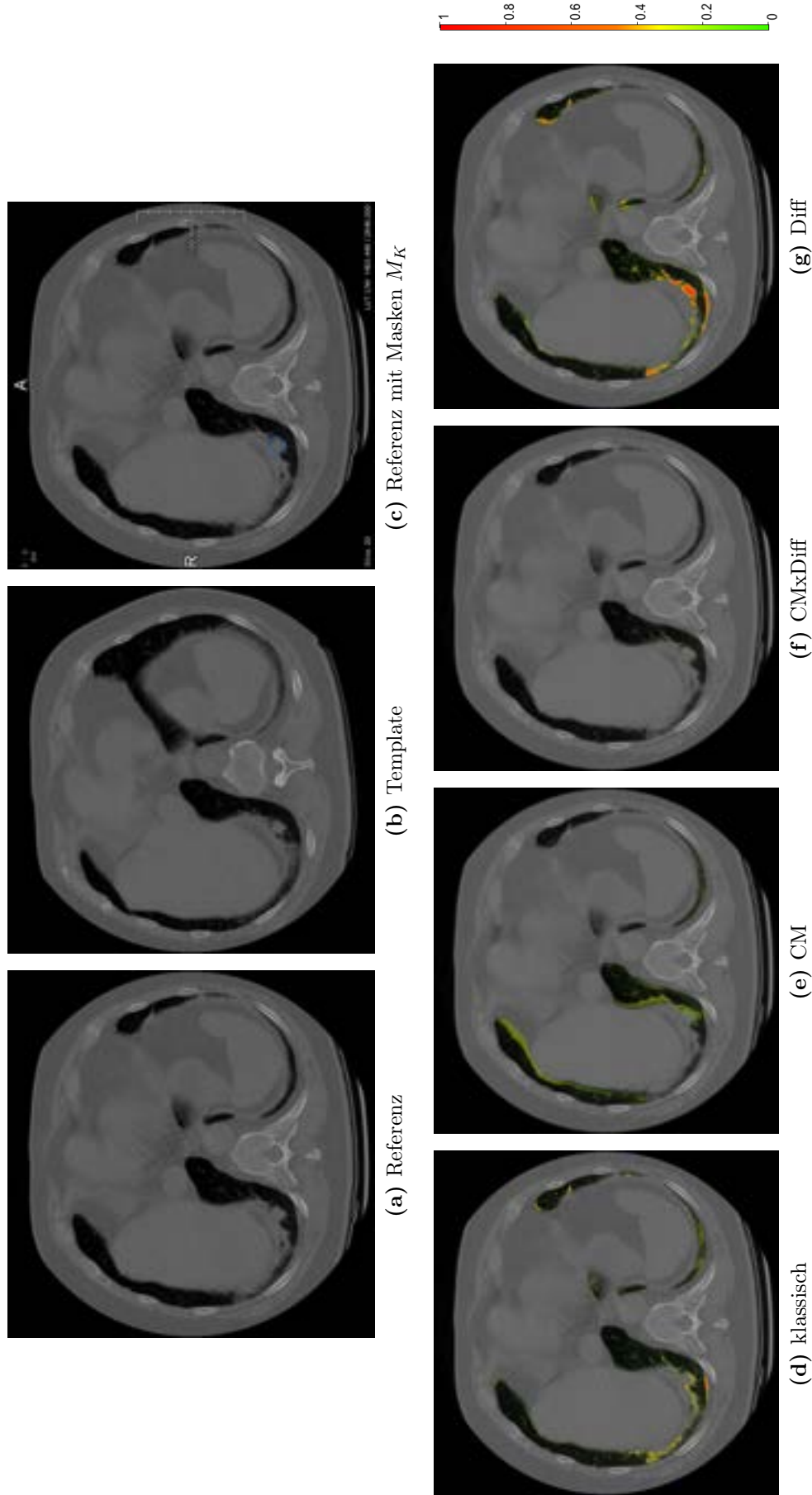
**Abb. A.2:** Visualisierung einer Schicht eines Bildpaares, die zwei NLST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_R$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihen werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Durch die verschiedenen Veränderungen der Segmentierungsmasken der größeren Läsion wird bei den Deep Learning-basierten Veränderungskarten eine Veränderung angezeigt. Die kleinere Läsion ist jedoch nicht markiert bzw. bei Diff nur durch eine Markierung von anderen Veränderungen. Bei klassisch sind in beiden Fällen die Masken nahezu identisch, sodass keine Veränderung für die kleiner Läsion und nur ein kleiner Rand für die größere Läsion angezeigt wird.



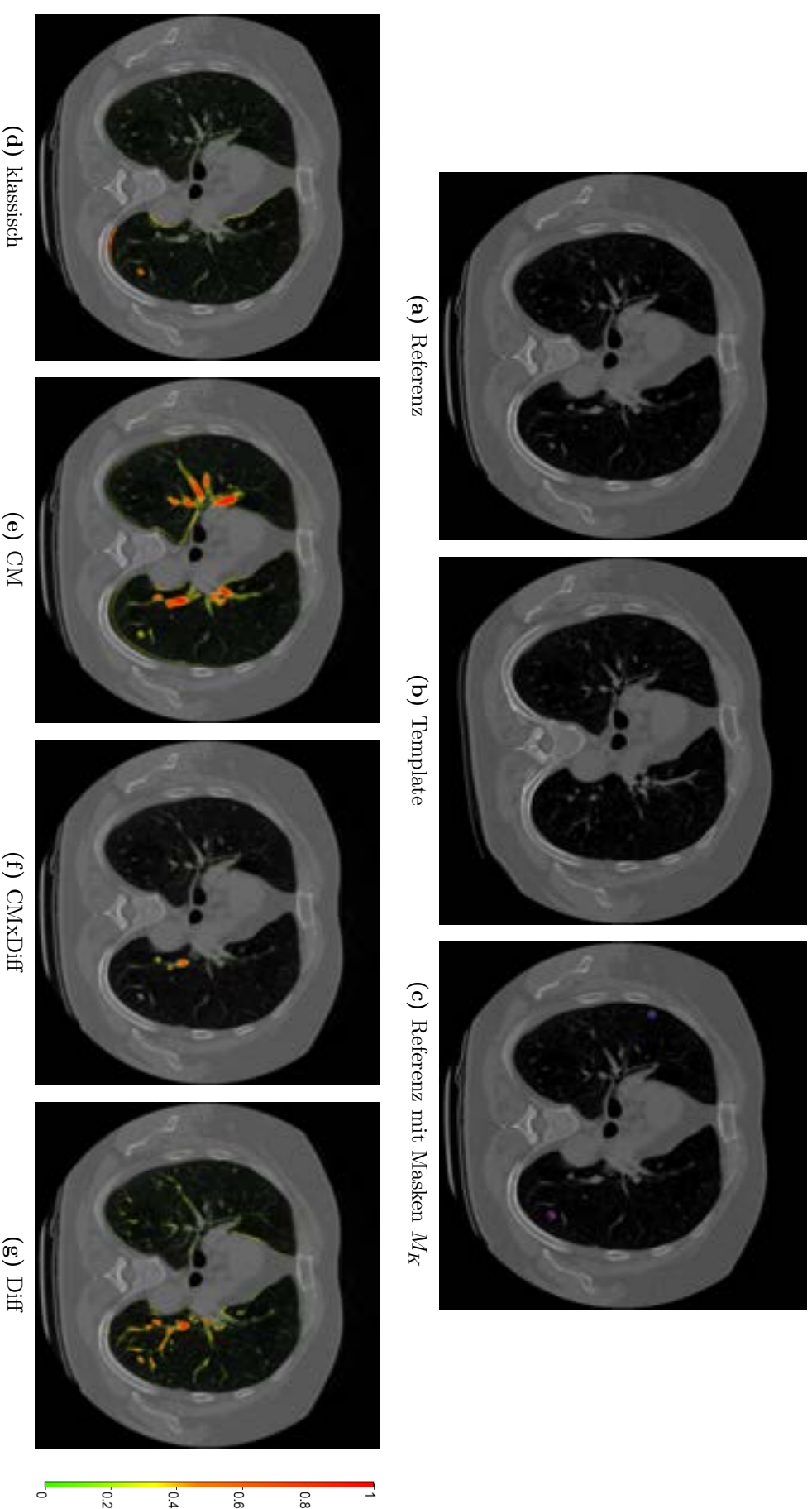
**Abb. A.3:** Visualisierung einer Schicht eines Bildpaares, die eine N1ST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die Läsionsveränderung ist nur in der Veränderungskarte von Diff eindeutig zu erkennen. Die Masken  $M_K$  stimmen vermehrt überein, sodass es für klassisch nur eine kleine Veränderung gibt. Diese wird in einem hellen Grünton anzeigt, sodass sie schwer zwischen den anderen markierten Veränderungen zu erkennen ist.



**Abb. A.4:** Visualisierung einer Schicht eines Bildpaares, die eine synthetisch eingefügte Läsion aus dem LIDC-Datensatz beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_R$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die synthetische Läsion hat im Referenz- und Templatebild eine unterschiedliche Größe. Für die sequenziellen Ansätze klassisch und Diff lässt sich dieser Unterschied erkennen, während in CM die vollständige Läsion des Referenzbildes markiert ist.

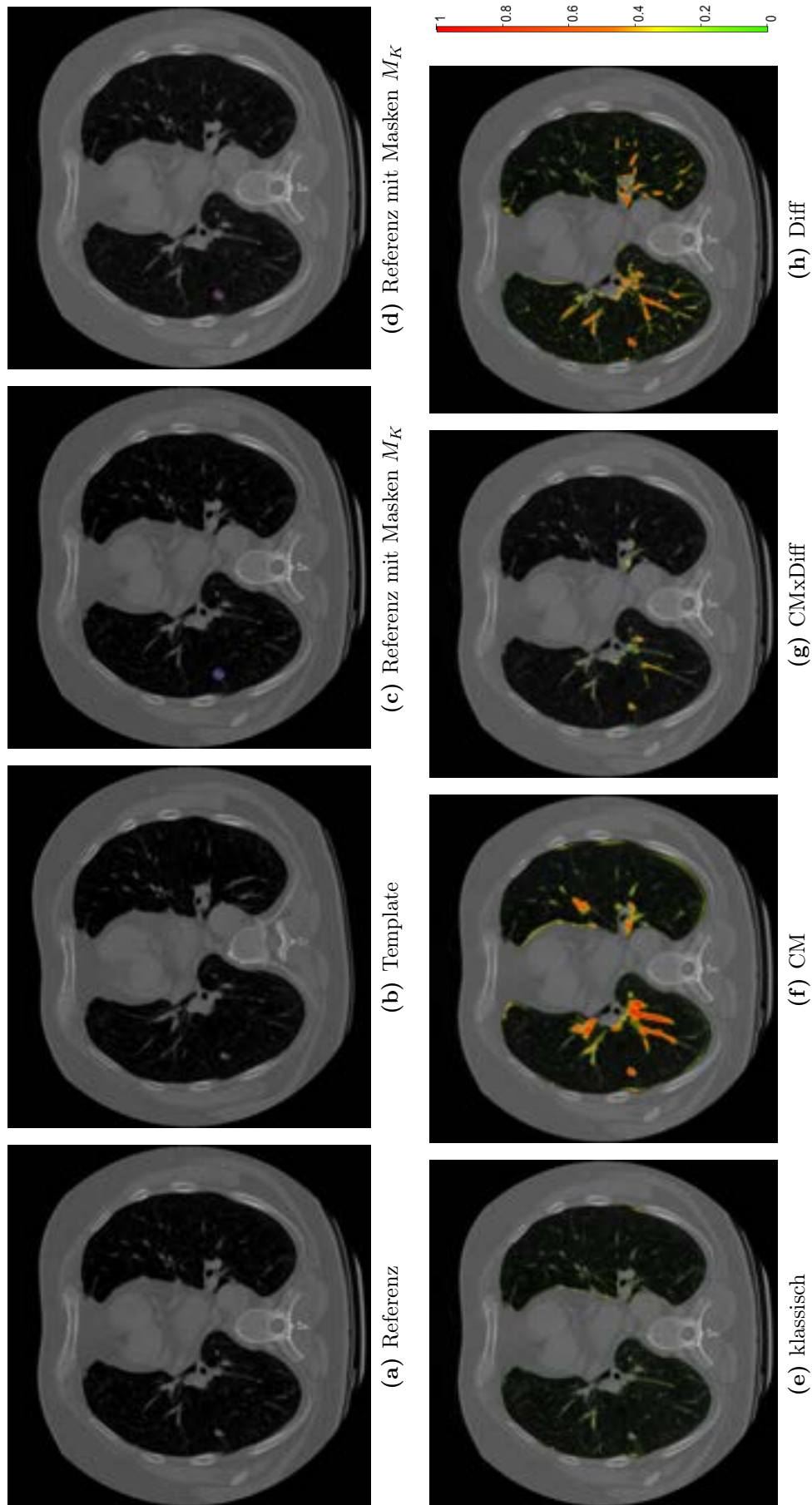


**Abb. A.5:** Visualisierung einer Schicht eines Bildpaares, die eine NLST-Läsion im Referenzbild beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die Läsion wurde nur für das Referenzbild annotiert, jedoch sind schon Strukturen dieser im Templatebild zu erkennen. Nur in Diff ist eine starke Markierung der vollständigen Läsion in der Veränderungskarte zu beobachten, während klassisch und CM nur am Rand markieren. Bei CMxDiff ist keine Markierung zu erkennen.

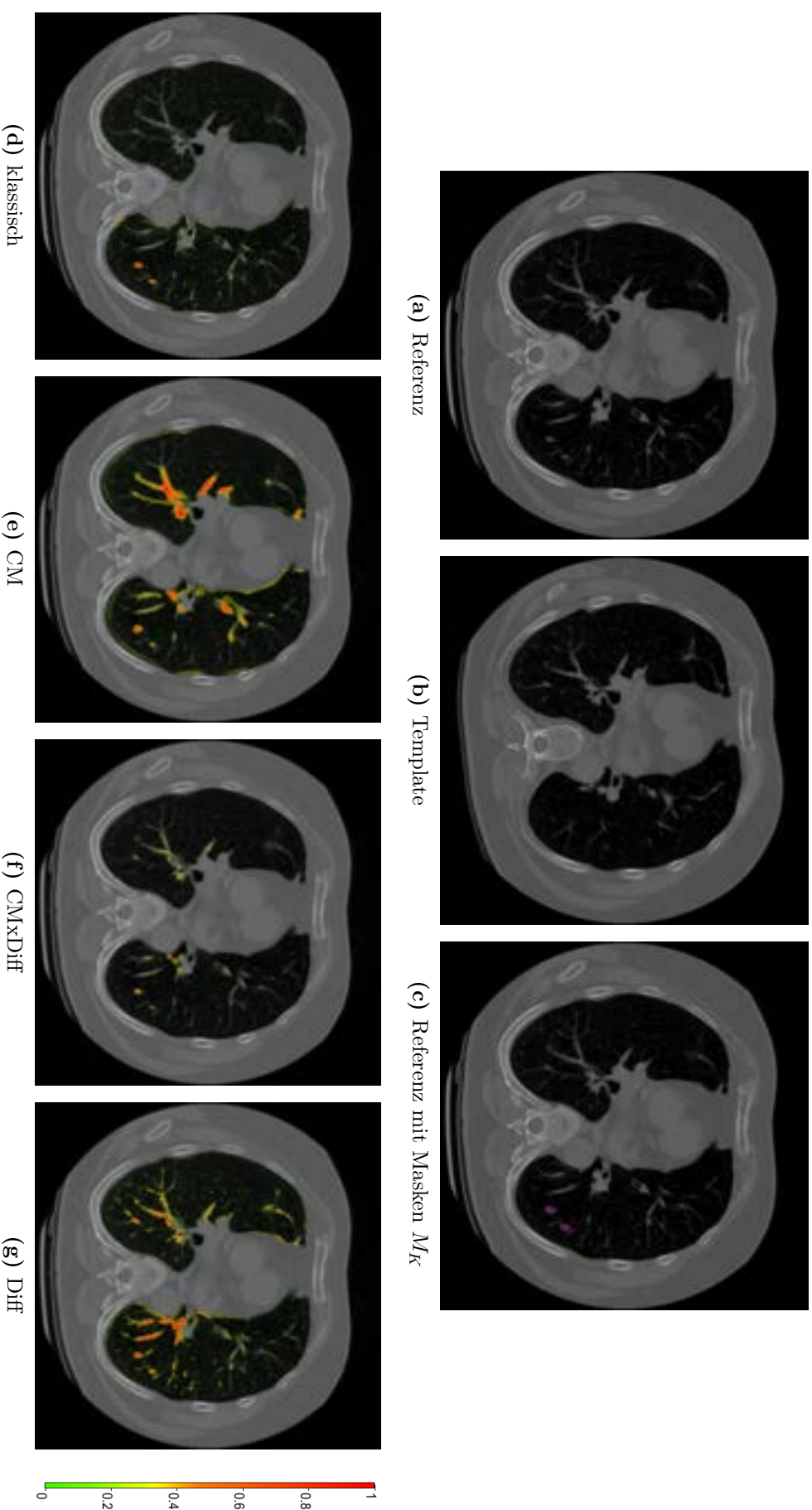


**Abb. A.6:** Visualisierung einer Schicht eines Bildpaares, die eine NIST-Läsion, sowie eine synthetisch eingefügte Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_R$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die synthetische Läsion (rechter Lungenflügel im Bild) tritt nur im Referenzbild auf und wurde nicht im Templatebild eingefügt. In allen Veränderungskarten der vier Methoden wird die vollständige Rezession markiert. Die Segmentierungsmasken der NIST-Läsion (linker Lungenflügel im Bild) stimmen nahezu überein, sodass in keiner Veränderungskarte eine deutliche Markierung vorhanden ist.

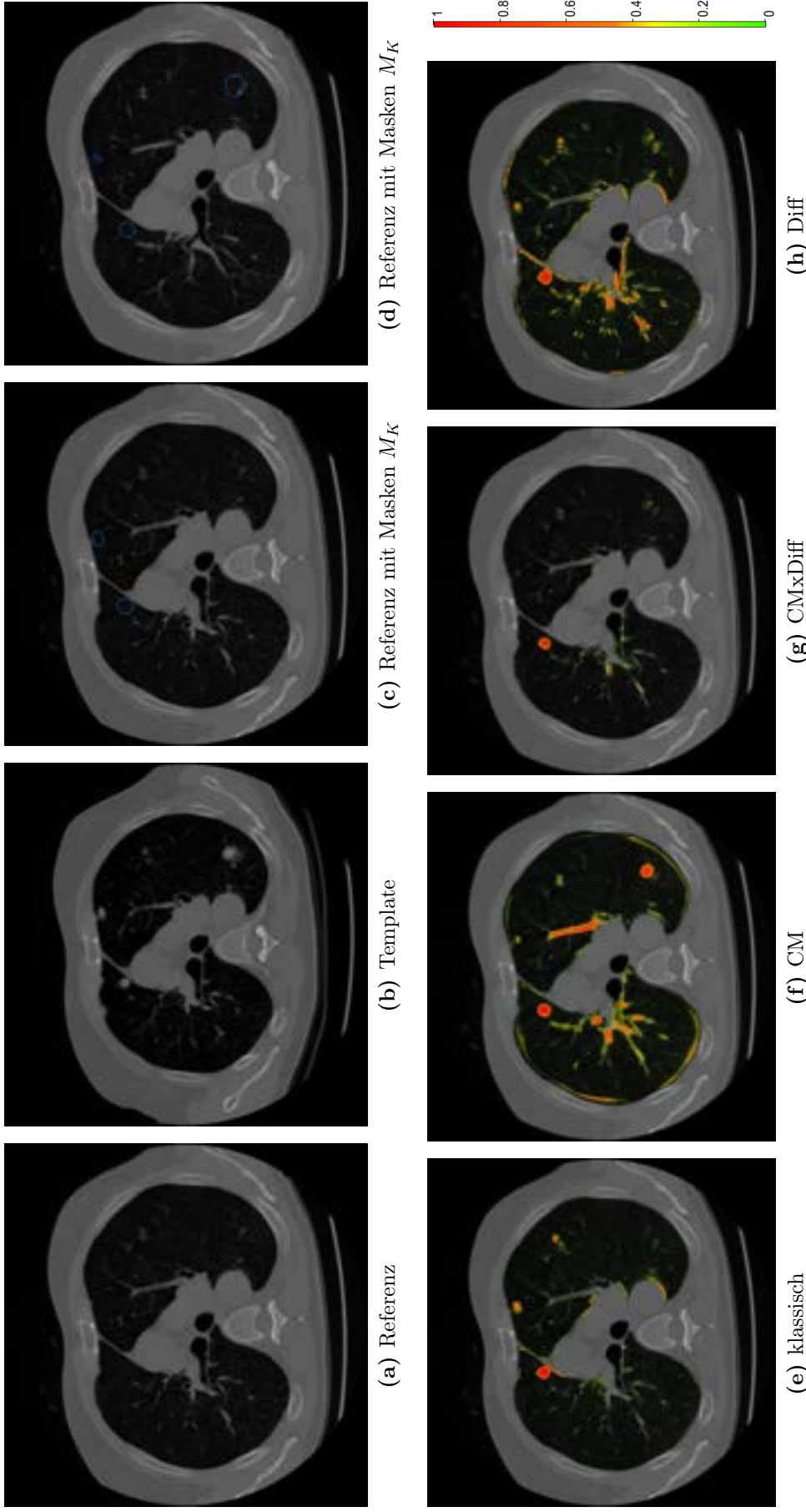




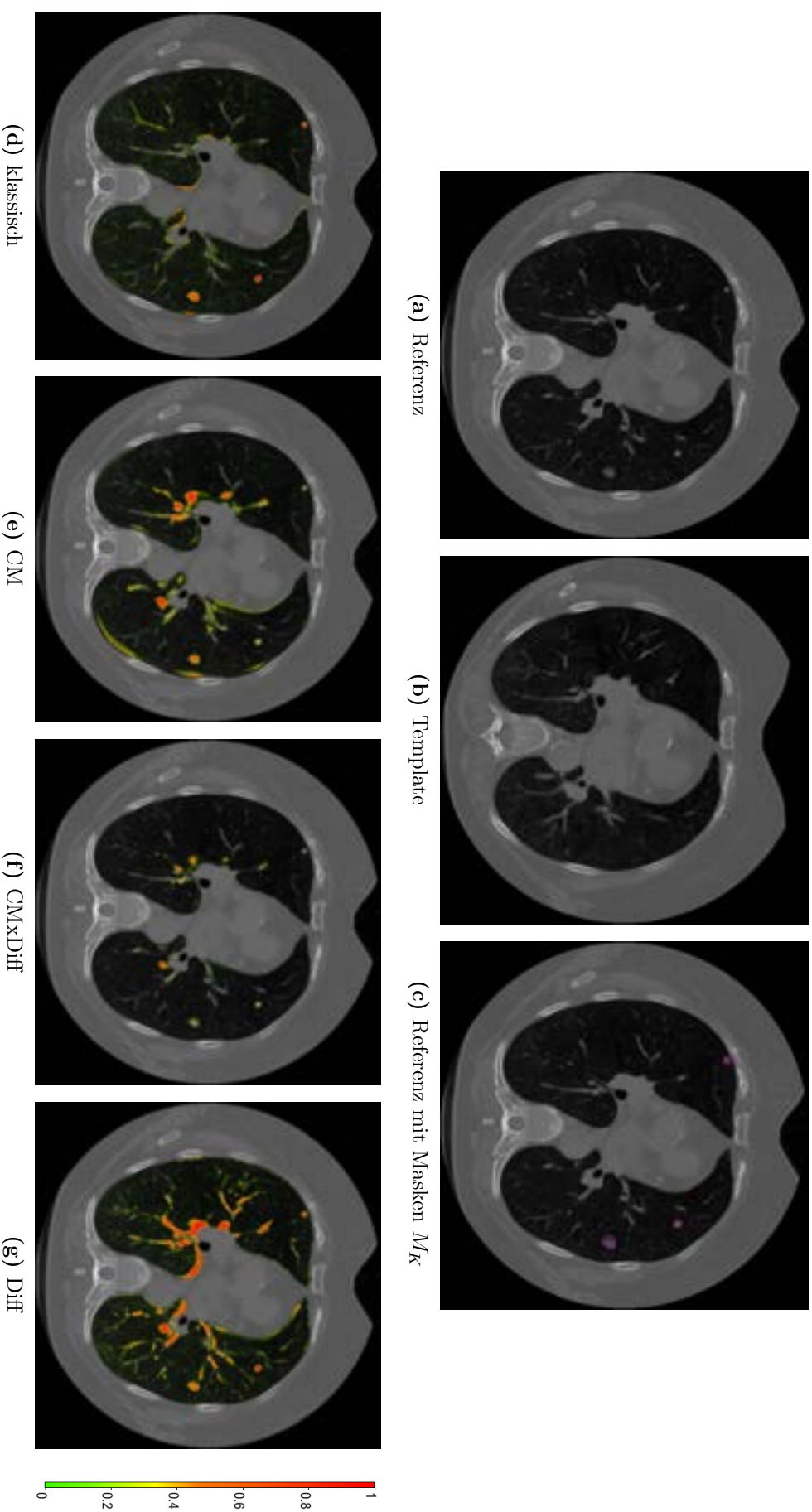
**Abb. A.7:** Visualisierung einer Schicht eines Bildpaares, die eine N1ST-Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die Masken  $M_{DL}$  unterscheiden sich jedoch. In der angezeigten Schicht ist nur die Segmentierungsmaske der Läsion in Referenzbild abgebildet, während die des deformierten Templatebildes erst in späteren Schichten auftritt. In den Veränderungskarten von CM, Diff und CMxDiff ist der Bereich der Läsion markiert. Zudem sind Bereiche des Bronchialbaums markiert. Die vollständigen Veränderungskarten sind unter A.8 zu finden.



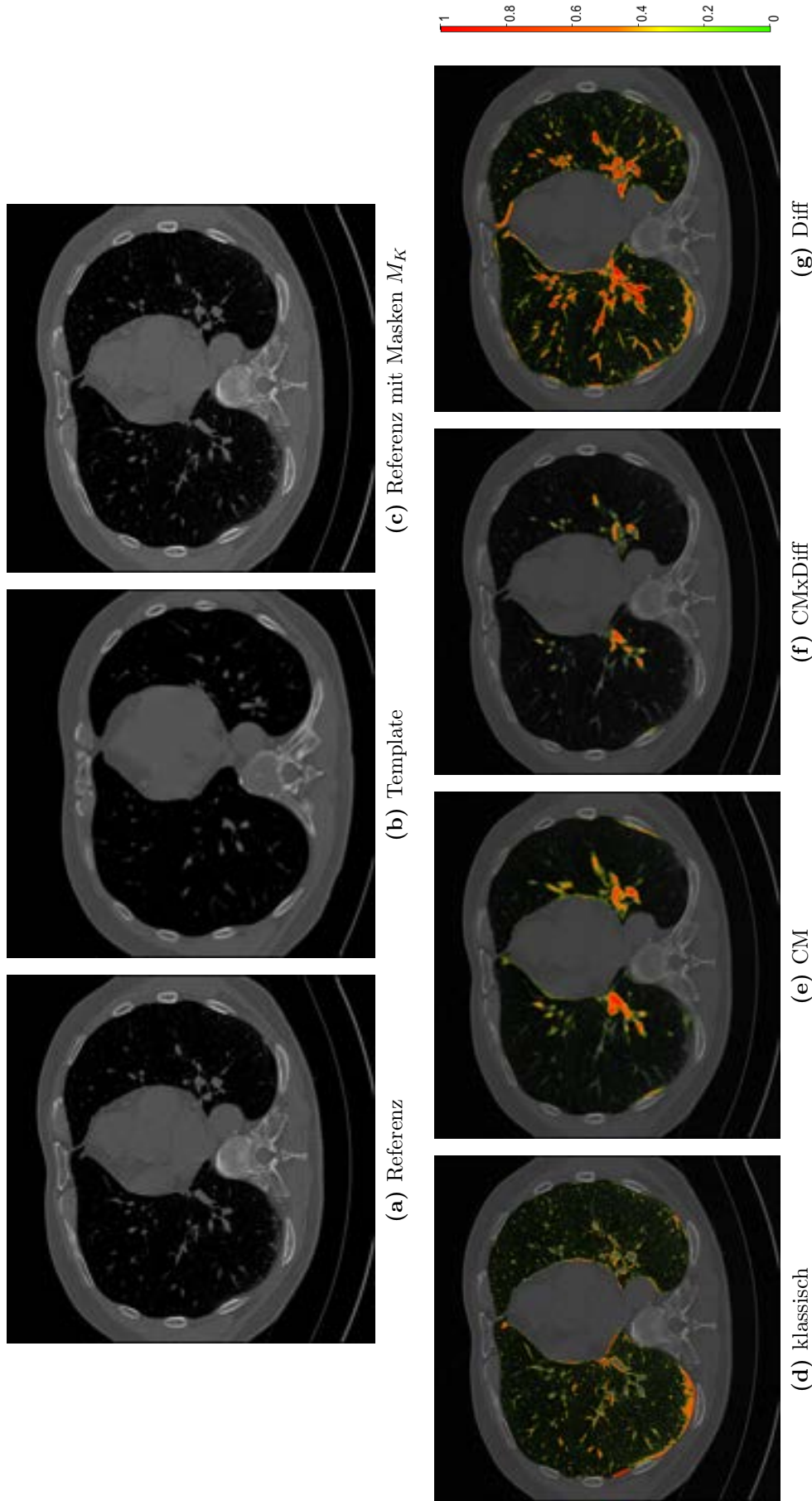
**Abb. A.8:** Visualisierung einer Schicht eines Bildpaares, die zwei synthetisch eingefügte Läsion beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihen werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die beiden Läsionen treten nur im Referenzbild auf, weshalb eine vollständige Markierung dieser zu erwarten ist. In klassisch werden beide Läsionen gut hervorgehoben und kaum andere Veränderungen markiert. Auch Diff markiert die Veränderungen eindeutig. Jedoch sind auch Teile des Bronchialbaums markiert, weshalb die Veränderungen weniger auffallen. Das Gleiche ist bei CM zu beobachten. Zudem wird die kleinere der beiden Läsionen schwächer farblich markiert. Dadurch ist diese ebenfalls bei CMxDiff schwächer markiert. CMxDiff weist zu weniger Markierungen am Bronchialbaum auf.



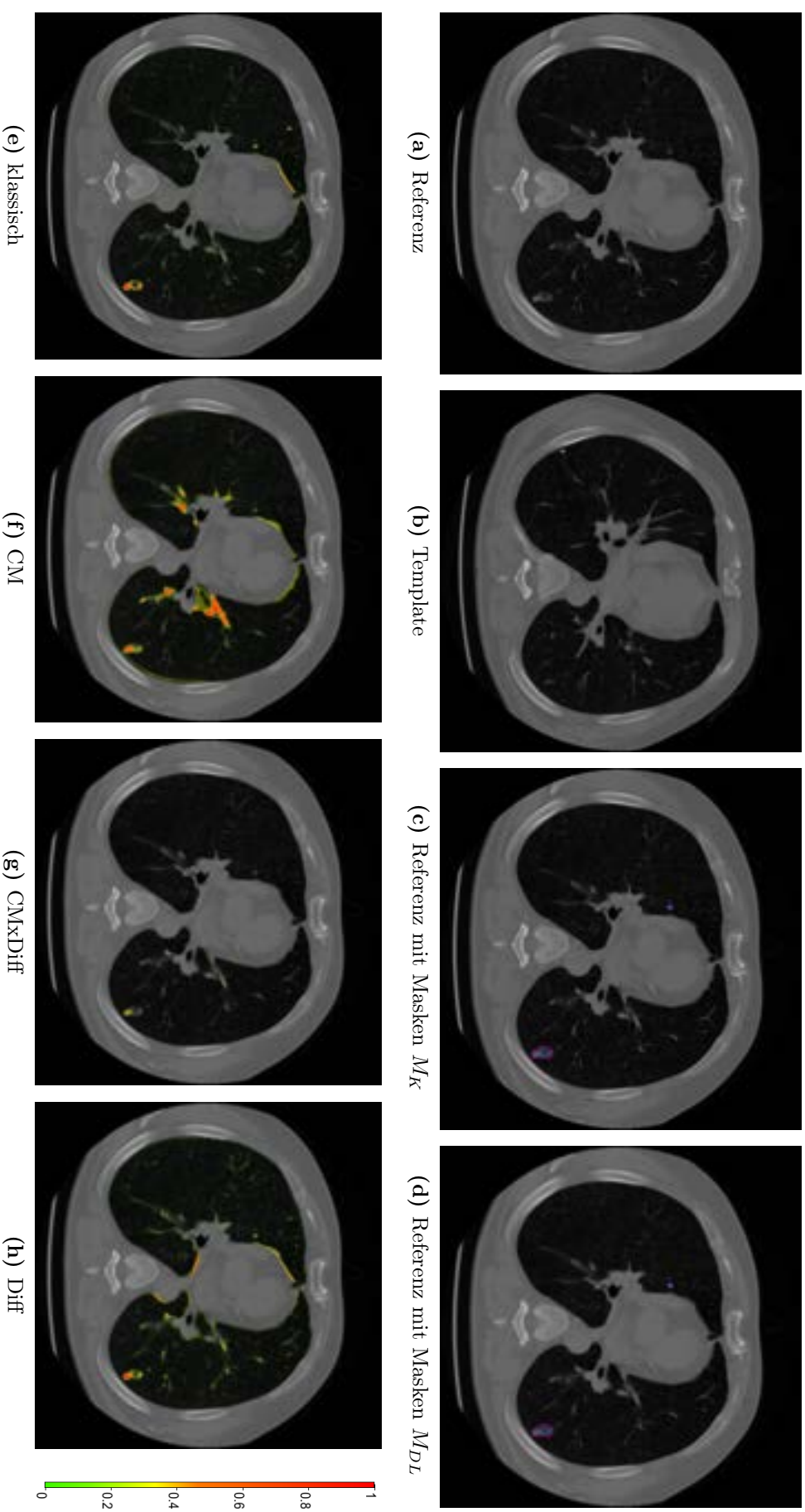
**Abb. A.9:** Visualisierung einer Schicht eines Bildpaares, die zwei bzw. drei synthetisch eingefügte Läsionen beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}(d)$  in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die Läsionen treten nur im Templatebild auf, weshalb sie in der Veränderungskarte vollständig markiert werden müssten. Bei klassisch sind die Veränderungen gut zu erkennen und es wurden wenig andere Veränderungen markiert. Auch bei Diff werden beide gut markiert, es werden jedoch auch vermehrt andere Veränderungen markiert. Bei CM ist eine Läsion im rechten Lungenflügel (links im Bild) sehr gut dargestellt. Die Läsion im linken Lungenflügel ist schlechter zu erkennen. Zusätzlich wurde im linken Lungenflügel eine andere Läsion stark hervorgehoben. Diese ist nur bei den Deep Learning-basierten Veränderungskarten zu sehen, da sie bei klassisch in einer anderen Schicht liegen. Auch bei Diff ist diese Veränderung markiert, jedoch deutlich geringer. Zusätzlich ist in allen vier Veränderungskarten mittig im linken Lungenflügel eine Veränderung markiert.



**Abb. A.10:** Visualisierung einer Schicht eines Bildpaares, die drei synthetische eingefügte Läsionen beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihen werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Alle 3 Läsionen kommen nur im Referenzbild vor, weshalb sie vollständig markiert werden müssten. Für klassisch sind alle sehr gut hervorgehoben und es sind wenige andere Veränderungen markiert. Auch bei Diff sind sie farblich gut markiert, gehen doch durch die anderen Markierungen von Veränderungen am Bronchialbaum unter. Bei CM sind im Vergleich zu Diff weniger und im Vergleich zu klassisch mehr Veränderungen markiert. Alle drei Läsionen sind markiert, aber es fällt auf, dass je kleiner die Läsion, desto geringer der Wert und folglich schwächer die Farbe.



**Abb. A.11:** Visualisierung einer Schicht eines Bildpaares, die keine Läsion, aber eine andere Veränderung beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_K$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihen werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Es gibt keine bekannten Läsionen in dieser Schicht, jedoch ist eine andere Veränderung im Referenzbild zu erkennen. Die Veränderung wurde in den Veränderungskarten von klassisch und Diff farblich stark hervorgehoben. Jedoch sind auch Veränderungen am Bronchialbaum markiert, sodass die Markierung der Veränderung weniger auffällt. Bei CM sind insgesamt weniger Markierungen. Im Bereich der genannten Veränderung ist eine leichte Markierung zu erkennen, jedoch nicht deutlich, wie bei den anderen beiden Methoden. Auch bei CMxDiff ist diese leicht markiert. Gleichzeitig sind weniger Veränderungen am Bronchialbaum markiert.



**Abb. A.12:** Visualisierung eines Ausschnitts einer Schicht eines Bildpaares, die eine synthetisch eingefügte Läsion des LIDC-Datensatzes beinhaltet. Dafür wird das Templatebild (b) sowie das Referenzbild (a) und das Referenzbild mit der Darstellung der Segmentierungsmasken der Läsionen aus der klassischen Registrierung  $M_R$  (c) und der Deep Learning-basierten Registrierung  $M_{DL}$  (d) in der oberen Reihe dargestellt. Dabei ist die Segmentierungsmaske der Läsionen im Referenzbild in lila und die Segmentierungsmaske der Läsionen des deformierten Templatebildes in blau dargestellt. In der unteren Reihe werden die Veränderungskarten der verschiedenen Methoden als Überlagerung auf dem Referenzbild, sowie die Legende für die Farbgebung angezeigt. Die synthetische Läsion hat im Referenz- und Templatebild eine unterschiedliche Größe. Für die sequenziellen Ansätze klassisch und Diff lässt sich dieser Unterschied erkennen, während in CM die Veränderungskarte einen glatteren Farbverlauf hat und die vollständige Läsion des Referenzbildes markiert ist. Die vollständigen Veränderungskarten sind unter A.4 zu finden.