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

Temke Kohlbrandt[1], Jan Moltz[1], Stefan Heldmann[1], Alessa Hering[1,3], Jan Lellmann[2]

[1]Fraunhofer Institute for Digital Medicine MEVIS, Bremen/Lübeck, Germany
[2]Institute of Mathematics and Image Computing, University of Lübeck, Germany
[3]Department of Imaging, Radboud University Medical Center, Nijmegen, The Netherlands
temke.kohlbrandt@mevis.fraunhofer.de

Abstract. Intuitive visualization of relevant changes between radiological image pairs in the form of change maps has the potential to not only increase efficiency in diagnostic reading, but also to decrease the number of missed abnormalities. Classically, change maps are created from difference images after an image registration step, which requires a careful balance in order to neither generate artifacts nor disguise relevant changes. We propose jointly learning registration and change map in order to address these limitations. As a proof of concept, the method was tested on NLST lung CT images and synthetically generated data, and shows comparable results to the conventional approach. In a reader study, the use of change maps resulted in a 23% reduction in reading time while maintaining similar recall.

1 Introduction

Medical imaging such as CT and MRI is important in the management of almost all cancer types. Imaging is used from detection, and diagnosis, to treatment planning, and in long-term follow-up. For metastatic cancer, imaging is particularly important because the sites where metastases have grown must be found on the CT scans and followed up during treatment to assess therapeutic success. At the same time, the scan must be examined for other additional abnormalities, so-called incidental findings.

The raising number of cancer cases results in a growing caseload for radiologists. An additional challenge lies in the 'inattentional blindness', when radiologists miss changes or unusual signs because they are concentrating on a specific task [1]. To reduce these problems, methods are being developed to assist radiologists; for example, segmentation and detection of anatomical structures. Image registration can assist radiologists in finding corresponding points in image pairs or in visualizing changes by subtracting the registered images.

Sieren et al. show that the use of difference images generated by a conventional image registration method as an overlay on the original CT images facilitates the evaluation for radiologists by reducing reading time and increasing sensitivity [2]. One problem with creating change maps with a conventional image registration pipeline is overcompensation by the registration, which leads to underestimation of the changes of interest, such as lesion changes. To overcome this limitation, Dufrese et al. [3] jointly estimates deformable registration and lesion change by assigning a weight to the data term in the registration minimization problem based on the changes in brain MR images.

A. Maier et al. (Hrsg.), *Bildverarbeitung für die Medizin 2024*,
Informatik aktuell, https://doi.org/10.1007/978-3-658-44037-4_15

In non-medical contexts, change maps are also of relevance. In [4], binary change maps from previously registered optical aerial images were learned using a siamese neural network. In recent years, an increasing number of methods based on deep learning have been developed for image registration [5]. The advantage of deep-learning-based image registration is the capability to incorporate additional information and tasks, such as joint learning of registration and segmentation [6].

In this work, we introduce a method that jointly learns to establish spatial correspondences and detect diagnostically relevant changes between longitudinal images to create continuous change maps. The aim is to differentiate between natural variations and those of clinical relevance. For comparison, we generate change maps using a conventional approach and test both methods on lung CT images.

2 Materials and methods

2.1 Creating change maps

A straightforward approach to generating change maps involves a sequential method. We use a variational registration approach specialized for lung CT images [7] (Fig. 1 (left)). For each pair of reference and template image $R, T : \mathbb{R}^3 \rightarrow \mathbb{R}$ and the associated lung segmentation masks ($M : \mathbb{R}^3 \rightarrow \{0, 1\}$) and keypoints ($K \in \mathbb{R}^{n,3}$), the cost function

$$\begin{aligned} \mathcal{I}(y) :=& \beta_0 D^{\mathrm{NGF}}(R, T(y)) + \beta_1 D^{\mathrm{key}}(y(K^R), K^T) + \beta_2 D^{\mathrm{mask}}(M^R, M^T(y)) + \\ & \beta_3 R^{\mathrm{curv}}(y) \end{aligned} \tag{1}$$

is minimized individually to find an optimal deformation $y : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. In addition to the normalized gradient field (D^{NGF}) as distance measure, the squared distance of the lung segmentation masks (D^{mask}) and corresponding keypoints, determined by the Förstner operator [7], (D^{key}) are used. The weights β_i were determined empirically using a partial data set. The registration is performed in a coarse-to-fine approach over 3 levels. The conventional change map $\mathrm{CM}_{\mathrm{seq}}$ is obtained by computing the difference image between R and $T(y)$, and normalizing the values to $[0, 1]$.

We propose a joint deep-learning-based method that produces a deformation field and change as its output (Fig. 1 right). We use a U-Net architecture with one head each for the deformation field y (3 channels) and the learned change C (2 channels, 1-hot

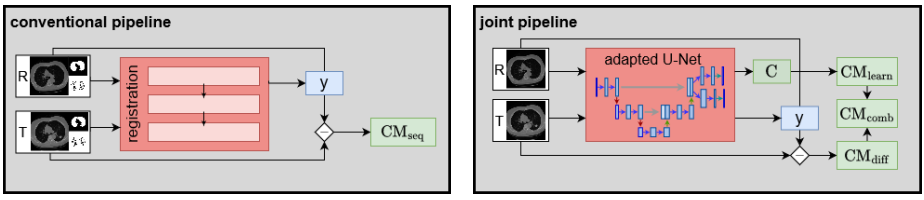


Fig. 1. The conventional sequential registration pipeline *(left)* takes images and their corresponding lung masks and keypoints as input and outputs the deformation y. The joint registration pipeline *(right)* takes only images as input and outputs the change C besides the deformation.

encoded). To ensure that both outputs are well-correlated, the feature map is only split into two separate convolutions in the final layer. Our U-Net is limited to a depth of 2 and an input resolution of $256 \times 256 \times 128$ by the available amount of VRAM (12GB).

To learn the registration with respect to the anatomical conditions of the lung, we employ the same distance measures and regularizers as in Equation 1 using the lung segmentation mask and keypoints. In addition to the curvature regularizer R^{curv}, we use a volume regularizer R^{vol} [8] in the area of the lesions to reduce overcompensation. For this purpose, segmentation masks of the lesions ($L : \mathbb{R}^3 \rightarrow \{0, 1\}$) are used, which are required only for training and not for inference.

For learning the change we use the recall loss L^{recall} from [9]. This loss function reweights the cross-entropy for each epoch, aiming to achieve a balanced precision-recall ratio, particularly in imbalanced datasets. The complete loss function for training is

$$\begin{aligned}\mathcal{L}(y, C) :=& \alpha_0 D^{\text{NGF}}(R, T(y)) + \alpha_1 R^{\text{curv}}(y) + \alpha_2 D^{\text{key}}(y(K^R), K^T) + \\ & \alpha_3 D^{\text{mask}}(M^R, M^T(y)) + \alpha_4 R^{vol}(L^R, L^T(y)) + \alpha_5 L^{\text{recall}}(C, |L^R - L^T(y)|)\end{aligned} \tag{2}$$

The weights α_i were determined empirically using a partial data set. While training relies on the additional inputs K, M, and L, inference, only requires the R and T.

With the outputs of the network, we create 3 different change maps. The learned change map $\text{CM}_{\text{learn}}$is obtained from the change output of the network after a softmax operation. Additionally, we create a change map CM_{diff} from the learned deformation y in the same way as in the conventional sequential approach. To highlight strong changes marked by both change maps, we created a third change map CM_{comb} by pointwise multiplication of the two change maps.

2.2 Data and training

For evaluation, we used the national lung screening trial (NLST) dataset [10] and selected data sets from patients who have imaging data at two time points and at least one lesion. We utilized 250 image pairs for which we also have information regarding the center of gravity and axial diameter to estimate the volume of the lesion. To increase the variability of our data, we generated synthetic images with lesions based on the lung image database consortium (LIDC) dataset [11], which contains images with lesions and associated segmentation masks. Based on the masks, lesions were extractedand

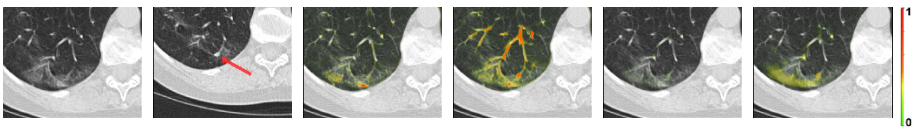


Fig. 2. Exemplary change map visualizations: Sub-images from the reference (a) and template image with an arrow to illustrate the location of the lesion (b). The different change maps are represented as color overlays on the reference image and the color legend: conventional (c), difference (d), combined (e), learned (f).

then inserted into a NLST image, followed by smoothing the boundaries. To model volume changes, we changed the size of the extracted lesions through erosion and inserted them at the corresponding location in the second image of the image pair.

In the conventional image registration pipeline, the images have an average resolution of $512 \times 512 \times 167$ with a voxel size of $0.66 \times 0.66 \times 2$mm.We trained for 50 epochs with a 160/40 pairs split for training and validation; 50 pairs were reserved for evaluation.

2.3 Experiments and evaluation

For evaluation, we binarized the change map using different thresholds on a logarithmic scale, discarded regions with a size of less than 5 voxels, and counted each remaining region as a true positive if it had an overlap with the ground truth region of a lesion. For each threshold, we computed the recall for the synthetic lesions. To determine what size of changes are detected, we created the change maps for each method with the threshold 0.68 and considered the proportion of detected synthetic lesions with respect to the different diameter differences of the lesions. For this purpose, the differences were sorted by magnitude and divided into 10 groups of equal size.

In order to quantify the impact on diagnostic evaluation, we conducted a reader study. Each reader was presented with the same 5 reference images and a rigidly pre-registered template image, and asked to find lung lesions. The study was split into three parts separated by one week. In the first part, no further assistance was given. In the later parts, readers were additionally provided a color-coded overlay of the conventional (second part) and combined learned change map (third part). The reader study was evaluated in terms of reading time as well as precision and recall. The Wilcoxon rank-sum test ($\alpha = 0.05$) was used to examine if there were significant differences between the study parts. We refer to [12] for a more comprehensive discussion of the study.

3 Results

Change maps were created for 50 pairs of images, containing 281 lesions. An example of the colored change maps as an overlay in the reference image is given in Figure 2. In addition, the reference and template image are shown. In CM_{diff} (Fig. 2d), more alterations are delineated due to the less precise registration compared to the conventional approach (Fig. 2c). In this example, a change in lung tissue occurs in addition to a lesion. Both were marked in the difference images as well as in the learned change map CM_{learn} (Fig. 2f). This indicates that other changes besides the known lesions are also learned

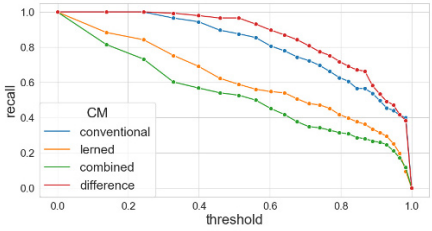


Fig. 3. Recall curves for the change maps (CM) of the different methods across synthetic lesions of all images. The continuous change maps were transformed into binary change maps using various threshold values.

Tab. 1. Mean [minimum; maximum] reading time for the five images in the 3 study parts.

	Reader 1	Reader 2	Reader 3
w/o CM	235.6 s [137.0; 329.5]	348.0 s [298.8; 450.4]	422.5 s [324.3; 617.7]
CM_{seq}	173.6 s [112.6; 267.5]	268.3 s [175.9; 360.5]	331.8 s [226.3; 392.5]
CM_{comb}	258.6 s [221.2; 293.6]	215.3 s [149.5; 279.4]	286.7 s [248.2; 305.8]

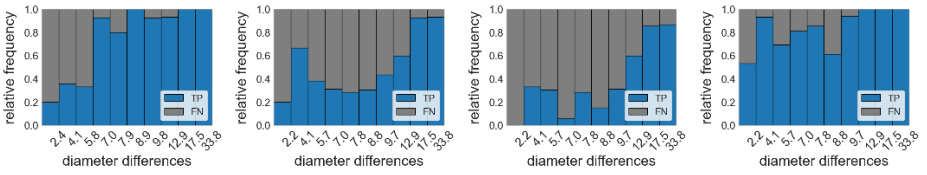
by our approach. Figure 3 shows the recall curves across the 141 synthetic lesions for all four change maps. The recall is higher for the change maps created through a difference image (CM_{seq},CM_{diff}).

When considering the proportion of detected lesions for varying diameter differences, it can be observed that larger changes in lesions can be detected more effectively (Fig 4). For diameter changes smaller than 5.8 mm, the learned methods (CM_{learn},CM_{diff}) exhibit a higher detection rate of lesions. The average diameter difference for synthetic lesions is 8.99 mm, differing compared to NLST (2.11 mm) and overall lesions (5.71 mm).

For the reader study, overlays were provided of the conventional change map (second part) and the learned change map (third part). Three participants, including a radiologist with 6 years of experience (Reader 1) and two individuals with expertise in medical image processing (Reader 2 and 3), took part in the study. No significant changes in case of mean recall of lesions marked by readers were observed between the study parts (0.29/0.25/0.23). However, a improvement in reading time was observed for Reader 2 and 3 between the first and third parts of the study (each p=0.0079). With each part of the study, Reader 2 and 3 became faster, and the range of reading time decreased across all three readers (Tab. 1). The reading time of all readers decreased by more than 23%.

4 Discussion

In this paper, we introduce an innovative methodology for the generation of change maps, by simultaneous learning deformation fields and change detection. Despite the lower resolution of the images used in the new approach, the results were comparable to the conventional approach. Both methods face challenges in detecting small changes, which could be attributed to the high voxel size and smoothing. However, with our new



(a) conventional **(b)** learned **(c)** combined **(d)** difference

Fig. 4. The proportion of detected (TP) and undetected (FN) synthetic lesions with respect to the occurring diameter differences. The boundaries for the 10 groups of diameter differences are indicated on the x-axis.

approach, smaller lesions could be detected more effectively. Additionally, it has been demonstrated that our network learns other changes besides the known lesions. Due to the small number of cases in the reader study, no significant changes in recall can be observed. But we observed an improvement in reading time while maintaining the same precision and recall values. Therefore it should be taken into account that getting accustomed to the task can also affect the processing time. Since slight improvements were achieved with our new approach, it can be assumed that, as shown in [2], significant improvements by using a change map can be demonstrated with a larger number of cases.

In the next step, the network will be expanded with a coarse-to-fine strategy in order to improve the registration, and semi-/self-supervised learning methods, such as the contrastive loss, will be applied to learn various changes even without existing segmentation. This study focuses on CT images of the lungs and changes in the lung region and will be extended to other organs. Overall, our proof of concept demonstrated that joint learning of a change map and a deformation field is possible and can provide comparable results to conventional registration in application.

References

1. Drew T, Võ MLH, Wolfe JM. The invisible gorilla strikes again: sustained inattentional blindness in expert observers. Psychol Sci. 2013;24(9):1848–53.
2. Sieren M, Brenne F, Hering A, Kienapfel H, Gebauer N, Oechtering T et al. Rapid study assessment in follow-up whole-body computed tomography in patients with multiple myeloma using a dedicated bone subtraction software. Eur Radiol. 2020;30(6):3198–209.
3. Dufresne E, Fortun D, Kumar B, Kremer S, Noblet V. Joint registration and change detection in longitudinal brain MRI. Proc IEEE ISBI. IEEE. 2020:104–8.
4. Wang Z, Peng C, Zhang Y, Wang N, Luo L. Fully convolutional siamese networks based change detection for optical aerial images with focal contrastive loss. Neurocomputing. 2021;457:155–67.
5. Hering A, Hansen L, Mok TC, Chung AC, Siebert H, Häger S et al. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Trans Med Imaging. 2022;42(3):697–712.
6. Li B, Niessen WJ, Klein S, Groot Md, Ikram MA, Vernooij MW et al. A hybrid deep learning framework for integrated segmentation and registration: evaluation on longitudinal white matter tract changes. Proc MICCAI. Springer. 2019:645–53.
7. Rühaak J, Polzin T, Heldmann S, Simpson IJ, Handels H, Modersitzki J et al. Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. IEEE Trans Med Imaging. 2017;36(8):1746–57.
8. Hering A, Häger S, Moltz J, Lessmann N, Heldmann S, Ginneken B van. CNN-based lung CT registration with multiple anatomical constraints. Med Image Anal. 2021;72:102139.
9. Tian J, Mithun NC, Seymour Z, Chiu HP, Kira Z. Striking the right balance: recall loss for semantic segmentation. Proc IEEE ICRA. 2022:5063–9.
10. Team NLSTR. Data from the national lung screening trial (NLST) [Data set]. 2013.
11. Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF. Data From LIDC-IDRI. 2015.
12. Kohlbrandt T. Gemeinsames Lernen von Bildregistrierung und Veränderungskarten für Lungen-CT-Bilder. Masters Thesis, University of Lübeck. 2023.