Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search

Mattias P. Heinrich and Lasse Hansen

Institute of Medical Informatics, Universität zu Lübeck
heinrich@imi.uni-luebeck.de

Abstract. Learning-based registration, in particular unsupervised approaches that use a deep network to predict a displacement field that minimise a conventional similarity metric, has gained huge interest within the last two years. It has, however, not yet reached the high accuracy of specialised conventional algorithms for estimating large 3D deformations. Employing a dense set of discrete displacements (in a so-called correlation layer) has shown great success in learning 2D optical flow estimation, cf. FlowNet and PWC-Net, but comes at excessive memory requirements when extended to 3D medical registration. We propose a highly accurate unsupervised learning framework for 3D abdominal CT registration that uses a discrete displacement layer and a contrast-invariant metric (MIND descriptors) that is evaluated in a probabilistic fashion. We realise a substantial reduction in memory and computational demand by iteratively subdividing the 3D search space into orthogonal planes. In our experimental validation on inter-subject deformable 3D registration, we demonstrate substantial improvements in accuracy (at least ≈10% points Dice) compared to widely used conventional methods (ANTs SyN, NiftyReg, IRTK) and state-of-the-art U-Net based learning methods (VoxelMorph). We reduce the search space 5-fold, speed-up the run-time twice and are on-par in terms of accuracy with a fully 3D discrete network.

Keywords: Deformable Registration · Deep Learning · Discrete Optimisation.

1 Motivation and Related Work

Medical image registration aims to align two or more 3D volumes of different patients, time-points or modalities. In many practical cases the transformation is highly deformable, which poses a complex task of regressing a continuous displacement field. Conventional registration methods aim to capture larger deformations with multi-resolution strategies iterative warping, discrete displacement search or a combination of these. Many clinical tasks that could benefit from automatic and accurate image registration are time-sensitive making many conventional algorithms inapt for practical use. Deep learning based image registration (DLIR) has the promise to reduce computation times from minutes to

sub-seconds as already realised in the field of image segmentation (exemplified by the wide adaptation of the U-Net [21]). In particular, unsupervised methods [2] that minimise a classical cost function are of interest, because they can reduce inference run-times without relying on extensive manual annotations of e.g. corresponding landmarks. However, up to date no boost in accuracy has been achieved in DLIR and its application remains restricted to less complex tasks. For intra-patient lung 4D-CT registration conventional approaches yield highly accurate motion estimation [22] with target registration errors of less than 1 mm (on the DIRLAB dataset), while reaching sub-second computation time when parallelised on GPU hardware [4], whereas state-of-the-art DLIR methods still exhibit large residual errors of $\approx$ 3.5mm [1,2,9]. This is most likely due to the limited or at least ineffective capture range of multi-scale feed forward architectures (U-Nets) that were designed for segmentation. Multi-resolution strategies [15], multi-stage networks [28] or progressive hierarchical training [8] can alleviate the limitations to some degree and reduce errors for 4DCT to $\approx$ 2.5 mm.

Discrete displacements: Exploring multiple potential displacements that are quantised as labels in MRF-based discrete optimisation has seen great success in image registration with large deformations, both for 2D images and medical volumes [10,12,18,24]. This idea is also reflected in the top-performing 2D DLIR methods for optical flow, among others FlowNet [7] and PWC-Net [26]. Instead of directly regressing a continuous displacement value, discrete methods quantise the range of expected displacements and estimate probabilities (or costs) for a (spatially regularised) label assignment. Therefore, discrete methods achieve remarkable results even when they are employed without iterative or multi-resolution strategies. Learning probabilistic and uncertainty-aware registration models has been studied in [6,17] yet in a continuous variational setting very different to discrete probabilistic methods.

In [11] a discrete probabilistic dense displacement network was presented that was trained with label supervision and substantially exceeded the accuracy of continuous (U-Net based) DLIR methods for 3D CT registration. It used a large number of control points and a densely quantised 3D displacement space. This approach comes at the cost of substantial memory use, since each feature tensor within the network requires >1200 MByte of GPU memory per channel - in our experiments it required $\approx$22 GByte GPU RAM (which is only available in few high-end cards). The authors, hence, restricted the network to few trainable weights, single channel features within their mean-field inference and made extensive use of checkpointing. We expect that a fully sampled 3D displacement space is not necessary to reach state-of-the-art results and hence a reduction in labels and subsequently memory will lead to wider adoption of discrete DLIR.

Subdivision of label spaces: Exploding numbers of parameters have been an issue in many conventional MRF-based registration methods and strategies to reduce the computational burden include: simplified graph models (e.g. minimum-spanning trees [27]), fewer nodes in the transformation model [10] and reduction from the dense 2D displacement space to a sparse setting where only displacements along the orthogonal 1D axes are sampled [24]. This subdivision

of displacement labels, which enables a decoupled MRF-regularisation, was also used in the popular SIFT flow approach [18] and the 3D medical MRF registration of [10]. However, as shown in [12] this very sparse sampling of a 3D space leads to a significant reduction of registration accuracy. Applying CNNs to 1D signals is also expected to limit their ability to learn meaningful patterns within the displacement space. Here, we propose an intermediate strategy that decomposes the 3D space into three orthogonal 2D planes and thereby more accurately approximates the full space. A related strategy was considered for linear 3D registration in a discrete setting [32], where 2D sub-spaces of a 12 parameter affine transform where optimised using an MRF.

2.5D approaches in deep learning: Several recent segmentation and classification networks for medical volumes consider multiple 2D views of the 3D input data. In [25] a 2.5D input view is created by extracting orthogonal planes of the 3D patches and representing them as RGB colour channels for a 2D CNN, which enabled transfer learning from ImageNet pretrained models. Multiview fusion of multiple 2D CNN classifiers was proven effective for pulmonary nodule classification of 3D CT in [23]. Large 2.5D planes were used to extract context features for self-supervised learning in [3]. Many fully-convolutional segmentation models decouple the prediction for axial, coronal and sagittal planes and fuse the resulting scores or directly employ 3D separable convolutions [5]. In learning-based image registration, [16] decouples the motion prediction in three orthogonal spatial planes and fuses the continuous regression values and achieves comparable accuracy to 3D U-Nets for weakly-supervised DLIR. This is fundamentally different to our idea to decompose the 3D displacement space in a discrete registration setting.

Contributions: 1) We are the first to propose an unsupervised discrete deep learning framework for 3D medical image registration that leverages probabilistic predictions to improve the guidance of the metric loss. **2)** Furthermore, we propose a highly efficient 2.5D approximation of the quantised 3D displacement space that substantially reduces the memory burden for training and computational complexity for inference of discrete DLIR. **3)** We demonstrate that the high accuracy of a full 3D search space can be matched using two iterations of sampling three orthogonal 2D displacement maps in combination with on-the-fly instance optimisation. **4)** The method is evaluated on a challenging public inter-subject abdominal CT dataset and the source code is already released anonymously.

2 Methods

3D pdd-net: We build our efficient 2.5D discrete registration framework upon the 3D pdd-net (probabilistic dense displacement network) proposed in [11], which is briefly summarised below. A pair of fixed and moving scans I_F and I_M, for which we seek a spatial alignment φ, are fed through a feature extractor that uses deformable convolutions [13] and outputs a 24-channel feature map (4D tensor) with a stride of 3 voxels. Next, a B-spline transformation model

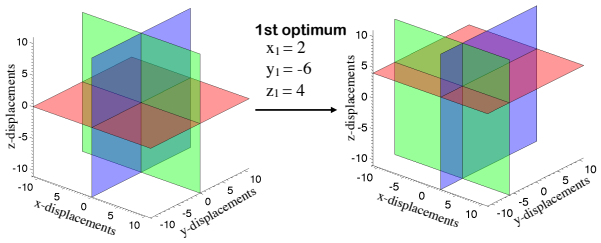

Fig. 1. Concept of 2.5D decomposition of discrete 3D displacement space. To reduce the computational complexity, displacements are only sampled along three orthogonal planes (reducing e.g. the number of quantised position by $5\times$ when using $|\mathcal{L}_{3D}| = 15^3 = 3375$. To compensate for the approximation error, a second pass of the same regularisation network (without image warping or additional feature computation) is performed that roots the planes at the previous optima.

with a set of $|K| \in \mathbb{R}^3$ control points on a coarser grid and a quantised 3D displacement space with linear spacing $\mathcal{L} = q \cdot \{-1, -\frac{6}{7}, -\frac{5}{7}, \ldots, +\frac{5}{7}, +\frac{6}{7}, +1\}^3$ are defined. Here q is a scalar that defines the capture range and the cardinalities are $|K| = 29^3 = 24389$ and $|\mathcal{L}| = 15^3 = 3375$. A correlation layer (cf. [7]) without any trainable weights is used to compute the matching cost of a fixed scan feature vector with all discretely displaced moving feature vectors that are within the search space spanned by $\mathcal{L}$. The second part of the network comprises several 3D max- and average pooling operations (with stride=1) that act in alternation on either the three spatial or the three displacement dimensions and model two iterations of MRF-regularisation (approx. mean-field inference as found in [31]).

Decoupled 2.5D subspaces: As mentioned above these dense 3D displacement computation enable highly accurate registration, but come with large memory and computational costs. We thus approximate the dense 3D space with three 2.5D subspaces and define $\mathcal{L}_{2D} = \{\mathcal{L}_{xy}, \mathcal{L}_{xz}, \mathcal{L}_{yz}\}$ where each subspace comprises a planar grid of 2D displacements: $\mathcal{L}_{xy} = q \cdot \{-1, -\frac{6}{7}, -\frac{5}{7}, \ldots, +\frac{5}{7}, +\frac{6}{7}, +1\}^2$ and a constant value z_1 for the third dimension. This step greatly reduces the memory requirements for learning a better feature extraction in the deformable convolution part, since $|\mathcal{L}_{2D}|$ is now only $3 \cdot 15 \cdot 15 = 625$, five-fold smaller than a full 3D space. The feature dissimilarities computations within the correlation layer are reduced from 4 GFlops to 790 MFlops. The spatial smoothing remains to operate on 3D (yet with a much smaller number of channels 625 instead of 3375), while the operations that regularise the displacement dimensions are now in 2D. In order to estimate a 3D field, the output of the network is converted into three 2D pseudo-probability maps (using the softmax) for each control point (see Fig. 2). The 3D vectors φ are then found by multiplying the probabilities with the displacement mesh-grid and averaging between the two non-zero elements of all three intersecting maps. A diffusion regularisation, $\lambda \cdot (|\nabla\varphi_1|^2 + |\nabla\varphi_2|^2 + |\nabla\varphi_3|^3)$, is added to promote plausible deformations.

Two-step instance optimisation with gradient descent: The approximation accuracy of three 2D planes of the full 3D displacement space depends on the closeness of their intersection points to the true optimum. A single pass of the regularisation part of the network on the initial 2.5D subset might yield inaccuracies in both training and inference. Therefore, a two-step approach is proposed that to alleviate limited capabilities of a feed-forward network to find the optimal compromise between metric loss and diffusion regularisation.Thus an iterative on-the-fly instance optimisation on the intermediate 2.5D displacement probabilities is performed similar to [2] and [11]. For this purpose a continuous 3D B-spline transformation model is considered and optimised for an improved $\varphi^* := (x + \Delta x^*, y + \Delta y^*, z + \Delta z^*)$ per instance (test registration pair) using Adam. We start with the feed-forward predicted discrete cost tensor $\mathcal{C} \in H \times W \times D \times 15 \times 15 \times 3$, where the last dimensions describes the three subplanes (2.5D) of a full 3D displacement search region. We minimise the following loss function: $L_{\text{instance}} = L_{xy} + L_{xz} + L_{yz} + L_{\text{diff.-reg}}$. For each sub-dimension we define $L_{xy} = \mathcal{C}(x, y, z, \Delta x^*, \Delta y^*, 0)$, where differentiable bilinear sampling is used for $\Delta x^*, \Delta y^*$. We, thus iteratively update the deformation and optimise a related cost function (the sum over the three 2D displacement metric values and a diffusion regularisation) in a continuous manner.

A similar principle to combine the complementary strengths of discrete and continuous optimisation was used in [22] and [7]. Once the instance optimisation is completed, the second pass of the discrete network with more accurate initial placement of 2.5D subplanes is performed (see Fig. 1 right for visualisation), followed again by a continuous refinement.

Non-local metric loss: The approach in [11] was restricted to weakly-supervised learning with segmentation labels. This may limit the applicability to learn from large unlabelled datasets and introduce an bias towards the chosen labels, cf. [2]. We thus introduce a novel unsupervised non-local metric loss for discrete DLIR. Due to its high performance in other CT registration tasks e.g. [30] modality independent neighbourhood descriptors (MIND) are extracted with self-similar context as proposed in [14] yielding 12-channel tensors. Instead to directly employing a warping loss as done in [2] and most related DLIR methods, we make full use of the 2.5D probabilistic prediction and compute the warped MIND vectors of the moving scan implicitly by a weighted average of the underlying features within the search region (averaged again for the three orthogonal probability maps): $\text{MIND}_{\text{warped}} = \frac{1}{3}\text{MIND}_{\text{xy}} + \frac{1}{3}\text{MIND}_{\text{xz}} + \frac{1}{3}\text{MIND}_{\text{yz}}$.

Here the probabilistic displacements are defined as $\mathcal{P}_{xy}(x, y, z, \Delta x', \Delta y') = \frac{\exp(-\alpha \mathcal{C}(x,y,z,\Delta x',\Delta y'))}{\sum_{\Delta x', \Delta y'} \exp(-\alpha \mathcal{C}(x,y,z,\Delta x',\Delta y'))}$ and the discretely warped MIND features (single channel for brevity) as: $\text{MIND}_{\text{xy}} = \sum_{\Delta x', \Delta y'} \mathcal{P}_{xy}(x, y, z, \Delta x', \Delta y') \cdot \text{MIND}(x + \Delta x', y + \Delta y', z)$, where $\Delta x', \Delta y'$ are local coordinates.

We also compare the benefits of this discrete probabilistic loss with a traditional warping loss (denoted as "w/o NL" in Table 1). The source code of our complete implementation is publicly available at [1].

[1] `https://github.com/multimodallearning/pdd2.5/`

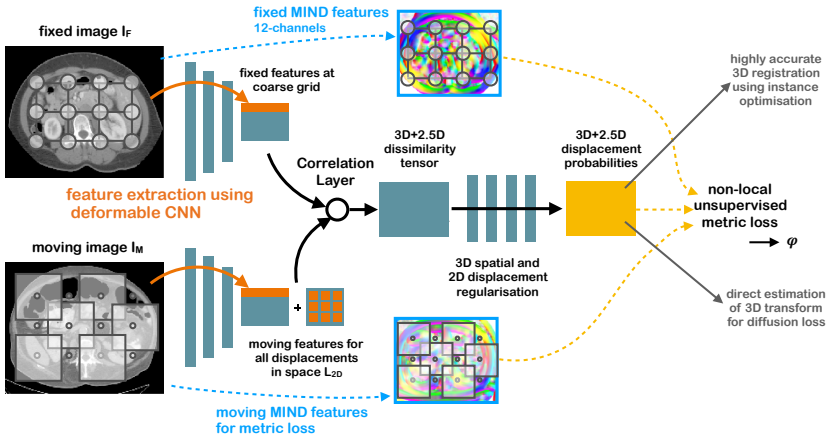

Fig. 2. Novel concept of 2.5D unsupervised dense displacement network. Deformable convolution layers firstly extract features for both fixed and moving image. Secondly, the correlation layer evaluates 2.5D displacement space for each 3D grid point yielding three 5D dissimilarity maps. These maps are spatially smoothed by filters acting on dimensions 1-3 and regularity within dimensions 4-5 (2D displacement planes) is obtained using approx. min-convolutions. The learning is supervised without any annotations using our proposed non-local MIND metric loss. The 2.5D probabilistic prediction is obtained using a softmax (over each 2D displacement plane) and either converted to continuous 3D displacements for a diffusion regularisation or further refined using instance optimisation to warp scans (or label images) with high accuracy.

3 Experiments and Results

Many state-of-the art DLIR registration methods have been evaluated on private datasets with unknown complexity. The public "beyond the cranial vault" abdominal CT dataset described in [30] us used here, which was used in a MICCAI 2015 challenge and is available for download. We pre-process the scans using the following steps: First, resampling to isotropic resolution of 2mm and automatic cropping to a similar field-of-view. Second, affine pre-registration was performed between all pairs of scans using the discrete registration tool linearBCV [2] an unbiased mean transform estimated to bring all scans into a canonical space (the run-time per registration is $\approx$ 2 sec.)[3]. Despite these reasonable efforts they are still very complex and challenging deformation to be compensated as evident from the low initial average Dice overlap of 28.1% (see Table 1).

We evaluate the registration accuracy using manual segmentations for ten scans (90 registrations) that were not used during unsupervised training and compute the Dice score for all 13 labels (see Fig. 3 and Table 1 for details). We compare the state-of-the-art unsupervised Voxelmorph network [2] and use the same MIND implementation as similarity metric that yields a Dice of 34.0%

[2] `https://github.com/mattiaspaul/deedsBCV`
[3] `https://learn2reg.grand-challenge.org/Dataset/` (Task 3)

Table 1. Quantitative evaluation using Dice overlap in % for 90 pair-wise inter-subject registrations of unseen 3D abdominal CT scans. Our proposed pdd 2.5D method compares favourable to the state-of-the-art DLIR method Voxelmorph and the conventional multiresolution, iterative approach NiftyReg. Hausdorff (95th percentile) and complexity of deformations (stddev of Jacobian determinants, smaller is better) are reported for a subset of methods. A representative, clinically relevant set of 9 out of the considered 13 anatomies are also evaluated individually: spleen ■, right kidney ■, left kidney ■, gallbladder ■, esophagus ■, liver ■, stomach ■, aorta ■ and pancreas ■.

Method	■	■	■	■	■	■	■	■	■	avg(9)	**avg(13)**	HD95	detJ	memory	infer.	time
affine pre-reg.	36	36	34	6	32	65	28	39	16	33.2	28.1±8.3	14.6		train	GPU	CPU
Voxelmorph (MIND)	49	44	44	8	32	77	34	48	19	40.0	34.0±9.2	12.8	0.66	6 GB	0.12 s	60 s
pdd 2.5D (w/o NL)	45	49	52	5	38	69	41	44	21	41.1	34.4±5.8			9 GB	0.54 s	29 s
pdd 3D (MIND)	60	65	67	12	43	81	50	56	35	52.6	44.7±4.6			22 GB	0.73 s	73 s
pdd 2.5D (MIND)	60	64	65	14	41	81	50	57	34	52.2	**44.8±4.9**	10.4	0.57	9 GB	0.54 s	29 s
NiftyReg	62	50	54	3	36	78	62	34	17	44.5	35.0					117 min

(results using the default MSE loss are ≈4% points worse and the NCC loss failed to converge). A small ablation study of variants of our proposed method is performed using either fully 3D displacement spaces (pdd 3D, 44.7%), the novel 2.5D subdivision (**pdd 2.5D**, 44.8%) and a conventional warping loss (pdd 2.5D (w/o NL), 34.4%). In addition, we include results of NiftyReg [20] from [30] with an average Dice of 35.0% that performed substantially better than IRTK and ANTs SyN (28% and 27%) and only slightly worse than deeds (49%). Note, that a larger subset of test pairs was used for these methods and their initial affine alignment was likely worse. Our networks were each trained with AdamW [19] (weight decay = 0.01, initial learning rate = 0.005 and exponential decay with $\gamma = 0.99$) for 250 epochs (1000 iterations with a mini-batch size of 4), using affine augmentation and a weighting $\lambda = 0.025$ for the diffusion regularisation loss (a higher λ was employed during a warm-up phase for 100 iterations to stabilise the training) within ≈25 minutes. The instance optimisation uses Adam with learning rate 0.02, $\lambda = 5$ and 30 iterations. It was also employed during training (without gradient tracking) to enable the sampling from 2D displacement planes that are not rooted at the origin (central voxel) and thus increasing the coverage for feature learning. We repeated the training three times and report the accuracies from the worst run (Voxelmorph was trained once for 50000 iterations requiring about 8 hours).

Clear improvements of 10% points accuracy gains are achieved with the discrete setting compared to the state-of-the-art in unsupervised DLIR . The gains over Voxelmorph are most visible for medium-sized organs (kidneys ■, ■) and highly deformable anatomies (stomach ■ and pancreas ■). The sorted Dice scores in Fig. 3 g) show that our 2.5D network compensates larger deformations (small initial Dice values) especially well. Our results without using the non-local MIND loss are similar to Voxelmorph, highlighting the fact that a meaningful probabilistic prediction is achieved. The proposed 2.5D approximation matches the quality of the full 3D search space (5× bigger) and reduces the memory use for the correlation layer and regularisation part during training from 10.2 to 1.7

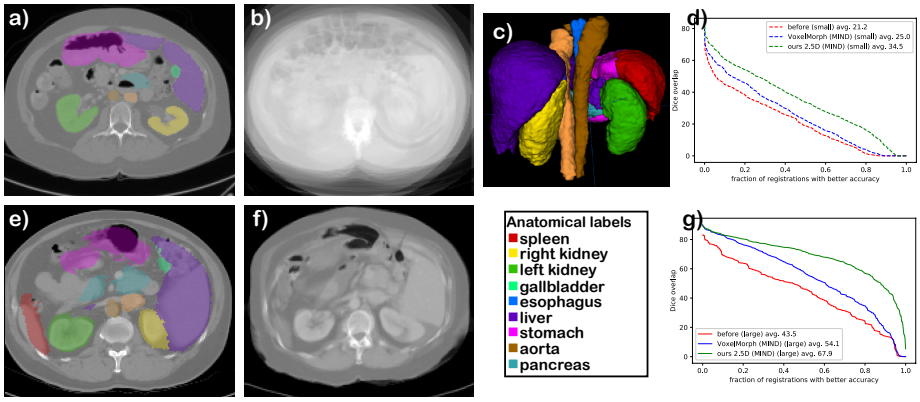

Fig. 3. Visual and numerical results for registration of unseen test scans with proposed **pdd 2.5D** method. a) random moving scan with ground truth annotation b) intensity mean of all test scans (after affine pre-registration) c) automatically propagated segmentation from training scans (stapled [29]) d) resulting Dice scores for four larger organs (spleen, kidneys, liver) e) propagated segmentation from other test scan f) intensity mean after our pair-wise registration g) resulting Dice scores for 8 smaller organs (gallbladder, esophagus, stomach, aorta, inferior and portal veins, glands)

GByte (the total memory usage is higher due to the feature extractor and non-local loss). The CPU runtime at inference is reduced by 2.5× to only 29 seconds (note, this includes the two-step instance optimisation, without the second step a CPU runtime of 17 seconds is obtained with a moderate decrease of 2% points in Dice accuracy). The complexity of the transformations estimated using the standard deviation of Jacobian determinants was 0.567 on average, 2.5% of voxels incurred a negative Jacobian (folding), indicating a reasonable smoothness for abdominal inter-subject registration with large topological differences. This can be further improved by increasing the parameters for instance optimisation to $\lambda = 15$ and 50 iterations, yielding std$(Jac) = 0.38$, 0.4% negative Jacobians and a similar Dice accuracy of 44.0%. The visual results in Fig. 3 demonstrate highly accurate alignment of unseen test images, yielding sharp intensity mean images (f) and convincing automatic 3D segmentations (c, stapled [29]).

4 Conclusion and Outlook

In summary, the novel 2.5D subdivision of displacement spaces for discrete deep learning based image registration (DLIR), in combination with a fast instance-optimisation, advances the state-of-the-art for highly complex abdominal inter-subject registration, while limiting the computational burden in comparison to 3D networks (that have to act on 6D tensors). A suitable non-local MIND metric loss is proposed that leverages the probabilistic predictions for unsupervised learning and enables fast training. This idea can further unleash its potential for

new developments in DLIR that overcome the current limitations in accuracy and enable new clinical applications of registration for image-guided interventions, radiotherapy and diagnostics. The advantages of a discrete search space and probabilistic predictions are demonstrated qualitatively and quantitatively in terms of highly accurate automatic propagation of segmentations between unseen images (with improvements of 10% points Dice over state-of-the-art), lower run times and by robustly capturing of large deformations. Future work could yield further gains of the memory-efficient 2.5D displacement space by employing more powerful regularisation networks and by incorporating contextual loss terms.

References

1. Anonymous: Tackling the problem of large deformations in deep learning based medical image registration using displacement embeddings. Medical Imaging with Deep Learning pp. 1–5 (2020), `https://openreview.net/pdf?id=kPBUZluVq`, under review
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans Med Imag (2019)
3. Blendowski, M., Nickisch, H., Heinrich, M.P.: How to learn from unlabeled volume data: Self-supervised 3d context feature learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 649–657. Springer (2019)
4. Budelmann, D., König, L., Papenberg, N., Lellmann, J.: Fully-deformable 3d image registration in two seconds. In: Proc. BVM, pp. 302–307. Springer (2019)
5. Chen, W., Liu, B., Peng, S., Sun, J., Qiao, X.: S3D-UNet: separable 3D U-Net for brain tumor segmentation. In: MICCAI Brainlesion. pp. 358–368. Springer (2018)
6. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 729–738. Springer (2018)
7. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proc. ICCV. pp. 2758–2766 (2015)
8. Eppenhof, K.A., Lafarge, M.W., Veta, M., Pluim, J.P.: Progressively trained convolutional neural networks for deformable image registration. IEEE Trans Med Imag (2019)
9. Eppenhof, K.A., Pluim, J.P.: Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans Med Imag **38**(5), 1097–1105 (2018)
10. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical image analysis **12**(6), 731–741 (2008)
11. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 50–58. Springer (2019)

12. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans Med Imag **32**(7), 1239–1248 (2013)
13. Heinrich, M.P., Oktay, O., Bouteldja, N.: OBELISK-Net: Fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions. Medical image analysis **54**, 1–9 (2019)
14. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 187–194. Springer (2013)
15. Hering, A., van Ginneken, B., Heldmann, S.: mlVIRNET: multilevel variational image registration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 257–265. Springer (2019)
16. Hering, A., Kuckertz, S., Heldmann, S., Heinrich, M.P.: Memory-efficient 2.5 d convolutional transformer networks for multi-modal deformable registration with weak label supervision applied to whole-heart CT and MRI scans. Int J Comp Assist Radiol Surgery **14**(11), 1901–1912 (2019)
17. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE Trans Med Imag **38**(9), 2165–2176 (2019)
18. Liu, C., Yuen, J., Torralba, A.: SIFT flow: Dense correspondence across scenes and its applications. IEEE Trans Patt Anal Mach Intell **33**(5), 978–994 (2010)
19. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
20. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer methods and programs in biomedicine **98**(3), 278–284 (2010)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
22. Rühaak, J., Polzin, T., Heldmann, S., Simpson, I.J., Handels, H., Modersitzki, J., Heinrich, M.P.: Estimation of large motion in lung ct by integrating regularized keypoint correspondences into dense deformable registration. IEEE Trans Med Imag **36**(8), 1746–1757 (2017)
23. Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S.J., Wille, M.M.W., Naqibullah, M., Sánchez, C.I., van Ginneken, B.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imag **35**(5), 1160–1169 (2016)
24. Shekhovtsov, A., Kovtun, I., Hlaváč, V.: Efficient MRF deformation model for non-rigid image matching. Comput Vis Image Und **112**(1), 91–99 (2008)
25. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans Med Imag **35**(5), 1285–1298 (2016)
26. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proc. CVPR. pp. 8934–8943 (2018)
27. Veksler, O.: Stereo correspondence by dynamic programming on a tree. In: Proc. CVPR. vol. 2, pp. 384–390. IEEE (2005)
28. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis **52**, 128–143 (2019)

29. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imag **23**(7), 903–921 (2004)
30. Xu, Z., Lee, C., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., Abramson, R.G., Landman, B.: Evaluation of 6 registration methods for the human abdomen on clinically acquired CT. IEEE Trans Biomed Eng **63**(8), 1563–1572 (2016)
31. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proc. ICCV. pp. 1529–1537 (2015)
32. Zikic, D., Glocker, B., Kutter, O., Groher, M., Komodakis, N., Kamen, A., Paragios, N., Navab, N.: Linear intensity-based image registration by markov random fields and discrete optimization. Medical image analysis **14**(4), 550–562 (2010)