SAME: Deformable Image Registration based on Self-supervised Anatomical Embeddings

Fengze Liu[1]†, Ke Yan[2]†, Adam Harrison[2], Dazhou Guo[2], Le Lu[2], Alan Yuille[1], Lingyun Huang[3], Guotong Xie[3], Jing Xiao[3], Xianghua Ye[4], and Dakai Jin[2]

[1] Johns Hopkins University, Baltimore, MD, USA
[2] PAII Inc., Bethesda, MD, USA
[3] Ping An Technology, ShenZhen, China
[4] The First Affiliated Hospital, Zhejiang University, Hangzhou, China

Abstract. In this work, we introduce a fast and accurate method for unsupervised 3D medical image registration. This work is built on top of a recent algorithm self-supervised anatomical embedding (SAM), which is capable of computing dense anatomical/semantic correspondences between two images at the pixel level. Our method is named SAM-enhanced registration (SAME), which breaks down image registration into three steps: affine transformation, coarse deformation, and deep deformable registration. Using SAM embeddings, we enhance these steps by finding more coherent correspondences, and providing features and a loss function with better semantic guidance. We collect a multi-phase chest computed tomography dataset with 35 annotated organs for each patient and conduct inter-subject registration for quantitative evaluation. Results show that SAME outperforms widely-used traditional registration techniques (Elastix FFD, ANTs SyN) and learning based VoxelMorph method by at least 4.7% and 2.7% in Dice scores for two separate tasks of within-contrast-phase and across-contrast-phase registration, respectively. SAME achieves the comparable performance to the best traditional registration method, DEEDS (from our evaluation), while being orders of magnitude faster (from 45 seconds to 1.2 seconds).

Keywords: Deformable Registration · Affine Registration · Unsupervised · Self-supervised Anatomical Embedding · Deep Learning.

1 Introduction

Deformable image registration is a fundamental task in medical image analysis [16]. Traditional registration methods solve an optimization problem and iteratively minimize a preset similarity measure to align a pair of images. Recently,

† equal contribution.

learning-based deformable registration, using deep networks, have been investigated [2,10,19,13,12]. Compared with their conventional counterparts, learning-based methods can incorporate more flexible losses, integrate other computing modules and are much faster in inference. VoxelMorph was a representative work [2] that learns a parameterized registration function using a convolutional neural network (CNN). Many recent methods focus on designing more sophisticated networks using pyramid [13] or cascaded structures [10,19], or connecting registration to pipelines that include synthesis and segmentation [12]. Ideally, registration should focus on aligning semantically similar/coherent voxels, e.g., the same anatomical locations. This semantic information can come in the form of extra manual annotations (e.g. organ masks) [2], but requiring prohibitive labor costs from professionals. Existing unsupervised methods instead optimize similarity measures describing local intensities as a proxy of the semantic information, such as the mean squared error (MSE) or normalized cross correlation (NCC). However, these are less reliable in settings with large deformations, complex anatomical differences, or cross-modality/cross-phase imagery.

In this paper, we exploit incorporating a novel form of semantic information in registration. Self-supervised anatomical embedding (SAM) is a recent work as a means to produce pixel-wise embeddings in radiological images by encoding anatomical semantic information [18]. It requires no annotations in training. SAM can match corresponding points between two images, which is exactly the fundamental goal of image registration. The most simple and straightforward way to register two images with SAM is to extract SAM embeddings from both fixed and moving images, match each moving pixel to the closest fixed pixel in SAM space, and calculate the corresponding coordinate offsets to generate a deformation field. However, this approach is highly inefficient, as there are millions of pixels in a typical 3D computed tomography (CT) scan. Besides, SAM would not incorporate spatial smoothness constraints [2], which is useful when the correspondences predicted by SAM contain noises.

We propose SAM-enhanced registration (SAME) to address these issues. SAME is comprised of three consecutive steps. (1) **SAM-affine**, which uses correspondence points generated from SAM on a sparse grid to compute the affine transformation matrix. Affine registration [11] has been widely used either alone or as an initialization of deformable methods [2,9]. (2) **SAM-coarse**, which uses a coarse correspondence grid to directly produce a coarse-level deformation field. These first two steps are efficient, require no additional training, and can provide a good initialization for the final step. (3) Lastly, **SAM-VoxelMorph** enhances the deep learning-based VoxelMorph registration method [2], using SAM-based correlation features [4] and a newly formulated SAM similarity loss. SAME is evaluated on a multi-phase chest CT dataset for inter-subject registration with 35 thoracic organs annotated. Quantitative experimental results show that SAM-affine significantly outperforms traditional optimization-based affine registration in both accuracy and speed. The complete SAME consistently outperforms traditional approaches [15,1] and VoxelMorph [2] in both within-contrast-phase and across-contrast-phase tasks by average Dice scores of 4.7%

and 2.7%, respectively. SAME matches DEEDS [9], as the state-of-the-art in CT registration [17], while being orders of magnitude faster (1.2 sec vs. 45 sec).

2 Method

In this section, we present the details of the proposed SAME for deformable registration and describe how SAM is integrated in each of the three steps.

2.1 Self-supervised anatomical embedding (SAM)

SAM is recently proposed by [18], as a novel pixel-level contrastive learning framework with a coarse-to-fine network and a hard-and-diverse negative sampling strategy. In an unsupervised manner, it predicts a global and a local embedding vector with semantic meanings per pixel in a CT volume—the same anatomical location in different images expressing similar embeddings. SAM is readily used to find correspondences between images, providing a means to solve the registration problem from a new perspective. Let $X_f, X_m \in \mathbb{R}^{D \times H \times W}$ be the fixed and moving images to be registered. For each image, we extract the global and local SAM embedding volumes and concatenate them in the channel dimension, resulting in $S_f, S_m \in \mathbb{R}^{C \times D \times H \times W}$ (C is the concatenated channel dimension). Given a point $p_f = (x, y, z)$ in X_f, we take its embedding vector $S_f(:, z, y, x)$ and convolve it with S_m to get a similarity heatmap volume. The point with the highest similarity score becomes the matched point in the moving image. Results show that matching for a single point only consumes 0.2 sec on a common chest CT scan [18].

2.2 SAM-affine and SAM-coarse

Matched SAM correspondences can be directly employed to estimate an affine transformation matrix [11, 9, 2]. First, we select a set of points on X_f for matching. Intuitively, evenly distributed points on the image may lead to a better estimation. Therefore, we use the points on a regular grid on X_f, see Fig. 1. It would be more precise to run point matching on every pixel (instead of a coarse grid) and directly generate a fine deformation field, but that would consume 0.5h for a CT with 200 slices. To balance accuracy and speed, we use a grid with stride 8. Since SAM is only designed for points inside the body, we segment the body mask of X_f using intensity thresholding and morphological post processing, and then remove grid points outside the mask. When doing point matching, we downsample S_m with spatial stride of 4 to reduce computation. After the corresponding points in X_m are located, we need to filter out low-quality matches. We examine their similarity scores and discard those lower than a threshold θ. After that, we can get k matched points in X_f, X_m, which can be represented by $3 \times k$ matrices: $\mathbf{P}_f$ and $\mathbf{P}_m$, respectively. We pad them with 1s to create

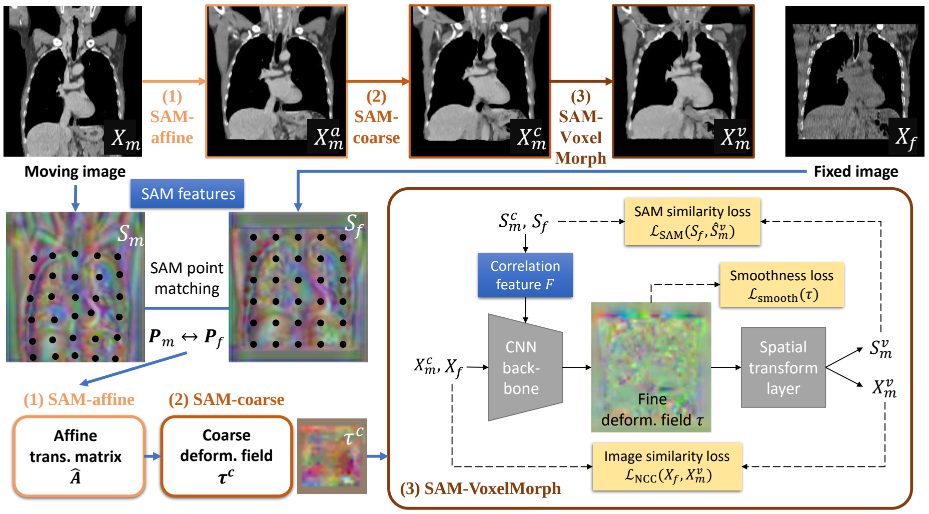

Fig. 1: SAM-enhanced registration (SAME) framework. The moving image is warped by three consecutive steps: SAM-affine, SAM-coarse, SAM-VoxelMorph, gradually approaching the fixed image. Variables X, S, and P denote the image, SAM embedding, and point coordinates, respectively. Subscripts m, f stand for moving or fixed, respectively. Superscripts a, c and v indicate the variable is generated after each of the three steps (affine, coarse deform, or VoxelMorph).

homogeneous versions of the matched points coordinates, $\tilde{\mathbf{P}}_f, \tilde{\mathbf{P}}_m \in \mathbb{R}^{4\times k}$, and estimate the affine matrix $\hat{\mathbf{A}} \in \mathbb{R}^{4\times 4}$ by a simple least squares fitting:

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathbf{A}\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_f\|_F^2. \tag{1}$$

Next, we transform X_m with $\hat{\mathbf{A}}$ to obtain X_m^a and extract new SAM embeddings S_m^a from it. Then, points in $\mathbf{P}_f$ are matched again on X_m^a to get $\mathbf{P}_m^a$. $\mathbf{P}_m^a$ and $\mathbf{P}_f$ actually represent a mapping from X_m^a to X_f on k sparse points. We can compute their difference $\varDelta = \mathbf{P}_f - \mathbf{P}_m^a$, and map each point in $\varDelta$ back to the original coordinates of the image to get $\tau^c \in \mathbb{R}^{3\times D\times H\times W}$. Note, there are only k deformation in $\varDelta$ that are not necessarily uniformly spaced. Thus values in τ^c are filled in using linear interpolation. This gives us the final coarsely estimated deformation map, which is applied to warp (X_m^a, S_m^a) to (X_m^c, S_m^c). Although coarsely estimated (on only k points), τ_c can effectively reduce the difference between the moving and the fixed images. Compared to a global affine alignment, this provides local warps that can serve as a better initialization for a final learning-based deformable registration step. One question is that whether we could omit SAM-affine and compute τ^c directly. We observed that before affine registration, the two images may have significant offsets, so τ^c is potentially large in magnitude, which will magnify the noises in the matched points. Thus, we first perform affine registration to reduce the magnitude of deformations.

2.3 SAM-VoxelMorph

The objective of the final step is to predict a fine deformation map $\tau \in \mathbb{R}^{3 \times D \times H \times W}$, which is a spatial transformation function that can warp the moving image to best match the fixed one. Following the framework of VoxelMorph [2], we learn a function $\varPhi : (X_f, X_m^c) \rightarrow \tau$ with a CNN. The original VoxelMorph uses pure pixel intensity-based features and similarity losses. We improve them by leveraging the semantic information contained in SAM embeddings using SAM correlation features and a SAM loss (see Fig. 1).

The loss function in VoxelMorph and follow-up works includes two parts, an image similarity loss and a smoothness loss. We use the local normalized cross-correlation (NCC) loss [2] for the former, while the latter is defined as

$$\mathcal{L}_{smooth}(\tau) = \frac{1}{|\varOmega|} \sum_{\mathbf{u} \in \varOmega} ||\nabla \tau_{\mathbf{u}}||^2, \tag{2}$$

where $\varOmega$ is the set of all pixels within the body mask. However, the NCC loss only compares local image intensities, which may not be robust under CT contrast injection, pathological changes, and large or complex deformations in the two images. On the other hand, the SAM embeddings can uncover semantic similarities between two pixels. Thus, we add a proposed SAM loss:

$$\mathcal{L}_{SAM}(S_f, S_m^v) = \frac{1}{|\varOmega|} \sum_{\mathbf{u} \in \varOmega} \langle S_f(\mathbf{u}), S_m^v(\mathbf{u}) \rangle, \tag{3}$$

where the superscript v indicates the feature map has been warped by τ predicted by SAM-VoxelMorph. The final loss is

$$\mathcal{L} = \mathcal{L}_{NCC}(X_f, X_m^v) + \lambda \mathcal{L}_{SAM}(S_f, S_m^v) + \gamma \mathcal{L}_{smooth}(\tau). \tag{4}$$

While the SAM loss is an effective means to more semantically align images, the *features* extracted in standard VoxelMorph still lack semantic information, which may be needed to better guide predictions. The correlation feature was originally proposed in FlowNet [4] to manage this problem for optical flow. It was also used in [7] for registration. Briefly, it computes the similarity of pixel $\mathbf{u}$ on X_f and pixel $\mathbf{u} + \mathbf{d}$ on X_m, where $\mathbf{d}$ is a small displacement. This similarity is computed for each pixel and for n possible displacement values to generate an n-channel feature map, which is then concatenated to the original feature map at some point in the network. When using SAM, the semantic similarity of two pixels can be simply computed as the inner product of two SAM vectors, $F(\mathbf{u}) = \langle S_f(\mathbf{u}), S_m^c(\mathbf{u} + \mathbf{d}) \rangle$. We empirically find that using 27 displacement values $\mathbf{d} \in \{-2, 0, 2\}^3$ yields good results. Injecting the SAM correlation features provides improved cues to the network when predicting deformations, thus brings further boosts in accuracy.

3 Experiments

Dataset and task. To evaluate SAME, we collected a chest CT dataset containing 94 subjects, each with a contrast-enhanced (CE) and a non-contrast (NC)

Table 1: Comparison of different registration methods. We show the average Dice score (%) of two tasks: CE-to-CE and CE-to-NC registration. VM: VoxelMorph. Best and second best performance is shown in bold and gray box, respectively.

Methods	CE-to-CE	CE-to-NC	Inference time (s)	std of $\lvert J_\phi \rvert$
Elastix-affine [11]	28.44	27.96	3.38	-
MIND-affine [8]	28.24	27.91	7.86	-
SAM-affine (SA)	33.80	33.77	0.48	-
SAM-coarse (SC)	44.67	43.68	0.78	-
SA + SC	46.76	45.67	1.05	0.40
SA + VM [2]	48.79	47.35	0.78	0.38
SA + SAM-VM	51.99	49.90	0.84	0.36
SA + SC + VM	54.12	50.64	1.13	0.68
SA + SC + SAM-VM (ours)	**54.42**	50.96	1.16	0.66
SyN [1]	49.75	47.95	74.34	-
FFD [15]	49.36	48.22	93.51	0.51
DEEDS [9]	52.72	**51.15**	45.35	0.40

*Paired t-tests show SAME significantly outperforms all other methods ($p < 10^{-4}$), except for DEEDS in the CE-to-NC setting. SAM-VM significantly outperforms VM ($p < 10^{-7}$).

**The average surface distance (ASD) in CE-to-CE: FFD 4.6mm, SA+VM 4.1mm, DEEDS 4.0mm, SA+SAM-VM 3.9mm, SA + SC + SAM-VM 3.8mm.

scan. We randomly split the patients to 74, 10, and 10 for training, validation, and testing. Each image has manually labeled masks of 35 organs (including lung, heart, airway, esophagus, aorta, bones, muscles, arteries and veins) [5]. For the validation and test sets, we construct 90 image pairs for inter-subject registration and calculate an atlas-based segmentation accuracy on the 35 organs. Performances of two tasks are evaluated: intra-phase registration (CE-to-CE) and cross-phase registration (CE-to-NC). Every image is resampled to an isotropic resolution of 2mm and cropped to $208 \times 144 \times 192$ by clipping black borders. The image intensity is normalized to $(-1, 1)$ using a window of $(-800, 400)$ HU.

Implementation details. Our method was developed using PyTorch 1.5. It was run on a Ubuntu server with 12 CPU cores of 3.60GHz. It requires one NVIDIA Quadro RTX 6000 GPU to train and test. We trained a SAM model using the training set of the chest CT dataset. Its structure is identical with the one in [18], which outputs a 128D global embedding and a 128D local one for each pixel. This model is fixed and applied in all three steps of SAME. In SAM-affine and SAM-coarse, the similarity threshold θ is set to 0.7 to select high-confidence matches. In SAM-VoxelMorph, we use a 3D progressive holistically-nested network (P-HNN) [6] as the backbone and concatenate the correlation feature before the third convolutional block. We also tried 3D U-Net [3] but observed no significant accuracy gains. The loss weights in Eq. 4 are empirically set to $\lambda = 1, \gamma = 0.5$. We train SAM-VoxelMorph using the Adam optimizer with a learning rate of 0.001 for 10 epochs. Each training batch contains 2 image

Table 2: Ablation study for different settings on incorporating SAM to VoxelMorph (VM). The average Dice score (%) is reported. All methods are initialized by SAM-affine without SAM-coarse.

Methods	SAM loss	SAM correlation feature	CE-to-CE	CE-to-NC
VM [2]	×	×	48.79	47.35
SAM-VM	✓	×	50.43	48.24
	×	✓	51.37	48.99
	✓	✓	**51.99**	**49.90**

pairs with random contrast phases (CE or NC). We evaluate the registration results using average Dice score over 35 organ masks. The organ masks are not used during training.

Quantitative results. From Table 1 we can see that **SAM-affine** outperforms the traditional affine registration method in Elastix [11] by 5-6%, meanwhile being 6 times faster. It is also better than affine registration with the MIND [8] robust descriptor. This is because SAM can match corresponding anatomical locations between two images accurately and efficiently. Compared with other methods that iteratively optimizes the affine parameters, SAM-affine directly calculates affine matrix by least squared fitting. **SAM-coarse** surpasses SAM-affine by 10% since it allows for locally deformable warping with more degrees of freedom. Cascading these two steps further boosts the accuracy. VoxelMorph pre-aligned by SAM-affine outperforms SAM-affine + SAM-coarse moderately since the latter can only perform a coarse deformable transformation. However, note that the former is a learning-based dense registration method, while the latter does not require any extra training. It only utilizes the matching result of a pretrained SAM model on grid points. The 2% small gap demonstrates the capability of our proposed SAM-coarse.

SAM-affine + SAM-coarse can provide a good initialization to the learning-based VM in the third step, allowing it to better perform. From the 4 rows in the middle block of Table 1, we also observe consistent improvement by replacing the original VoxelMorph [2] with **SAM-VM**. The SAM embeddings contain more semantic information than the raw pixel intensities, which is incorporated to SAM-VM by the SAM-based correlation feature and SAM loss. An ablation study of SAM-VM is shown in Table 2, where the best result is achieved when both the correlation feature and SAM loss are used. On one hand, explicitly inputting the correlation feature calculated by SAM provides extra guidance for determining the deformation fields. On the other hand, the SAM loss provides a more semantically informed supervisory signal.

In the bottom block of Table 1, we evaluate several widely-used non-rigid registration methods including FFD [15], SyN [1], and DEEDS [9]. FFD was implemented using Elastix [10], where parameters matched the best performing FFD method in EMPIRE10 Challenge [14]. The only modification was an extra bending energy term with weight 0.01 to regularize the smoothness. For

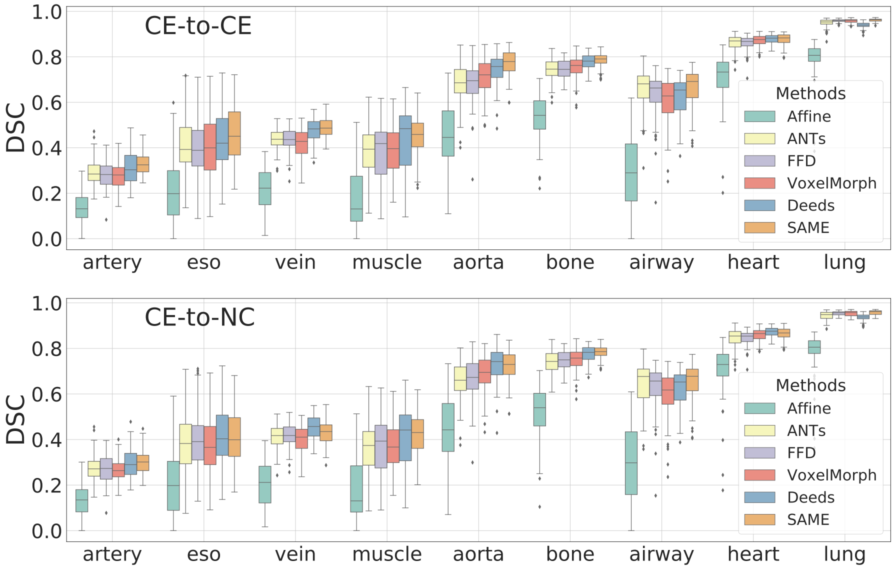

Fig. 2: Comparison of registration methods on all organ groups. Eso: esophagus.

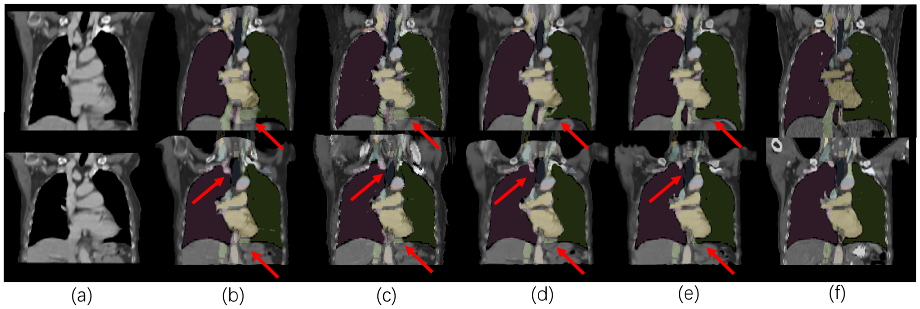

Fig. 3: Visualization of registration results from different methods. From left to right is (a) the moving image, (b) warped moving image of ANTs, (c) DEEDS, (d) SAM-affine + VoxelMorph, (e) SAME, and (f) the fixed image.

SyN (implemented in ANTS) and DEEDS (implemented by the original author), parameters were set according to those used in [17]. For affine transform, the default implementation in each package was used. The proposed SAME (combination of three steps) achieves markedly better results than SyN and FFD. Compared with the best traditional method (DEEDS), it performs better in the within-phase setting and comparably in the cross-phase setting, meanwhile is 38 times faster. Cross-phase registration is more difficult because the brightness and appearance of contrast-enhanced and non-contrast CTs can be very different (see X_m and X_f in Fig. 1), and DEEDS has explicitly designed the modality independent features in its registration. SAME takes a different approach that uses the modality invariant SAM embeddings to align images.

We have computed the standard deviation of Jacobian determinants to measure the smoothness of the deformation field. In Table 1, it is observed that SAME achieves the best Dice with a certain degree of sacrifice in smoothness. This is mainly because SAME cascades two deformable methods, SAM-coarse (SC) and SAM-VM. The smoothness of SAM-VM alone is slightly better than the original VM (0.36 vs. 0.38), but SC itself brings more non-smoothness (0.40). SC generates a deformation field by directly differentiating two sets of coordinates without any constraint. This approach gives SC more flexibility to model large deformation but may also produce less smoothed results. We will study on adding constraints to improve the smoothness of SC in the future. On the other hand, if SC is not used, SA + SAM-VM can also achieve competing accuracy (52.0% Dice score) with good smoothness (0.36), where the overall performance is still comparable to DEEDS (52.7%, 0.40) while significantly better than FFD (49.4%, 0.51), and SA+VM (50.8%, 0.38).

Organ-specific results are shown in Fig. 2. For the sake of conciseness, we divide the 35 organs in our dataset into 9 groups and calculate the median and inter-quartile range of Dice score within each group. The affine in Fig. 2 is from Elastix [11], whereas the VoxelMorph refers to SAM-affine + VM [2] in Table 1. The results of SAME surpass DEEDS on 8 out of 9 groups except heart in the within-phase condition. In the cross-phase setting, SAME outperforms DEEDS on the artery, bone, airway and lung organs. In other organs, like esophagus and muscle, SAME shows results with smaller variance and comparable median performance with DEEDS. Organ groups such as artery, esophagus, vein, and muscle display lower Dice scores for all methods because they are typically small and can be confused with surrounding tissues. Qualitative examples are illustrated in Fig. 3. Manual organ masks of the fixed images are overlaid to show whether the warped moving images align well with the fixed image. Arrows pointed to regions where SAME works better than other methods.

4 Conclusion

In this paper, we propose SAME, a fast and accurate framework for unsupervised medical image registration. We expect SAM-affine and SAM-coarse to be promising alternatives of traditional optimization-based methods for registration initialization. The SAM correlation feature and SAM loss may also be combined with other learning-based algorithms [12, 19] for further accuracy improvement.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008). https://doi.org/10.1016/j.media.2007.06.004, www.itk.org

2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019). https://doi.org/10.1109/TMI.2019.2897538, http://voxelmorph.csail.mit.edu.
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: MICCAI. vol. 9901 LNCS, pp. 424–432 (2016)
4. Dosovitskiy, A., Fischery, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Smagt, P.V.D., Cremers, D., Brox, T.: FlowNet: Learning optical flow with convolutional networks. In: ICCV. vol. 2015 Inter, pp. 2758–2766 (2015). https://doi.org/10.1109/ICCV.2015.316
5. Guo, D., Ye, X., Ge, J., Di, X., Lu, L., Huang, L., Xie, G., Xiao, J., Lu, Z., Peng, L., Yan, S., Jin, D.: DeepStationing: Thoracic Lymph Node Station Parsing in CT Scans using Anatomical Context Encoding and Key Organ Auto-Search . In: MICCAI. vol. LNCS (2021)
6. Harrison, A.P., Xu, Z., George, K., Lu, L., Summers, R.M., Mollura, D.J.: Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In: MICCAI. vol. 10435 LNCS (2017), https://adampharrison.gitlab.io/p-hnn/
7. Heinrich, M.P., Hansen, L.: Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5 D displacement search. In: MICCAI (2020)
8. Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, S.M., Schnabel, J.A.: MIND: Modality independent neighbourhood descriptor for multimodal deformable registration. Medical Image Analysis **16**(7), 1423–1435 (2012). https://doi.org/10.1016/j.media.2012.05.008, http://users.ox.ac.uk/ shil3388/
9. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: MICCAI. vol. 7512 LNCS, pp. 115–122 (2012)
10. Hu, X., Kang, M., Huang, W., Scott, M.R., Wiest, R., Reyes, M.: Dual-stream pyramid registration network. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing (2019)
11. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: A toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2010). https://doi.org/10.1109/TMI.2009.2035616, http://elastix.isi.uu.nl/wiki.php
12. Liu, F., Cai, J., Huo, Y., Cheng, C.T., Raju, A., Jin, D., Xiao, J., Yuille, A., Lu, L., Liao, C., Harrison, A.P.: Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) Computer Vision – ECCV 2020. pp. 257–274. Springer International Publishing, Cham (2020)
13. Mok, T.C.W., Chung, A.C.S.: Large deformation image registration with anatomy-aware laplacian pyramid networks. Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data: MICCAI 2020 Challenges, ABCs 2020, L2R 2020, TN-SCUI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings **12587**, 61–67 (Feb 2021)
14. Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al.: Evaluation of registration

methods on thoracic ct: the empire10 challenge. IEEE transactions on medical imaging **30**(11), 1901–1920 (2011)

15. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Trans. Med. Imaging **18**(8) (1999)
16. Rueckert, D., Schnabel, J.A.: Medical Image Registration, pp. 131–154. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
17. Xu, Z., Lee, C.P., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., Abramson, R.G., Landman, B.A.: Evaluation of six registration methods for the human abdomen on clinically acquired ct. IEEE Transactions on Biomedical Engineering **63**(8), 1563–1572 (2016)
18. Yan, K., Cai, J., Jin, D., Miao, S., Harrison, A.P., Guo, D., Tang, Y., Xiao, J., Lu, J., Lu, L.: Self-supervised learning of pixel-wise anatomical embeddings in radiological images (2020), https://arxiv.org/abs/2012.02383
19. Zhao, S., Dong, Y., Chang, E., Xu, Y.: Recursive cascaded networks for unsupervised medical image registration. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10599–10609 (2019). https://doi.org/10.1109/ICCV.2019.01070