This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

BlendNet: on the Importance of Accurate Local Pre-Registration in the Image Registration Task

Gabriel Ravelomanana[0009-0003-4313-0569]*, Julien Pansiot[0000-0002-6929-3441], and Sergi Pujades[0000-0002-9604-7721]

Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France
nampoina.ravelomanana@inria.fr

Abstract. In this work, we address the problem of pre-registering images before running any registration model. This step is often considered solved by computing a global affine transformation between the fixed and moving images. However, we claim that accurate local pre-registration of images has a major impact on the quality of the final registration. Our approach introduces BlendNet, a shape-aware network that predicts blending weights to combine multiple locally optimized transformations into a single, dense, and spatially smooth transformation. BlendNet leverages as input fixed and moving images, as well as their corresponding segmentation masks, which provide explicit anatomical shape context and are obtained through manual annotation or a foundational segmentation model. BlendNet is trained with local anatomy and shape-aware supervision and regularization, which enforce a physically plausible and accurate pre-registration. The proposed framework is plug-and-play, as it can be used as a pre-processing step before any other registration method. Through extensive experiments, we demonstrate the benefits of BlendNet. Existing registration methods benefit from a systematic performance boost in alignment accuracy with greater improvements observed for high quality segmentation masks. However, this increase in alignment accuracy may be associated with lower deformation regularity, depending on the used registration method. The learned models and code are made available for research purposes.

Keywords: Registration · Pre-registration · Blending Affine Transforms

1 Introduction

The registration of two volumetric images is a foundational problem with high relevance in many clinical applications. For example, it allows for comparing the evolution of the same patient at different time steps, or mapping a subject into a chosen segmented reference template. Registration algorithms take as input a *moving* image and a *target* image and output a 3D transformation (or 3D displacement field) that best aligns the content of both images. To initialize the registration process, standard toolkits, such as Elastix [13], ITK [15], or FSL

* Corresponding author.

FLIRT [12] are used to compute a "pre-registration", usually a global affine transformation based on image features. Most MICCAI registration Challenges, such as Learn2Reg [8] or Brats [1], provide as input globally pre-registered images. In general, the community considers the pre-registration problem *solved*.

Indeed, a global transformation may correctly position two anatomies globally, but local anatomic structures may remain poorly aligned: each anatomic structure may have a different local orientation and morphology. In theory, these local misalignments should be negligible, and deformable registration methods should be able to solve them. But in practice, pixel-wise accurate registration remains a difficult problem as simultaneously computing small local non-linear displacements and potentially long-range displacements still represents a major challenge for existing algorithms. As of today, local anatomic structures can automatically be identified in two images with foundational segmentation models. And with the obtained paired segmentation masks, one can compute local independent transformations between them. However, defining how the surrounding tissue should continuously deform poses a non-trivial problem.

In our work, we present a novel framework to compute a global, yet locally linear, pre-registration transformation between two images. Our core idea is to learn how to combine locally affine transformations, extracted from anatomic segmentations, into a coherent global transformation. We propose to learn BlendNet, a neural network that takes two volumetric images and their corresponding segmentation masks as input, and outputs pixel-wise blending weights, which are used to compose the global transformation. Our pre-processing can be used before any other deformable registration method, and we show that, systematically, all tested approaches benefit from a more accurate locally affine pre-registration.

Regarding the related work on image registration, since the pioneering image registration work using landmarks [4], methods have evolved from optimization-based to deep learning-based approaches. For example, Deeds [7] is a non-learning-based algorithm using a multi-scale approach to match image features and obtain a dense displacement field. While robust to many modalities and anatomic regions, it remains computationally expensive. The advent of deep learning models has unlocked fast registration times. Early models, such as VoxelMorph [3,2], provided fast and accurate inferences, but are relatively sensitive to domain shifts. Follow-up approaches have explored ways to become more robust to the domain shift, either with automatic hyperparameter selection [10,11] or leveraging synthetic data [9]. Most interestingly, recent foundational registration models, such as UniGradICON [17], trained on diverse large datasets, have shown a high capacity to generalize across organs, modalities, contrasts, and anatomies with minimal fine-tuning. All these works assume that a pre-registration is performed on the images they take as input.

As mentioned earlier, other than toolkits providing affine global registrations, very few works have specifically addressed the pre-registration step. One of them, most related to our work, is Polaffini [14], an elegant training-free approach. In it, similar to us, segmentation masks of the moving and fixed images are extracted with the help of a segmentation model, and, for each detected instance, a local affine transformation is computed. To merge the local transformations, Polaffini uses a blending scheme based on weights computed with a smooth kernel function. This approach has a main drawback: it is very sensitive to the empirical choice of the smoothing parameter σ. In contrast, our method proposes to learn the weighting function by considering not only the segmentation masks, but also the input images.

In summary, our work makes the following contributions: (i) We formulate the pre-registration of two volumes as a learnable weighted average of multiple locally affine transformations. Our approach leverages the input images and the segmentation masks to learn BlendNet, a network that predicts smooth blending weights that allow for combining multiple locally optimized transformations into a single, dense, and spatially smooth transformation; (ii) We introduce a training objective that combines local anatomic-aware supervision, a smoothness regularization term, and negative folding losses to prevent non-plausible deformations; (iii) Our approach is plug-and-play with respect to all existing registration methods: it can be used as a pre-processing before any other registration method; (iv) We present extensive experiments on the Learn2Reg challenge data, in which the benefits of BlendNet are shown. Existing registration methods, such as Deeds, VoxelMorph, and UniGradICON, exhibit a systematic boost in performance, particularly, when high-quality segmentation images are available. (v) The learned models and code are made available for research purposes.

2 Method

In Fig. 1 we present the overview scheme of our method. The inputs to our method are a pair of volumetric images and their corresponding segmentation masks. We note $I_f \in \mathbb{R}^{W\times H\times D}$, with $W, H, D \in \mathbb{N}$, the fixed image defining the target of the registration, and $I_m \in \mathbb{R}^{W\times H\times D}$ the moving image to be deformed. We define the set of all voxel indices of the image as Ω. Analogously, we note the fixed and moving segmentation masks as S_f and S_m, defined in $\mathbb{B}^{W\times H\times D\times C}$, where each channel is the binary segmentation of the anatomic class c, with $c \in [1, C]$ and $C \in \mathbb{N}$ is the number of segmented classes. We note S_f^c and S_m^c the channel corresponding to the anatomic class c.

Our approach starts by using each pair of segmentation masks S_f^c and S_m^c to compute C linear affine transformations $\{T^c(S_f^c, S_m^c)\}_{c=1,\ldots,C}$, by optimizing the alignment of each pair of segmented objects using a variant of the Iterative Closest Points algorithm [5]. In parallel, we define the BlendNet network $\mathcal{B}$, that takes as input the images and all channels of the segmentation masks, and outputs spatial, per segmented class, weights, i.e.

$$\mathcal{B}(I_f, I_m, S_f, S_m) = \omega \in \mathbb{R}^{W\times H\times D\times C} \tag{1}$$

where $\omega(i, j, k, c) \geq 0$, and $\sum_{c=1}^{C} \omega(i, j, k, c) = 1, \forall (i, j, k) \in \Omega$. These weights are used to compute, on each voxel, the pre-registration transformation

$$\phi(i, j, k) = \sum_{c=1}^{C} \omega(i, j, k, c)\ T^c(S_f^c, S_m^c) \tag{2}$$

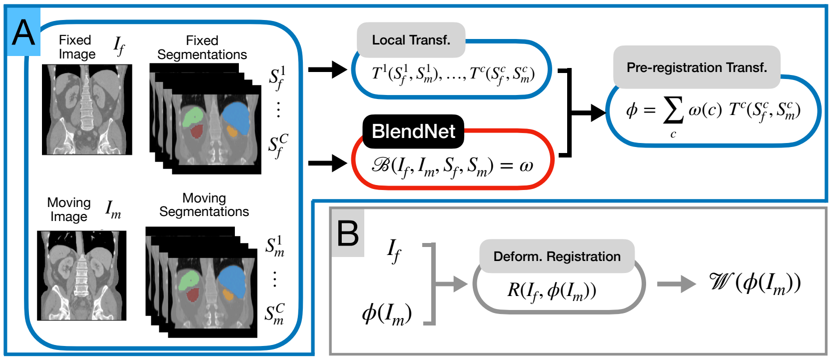

Fig. 1. A: Pre-registration pipeline. The fixed and moving images I_f, I_m and their associated segmentations $S_f^{1,\ldots,C}$ and $S_m^{1,\ldots,C}$ are used as input. From the segmentation masks, C local affine transformations are computed. From the images and segmentation masks, BlendNet predicts optimal blending weights ω used to merge the local transformations into the pre-registration transformation ϕ. B: Any registration method R, can be used with I_f and $\phi(I_m)$ to obtain the final warped moving image $(W)(\phi(I_m))$.

as a linear combination of the individually computed transformations. The resulting transformation ϕ is applied to the moving image I_m and the segmentation mask S_m, and we obtain the warped image $\phi(I_m)$, which we name the *pre-registered image*, and its associated warped segmentation $\phi(S_m)$.

Then, given an arbitrary non-rigid registration algorithm $\mathcal{R}(.,.)$, we use it with the original fixed image I_f and the pre-registered image $\phi(I_m)$, i.e. we compute $\mathcal{R}(I_f, \phi(I_m))$, and obtain the final warped image that we note $\mathcal{W}(\phi(I_m))$.

2.1 Local Transformations

To compute local transformations $\{T^c(S_f^c, S_m^c)\}_{c=1,\ldots,C}$, we use a variant of the Iterative Closest Points algorithm [5]. Between each pair of $(S_f^c,\ S_m^c)$, we optimize an affine transformation with a symmetric ICP: at each iteration, we build bidirectional matches between target and source surfaces with a nearest-neighbor approach, and optimize the alignment of these correspondences. To avoid degenerate transformations, we first compute a rigid alignment, before estimating the affine transformation.

2.2 BlendNet

The key component of our approach is the BlendNet network, taking as input the images and segmentation masks and predicting the blend weights of the anatomic local transformations, i.e. $\mathcal{B}(I_f, I_m, S_f, S_m) = \omega$. The architecture of our BlendNet is a 3D U-Net [6] with skip connections between encoder and decoder layers, and a final softmax layer. It is composed of 3 encoder layers, 1 bottleneck, and 3 decoder layers. Spatial resolution is reduced by a factor of 8, and feature channels size are 16, 32, 64, and 128. In total, our U-Net has around 1.41 M trainable parameters.

To train our BlendNet we use a multi-objective loss composed of four terms: i) an image similarity loss ii) a Dice loss, iii) a smoothness regularization loss, and iv) a negative folding penalty loss. In addition, to avoid instability, we define the weight inside a segmented class to be equal to one, i.e., the transformation matrix of the voxels inside the class c is $T_c(i, j, k)$. Each loss is weighted with an empirically chosen $\lambda \in \mathbb{R}$, so that the contribution of all losses is balanced in the final loss.

Image similarity loss. To encourage global intensity-based alignment, we propose a normalized mutual information loss between the fixed image I_f and the warped CT image $\phi(I_m)$:

$$\mathcal{L}_{sim}(I_f, I_m, \phi) = 1 - \mathrm{NMI}(I_f, \phi(I_m)) \tag{3}$$

where $NMI(A, B) = \frac{2I(A,B)}{H(A)+H(B)}$ with H(.) the marginal entropy of the intensity distribution; and I(. , .) the mutual information, computed with the joint and marginal entropies.

Dice loss. We use the soft Dice loss [16] to encourage overlap between the pre-registered moving segmentation $\phi(S_m)$ and the target segmentation S_f, i.e. $\mathcal{L}_{dice}(\phi(S_m), S_f)$. The loss is individually computed for each anatomic class and then averaged over all classes.

$$\mathcal{L}_{dice}(\phi(S_m), S_f) = 1 - \frac{1}{C}\sum_{c=1}^{C} Dice_c(S_f^c, \phi(S_m^c)) \tag{4}$$

It is worth noting that the dice loss is computed after the pre-processing step, i.e., we use $\phi(S_m)$ and compare it to S_f. Although computing the dice loss after the pre-processing is conceptually not accurate, as the warped image with a combination of linear transformations can not fully match the target segmentation, a key advantage of this approach is that the learned BlendNet network is generic with respect to the deformable registration method $\mathcal{R}(., .)$ used afterwards.

Smoothness regularisation loss. To enforce spatial smoothness of the combined transformation, we enforce smoothness between the neighbouring predicted weights in ω with a quadratic penalty loss

$$\mathcal{L}_{smooth}(\omega) = \frac{1}{|\Omega|}\sum_{(i,j,k)\in\Omega}\sum_{c=1}^{C}\sum_{e\in\mathcal{N}}\left(\omega_c((i, j, k) + e) - \omega_c(i, j, k)\right)^2, \tag{5}$$

where $\mathcal{N}$ is the set of 6-direction vectors covering the 1-connected 3D neighbourhood, i.e. $\{[-1, 0, 0], [1, 0, 0], [0, -1, 0], [0, 1, 0], [0, 0, -1], [0, 0, 1]\}$. Although one could enforce this smoothness supervision on the resulting warping fields ϕ defined in Eq. 2, note that, as the transformations T are linear, if the weights ω are smooth, then the warping field ϕ is also smooth.

Negative folding loss. To prevent foldings and non-manifold transformations, we introduce the hinge loss

$$\mathcal{L}_{\text{fold}}(\phi) = \frac{1}{|\Omega|} \sum_{(i,j,k)\in\Omega} \max(0, -\det \mathbf{J}_{\phi}(i,j,k)) \tag{6}$$

where $\mathbf{J}_{\phi}$ is the determinant of the Jacobian of ϕ computed with finite differences.

3 Experiments

3.1 Dataset and metrics

To evaluate our method, we use the Abdomen CT-CT dataset and associated metrics from the Learn2Reg challenge [8]: Dice (DSC), Dice 30% quantile (DSC30), 95th Hausdorff distance (HD95), and the standard deviation of the logarithm of the Jacobian determinant (SDlogJ). DSC, DSC30 and HD95 are computed without label weighting, with each anatomical structure contributing equally to the reported scores. SDlogJ is computed over the full image domain, reflecting global deformation regularity rather than only within labeled anatomical structures.

The scans were pre-processed by organizers to an isotropic 2 mm resolution, affinely registered to a reference, and cropped to the dimension ($192 \times 160 \times 256$). Each CT is annotated with Ground Truth (GT) segmentation masks S^{GT} for 13 organs. We use TotalSegmentator (TS) [18] to compute segmentation masks, that we name S^{TS}. The original dataset is composed of 50 scans, but the segmentations of the 20 test cases are not publicly available. Thus, we use the 30 public cases and split them into $N_{tr} = 15$ train cases, $N_v = 5$ for validation, and $N_t = 10$ to test. By considering all pairwise cases, each set has $(N_* - 1)^2/2$ registration cases.

3.2 Experimental setting and baselines

To explore the impact of the pre-registration step on different registration approaches, we perform our experiments with state-of-the-art methods: Deeds [7], VoxelMorph [3], and the foundational registration model UniGradICON [17].

To quantify the impact of the proposed framework, we compare to two state-of-the art pre-registration approaches: the classical global affine transformation (G. Aff.), and Polaffini [14]. As the used dataset was already preliminary affinely aligned by the organisers, we don't need to perform any treatment to the images for G.Aff. We run Polaffini using its publicly available implementation and its default parameters. The smoothness parameter controlling the Gaussian blending of local affine transformations was computed using the authors's implementation of the Silverman's rule of thumb.

For BlendNet, we train by alternatively using S^{GT} and S^{TS} for S_f and S_m, so that the network learns some robustness to the segmentation masks errors. It is trained using Adam with a learning rate of 10^{-4}. The training objective

Table 1. Registration results on the left out test data using UniGradICON, Deeds and VoxelMorph with different pre-registration methods: G. Aff., Polaffini, and Ours. The GT suffix in the method name indicates that ground truth segmentation masks were used as input. (∗) signifies that a method is statistically better than G. Aff and Polaffini using Wilcoxon signed-rank tests ($\alpha = 0.05$).

Method	Init	DSC↑	DSC30↑	HD95↓	SDlogJ ↓
UniGrad	G. Aff.	0.55 ± 0.10	0.51	12.62 ± 3.52	0.97 ± 0.48
	Polaffini	0.53 ± 0.08	0.50	13.51 ± 3.73	1.27 ± 0.56
	Ours	0.63 ± 0.05 (∗)	0.61	14.03 ± 4.10	0.66 ± 0.28
	PolaffiniGT	0.56 ± 0.10	0.52	13.54 ± 4.30	1.56 ± 0.59
	OursGT	0.71 ± 0.04 (∗)	0.70	8.97 ± 2.02(∗)	0.75 ± 0.32
Deeds	G. Aff.	0.49 ± 0.09	0.46	14.10 ± 6.18	1.21 ± 0.74
	Polaffini	0.48 ± 0.10	0.44	14.56 ± 4.79	2.40 ± 1.06
	Ours	0.56 ± 0.06	0.52	14.85 ± 5.71	1.55 ± 0.66
	PolaffiniGT	0.48 ± 0.07	0.41	15.36 ± 6.63	2.42 ± 1.08
	OursGT	0.58 ± 0.07 (∗)	0.54	10.92 ± 2.82	1.52 ± 0.68
VoxelMorph	G. Aff.	0.29 ± 0.08	0.25	20.70 ± 5.73	0.16 ± 0.06
	Polaffini	0.32 ± 0.09	0.28	21.12 ± 7.00	0.33 ± 0.19
	Ours	0.49 ± 0.07 (∗)	0.45	15.10 ± 3.40	0.61 ± 0.24
	PolaffiniGT	0.33 ± 0.09	0.28	19.95 ± 6.15	0.32 ± 0.20
	OursGT	0.52 ± 0.08 (∗)	0.48	11.54 ± 2.04 (∗)	0.66 ± 0.25

combines mutual information, dice, smoothness regularization and negative folding losses, respectively weighted by $1.0, 1.0, 0.5$, and 0.05. We evaluate BlendNet with the TotalSegmentator [18] masks ($S_f := S_f^{TS}, S_m := S_m^{TS}$) and report these values as *Ours*. To mimic a case in which the moving and fixed images do have curated segmentation masks, we also evaluate BlendNet with the ground truth segmentation masks ($S_f := S_f^{GT}, S_m := S_m^{GT}$) and report these values as *OursGT*. Analogously, we run Polaffini with the TotalSegmentator and GT segmentation masks as input and we report these results as *Polaffini* and *PolaffiniGT* respectively.

3.3 Impact of pre-registration on non-linear registration

In Table 1, we report the obtained results by the three registration methods using the different pre-registration strategies. We performed Wilcoxon signed-rank tests ($\alpha = 0.05$) [19] and report (*) whenever a method is statistically better than G. Aff and Polaffini. Fig. 3 shows the 13-per-organ distributions of DSC obtained using UniGradICON with G. Aff. (blue) and BlendNet (orange).

As seen in Table 1, our approach leads to: i) a significant improvement in DSC and DSC30 accuracy; ii) a significant improvement or on par HD95 results, and iii) a negative effect on the SDlogJ metric, when compared to G.Aff. . These results are systematic for the three tested registration methods. Methods using GT masks (OursGT and PolaffiniGT) are clearly superior to the ones using TS masks (Ours and Polaffini), highlighting the fact that high-quality segmentation masks of the shape of the anatomy provide a quantitative advantage to the

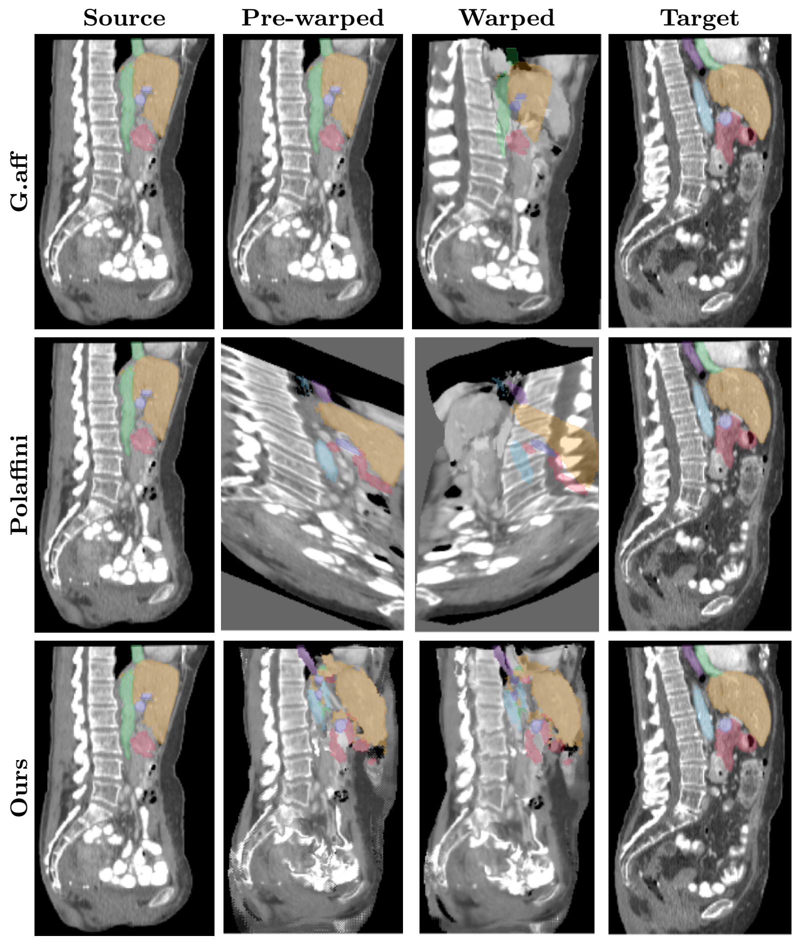

Fig. 2. Qualitative comparison of the registration results obtained with G.aff (1st row), Polaffini (2nd row), and our method (3rd row). Columns show the source image, the pre-warped image, the final warped image with UniGradICON, and the target image.

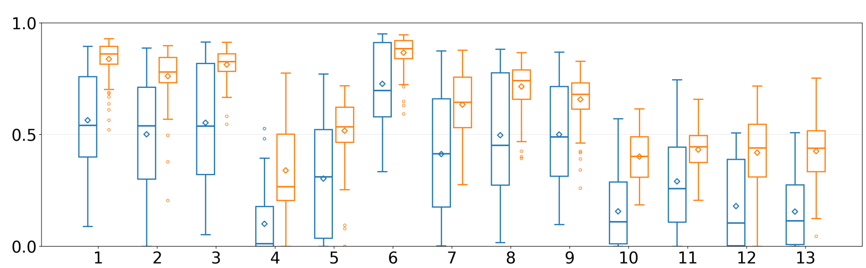

Fig. 3. Dice scores distributions of organ labels 1 to 13 (we use the Learn2Reg index convention) for UniGradICON, using as initialization the G. Aff. (in blue) and the BlendNet scheme (in orange). Legend - Mean: ⋄; Median: −; Distribution outliers: ○.

pre-registration process. The DSC distributions shown in Fig. 3 clearly highlight the benefit of BlendNet for each and every organ in the registration task. Distribution medians and means clearly improved using our BlendNet scheme initialization. We can see up to 0.4 DSC improvement, while the smallest improvement remains around 0.2. Although a few outliers are still present, the overall DSCs distributions are higher.

Regarding the comparison to Polaffini [14], Tab. 1 shows that our approach yields, in both TS and GT settings, significantly superior performance in terms of DSC and SDlogJ, compared to Polaffini, except for VoxelMorph. In terms of HD95, our methods achieve comparable performance to Polaffini in TS settings and outperform it in GT settings. It is worth noting that Polaffini DSC performance, compared to G. Aff., is relatively low. Two elements could be responsible for this low performance. First, as explained in their paper, the choice of their smoothing parameter is critical. To obtain the results of Tab. 1, we used their open implementation that uses the "Silverman's rule" to automatically choose the adequate parameter. Manually fine-tuning this parameter could improve its performance. A second source of explanation could arise by the fact that Polaffini only focuses on the segmentation masks and does not use the image content. In Fig. 2, we present qualitative results of the pre-registration and final registration for G. Aff, Polaffini and ours. Indeed, Polaffini's pre-registration, estimated solely by the segmentations masks, severely deforms the image in the other locations, making it difficult for the non-linear registration to obtain correct correspondences between the images. Our proposed pre-registration step, aligns the shapes of the local structures while preserving the overall shape of the image.

Another noteworthy result observed in Table 1, is the improvement of the metrics when using the GT segmentation masks instead of the TS ones. To further understand this difference, we compared both segmentations for all dataset samples by computing the DSC between the TS and GT masks. For all organs, the mean (and median) DSC range is [.66,.96] (and [.68,.96]), indicating an overall good accuracy. However, for some organs, the minimum DSC values drop to [0.03- 0.06], indicating that in some cases the organs are poorly segmented, and the performance drops. As a takeaway message, registration users should consider the benefit of spending time in the manual segmentation of a few important anatomic regions, which can significantly boost the registration accuracy, with a mean gain as high as of .14 Dice points for UniGradICON.

3.4 Ablation studies

We next perform three experiments to validate our architectural choices, namely the folding penalty, the choice of affine transformations, as well as the use of the Mutual Information loss.

We validated the relevance of the folding penalty in Eq. 6, by running our approach with the corresponding lambda weights equal to zero. We observed stable DSC and HD metrics with a major increase in SDlogJ.

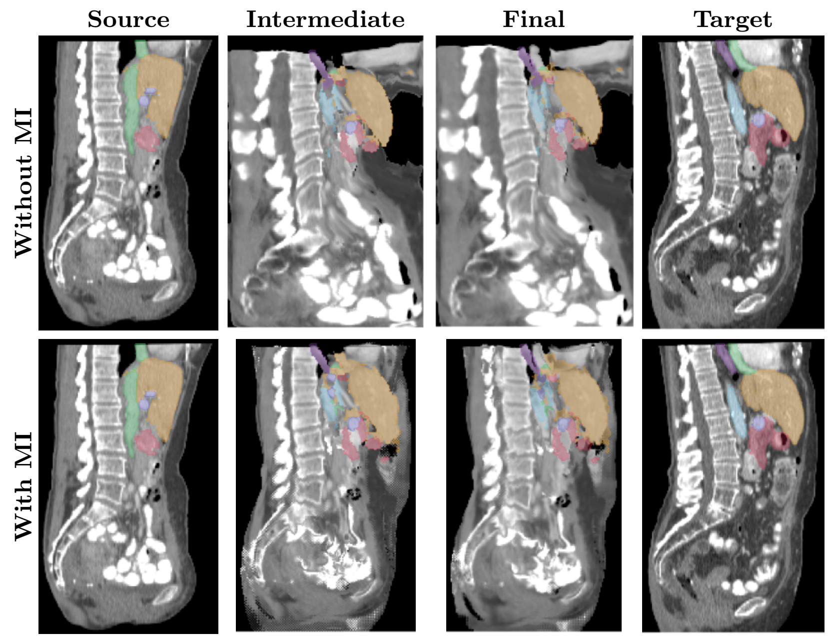

Fig. 4. Qualitative comparison of BlendNet warping with and without mutual information (MI). Without MI, we only constrain the warping of the segmented structures; and thus, BlendNet is allowed to deform the soft tissue without any alignment objective, resulting in distorded images.

We also experimented using Rigid Transformations instead of Affine Transformations for all settings (in Table 1), but all results were systematically worse, thus highlighting the benefit to use Affine transformations for the pre-processing step.

We also experimented by setting the lambda weight to zero of the Mutual Information loss term (Eq. 3). In Fig 4 the absence of the Mutual Information loss leads to visibly unrealistic deformations in the BlendNet pre-warped image, showing that this loss plays an important role in preserving global image consistency. Most interestingly, this visual failure is not reflected by the standard quantitative metrics: all numerical results were very close to those reported in the *Ours* and *OursGT* lines in Table 1. As the Learn2Reg metrics are focused on ROI overlap and deformation plausibility, the evaluation is biased toward the segmented regions of interest and may overlook severe distortions occurring in the image. This suggests that good metric performance within selected ROIs, along with deformation plausibility, do not necessarily guarantee anatomically plausible or globally accurate registration; opening a lead for future registration evaluation protocols.

4 Conclusion

We presented a framework to pre-register volumetric images by leveraging the semantic segmentation of anatomic structures. These segmentations are relatively easy to obtain, either by manual annotation or by leveraging segmentation foundational models. Leveraging the shape of each individual anatomic structure we compute local affine transformations, and we learn BlendNet to predict optimal blending weights to merge them into a global, continuous, smooth function. Our pre-registration approach is plug&play and can be used before any existing registration model. Extensive experiments show how all tested registration approaches systematically benefit from our more accurate pre-alignment, with important performance boosts, and a possible trade-off in deformation regularity. For the recent foundational registration model UniGradICON [17], they reach .16 points Dice on average if high-quality segmentation masks are available.

BlendNet also comes with some limitations. Although BlendNet is agnostic to the input modality and the anatomic region; to train BlendNet, one has to fix the number of segmentation classes in the input. This implies that if the number of anatomic regions varies, one needs to retrain BlendNet. A lead to overcome this limitation would be to use paired landmarks, i.e., a list of correspondences between both images, which could better generalize from one anatomic region to another. Another lead for future work would be to learn fine-tuned versions of BlendNet. BlendNet is supervised with an intermediate dice score, but if the used registration approach is end-to-end differentiable with respect to the images, one could use the final segmentation masks as supervision and fine-tune BlendNet to a specific model (such as UniGradICON). In terms of evaluation, in this paper, we evaluated BlendNet on the CT-CT registration case. As BlendNet is agnostic to the input modality, it would be interesting to evaluate its performance in other modalities (MRI-MRI, CBCT-CBCT) and cross-modality settings (CT-MRI, MRI-US), as the obtained results clearly point out that an accurate local pre-registration step plays an important role in the accuracy of the image registration task.

Acknowledgments. Authors' work was funded by the ANR PRC INORA project with the grant number ANR-22-CE19-0004.

Data Statement. This study relies exclusively on the publicly available Abdomen-CT dataset from the Learn2Reg Challenge, which is fully de-identified. As no new data involving human participants were collected, no additional ethical approval or participant consent was required.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baheti, B., Chakrabarty, S., Akbari, H., Bilello, M., Wiestler, B., Schwarting, J., Calabrese, E., Rudie, J., Abidi, S., Mousa, M., Villanueva-Meyer, J., Fields,

B.K.K., Kofler, F., et al.: The brain tumor sequence registration (BraTS-Reg) challenge: Establishing correspondence between pre-operative and follow-up MRI scans of diffuse glioma patients. arXiv preprint arXiv:2112.06979 (2021), https://arxiv.org/abs/2112.06979

2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9252–9260 (2018)
3. 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
4. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(6), 567–585 (1989). https://doi.org/10.1109/34.24792
5. Chen, Y., Medioni, G.: Object modelling by registration of multiple range images. Image and Vision Computing **10**(3), 145–155 (1992). https://doi.org/10.1016/0262-8856(92)90066-C
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, vol. 9901, pp. 424–432. Springer (2016). https://doi.org/10.1007/978-3-319-46723-8_49
7. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transactions on Medical Imaging **32**(7), 1239–1248 (2013). https://doi.org/10.1109/TMI.2013.2246577
8. Hering, A., Hansen, L., Mok, T.C.W., Chung, A.C.S., Siebert, H., Häger, S., Lange, A., Kuckertz, S., Heldmann, S., Shao, W., et al.: Learn2Reg: Comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. IEEE Transactions on Medical Imaging **42**(3), 697–712 (2023). https://doi.org/10.1109/TMI.2022.3213983
9. Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: SynthMorph: Learning contrast-invariant registration without acquired images. IEEE Transactions on Medical Imaging **41**(3), 543–558 (2022). https://doi.org/10.1109/TMI.2021.3116879
10. Hoopes, A., Hoffmann, M., Fischl, B., Guttag, J., Dalca, A.V.: HyperMorph: Amortized hyperparameter learning for image registration. In: Information Processing in Medical Imaging – IPMI 2021. Lecture Notes in Computer Science, vol. 12729, pp. 3–17. Springer (2021). https://doi.org/10.1007/978-3-030-78191-0_1
11. Hoopes, A., Hoffmann, M., Greve, D.N., Fischl, B., Guttag, J., Dalca, A.V.: Learning the effect of registration hyperparameters with HyperMorph. Machine Learning for Biomedical Imaging **1**, 1–30 (2022). https://doi.org/10.59275/j.melba.2022-74f1
12. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage **17**(2), 825–841 (2002). https://doi.org/10.1006/nimg.2002.1132
13. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: 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

14. Legouhy, A., Callaghan, R., Azadbakht, H., Zhang, H.: POLAFFINI: Efficient feature-based polyaffine initialization for improved non-linear image registration. In: Information Processing in Medical Imaging – IPMI 2023. Lecture Notes in Computer Science, vol. 13939, pp. 614–625. Springer (2023). https://doi.org/10.1007/978-3-031-34048-2_47
15. McCormick, M., Liu, X., Jomier, J., Marion, C., Ibáñez, L.: ITK: Enabling reproducible research and open science. Frontiers in Neuroinformatics **8**, 13 (2014). https://doi.org/10.3389/fninf.2014.00013
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016). https://doi.org/10.1109/3DV.2016.79
17. Tian, L., Greer, H., Kwitt, R., Vialard, F.X., San José Estépar, R., Bouix, S., Rushmore, R., Niethammer, M.: uniGradICON: A foundation model for medical image registration. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Lecture Notes in Computer Science, vol. 15002, pp. 749–760. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-72069-7_70
18. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence **5**(5), e230024 (2023). https://doi.org/10.1148/ryai.230024
19. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin **1**(6), 80–83 (1945). https://doi.org/10.2307/3001968