

# VOTENET++: REGISTRATION REFINEMENT FOR MULTI-ATLAS SEGMENTATION

Zhipeng Ding, Marc Niethammer

Department of Computer Science, UNC Chapel Hill, USA

## ABSTRACT

Multi-atlas segmentation (MAS) is a popular image segmentation technique for medical images. In this work, we improve the performance of MAS by correcting registration errors before label fusion. Specifically, we use a volumetric displacement field to refine registrations based on image anatomical appearance and predicted labels. We show the influence of the initial spatial alignment as well as the beneficial effect of using label information for MAS performance. Experiments demonstrate that the proposed refinement approach improves MAS performance on a 3D magnetic resonance dataset of the knee.

**Index Terms**— multi-atlas segmentation, registration refinement

## 1. INTRODUCTION

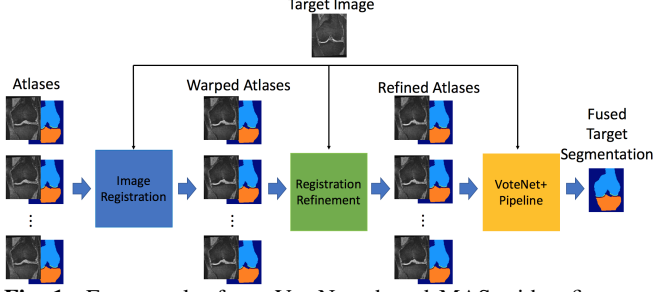
Even though deep segmentation networks [1] have been highly successful and hence have become very popular, multi atlas segmentation (MAS) benefits from its ability to maintain consistent anatomical structures [2]. Assuming label segmentation strongly associates with image appearance, atlas-based image segmentation uses one or multiple image registrations to transfer atlas labels from one or multiple images to a target image that is to be segmented. MAS obtains a consensus optimization via a label fusion step and has achieved great success in medical image segmentation due to its reliable performance [3]. In particular, as multiple atlases are used, individual registration failures are not critical. Nevertheless, reliable image registrations for label propagation are desirable to accurately align anatomical structures across target images. In consequence, better registration can lead to improved MAS performance [3].

However, registration will in general be inaccurate due to anatomical complexities and variations across different images. As pointed out in [4], registration errors (i.e., failure by the registration algorithm to correctly recover correspondences between images) are the principal source of error in multi-atlas segmentation approaches. Different strategies to compensate for these errors have been proposed. Patch-based methods [5, 4, 6, 7] search image neighborhoods at registered atlas positions to find the patches best matching the target image. Segmentation performance can then be moderately

improved by using the displaced patch when computing the consensus segmentation of the target image [4]. Such a local patch search technique can be viewed as refining the voxel-to-voxel correspondences computed by registration, while relaxing the regularization constraints that registration imposes on deformation fields [4]. The key drawback of this approach is that refinement without spatial regularization tends to produce unrealistic atlas images, thus breaking the consistent appearance of anatomical structures. Since atlas appearance is not directly used for the final label fusion, the consequences of unrealistic atlas registrations are not well studied. Fig. 2 shows an example of such an unrealistic atlas appearance after patch-based refinement.

Besides approaches to improve spatial correspondences, different label fusion strategies have been proposed for MAS, the simplest being plurality voting [8] and global or local weighted voting [9, 4]. Deep learning based approaches for label fusion have recently been proposed. For example, the VoteNet [2, 10] approaches use a deep convolutional neural network to locally select the most trustworthy atlases and achieve good performance even when followed by label fusion via simple plurality voting. The VoteNet approaches do not directly cope with registration errors, but instead aim at removing atlases with big registration errors from the label fusion step.

In this work, we integrate a registration refinement step into VoteNet+ [10]. Different from patch based methods [5, 4, 6, 7], we estimate a spatially smooth voxel-to-voxel correspondence to retain realistic anatomical atlas appearance after refinement. Specifically, we make use of regularized spatial transformation networks [11] to obtain a volumetric displacement field to refine atlas-to-target registrations. The refined results are input into VoteNet+ [10] to obtain a final consensus segmentation. In contrast to prior work on registration refinement for MAS [6], the key differences are: 1) we incorporate consistent predicted segmentations while [6] uses MAS fused segmentations iteratively to refine registrations; 2) we use a volumetric displacement field while [6] uses local patch search. Our approach can keep anatomical structures while local patch search would create unrealistic atlas appearance that can not be used in the VoteNet+ framework. 3) we base our approach on deep learning [10] while [6] uses a Bayesian approach for label fusion. Since [2] already demonstrated that VoteNet can outperform a patch based approach [6] and local



**Fig. 1:** Framework of our VoteNet+ based MAS with refinement step. First, images are registered to target images via an image registration method. Second, a refinement step improves registrations. Lastly, VoteNet+ uses the refined atlases and the target image to obtain a fused consensus target segmentation.

patch search would create unrealistic atlas appearance that can not be used in the VoteNet+ [10], we omit direct comparisons with these approaches and instead focus on refining registrations under the VoteNet+ framework.

Our main **contributions** are as follows: 1) *Novel Registration Refinement in MAS*: We propose using anatomical appearance and predicted labels to refine atlas-to-target registrations via a volumetric displacement field and produce smooth and accurate registration refinements. (2) *Comprehensive analysis*: We thoroughly examine the effect of registration refinement for MAS with different registration accuracy.

## 2. METHODOLOGY

Fig. 1 illustrates our framework. For completeness, Sec. 2.1 briefly introduces the VoteNet+ approach. Sec. 2.2 describes our proposed refinement approach. Necessary preregistrations are performed via deep registration networks for speed (see Sec. 3.1 for details).

### 2.1. MAS Overview

Let  $T_I$  represent the target image that needs to be segmented. Denote the pairs of  $n$  atlas images,  $A_I^i$ , and their corresponding manual segmentations,  $A_S^i$ , as  $A^1 = (A_I^1, A_S^1)$ ,  $A^2 = (A_I^2, A_S^2)$ , ...,  $A^n = (A_I^n, A_S^n)$ . MAS first employs a reliable deformable image registration method to warp all atlas images into the space of the target image  $T_I$ , obtaining the registered atlas images and their segmentations  $\tilde{A}^i = (\tilde{A}_I^i, \tilde{A}_S^i)$ ,  $i = 1, \dots, n$ . Each  $\tilde{A}_S^i$  is considered as a candidate segmentation for  $T_I$ . A label fusion method  $\mathcal{F}$  produces the final estimated segmentation  $\hat{T}_S$  for  $T_I$ :

$$\hat{T}_S = \mathcal{F}(\tilde{A}^1, \tilde{A}^2, \dots, \tilde{A}^n, T_I). \quad (1)$$

Furthermore, each atlas may exhibit different anatomical variations. Thus, it is sensible to assign different weights to different atlases [3, 9, 4] to focus on atlases which are the best fits for the target images. Weighting can be global or local. VoteNet+ [10] uses deep learning to provide voxel-wise

weight assignments and label fusion. Given a warped atlas image  $\tilde{A}_I$  and a target image  $T_I$  as inputs, VoteNet+ predicts the probability of  $p(\tilde{A}_S = T_S | \tilde{A}_I, T_I)$  at each voxel location. These probabilities are then used for *joint label fusion* (JLF) [4]. JLF accounts for the possibility that atlases may make correlated errors and computes the weights of atlases via error minimization. Details can be found in [10]. Finally, the segmentation for target image  $T_I$  is obtained by solving

$$\hat{T}_S(x) = \arg \max_{l \in \Omega} \sum_{i=1}^n w_x^i \cdot \mathbb{1}[\tilde{A}_S^i(x) = l], \quad (2)$$

where  $l \in \Omega = \{0, \dots, K\}$  is the set of labels ( $K$  structures; 0 indicating background),  $\mathbb{1}[\cdot]$  is the indicator function, and  $w_x^i$  is the  $i$ -th atlas weight obtained from JLF.

### 2.2. Refining Registrations by Numerical Optimization

Having introduced the basic background of VoteNet+ based MAS, this section discusses how we refine the atlas-to-target registrations via a volumetric displacement field to obtain more accurate MAS consensus segmentation results.

**Objective Function.** Our goal is to optimize a displacement field to refine the alignment of an already pre-aligned atlas with respect to the target image that we want to segment. We hypothesize that better label fusion by VoteNet+ can be achieved by refining this alignment. We use three different losses for this refinement: an image dis-similarity loss,  $\mathcal{L}_{img}$ ; a segmentation dis-similarity loss,  $\mathcal{L}_{seg}$ ; and a loss encouraging spatial regularity  $\mathcal{L}_{reg}$ . This is essentially a standard registration setup with a similarity measure taking into account images and their segmentations. Note that we do not know the segmentation of the target image, as this is precisely what we are after; we will therefore use a simpler automatic segmentation method to obtain segmentation surrogates (see details below). Specifically, our optimized displacement field is

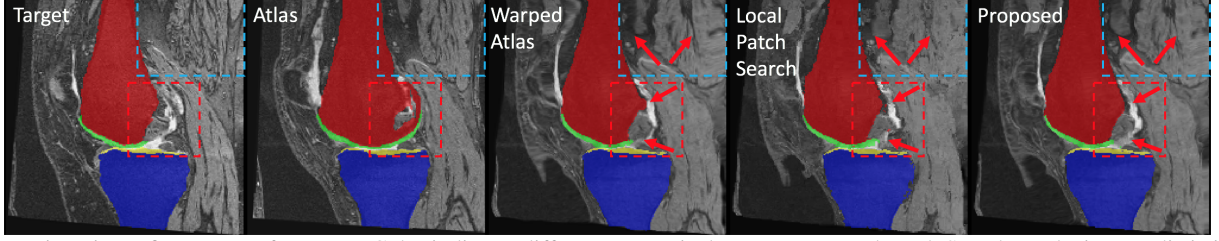
$$f_i^* = \arg \min_{f_i} \mathcal{L}_{img}(\tilde{A}_I^i \circ \phi_i^{-1}, T_I) + \alpha \mathcal{L}_{seg}(S_{src}^i \circ \phi_i^{-1}, S_{tar}) + \gamma \mathcal{L}_{reg}(f_i), \quad \text{s.t. } \phi_i^{-1}(x) = x + f_i(x), \quad (3)$$

where  $S_{src}$  and  $S_{tar}$  are predicted segmentations of  $\tilde{A}_S^i$  and  $T_I$  respectively;  $\alpha = 3, \gamma = 20,000$  in our experiment and  $\phi_i^{-1}$  is the spatial transformation map.  $\mathcal{L}_{img}$  encourages a warped atlas image to look similar to the target image;  $\mathcal{L}_{seg}$  incorporates a consistent segmentation prediction result into the registration in order to improve registration accuracy; and  $\mathcal{L}_{reg}$  encourages smoothness of the transformation.

We use normalized cross correlation (NCC) as the image similarity measure

$$\mathcal{L}_{img}(\tilde{A}_I^i \circ \phi_i^{-1}, T_I) = 1 - NCC(\tilde{A}_I^i \circ \phi_i^{-1}, T_I). \quad (4)$$

Different from [6] where the output consensus segmentation of MAS is used within the registration to improve registration accuracy, we use predicted segmentations (obtained



**Fig. 2:** Registration refinement performance. Color indicates different anatomical structures. Local Patch Search results in unrealistic images (blue dotted rectangle) and labels are not smooth at the structure boundaries (red dotted rectangle). Our proposed method maintains realistic anatomical structures and smooth boundaries. Best viewed zoomed.

via a deep network, e.g., a U-Net [1]) for  $\mathcal{L}_{seg}$ . The idea is that even if segmentations are imperfect they can be reliably used for registration as long as they are reasonably consistent. Specifically, we use a soft multi-class Dice loss

$$\mathcal{L}_{seg}(S_{ref}, S_{tar}) = 1 - \frac{1}{K} \sum_{k=1}^K \frac{\sum_x S_{ref}^{(k)}(x) S_{tar}^{(k)}(x)}{\sum_x S_{ref}^{(k)}(x) + \sum_x S_{tar}^{(k)}(x)}, \quad (5)$$

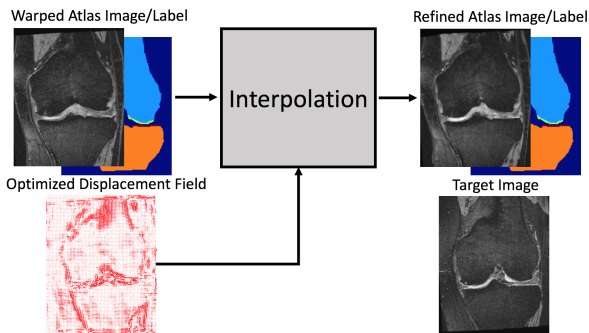
where  $k$  indicates a segmentation label (out of  $K$ ),  $x$  is the voxel position, and  $S_{ref} = S_{src}^i \circ \phi_i^{-1}$  is the refined predicted warped atlas segmentation.

We use the same smoothness term as in [12], which amounts to regularization based on a bending energy [13]. I.e.,

$$\mathcal{L}_{reg}(f_i) = \frac{1}{M} \sum_x \sum_{i=1}^d \|H(f_i(x))\|_F^2, \quad (6)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm,  $H(f_i(x))$  is the Hessian of the  $i$ -th component of  $f$  at position  $x$ , and  $d$  denotes the volumetric dimension ( $d = 3$  in our case).  $M$  denotes the number of voxels.

**Volumetric displacement field.** Fig. 3 illustrates the refinement process. Using the optimized displacement field from Eq. 3 via differentiable trilinear interpolation [11], we can refine the warped atlas image to be closer to the target image. Thus, the warped atlas segmentation is also expected to be closer to the true target segmentation. Note that we use the warped atlas segmentation here, not the predicted segmentations used in Eq. 5 to optimize the displacement field.



**Fig. 3:** Volumetric transformation. Warped atlas image (label) is transformed via a displacement field using interpolation.

### 3. EXPERIMENTAL RESULTS AND DISCUSSION

#### 3.1. Experimental Setting

**Dataset.** For our experiments we use a 3D knee MRI dataset from the Osteoarthritis Initiative (OAI), which includes four labels (i.e., femur, tibia, femoral cartilage and tibial cartilage) [14]. All images are affinely registered to an atlas built from training images. Images are resampled to size  $160 \times 200 \times 200$  with isotropic spacing of 0.7 mm and intensity normalized to  $[0, 1]$ . Our Train/Validate/Test split is 200/53/254 for both our registration networks and our U-Net [1] baseline segmentation network. To train VoteNet+, we randomly select 11 images (and their corresponding labels) from the training set as atlases and register these 11 atlases to the 200 training images to form the training dataset.

**Deep Networks.** For our pre-registration networks, we use the same architecture as in [12] to predict a displacement field, similar to [15, 16]. All images are initially affinely registered to an atlas. We train two models: one without label information as loss term (DI) and one with label information as loss term (DIS). We expect better accuracy from DIS as the network can benefit from stronger guidance via the given labels. We train using ADAM over 100 epochs with a multi-step learning rate. The initial learning rate is 0.0001 and reduced to 0.00001 after 50 epochs. We use soft multi-class Dice, normalized cross correlation, and the bending energy as loss terms. See [12] for details. We use a simple 3D U-Net [1] architecture for both U-Net segmentation and VoteNet+ probability prediction. For fair comparison, we train both models using ADAM with a cross entropy loss over 100 epochs with fixed learning rate 0.0001. VoteNet+ is trained based on the pre-registration results of the different registration models to account for different registration accuracies. The models we consider are: VoteNet<sub>0</sub>+ with only Affine pre-registration, VoteNet<sub>1</sub>+ with DI and VoteNet<sub>2</sub>+ with DIS. It is expected that VoteNet<sub>2</sub>+ is better than VoteNet<sub>1</sub>+ and VoteNet<sub>0</sub>+. All models are trained on an NVIDIA RTX 8000 GPU.

**Metrics.** We used 4 metrics to evaluate segmentation performance: average surface distance, average surface Dice score (i.e., a surface element is considered overlapping if it is within a certain distance ( $\leq 0.7$  mm) to the other surface), 95% maximum surface distance, and average volume Dice score.

Atlas-to-Target Registration	Method	ASD (mm) ↓		SD (%) ↑		95MD (mm) ↓		VD (%) ↑	
		Bone	Cartilages	Bone	Cartilages	Bone	Cartilages	Bone	Cartilages
–	U-Net (baseline)	0.24(0.13)	0.27(0.06)	96.00(1.87)	94.20(2.61)	0.93(0.95)	1.00(0.31)	98.07(0.34)	81.66(3.12)
Affine	PV	1.40(0.24)	1.22(0.32)	44.32(8.10)	55.61(11.80)	3.83(0.66)	3.66(0.75)	90.53(2.12)	33.93(10.95)
	VoteNet <sub>0</sub> +	0.38(0.12)	0.50(0.18)	88.73(6.11)	84.74(9.22)	1.68(0.73)	1.88(0.68)	97.09(0.99)	71.56(9.10)
	VoteNet <sub>0</sub> + (I)	0.42(0.23)	0.36(0.09)	91.98(3.55)	90.76(4.05)	2.08(1.65)	1.45(0.49)	97.49(0.71)	77.55(4.51)
	VoteNet <sub>0</sub> + (IS)	<b>0.28(0.06)</b>	<b>0.32(0.10)</b>	<b>93.38(2.92)</b>	<b>91.91(4.64)</b>	<b>1.20(0.37)</b>	<b>1.27(0.50)</b>	<b>97.79(0.44)</b>	<b>78.69(5.03)</b>
DI	PV	0.57(0.11)	0.47(0.12)	79.63(5.45)	86.46(5.20)	2.26(0.47)	1.95(0.60)	95.75(1.00)	73.20(5.20)
	VoteNet <sub>1</sub> +	0.25(0.04)	0.32(0.08)	94.58(2.68)	92.20(3.64)	0.93(0.19)	1.23(0.42)	97.86(0.42)	79.13(4.01)
	VoteNet <sub>1</sub> + (I)	<b>0.24(0.04)</b>	0.32(0.08)	94.72(2.65)	92.35(3.60)	0.92(0.20)	1.21(0.43)	97.88(0.41)	79.29(4.00)
	VoteNet <sub>1</sub> + (IS)	0.26(0.13)	<b>0.28(0.06)</b>	<b>95.62(2.36)</b>	<b>93.94(2.85)</b>	<b>0.90(0.13)</b>	<b>1.03(0.34)</b>	<b>98.01(0.34)</b>	<b>81.04(3.38)</b>
DIS	PV	0.31(0.05)	0.28(0.06)	91.22(3.19)	93.91(2.91)	1.18(0.22)	1.03(0.34)	97.42(0.46)	81.89(3.34)
	VoteNet <sub>2</sub> +	0.22(0.03)	0.26(0.06)	95.55(2.16)	94.56(2.71)	0.85(0.15)	0.96(0.31)	98.04(0.32)	82.26(3.29)
	VoteNet <sub>2</sub> + (I)	0.22(0.03)	0.26(0.06)	95.60(2.15)	94.56(2.71)	0.85(0.14)	0.96(0.31)	98.04(0.33)	82.30(3.28)
	VoteNet <sub>2</sub> + (IS)	<b>0.21(0.04)</b>	<b>0.25(0.05)</b>	<b>96.29(1.82)</b>	<b>94.72(2.64)</b>	<b>0.82(0.22)</b>	<b>0.94(0.30)</b>	<b>98.12(0.28)</b>	<b>82.54(3.16)</b>
DIS	VoteNet <sub>0</sub> +	0.24(0.03)	0.26(0.06)	94.64(2.21)	94.36(2.68)	0.90(0.16)	0.97(0.30)	97.89(0.33)	81.78(3.31)
	VoteNet <sub>1</sub> +	0.22(0.03)	0.26(0.06)	95.43(2.19)	94.29(2.73)	0.85(0.15)	0.98(0.31)	98.01(0.33)	81.97(3.22)

**Table 1:** Evaluation metrics for OAI segmentation performance. **ASD:** average surface distance. **SD:** average surface Dice score. **95MD:** 95 percentile of the maximum symmetric surface distance. **VD:** average volume Dice score. ↑ (↓) means the higher (lower) the better. **(I)** indicates registration refined using only  $\mathcal{L}_{img}$  and  $\mathcal{L}_{reg}$  while **(IS)** indicates also using  $\mathcal{L}_{seg}$ .

### 3.2. Performance Analysis

Tab. 1 shows our results. The plurality voting (PV) results show that the three registration models result in different segmentation performance; DIS is best, followed by DI, Affine shows the worst performance. All PV segmentation results are worse than the U-Net baseline. Using VoteNet+ for label fusion increases segmentation performance. In terms of volume Dice, we observe an average increase of 22% for VoteNet<sub>0</sub>+, 4% for VoteNet<sub>1</sub>+, and 0.5% for VoteNet<sub>2</sub>+ compared to PV. The other three metrics improve as well. Note that the performance of VoteNet<sub>2</sub>+ is already on par with that of the U-Net baseline and even slightly better than it.

To test our hypothesis that *a better registration model generally leads to the better VoteNet+ based MAS segmentation performance*, we use the DIS pre-registration results to test VoteNet<sub>0</sub>+ and VoteNet<sub>1</sub>+. Note that DIS is expected to be better than Affine and DI pre-registration. The bottom two entries of Tab. 1 show the results. We observe that, after changing to a better pre-registration model, DIS-VoteNet<sub>0</sub>+ increases by 5.5% volume Dice score on average over Affine-VoteNet<sub>0</sub>+, DIS-VoteNet<sub>1</sub>+ increases by 1.5% volume Dice score on average over DI-VoteNet<sub>1</sub>+. The other three metrics are also improve significantly. Note that DIS-VoteNet<sub>0</sub>+ and DIS-VoteNet<sub>1</sub>+ are comparable with the U-Net baseline, while the previous Affine-VoteNet<sub>0</sub>+ and DI-VoteNet<sub>1</sub>+ perform significantly worse than the U-Net baseline. This illustrates that registration refinement is promising in MAS.

To refine imperfect registrations, we mainly use two loss terms as described in Sec. 2.2.  $\mathcal{L}_{img}$  can be viewed as playing the same role as local patch search [5, 4, 6, 7], where  $\mathcal{L}_{reg}$  controls spatial regularity. Tab. 1 shows that it improves segmentation performance significantly if the pre-registration model is not accurate (e.g. Affine-VoteNet<sub>0</sub>+(I) compared to Affine-VoteNet<sub>0</sub>+(IS)); but it barely helps when the pre-registration model is accurate (e.g. DIS-VoteNet<sub>2</sub>+(I) compared to DIS-VoteNet<sub>2</sub>+(IS)). This may be because the remaining anatomical difference tends to be large for in-

accurate pre-registration algorithms and small for accurate pre-registration algorithms. Further, we introduce  $\mathcal{L}_{seg}$  to help the registration refinement. Previous literature [12] has demonstrated that label information can improve registration accuracy. Inspired by this observation, we use the baseline U-Net to provide segmentation predictions for the warped atlas images and the target image. Since the predicted segmentations are from the same algorithm, they are expected to be consistent (i.e. simultaneously correct or wrong for the same type of anatomical structures). For inaccurate pre-registrations (i.e. Affine, DI), the performance improvement from VoteNet+ to VoteNet+ (IS) is very significant. For example, volume Dice increases by 3.9% on average for Affine pre-registration and 1% for DI pre-registration. For accurate pre-registration (i.e. DIS) we still observe moderate improvements. The volume Dice score increases by 0.08% for bones and by 0.28% for cartilage. Fig. 2 demonstrates that our proposed refinement method makes registration more accurate and is better than local patch search.

Furthermore, we find that pre-registration is vital for MAS. First, our pre-registration network can save computation time, because refinement based on good pre-registration results can converge more quickly. Second, refinement after poor pre-registration tends to result in worse accuracy than when starting with a good pre-registration. For example, in Tab. 1, Affine-VoteNet<sub>0</sub>+(IS) is worse than DIS-VoteNet<sub>0</sub>+(IS) and DI-VoteNet<sub>1</sub>+(IS) is worse than DIS-VoteNet<sub>1</sub>+(IS). Hence, registration refinement is beneficial, but a good pre-registration is also important.

## 4. CONCLUSION

In this work, we studied registration refinement for multi-atlas segmentation. Specifically, we used predicted consistent segmentation information to improve atlas-to-target registration accuracy. We demonstrated that registration accuracy is vital for MAS and the predicted segmentation is beneficial for registration refinement.

## 5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access form by the Osteoarthritis Initiative (OAI). Ethical approval was not required as confirmed by the license attached with the open access data.

## 6. ACKNOWLEDGMENTS

Research reported in this work was supported by the National Institutes of Health (NIH) and the National Science Foundation (NSF) under award numbers NSF EECS-1711776 and NIH 1R01AR072013. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the NSF. The authors have no conflicts of interest.

## 7. REFERENCES

- [1] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *MICCAI*. Springer, 2016, pp. 424–432.
- [2] Zhipeng Ding, Xu Han, and Marc Niethammer, "Votenet: a deep learning label fusion method for multi-atlas segmentation," in *MICCAI*. Springer, 2019, pp. 202–210.
- [3] Juan Eugenio Iglesias and Mert R Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.
- [4] Hongzhi Wang, Jung W Suh, Sandhitsu R Das, John B Pluta, Caryne Craige, and Paul A Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 611–623, 2012.
- [5] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D Louis Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [6] Wenjia Bai, Wenzhe Shi, Declan P O'regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel Rueckert, "A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1302–1315, 2013.
- [7] Long Xie, Jiancong Wang, Mengjin Dong, David A Wolk, and Paul A Yushkevich, "Improving multi-atlas segmentation by convolutional neural network based patch error estimation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 347–355.
- [8] Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [9] Xabier Artaechevarria, Arrate Muñoz-Barrutia, and Carlos Ortiz-de Solórzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle," in *Medical Imaging 2008: Image Processing*. International Society for Optics and Photonics, 2008, vol. 6914, p. 69141W.
- [10] Zhipeng Ding, Xu Han, and Marc Niethammer, "Votenet+: An improved deep learning label fusion method for multi-atlas segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 363–367.
- [11] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al., "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.
- [12] Zhenlin Xu and Marc Niethammer, "Deepatlas: Joint semi-supervised learning of image registration and segmentation," in *MICCAI*. Springer, 2019, pp. 420–429.
- [13] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [14] Felix Ambellan, Alexander Tack, Moritz Ehlke, and Stefan Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Medical image analysis*, vol. 52, pp. 109–118, 2019.
- [15] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer, "Quicksilver: Fast predictive image registration—a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.
- [16] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca, "An unsupervised learning model for deformable medical image registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9252–9260.