Supplementary Material for
LT-Net: Label Transfer by Learning Reversible Voxel-wise Correspondence for One-shot Medical Image Segmentation

Shuxin Wang*1,2, Shilei Cao*2, Dong Wei*2, Renzheng Wang†4, Kai Ma2, Liansheng Wang†1, Deyu Meng3,4 and Yefeng Zheng†2

1 Xiamen University 2 Jarvis Lab, Tencent 3 Macau University of Science and Technology 4 Xi’an Jiaotong University
sxwang@stu.xmu.edu.cn, lswang@xmu.edu.cn, {eliasslcao, donwei, kylema, yefengzheng}@tencent.com,
wrzh@stu.xjtu.edu.cn, dy meng@mail.xjtu.edu.cn

This supplementary document provides more details about the basic framework for correspondence learning, in addition to the concise description in Section 3.

The image similarity and transformation smoothness losses: As shown in Fig. 1 to implement atlas-based segmentation with deep convolutional neural networks (DCNNs), a generator network $G_F$ is employed to learn the correspondences from the atlas to unlabeled images, and two unsupervised loss functions—the image similarity loss $\mathcal{L}_{\text{sim}}(u, \bar{u})$ and the transformation smoothness loss $\mathcal{L}_{\text{smooth}}(\Delta p_F)$—are used to supervise the learning process. Minimizing $\mathcal{L}_{\text{sim}}$ encourages $\bar{u}$ to approximate $u$, whereas minimizing $\mathcal{L}_{\text{smooth}}$ regularizes $\Delta p_F$ to be smooth.

To introduce robustness against global intensity variations in medical images caused by the differences in manufactures, scanning protocols, and reconstruction methods, we adopt a locally normalized cross-correlation loss [8,9] to formulate $\mathcal{L}_{\text{sim}}$ that encourages local coherence, which has been proven to be highly effective in correspondence-related tasks [8,9]. Let $f_u(t)$ and $f_{\bar{u}}(t)$ denote the functions to calculate local mean intensities of the unlabeled volume $u$ and deformed atlas $\bar{u}$: $f_u(t) = \frac{1}{n^3} \sum_{t_i} u(t_i)$ and $f_{\bar{u}}(t) = \frac{1}{n^3} \sum_{t_i} \bar{u}(t_i)$, where $t_i$ iterates over a $n^3$ cube around position $t$ in the volume, with $n = 9$ in our experiments (the same as [1]). Then $\mathcal{L}_{\text{sim}}(u, \bar{u})$ is defined as:

$$\mathcal{L}_{\text{sim}}(u, \bar{u}) = -\sum_{t \in \Omega} \left( \sum_{t_i} \left( u(t_i) - f_u(t) \right) \left( \bar{u}(t_i) - f_{\bar{u}}(t) \right) \right)^2 \left( \sum_{t_i} \left( \bar{u}(t_i) - f_{\bar{u}}(t) \right)^2 \right)^{-1}.$$ (1)

The smoothness constraint plays a key role in atlas-based segmentation methods [8,9]; it is also widely used in other correspondence learning problems, such as optical flow estimation [6,7] and stereo matching [5]. In addition, smoothness regularization can be considered as a strategy to alleviate the overfitting problem while encoding the anatomical prior. Here, $\mathcal{L}_{\text{smooth}}$ is formulated with the first-order derivative of $\Delta p_F$:

$$\mathcal{L}_{\text{smooth}}(\Delta p_F) = \sum_{t \in \Omega} \| \nabla (\Delta p_F(t)) \|_2,$$ (2)

where $t \in \Omega$ iterates over all spatial locations in $\Delta p_F$, and we approximate $\| \nabla (\Delta p(t)) \|_2$ with spatial gradient differences between neighboring voxels along $x, y, z$ directions [1]:

$$\| \nabla (\Delta p(t)) \|_2 = \frac{1}{3} \left( \| \nabla_x (\Delta p(t)) \|_2 + \| \nabla_y (\Delta p(t)) \|_2 + \| \nabla_z (\Delta p(t)) \|_2 \right).$$ (3)

The generative adversarial network (GAN) subnet: Besides $\mathcal{L}_{\text{sim}}(u, \bar{u})$ and $\mathcal{L}_{\text{smooth}}(\Delta p_F)$—which are pretty much the standard configuration in atlas-based segmentation problems [3] (e.g., they were used as the main losses in VoxelMorph [1]), we introduce a GAN [2] into our basic framework to offer additional supervision. The GAN subnet

*Equal contributions. Shuxin Wang contributed to this work during an internship at Tencent.
†Corresponding authors.
in our framework comprises $G_F$ and an additional discriminator network $D$ (see Fig. 1). A vanilla GAN would make the discriminator $D$ differentiate $\Delta p_F$ from the true underlying correspondence map. In practice, however, it is usually infeasible to obtain the true correspondence between a pair of clinical images. Instead, we make $D$ distinguish $\bar{u}$ from $u$. In this sense, $\bar{u}$ serves as a delegate of $\Delta p_F$, and $G_F$ is trained to generate $\Delta p_F$ that can be used to synthesize $\bar{u}$ authentically enough to confuse $D$; meanwhile, $D$ becomes more skilled at flagging synthesized images. This delegation strategy provides indirect supervision to $G_F$ and $\Delta p_F$, and allows the networks to be trained end-to-end with a large number of unlabelled images.

**References**

[1] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Trans. Medical Imaging*, 2019.

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[3] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, 2015.

[4] David N. Kennedy, Christian Haselgrove, Steven M. Hodge, Pallavi S. Rane, Nikos Makris, and Jean A. Frazier. CAN-Dis: A resource for pediatric neuroimaging data. *Neuroinformatics*, 10(3):319–322, Oct. 2011.

[5] Hsueh-Ying Lai, Yi-Hsuan Tsai, and Wei-Chen Chiu. Bridging stereo matching and optical flow via spatiotemporal correspondence. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1890–1899, 2019.

[6] Simon Meister, Junhua Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In *Proc. AAAI Conf. Artificial Intelligence*, pages 7251–7259, 2018.

[7] Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang, and Wei Xu. Occlusion aware unsupervised learning of optical flow. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 4884–4893, 2018.

[8] Zhelin Xu and Marc Niethammer. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In *Proc. Int’l Conf. Medical Image Computing and Computer Assisted Intervention*, pages 420–429, 2019.

[9] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V. Guttag, and Adrian V. Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 8543–8553, 2019.