Pure Mathematics 理论数学, 2021, 11(9), 1630-1638
Published Online September 2021 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2021.119180

Hans汉斯

基于改进的自引导网络的图像配准

张　弛

上海大学理学院，上海
Email: axpwyk@gmail.com

收稿日期：2021年7月31日；录用日期：2021年9月1日；发布日期：2021年9月8日

摘　要

图像配准问题关注的是如何通过空间的几何变换将两幅或多幅图像中相似的部分进行对齐。U型卷积神经网络在图像配准任务中有着十分成功的应用。本文首先分析了U型卷积神经网络自身结构的正则性对图像配准的影响，然后以此为出发点改进了一个图像去噪领域的卷积神经网络，将其应用在图像配准任务中。最后在手写体数字数据集的类内图像上进行了配准实验，结果证明本文所提出的方法在可视化效果和数字指标上均有所提升。本文还展示了两种网络的泛化能力，以及多种正则项对于配准数值结果和可视化效果的影响。

关键词

图像配准，卷积神经网络，自引导网络，正则性

Modified Self-Guided Network Based Image Registration

Chi Zhang

College of Sciences, Shanghai University, Shanghai
Email: axpwyk@gmail.com

Received: Jul. 31st, 2021; accepted: Sep. 1st, 2021; published: Sep. 8th, 2021

Abstract

Image registration is concerned with aligning similar parts of two or more images through spatial geometric transformations. U-shaped convolutional neural networks have been used successfully in image registration tasks. In this paper, we first analyze the effect of the regularity of the

文章引用: 张弛. 基于改进的自引导网络的图像配准[J]. 理论数学, 2021, 11(9): 1630-1638.
DOI: 10.12677/pm.2021.119180

U-shaped convolutional neural network structure on image registration, and then improve a convolutional neural network in the field of image denoising as a starting point to apply it to the image registration task. Finally, the registration experiments are conducted on intra-class images of a handwritten digital dataset, and the results demonstrate that the proposed method improves both visualization and numerical metrics. This paper also demonstrates the generalization ability of both networks and the effect of multiple regular terms on the numerical results and visualization of the registration.

Keywords

Image Registration, Convolutional Neural Networks, Self-Guided Network, Regularity

1. 引言

图像配准是通过全局或局部的空间几何变换把两张或多张图像中相似的部分匹配在一起的技术。待配准的图像可能是在不同的时间、不同的角度或使用不同的传感器对同一对象进行拍摄得到的，也可能是对不同对象的相似部分进行拍摄得到的。图像配准是图像处理中的一个基础而重要的问题，在众多领域中得到了广泛的应用。例如，遥感领域、医学图像领域等。

在本文中，将要发生形变的图像被称为移动图像，用 $m:\mathbb{R}^d \to [0,1]$ 来表示；形变的目标图像被称为固定图像，用 $f:\mathbb{R}^d \to [0,1]$ 来表示。图像配准任务即找到一个几何变换 $\phi:\mathbb{R}^d \to \mathbb{R}^d$，使得形变后的移动图像 $\hat{m} = m \circ \phi$ 与固定图像 f 在某种度量的意义下尽量接近。当 ϕ 是一个线性映射时，对应的图像配准被称为线性图像配准；而当 ϕ 是一个非线性映射时，对应的图像配准则被称为非线性图像配准。当 ϕ 可以被一组参数进行显式表示时，对应的图像配准被称为参数化图像配准；而当 ϕ 被某个微分方程所隐式表示时，对应的图像配准被称为非参数图像配准。参数化图像配准的一些工作有[1] [2] [3]，非参数图像配准的一些工作有[4] [5]。所有图像配准任务可以被分为传统方法和基于学习的方法两大类。传统方法较为依赖人类智能，需要人根据自己对图像配准问题的认识设计图像配准任务中每个环节的算法[6]；而基于学习的方法则把图像配准任务中的部分环节或全部环节交给机器，让机器通过算法自行从数据中发现模式。近年来，随着计算机算力的提升，基于学习的方法，尤其是深度学习方法在图像配准领域有着越来越多的应用，得到了研究者的广泛关注。

U 型卷积神经网络在图像配准任务中有着十分成功的应用[7] [8]，这主要得益于它的结构的多尺度特性。这种多尺度特性对于图像配准来说是十分必要的，因为仅仅在大尺度上配准图像可能会忽略细节，而只在小尺度上配准图像可能会使配准陷入局部最优。为了更好挖掘网络结构的多尺度正则性在图像配准任务中的作用，本文改进了一个图像去噪领域的经典工作——自引导网络(self-guided network, SGN) [9]，并将其应用在图像配准任务中。和以 U 型卷积神经网络为基础网络的图像配准模型相比，本文改进的网络在手写体数字数据集的类内图像配准任务上取得了更好的结果。

2. 模型

本文工作的主要框架如图 1 所示。该框架是一个无监督神经网络图像配准的经典范式，在[7] [8]等文

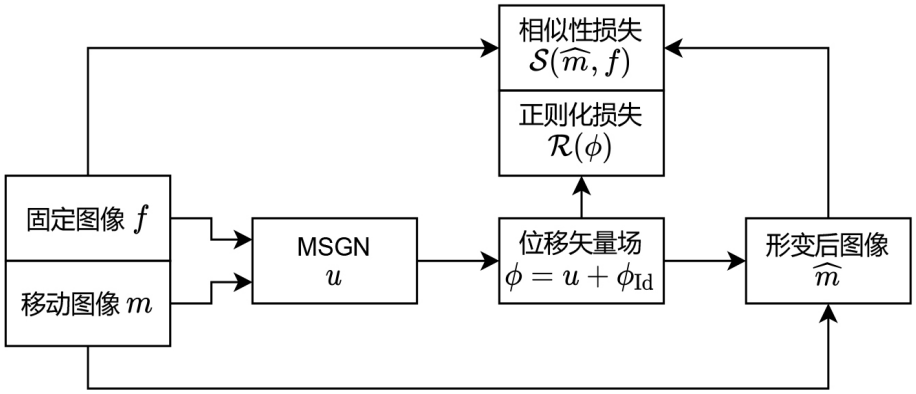


Figure 1. The framework of MSGN based image registration
图 1. 使用 MSGN 进行图像配准的框架

献中被广泛使用。改进后的自引导网络被称为 MSGN (modified SGN)，在这里被用于从固定图像 f 和移动图像 m 组成的图像对中预测形变场 u。形变场 u 和恒等变换 ϕ_{Id} 的和即为所求几何变换 ϕ。形变场 u 被一组参数 Θ 表示，训练神经网络 MSGN 的目的是找到一组最优的参数 Θ^*，使得在给定的数据集 $\{(f_1, m_1), (f_2, m_2), \cdots, (f_N, m_N)\}$ 上平均相似性损失和平均正则化损失的加权和达到最小：

$$\Theta^* = \arg\min_{\Theta} \frac{1}{N}\sum_{k=1}^{N}\left((1-\lambda)\mathcal{S}(\hat{m}_k, f_k) + \lambda\mathcal{R}(\phi_k)\right) \tag{1}$$

其中 $\phi_k = u(f_k, m_k) + \phi_{\mathrm{Id}}$，$\hat{m}_k = m_k \circ \phi_k$。相似性损失使用的是平方距离，两幅图像越相近，相似性损失越小：

$$\mathcal{S}(\hat{m}_k, f_k) = \int_{\mathcal{R}^d} |\hat{m}_k - f_k|^2 \,\mathrm{d}x \tag{2}$$

正则化损失在本文中有四种形式，它们是形变场梯度的 L_1 正则，形变场梯度的 L_2 正则，形变场梯度的 βTV 正则以及文献[10]提出的形变场梯度的 MTV 正则。这四种正则项的表达式分别为：

$$\mathcal{R}_{L_1}(\phi) = \sum_{i=1}^{d}\int_{\mathbb{R}^d}\sum_{j=1}^{d}\left|\frac{\partial u_i}{\partial x_j}\right| \mathrm{d}x \tag{3}$$

$$\mathcal{R}_{L_2}(\phi) = \sum_{i=1}^{d}\int_{\mathbb{R}^d}\sum_{j=1}^{d}\left|\frac{\partial u_i}{\partial x_j}\right|^2 \mathrm{d}x \tag{4}$$

$$\mathcal{R}_{\beta\mathrm{TV}}(\phi) = \sum_{i=1}^{d}\int_{\mathbb{R}^d}\sqrt{\sum_{j=1}^{d}\left|\frac{\partial u_i}{\partial x_j}\right|^2 + \beta}\,\mathrm{d}x \tag{5}$$

$$\mathcal{R}_{\mathrm{MTV}}(\phi) = \sum_{i=1}^{d}\int_{\mathbb{R}^d}\log\left(1 + \sum_{j=1}^{d}\left|\frac{\partial u_i}{\partial x_j}\right|^2\right)\mathrm{d}x \tag{6}$$

文献[9]中用作图像去噪的 SGN 如图 2 所示。其中 $\boldsymbol{C}$ 是输入图像，$\boldsymbol{F}$ 和 $\boldsymbol{R}$ 是一些卷积层生成的特征图像。SGN 使用了一种图像的像素重排操作一次性生成数个空间分辨率依次减半的输入图像。像素重排操作最早见于文献[11]，其中重排编码 shuffle/2 和重排解码 shuffle × 2 被用来降低和增加图像的空间分辨率，如图 3 所示。这样一种自顶向下的、高层特征逐渐汇入低层以引导低层卷积核进行学习的自引导机制有助于使网络更加有效地整合图像多尺度信息。SGN 原本被用于图像去噪任务，因此需

要对其进行一些修改，才能让它更好地处理图像配准任务。本文提出的模型MSGN其结构如图4所示。改进主要有如下三点：第一，起始部分的编码重排操作被替换为了带步长的卷积层。卷积核的大小逐级增大，步长也逐级增大。第二，SGN 每一级的基础卷积核数目是逐级倍增的，若 level 0 中所有卷积层的卷积核数目为 n，则 level 1 至 level 3 中所有卷积层的卷积核数目为 $2n, 4n$ 和 $8n$。本文将它们设定为四个可以自由确定的超参数 n_0, n_1, n_2 和 n_3。最后，在 MSGN 中，最外层的跳跃链接被删除。删除最外层的跳跃链接的原因是：在图像去噪任务中，SGN 的输出是去除噪声后的干净图像，跳跃链接的存在迫使网络学习噪声图像和干净图像的差，也就是噪声模式。但是在配准任务中，网络需要给出多尺度特征图像，并在这个多尺度特征图像后再次使用数个卷积层回归出图像重采样使用的坐标从而组成采样网格。配准网络的输入和输出并非处在相似的空间中，因此强迫网络学习输入和输出之间的某种残差是无意义的。

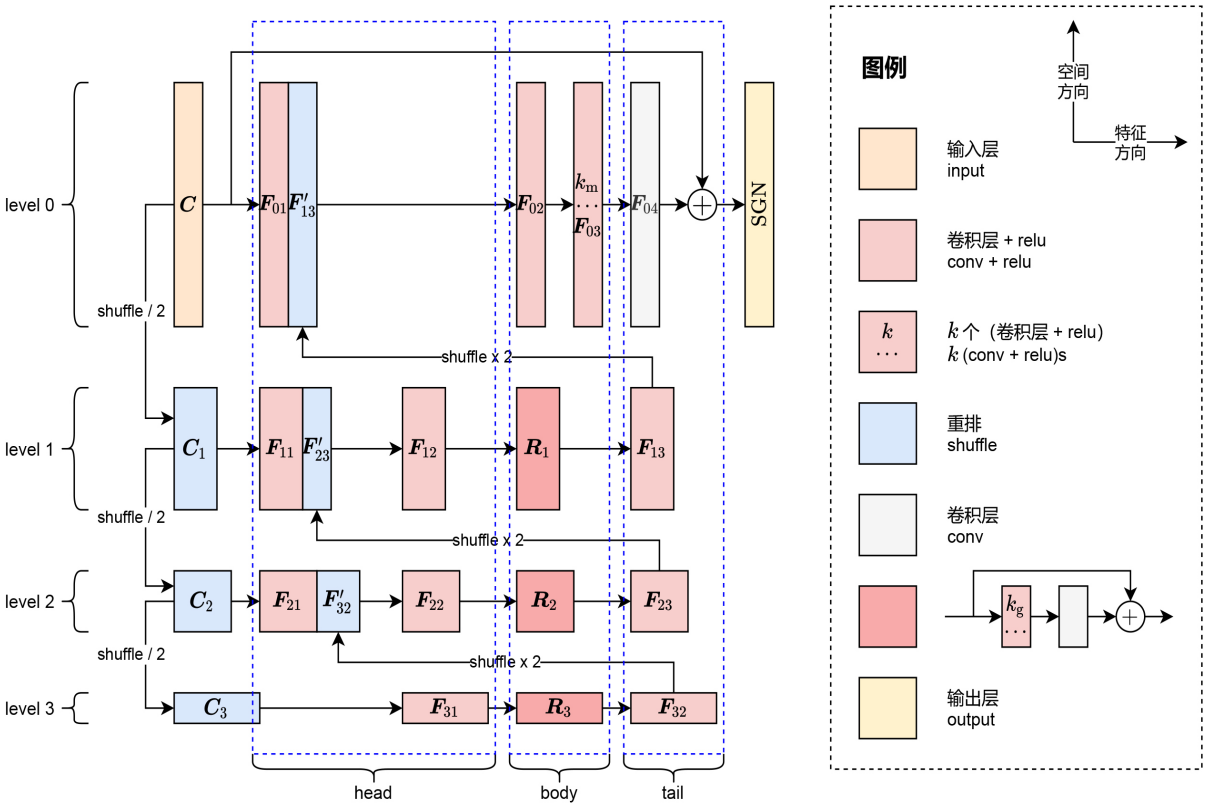


Figure 2. The framework of self-guided network
图 2. 自引导网络 SGN 的网络框架示意图

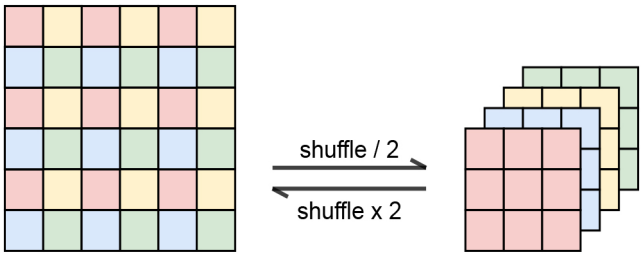


Figure 3. Pixel shuffle
图 3. 像素重排操作示意图

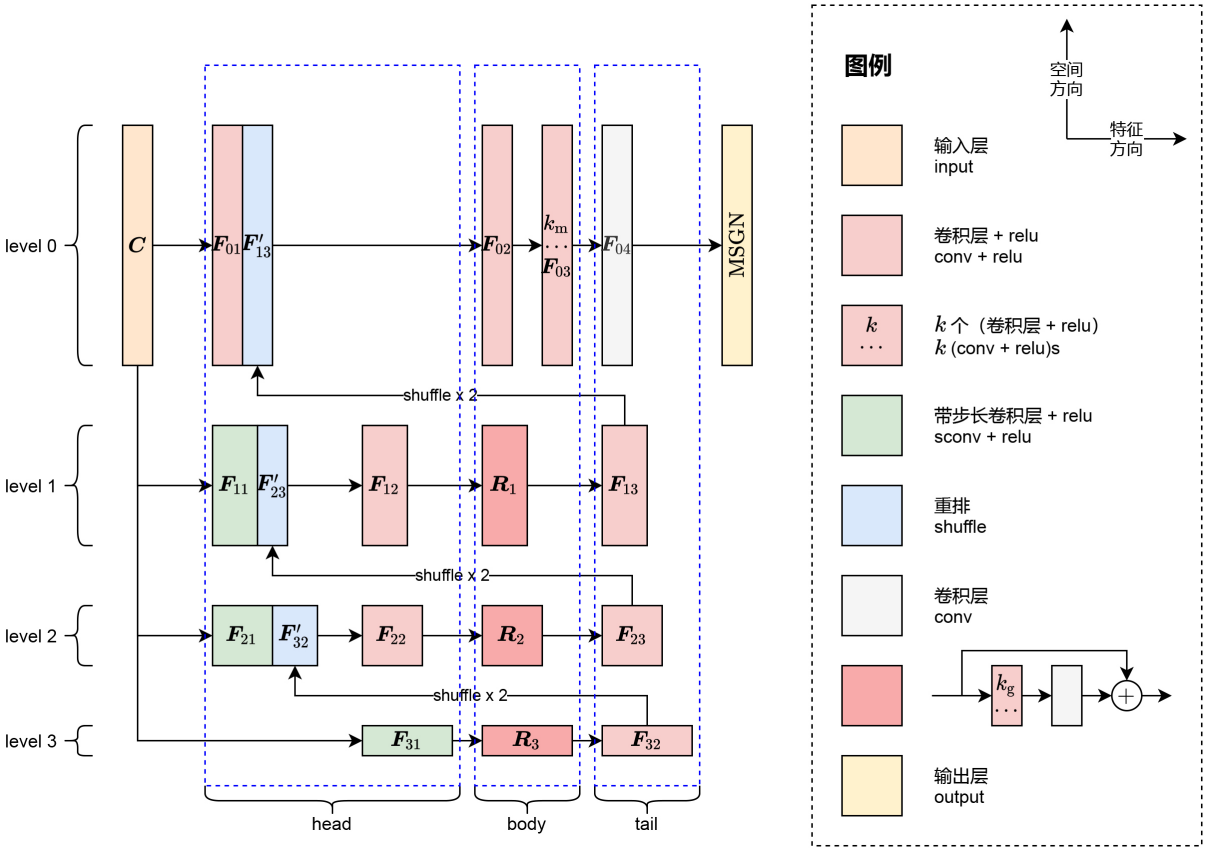


Figure 4. The framework of modified self-guided network
图 4. MSGN 的网络框架示意图

3. 实验

3.1. 实验设置

本文在手写体数字数据集 MNIST 上进行了多组实验，以对比我们所提出的 MSGN 与文献[7]中经典的 U 型卷积神经网络(图表中记为 UNet)在该数据集上配准的性能。MNIST 是机器学习领域一个经典的数据集，由 60,000 张训练图像和 10,000 张测试图像组成，图像大小均为 28 × 28。训练集中数字 0 到 9 的图像数量分别为 5923，6742，5958，6131，5842，5421，5918，6265，5851，5949 张；测试集中数字 0 到 9 的图像数量分别为 5923，6742，5958，6131，5842，5421，5918，6265，5851，5949 张。每张图像都有一个 one-hot 标签作为分类的监督信息，但由于本文使用 MNIST 研究的是图像配准任务，因此这些标签只在最开始用于划分数字类别以使配准在类内进行，在后续的训练中并不会使用这些标签。

实验中所有的数字图像都被零填充成 32 × 32 大小，并把训练集中 5918 张数字 6 的图像随机两两配对用于训练神经网络。使用单一数字 6 训练好的神经网络将会在测试集所有的数字类别上进行测试。我们使用峰值信噪比 PSNR 与 Dice 系数来评估实验结果，同时也会将数字对与形变场进行可视化，在此基础上做更多说明。Dice 系数的定义如下：

$$\mathcal{I}_{\text{Dice}}(\boldsymbol{A},\boldsymbol{B}) = \frac{2\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{A}(i,j)\boldsymbol{B}(i,j)}{\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\boldsymbol{A}(i,j)+\boldsymbol{B}(i,j)\right)} \tag{7}$$

其中 $\boldsymbol{A},\boldsymbol{B}$ 是两个矩阵，取值位于 $[0,1]$，是表示图像分割结果的矩阵。MNIST 数据集较为简单，初始的

分割结果由二值化给出。

在式(1)中，正则项平衡参数 λ 对配准结果影响较大。按照惯常的做法，需要对其进行线搜索以确定最优配准模型。本文将在 $0.0, 0.1, \cdots, 0.9$ 这 10 个点上独立地训练每个模型，并找出其中最优的一个。每个模型均使用大小为 32 的 batch 进行训练，使用 Adam 优化器，初始学习率为 1e−4，训练集被重复使用 250 次。U 型卷积神经网络和 MSGN 训练时的 λ -PSNR 曲线如图 5 所示，可见正则项超参数 λ 的最优值均为 0.1 左右。本文使用了 PyTorch 版 VoxelMorph 作为基础配准框架，使用了 PyTorch Lightning 训练神经网络。

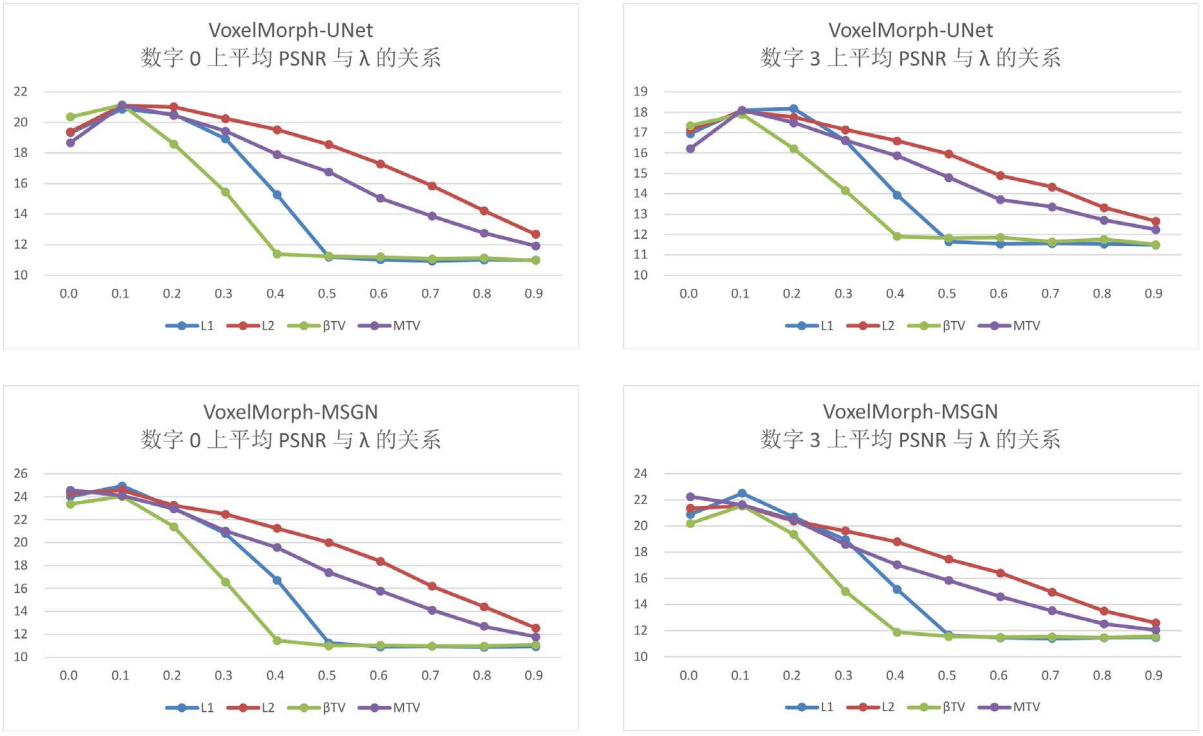


Figure 5. Relationship of PSNR and regularity hyperparameter λ
图 5. PSNR 和正则项超参数 λ 的关系

3.2. 实验结果

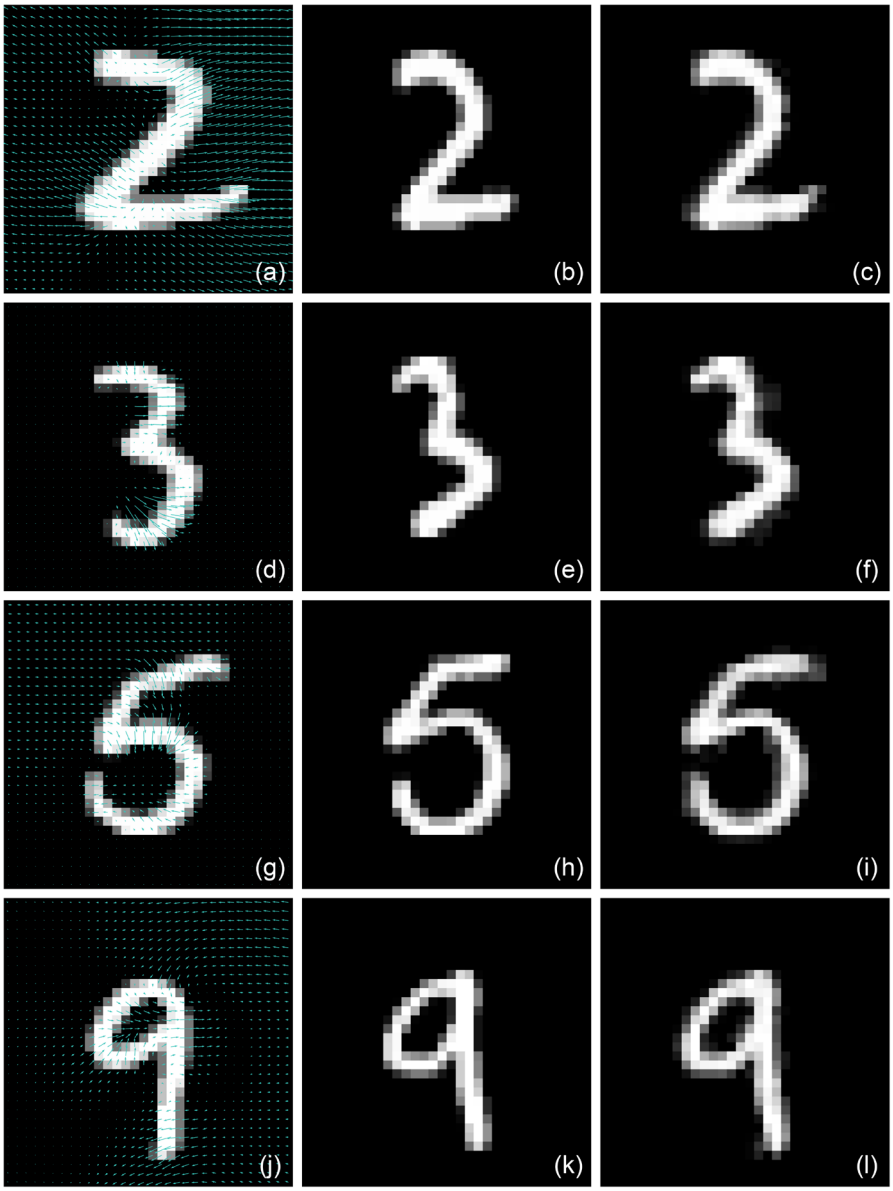
表 1 显示了当正则项平衡参数 λ 取 0.1 时 U 型卷积神经网络和本文提出的 MSGN 的配准结果。可以看出，在每个数字类别以及每种正则项下，本文提出模型都具有更佳的平均 PSNR 和平均 Dice 系数。此外对于 MSGN 来说，形变场梯度的 L_1 正则在所有数字类别上的结果都是四种正则项中最好的。而对于 U 型卷积神经网络来说，形变场梯度的 L_1 正则在大部分数字类别上取得最好结果，除了在数字 0 上形变场梯度的 βTV 正则取得了最优结果，在数字 3 上形变场梯度的 MTV 正则取得了最优结果，在数字 7 和 9 上形变场梯度的 L_2 正则取得了最优结果。对于 U 型卷积神经网络来说，Dice 系数的结果并不总是和 PSNR 呈正相关关系。值得注意的是，在数字 6 上学得的两种网络模型均具有一定的泛化能力，可以处理所有其他类别的手写体数字配准问题。图 6 显示了 MSGN 在几种不同的数字上的配准可视化结果。可以看出，四种正则项的作用均被网络捕捉到，形变场梯度的 L_2 正则为形变场带来了全局光滑性，形变场梯度的 L_1 正则为形变场带来了分片常值性以及稀疏性，形变场梯度的 βTV 正则为形变场带来了分片光滑性，而形变场梯度的 MTV 正则则介于 L_1 和 L_2 两种正则项的效果中间。

Table 1. Comparison between UNet based and MSGN based registration networks
表 1. U 型卷积神经网络和本文提出模型 MSGN 的实验结果比较

数字类别	正则项类型	平均 **PSNR (dB)**		平均 **Dice 系数**		平均预测时间(秒)	
		UNet	**MSGN**	**UNet**	**MSGN**	**UNet**	**MSGN**
0	L_1	20.8722	24.9410	0.9084	0.9571	0.002874	0.008182
0	L_2	21.1055	24.5693	0.9078	0.9497	0.002889	0.007690
0	βTV	21.1518	24.0317	0.9093	0.9512	0.002908	0.007765
0	MTV	21.1020	24.1004	0.9061	0.9479	0.002957	0.007637
1	L_1	26.3562	29.6574	0.9254	0.9637	0.002920	0.007700
1	L_2	26.1091	29.2715	0.9132	0.9544	0.002974	0.007822
1	βTV	25.2050	28.2628	0.9025	0.9537	0.002837	0.007574
1	MTV	25.7999	28.9726	0.9130	0.9549	0.002886	0.007766
2	L_1	17.7678	22.1875	0.8315	0.9278	0.002869	0.007874
2	L_2	17.3122	21.0214	0.8165	0.9040	0.003055	0.007611
2	βTV	17.1773	21.2615	0.8097	0.9149	0.002852	0.007839
2	MTV	17.0974	21.1883	0.8022	0.9088	0.002996	0.007713
3	L_1	18.1001	22.4949	0.8390	0.9259	0.002774	0.007919
3	L_2	18.0394	21.5393	0.8268	0.9072	0.002886	0.007455
3	βTV	17.8933	21.5698	0.8239	0.9156	0.003205	0.007815
3	MTV	18.1050	21.6235	0.8279	0.9125	0.002966	0.007629
4	L_1	18.8365	23.8939	0.8336	0.9352	0.002915	0.007908
4	L_2	18.3769	22.2510	0.8115	0.9048	0.002875	0.007603
4	βTV	18.1806	22.2781	0.8120	0.9144	0.002878	0.007705
4	MTV	18.0108	22.6919	0.7976	0.9144	0.002905	0.007826
5	L_1	18.0286	22.2075	0.8049	0.9119	0.002768	0.008085
5	L_2	17.3202	21.1838	0.7792	0.8860	0.002923	0.007660
5	βTV	17.9255	21.3773	0.7979	0.8988	0.002886	0.007895
5	MTV	17.3521	21.3097	0.7715	0.8916	0.002899	0.007777
6	L_1	22.3173	24.8509	0.9165	0.9517	0.002849	0.007904
6	L_2	21.9856	24.0270	0.9080	0.9394	0.002989	0.007572
6	βTV	20.9708	23.9028	0.8951	0.9432	0.002929	0.007851
6	MTV	21.5745	24.0386	0.9026	0.9400	0.002845	0.007703
7	L_1	19.5380	24.6984	0.8431	0.9385	0.002879	0.008023
7	L_2	19.8594	23.4645	0.8426	0.9148	0.002921	0.007671
7	βTV	19.5435	23.1148	0.8359	0.9186	0.002879	0.007836
7	MTV	19.1853	23.5580	0.8197	0.9197	0.002802	0.007831
8	L_1	18.5135	23.1229	0.8567	0.9382	0.002895	0.008076
8	L_2	17.8736	21.7187	0.8321	0.9160	0.002802	0.007747

Continued

8	βTV	17.9153	22.1014	0.8358	0.9266	0.002916	0.007784
8	MTV	17.6955	21.9336	0.8275	0.9200	0.002889	0.007810
9	L_1	19.5757	23.9752	0.8569	0.9378	0.002943	0.008058
9	L_2	19.5820	23.1028	0.8476	0.9193	0.002897	0.007418
9	βTV	19.2713	22.9673	0.8436	0.9231	0.002926	0.007632
9	MTV	19.3042	22.9723	0.8415	0.9207	0.002972	0.007719



(a) (b) (c)图是形变场梯度的 L_2 正则；(d) (e) (f)图是形变场梯度的 L_1 正则；(g) (h) (i)图是形变场梯度的 βTV 正则；(j) (k) (l)图是形变场梯度的 MTV 正则。其中 λ 均为 0.1

Figure 6. Best results of MSGN with 4 different types of regularity
图 6. MSGN 在四种正则项上的最优结果

4. 结论

本文分析了U型卷积神经网络在配准任务中得到成功应用的原因之一在于它的多尺度结构特性，这种特性正是配准任务所需要的。本文指出图像去噪领域的经典网络——自引导网络由于充分发掘了网络结构的多尺度特性，所以非常适合用于处理图像配准任务。通过对自引导网络进行了一些使其适用于图像配准任务的改进，我们在手写体数字的类内图像配准任务中取得了超过经典的U型卷积神经网络的配准结果。

参考文献

[1] Bookstein, F.L. (1989) Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 567-585. https://doi.org/10.1109/34.24792

[2] Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O. and Hawkes, D.J. (1999) Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, **18**, 712-721. https://doi.org/10.1109/42.796284

[3] Ying, S., Peng, J., Du, S. and Qiao, H. (2009) Lie Group Framework of Iterative Closest Point Algorithm for nD Data Registration. *International Journal of Pattern Recognition and Artificial Intelligence*, **23**, 1201-1220. https://doi.org/10.1142/S0218001409007533

[4] Beg, M.F., Miller, M.I., Trouvé, A. and Younes, L. (2005) Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision*, **61**, 139-157. https://doi.org/10.1023/B:VISI.0000043755.93987.aa

[5] Vercauteren, T., Pennec, X., Perchant, A. and Ayache, N. (2009) Diffeomorphic Demons: Efficient Non-Parametric Image Registration. *NeuroImage*, **45**, S61-S72. https://doi.org/10.1016/j.neuroimage.2008.10.040

[6] Zitova, B. and Flusser, J. (2003) Image Registration Methods: A Survey. *Image and Vision Computing*, **21**, 977-1000. https://doi.org/10.1016/S0262-8856(03)00137-9

[7] Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J. and Dalca, A.V. (2018) An Unsupervised Learning Model for Deformable Medical Image Registration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 9252-9260. https://doi.org/10.1109/CVPR.2018.00964

[8] Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J. and Dalca, A.V. (2019) VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, **38**, 1788-1800. https://doi.org/10.1109/TMI.2019.2897538

[9] Gu, S., Li, Y., Gool, L.V. and Timofte, R. (2019) Self-Guided Network for Fast Image Denoising. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 2511-2520. https://doi.org/10.1109/ICCV.2019.00260

[10] Chumchob, N. and Chen, K. (2010) A Variational Approach for Discontinuity-Preserving Image Registration. *East-West Journal of Mathematics*, 266-282.

[11] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z. (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1874-1883. https://doi.org/10.1109/CVPR.2016.207