

基于深度学习的无监督单模态医学图像配准算法

郑子涵, 谢颖华, 蒋学芹, 周树波, 潘 峰

东华大学信息科学与技术学院, 上海

收稿日期: 2022年12月12日; 录用日期: 2023年1月9日; 发布日期: 2023年1月17日

摘 要

图像配准在手术导航、肿瘤监测等临床医学图像分析领域有着重要应用。本文针对现有无监督单模图像匹配算法的配准精度不够高的问题, 提出了一种基于深度学习的无监督单模态医学图像配准算法。该方法引入短连接与长连接结合形成密集连接, 改进了U-Net的特征图连接方式, 解决相连接两特征图由于采样深度差距较大而产生较大语义差距的问题; 在U型网络的解码器上设计部署了通道注意力机制, 能够有效抑制噪音, 产生更加光滑的形变场, 从而进一步提高后续配准精度。在临床应用的单模态脑部核磁共振图像数据集上进行了训练及测试, 结果表明, 本文提出算法在配准精度上有了一定的提高。

关键词

医学图像配准, 深度学习, 卷积神经网络, 通道注意力

An Unsupervised Single-Modal Medical Image Registration Algorithm Based on Deep Learning

Zihan Zheng, Yinghua Xie, Xueqin Jiang, Shubo Zhou, Feng Pan

Information Science and Technology Academy, Donghua University, Shanghai

Received: Dec. 12th, 2022; accepted: Jan. 9th, 2023; published: Jan. 17th, 2023

Abstract

Image registration has important applications in clinical medical image analysis fields such as surgical navigation and tumor monitoring. Aiming at the problem that the registration accuracy of

the existing unsupervised single-modal image matching algorithm is not high enough, this paper proposes an unsupervised single-modal medical image registration algorithm based on deep learning. This method introduces the combination of short connections and long connections to form dense connections, improves the connection method of U-Net's feature maps, and solves the problem of a large semantic gap between two connected feature maps due to the large difference in sampling depth; the channel attention mechanism is designed and deployed on the decoder of U-shaped network, which can effectively suppress noise and generate a smoother deformation field, thereby further improving the subsequent registration accuracy. The training and testing were carried out on the single-mode brain MRI image data set for clinical application. The results show that the algorithm proposed in this paper has a certain improvement in the registration accuracy.

Keywords

Medical Image Registration, Deep Learning, Convolutional Neural Network, Channel Attention

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

利用计算机视觉进行医学图像处理在临床医学中有着举足轻重的作用，可以大大提高医生在疾病诊断中的效率，也能避免医生因视觉疲劳造成的误判现象。例如利用设备拍摄的医学图像进行手术导航，通过不同时期拍摄的医学图像观察肿瘤生长情况等临床应用[1]。对于这些图像的处理毫无例外，都要进行配准对齐，即两幅图像进行融合，使其在空间坐标上保持一致。

近年来，由于深度学习的方法能用于增强迭代、增强基于强度的配准性能，在医学图像配准领域得到了广泛的应用[2]。基于深度学习的配准分为监督学习[3]的方法和无监督学习[4] [5]的方法。二者都利用神经网络估计变换参数，相较于传统配准算法的泛化性更高。监督学习的方法指的是训练时需要数据样本的标签，即真实形变场，利用神经网络来执行配准过程。Yang 等人[6]首次提出直接基于图像外观的变换模型，逐一预测各个模块来实现医学图像的配准，其将固定图像和移动图像组成的图像对作为输入，通过 U 型网络编解码的初始动量来获得形变场，形变场用于将移动图像重新采样为参考固定图像的图像。Sokooti 等人[7]提出利用多尺度卷积神经网络(Convolutional Neural Network, CNN)来直接获取形变场的方法，其配准精度优于传统 B 样条的配准方法。基于无监督学习的方法可以利用原始图像进行训练，克服了对标签数据的依赖。Bob 等人[4]提出用于无监督仿射和可变形图像配准的深度学习图像配准(Deep Learning Image Registration, DLIR)框架，利用图像对之间的图像相似性来训练 CNN 而无需标签数据，通过分析图像对来学习预测转换参数来形成形变场。Balakrishnan 等人[8]使用了类似 U-Net [9]的 CNN 来获取形变场，并将该算法命名为 VoxelMorph，其配准速度和精度都有着可观的提升，在医学图像配准领域得到了广泛的认可。然而 VoxelMorph 在编解码结构处仍然使用的是类似于 U-Net 的长连接方式，相连接的两个特征图可能存在较大的语义差距，影响后续配准精度。

基于此，本文提出了基于深度学习的无监督单模态医学图像配准算法，在 U-Net 的基础上部署了短连接方式，与原有的长连接方式相结合，既保留长连接表征距离较远的特征之间关系的优点，也克服了相连接的特征之间语义差距较大的缺点。本文也在 U 型网络的解码器上设计了通道注意力机制，进一步

提升配准精度。

2. 相关知识

医学图像配准的目的是针对于固定图像 I_F 与浮动图像 I_M 找到最佳的空间对应关系，并通过损失函数进行反向更新迭代，这个损失函数来自传统配准方法的能量函数：

$$\hat{\varphi} = \arg \min_{\varphi} E(I_M, I_F, \varphi) \quad (1)$$

其中， φ 表示空间变换， $\hat{\varphi}$ 表示最优变换，配准目标即最小化损失函数来优化固定图像 I_F 与浮动图像 I_M 的空间对应关系，直至得到最优的配准后的图像 \hat{I}_F ($I_M \circ \varphi$)。最终得到的 \hat{I}_F 在结构上保留浮动图像 I_M 的特征，在外观上尽可能地与固定图像 I_F 相似。

在配准领域通常利用神经网络参数化空间映射关系，使用神经网络的好处是：通过最小化损失函数自主学习优化参数，从而找到图像间的匹配模型。对于配准问题，通常采用卷积神经网络生成形变场，特别是 U 型网络来对输入的图像进行下采样捕获图像对的空间对应关系，然后上采样进行图像重建。这样做的目的是找到图像中的关键点，去除噪音等无效特征。

对于空间变换 φ (即形变场)指的是，浮动图像向固定图像变换的矢量位移场，反映了浮动图像向固定图像配准需要的位移，后续利用形变场对浮动图像进行扭曲插值处理可以得到配准后的图像。

3. 基于深度学习的无监督单模态医学图像配准算法

3.1. 总体网络框架

本文提出的基于深度学习的无监督单模态医学图像配准算法的总体框架如图 1 所示。需要说明的是，本文提出的配准算法与图像维度无关，为方便演示，本文以二维脑部核磁共振图像(MR)为例，该算法在三维图像上依然适用。首先将固定图像 I_F 与浮动图像 I_M 作为一个图像对双通道输入到整体的网络中，卷积神经网络会对图像进行特征的提取，并生成一个估计的矢量位移场，即形变场 φ 。然后利用空间变换网络(Spatial Transformer Networks, STN) [10]对输入的浮动图像 I_M 施加形变场 φ 并进行插值处理，这里针对二维图像采用的是双线性插值方式，生成配准后的图像 \hat{I}_F 。利用固定图像 I_F 与配准后图像 \hat{I}_F 的相似性度量 L_{sim} 以及形变场 φ 的光滑程度 L_{smooth} 作为目标函数，对训练模型进行迭代，更新网络参数。

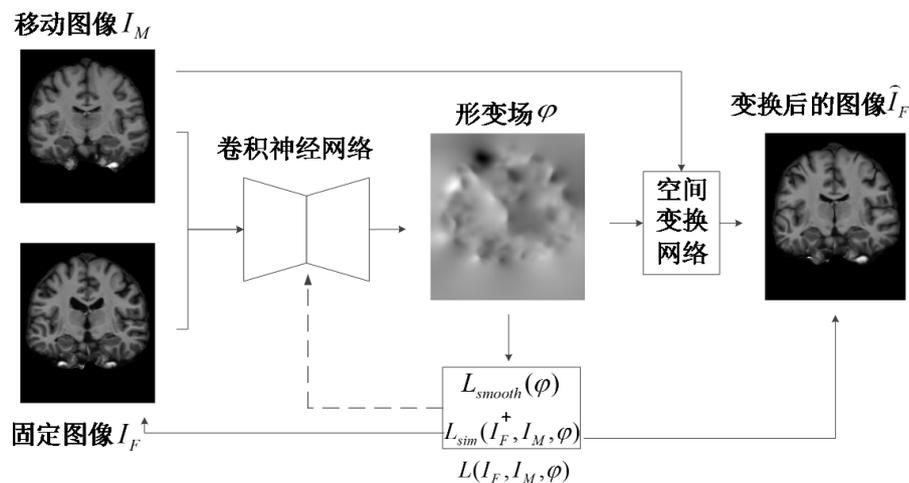


Figure 1. Overall framework

图 1. 总体框架

3.2. 网络结构

所提出的配准网络采用的是卷积神经网络，具体为类似于 U-Net 的一个 U 型网络，包括编码与解码两个阶段，对输入图像对进行特征提取以及变换，来生成形变场。本文还在解码器上部署了通道注意力机制，能够进一步提高由卷积神经网络生成的形变场在扭曲移动图像来生成变换后的图像的真实性，配准网络结构如图 2 所示。

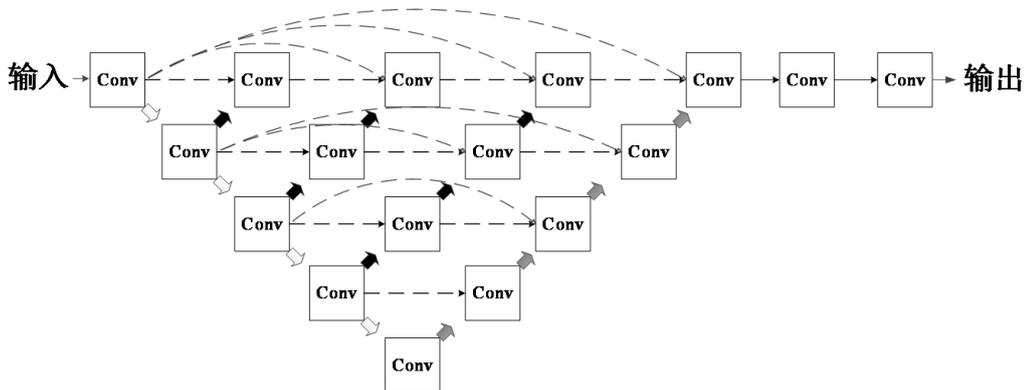


Figure 2. Registration network
图 2. 配准网络

其中，Conv 代表二维卷积以及 LeakyReLU 函数激活；白色箭头代表最大池化(MaxPooling)进行下采样；黑色箭头代表 Upsampling 进行上采样；灰色箭头在 Upsampling 后加入通道注意力(Channel Attention)；虚线箭头代表长短连接方式。

3.2.1. 密集连接的 U-Net

本文的卷积神经网络采用的是密集连接的 U-Net，是基于 U-Net 的改进，编码器进行下采样，捕获图像对的空间对应关系，解码器进行上采样，用于重建图像并获得矢量位移场[11]。一般情况下，U-Net 的编解码器之间直接采用了简单的长连接方式，能够表征相距较远的两个像素之间的关系，然而直接连接的方式会造成相连接的两个卷积层由于深度差距较大而产生较大的语义差异的情况，且这种语义差异会加大网络的学习难度，进而影响后续配准精度[12]。针对于此，本文提出的方法在 U-Net 的结构上部署短连接方式与上采样操作，减小相连接的深度差距较大的特征图之间的语义差距。

具体做法为，将同一模态的固定图像与浮动图像组成的图像对拼接成双通道，输入卷积神经网络，输入图像尺寸为 160×192 。编码器使用 4 个步长为 1、卷积核大小为 3 的卷积操作，每次卷积后跟 MaxPooling 完成下采样，下采样后的图像分辨率分别为原始图像的 $1/2$ 、 $1/4$ 、 $1/8$ 、 $1/16$ ，卷积的感受野也会逐渐扩大，然后使用 LeakyReLU 激活函数进行激活，捕获图像对的空间对应关系。解码器由与编码器相同的卷积操作后跟 Upsampling 进行上采样，使得图像得以重建为原来的分辨率。该网络在每次下采样后跟一次上采样操作，将相同分辨率的特征图拼接在一起，形成密集的短连接，同时，也部署由编码器到解码器的长连接，这样做既能缩小相连接的两个卷积层之间语义差距，也能表征相距较远的两个像素之间的关系。

3.2.2. 注意力模块

注意力机制被证明用在配准算法里可以提高配准性能[13]。基于此，本文受到 ECA-Net [14]的启发，在卷积神经网络上设计引入通道注意力 ECA 模块。其能在全局范围内，对有效特征进行强调，抑制噪音。

在提高后续配准精度的同时，只为网络模型引入了很少量的参数。

具体做法为，将 ECA 模块部署在 U 型网络的解码器上，首先对解码器上采样得到的特征图 $W \times H \times C$ 采用全局平均池化(Global Average Pooling, GAP)进行通道压缩，压缩后变为 $1 \times 1 \times C$ ，其中， W 、 H 表示特征图的尺寸， C 表示通道数，这里的 W 、 H 、 C 的值根据解码器每次上采样后的图像尺寸与通道数来确定；然后，通过卷积核大小为 K 的快速一维卷积后跟 Sigmoid 激活函数来生成各通道的权重信息，其中， K 代表了局部跨信道交互的覆盖率， K 的取值应该根据通道数 C 调节。 K 的取值表达式为：

$$K = \psi(C) = \left\lfloor \frac{\log_2(C)}{r} + \frac{b}{r_{odd}} \right\rfloor \quad (2)$$

其中， r 与 b 表示的都是系数。本文取值为： $r = 2$ ， $b = 1$ 。ECA 模块的示意如图 3 所示。

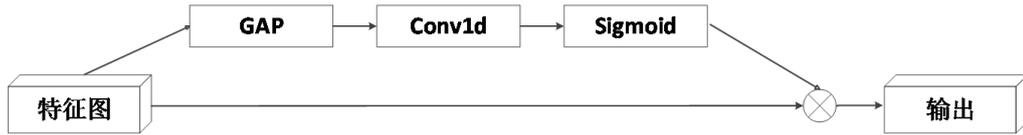


Figure 3. ECA module
图 3. ECA 模块

3.3. 损失函数

本文提出配准算法的损失函数由两部分构成：固定图像 I_F 与配准后图像 \hat{I}_F 的相似性度量 L_{sim} ，用来惩罚配准结果在外观上的差异；形变场 φ 的光滑程度 L_{smooth} 即正则化项，用来惩罚形变场的失真。损失函数的表达式为：

$$L(I_F, I_M, \varphi) = L_{sim}(I_F, I_M, \varphi) + \sigma L_{smooth}(\varphi) \quad (3)$$

其中， σ 为权重超参数，通常情况下，配准结果的相似性程度与正则化项成反比例关系，因此需要权重系数来权衡两者。

相似性度量 L_{sim} 采用平均平方误差(Mean Squared Error, MSE)函数，即 I_F 和 \hat{I}_F 之间像素差值的平方差的平均值，表达式为：

$$MSE(I_F, I_M \circ \varphi) = \frac{1}{|\Omega|} \sum_{p \in \Omega} [I_F(p) - [I_M \circ \varphi](p)]^2 \quad (4)$$

其中， p 为像素位置， Ω 表示图像域。

正则化项解决由于图片相似性度量 L_{sim} 鼓励 \hat{I}_F 尽可能逼近 I_F 而产生不光滑 φ 的问题，本文使用扩散正则化器来平滑 φ ，表达式为：

$$L_{smooth}(\varphi) = \sum_{p \in \Omega} \|\nabla m(p)\|^2 \quad (5)$$

其中， m 代表像素在 p 的位移量。

4. 实验与结果

4.1. 数据集

本文采用医学图像配准领域常用的数据集 OASIS (Open Access Series of Imaging Studies) [15]，并选取此数据集中 416 名年轻、中年、非痴呆和痴呆老年人的脑部横断面 MR 数据。

首先, 对数据集中的图像重采样为 160×192 的大小。然后, 使用 FreeSurfer 软件[16]对数据集中每张 MR 图像进行标准预处理步骤, 包括运动校正、颅骨剥离、仿射空间归一化和皮层下结构分割等。最后, 按 4:1 的比例将数据集随机分为训练集和测试集两部分。图 4 展示了原始图像和经过预处理后的图像。

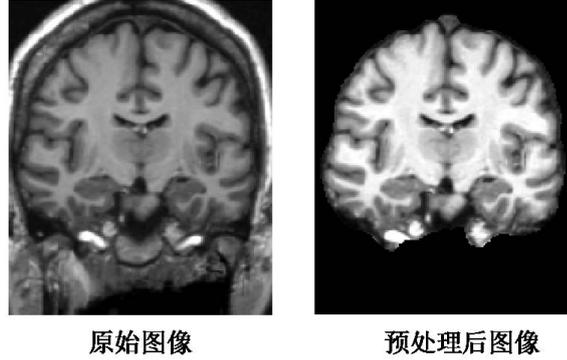


Figure 4. Images before and after preprocessing
图 4. 预处理前后的图像

4.2. 评价指标

对于配准结果的评估, 本文采用的是医学图像配准通用的评价指标戴斯相似性系数(Dice Similarity Coefficient, DSC)。 DSC 系数用于计算评估两个图像的对齐程度, 取值范围为 $[0,1]$, DSC 值越接近 1, 表明配准效果越好[17]。 DSC 系数的表达式为:

$$DSC(s_F^k, s_M^k, \varphi) = \frac{2|s_F^k \cap (s_M^k \circ \varphi)|}{|s_F^k| + |s_M^k \circ \varphi|} \quad (6)$$

其中, s_F , s_M 分别表示固定图像 I_F 及浮动图像 I_M 的分割标签; k 表示第 k 个标签; $s_M^k \circ \varphi$ 表示配准后的图像 \hat{I}_F 的分割标签。针对于 $s_M^k \circ \varphi$ 与训练时扭曲浮动图像的插值方式不同, 其采用最近邻插值重采样得到。

4.3. 实验设置

本文的实验运行在一张 NVIDIA RTX 3090 显卡上, 软件环境为 python 3.8, 使用 Keras 和 TensorFlow 框架实现所提出的无监督配准模型, 选用 Adam 优化器驱动网络优化。本文实验对数据集中的图像进行基于图谱的配准(Atlas-based Registration), 即所有浮动图像都向一个固定的图像上配准。初始学习率设置为 10^{-4} , 实验中设置 2000 个 epoch, 对于损失函数中权重超参数 σ 设置为 0.01, 批处理次数(Batch Size)为 8, 即在每次迭代时从训练集中随机选取 8 张浮动图像与 8 张参考图像组成 8 个图像对进行训练。

4.4. 实验结果

为评估本文提出的无监督配准算法的性能, 本文选择的对比算法为仿射变换 Affine [18]以及 VoxelMorph。值得注意的是, VoxelMorph 算法的卷进神经网络也采用了类似 U-Net 的 U 型网络来生成形变场, 本文实验分别在该算法所提出的 Vxm-1 与 Vxm-2 两个框架上进行对比实验, 通过在测试集上对每个图像的 24 个分割标签的 DSC 系数高低以及可视化结果来评估各个算法的配准效果。

表 1 展示了不同算法在测试集上随机抽取了 5 张浮动图片向同一张固定图片配准的结果, 可以看出本文方法在各组图像的 DSC 系数都高于对比算法, 也意味着本文方法的配准精度较于对比算法有较为明

显的优势。

Table 1. Comparison of DSC coefficients of different algorithms
表 1. 不同算法的 DSC 系数比较

测试组别	Affine	Vxm-1	Vxm-2	本文方法
1	0.5254 +/- 0.2752	0.7314 +/- 0.2308	0.7642 +/- 0.2071	0.7938 +/- 0.1464
2	0.5019 +/- 0.2767	0.7240 +/- 0.2291	0.7544 +/- 0.2116	0.7815 +/- 0.2041
3	0.5059 +/- 0.2459	0.7343 +/- 0.2647	0.7563 +/- 0.2321	0.7834 +/- 0.1681
4	0.5637 +/- 0.2355	0.7570 +/- 0.2260	0.7651 +/- 0.2246	0.8173 +/- 0.1548
5	0.5608 +/- 0.2840	0.7387 +/- 0.2097	0.7664 +/- 0.1918	0.8055 +/- 0.1868
均值	0.5315 +/- 0.2635	0.7371 +/- 0.2321	0.7613 +/- 0.2134	0.7963 +/- 0.1720

图 5 展示了不同算法在测试集上的可视化结果。由于仿射变换在数据集预处理时已经完成，图中的浮动图像即为仿射变换的结果。由白色框显示的解剖结构可以看出，本文方法得到的配准后的图像结构上与统一作为参考的固定图像最为相似。

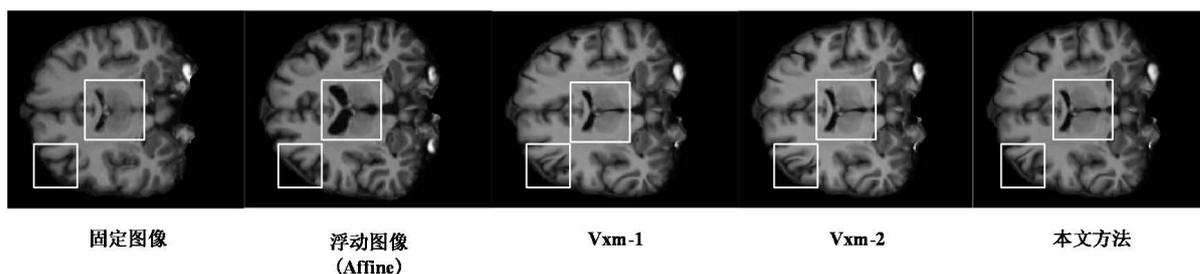


Figure 5. Registration result visualization
图 5. 配准结果可视化

5. 结束语

本文提出了一种基于深度学习的脑部单模态核磁共振图像的无监督配准方法。利用经形变场扭曲后的浮动图像与固定图像的差异来反向迭代优化卷积神经网络中的参数，省去人工标注标签的时间金钱成本。本文方法首先改进了卷积神经网络的连接方式，设计了密集连接的 U-Net，这是一种类似于 U-Net 的 U 型网络，在编码器与解码器之间部署短连接，可以解决相连接的两个卷积层因采样深度差异较大而造成的语义差距较大的问题，同时也保留了长连接能表征距离较远的两个像素之间关系的优势。本文方法还在解码器上采样阶段设计了通道注意力机制，通过特征重校准，能有效的强调有效特征，抑制图像重建时的噪音，进而在后续生成配准后图像时，提升配准效果。实验结果表明，本文提出的无监督图像配准算法在 DSC 系数上优于所对比的 Affine、Voxelmorph 方法。

参考文献

- [1] 谢姝婷. 基于深度学习的医学图像配准与分割研究[D]: [硕士学位论文]. 上海: 华东师范大学, 2022.
- [2] Fu, Y., Lei, Y., Wang, T., et al. (2020) Deep Learning in Medical Image Registration: A Review. *Physics in Medicine & Biology*, 65, 20TR01. <https://doi.org/10.1088/1361-6560/ab843e>
- [3] Rohé, M.-M., Datar, M., Heimann, T., et al. (2017) SVF-Net: Learning Deformable Image Registration Using Shape

- Matching. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec City, 11-13 September 2017, 266-274. https://doi.org/10.1007/978-3-319-66182-7_31
- [4] De Vos, B.D., Berendsen, F.F., Viergever, M.A., *et al.* (2019) A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration. *Medical Image Analysis*, **52**, 128-143. <https://doi.org/10.1016/j.media.2018.11.010>
- [5] Li, H. and Fan, Y. (2018) Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data. *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington DC, 4-7 April 2018, 1075-1078. <https://doi.org/10.1109/ISBI.2018.8363757>
- [6] Yang, X., Kwitt, R. and Niethammer, M. (2016) Fast Predictive Image Registration. In: *Deep Learning and Data Labeling for Medical Applications*, Springer, Berlin, 48-57. https://doi.org/10.1007/978-3-319-46976-8_6
- [7] Sokooti, H., Vos, B.D., Berendsen, F., *et al.* (2017) Nonrigid Image Registration Using Multi-Scale 3D Convolutional Neural Networks. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Quebec City, 11-13 September 2017, 232-239. https://doi.org/10.1007/978-3-319-66182-7_27
- [8] Balakrishnan, G., Zhao, A., Sabuncu, M.R., *et al.* (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] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [10] Jaderberg, M., Simonyan, K. and Zisserman, A. (2015) Spatial Transformer Networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Volume 2, 2017-2025.
- [11] Tian, F., Gao, Y., Fang, Z., *et al.* (2021) Depth Estimation Using a Self-Supervised Network Based on Cross-Layer Feature Fusion and the Quadtree Constraint. *IEEE Transactions on Circuits and Systems for Video Technology*, **32**, 1751-1766. <https://doi.org/10.1109/TCSVT.2021.3080928>
- [12] Gao, Y., Liu, X., Li, J., *et al.* (2022) LFT-Net: Local Feature Transformer Network for Point Clouds Analysis. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2022.3140355>
- [13] Oh, D., Kim, B., Lee, J., *et al.* (2021) Unsupervised Deep Learning Network with Self-Attention Mechanism for Non-Rigid Registration of 3D Brain MR Images. *Journal of Medical Imaging and Health Informatics*, **11**, 736-751. <https://doi.org/10.1166/jmih.2021.3345>
- [14] Wang, Q., Wu, B., Zhu, P., *et al.* (2019) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 11532-11542. <https://doi.org/10.1109/CVPR42600.2020.01155>
- [15] Marcus, D.S., Wang, T.H., Parker, J., *et al.* (2007) Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, **19**, 1498-1507. <https://doi.org/10.1162/jocn.2007.19.9.1498>
- [16] Fischl, B. (2012) FreeSurfer. *Neuroimage*, **62**, 774-781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- [17] Collins, D.L., Zijdenbos, A.P., Kollokian, V., *et al.* (1998) Design and Construction of a Realistic Digital Brain Phantom. *IEEE Transactions on Medical Imaging*, **17**, 463-468. <https://doi.org/10.1109/42.712135>
- [18] Avants, B.B., Tustison, N.J., Song, G., *et al.* (2011) A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration. *Neuroimage*, **54**, 2033-2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>