

DOI:10.19651/j.cnki.emt.2105857

# 基于 VV-Net 的三维医学图像配准

李姗姗 张娜娜 张媛媛 丁维奇

(昆明理工大学 信息工程与自动化学院 昆明 650500)

**摘要:** 三维医学图像配准算法被广泛应用于科学研究和随访等医学场景,提高其配准精度具有重要的意义。针对医学图像配准问题,提出一种基于 V-Net 的 V 形网络(VV-Net),该配准模型可以通过堆叠 V-Net 进行端到端的训练。具体的说,移动图像经过两个 V-Net 依次进行扭曲,使用额外的 V-Net 为前两个 V-Net 提供补充信息,共同构成 V 形网络,使移动图像与固定图像更好的对齐。同时,对提出的模型增加深度监督辅助分支防止过拟合。采用上述渐进配准与信息补充提高配准对之间的配准精度。通过 ADNI、ABIDE、ADHD200 和 OASIS 四个数据集评估模型性能。以 ADNI 数据集为例,所提出的配准方法与仿射变换(Affine)、对称归一化(SyN)以及体素变形(VoxelMorph)比较分别获得 24.7%、13.2%以及 1.3%的精度提升。实验结果表明,VV-Net 在医学图像配准领域取得了良好效果。

**关键词:** 深度学习;医学图像;图像配准;改进 V-Net

中图分类号: TP391 文献标识码: A 国家标准学科分类代码: 520.6040

## 3D medical image registration based on VV-Net

Li Shanshan Zhang Na'na Zhang Yuanyuan Ding Weiqi

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** 3D medical image registration algorithm is widely used in scientific research, follow-up and other medical fields, so it is of great significance to improve its registration accuracy. Aiming at the problem of medical image registration, a V-shape network based on V-Net(VV-Net) is proposed. The registration model can be trained end-to-end by stacking V-Net. Specifically, the moving image is distorted by two V-Nets in turn, and the additional V-Net is used to provide supplementary information for the first two V-Nets to form a V-shaped network, so that the moving image can be better aligned with the fixed image. At the same time, the depth supervision auxiliary branch is added to the proposed model to prevent over fitting. The accuracy of registration is improved by using the progressive registration and information supplement. The performance of the model is evaluated by ADNI, ABIDE, ADHD200 and OASIS data sets. Compared with Affine transformation, symmetric normalization (SyN) and VoxelMorph, the proposed registration method achieves 24.7%, 13.2% and 1.3% accuracy improvements, respectively. The experimental results show that VV-Net has achieved good results in the field of medical image registration.

**Keywords:** deep learning; medical image; image registration; improved V-Net

## 0 引言

三维医学图像配准在术中规划以及疗效评估等方面有着重要的应用价值。目前为止,三维医学图像配准已经得到了大量的研究,传统的配准方法有文献[1-2]等。随着深度学习在计算机视觉中的应用,基于深度学习的医学图像配准方法也取得了诸多令人印象深刻的成果。在基于深度学习的方法中,文献[3-5]中提出的方法属于有监督的医学图像配准方法,人工标注的真实变形场的质量以及人力消耗是该类方法需要攻克的难题。文献[6-11]中提出的方法

在训练的过程中无需人工标注的真实变形场,但是部分方法如文献[6-7]在训练的过程中只支持小的变换,影响配准的精度。同时,医学图像配准的最新神经网络模型是编解码结构的变体,如 FCN<sup>[12]</sup>和 U-Net<sup>[13]</sup>。然而这些模型仅仅是进行简单的预测,在处理复杂变形的过程中具有一定的局限性,因此配准的精度有待提高。

基于此,提出一种基于 V-Net 的 V 形网络模型用于端到端的三维医学图像配准,该配准模型在训练的过程中无需人工标注的真实变形场,可以一次性预测整个位移向量场。该模型中使用第 1 个 V-Net 网络进行粗略的配准,

收稿日期:2021-02-28

• 117 •

第 2 个 V-Net 模型进行精细的配准,并通过延伸第 1 个 V-Net 模型的深度提取更深层次的特征为第 2 个 V-Net 提供补充信息,通过渐进配准的方式提高配准的精度。同时,通过添加深度监督辅助分支来防止网络过拟合,监督网络进行更好的优化学习。

### 1 方法与实验

所提出的三维医学图像配准方法的整体框架如图 1 所示,移动图像  $M$  和固定图像  $F$  作为整个配准模型的输入,通过配准网络 VV-Net 对参数函数  $g_\theta(F, M) = \varphi$  进行建模,其中  $\varphi$  是预测的位移向量场,  $\theta$  是函数  $g$  的可学习参数。在训练的过程中,通过最小化损失函数  $L$ ,即通过最小化扭曲图像  $M(\varphi)$  和固定图像  $F$  之间的相似性损失以及位移向量场  $\varphi$  的正则化损失更新  $\theta$ 。公式如式(1)所示。

$$\bar{\theta} = \arg \min_{\theta} (L_{sim}(F, M(g_\theta(F, M))) + \lambda L_{smooth}(g_\theta(F, M))) \tag{1}$$

其中,对于每个体素  $p \in R^3$ ,  $\varphi(p)$  使  $F(p)$  和  $M(\varphi(p))$  定义相似的解剖位置。  $M(\varphi)$  为通过空间变换网络(STN)对移动图像  $M$  施加位移向量场  $\varphi$  扭曲后的图像,  $L_{sim}(\cdot, \cdot)$  度量扭曲图像  $M(\varphi)$  和固定图像  $F$  之间的相似性,  $L_{smooth}(\cdot)$  表示对变形场  $\varphi$  施加正则化,  $\lambda$  是正则化参数。当给定一组新的待配准的医学图像对,可以通过学习到的函数参数  $\theta$  快速的预测位移向量场  $\varphi$ 。

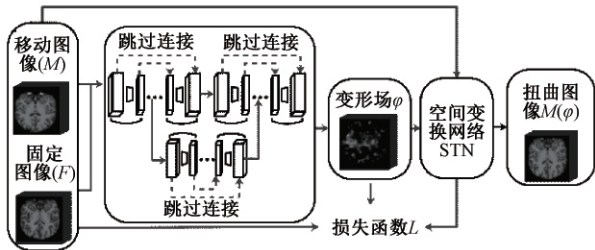


图 1 三维医学图像配准的整体架构

#### 1.1 VV-Net 模型

本文使用基于 V-Net 的 V 形网络模型作为可变形配准网络。整个模型的输入为三维医学图像对(包括 1 个固定图像  $F$  和 1 个移动图像  $M$ ),输入图像的大小为  $160 \times 192 \times 224$ ,模型的最终输出为预测的位移向量场  $\varphi$ ,即一个具有 3 个通道( $x, y, z$  位移)的三维特征图,其大小与输入相同。如图 2 所示, VV-Net 中共包含 3 个两层的 V-Net。每个 V-Net 都由 1 个具有提取特征的编码器和 1 个用于重建图像的解码器组成。固定图像  $F$  和移动图像  $M$  级联作为第 1 个 V-Net 的输入,编码器的第 1 层包含 1 个标准的  $3 \times 3 \times 3$  卷积层和 1 个步长为 2 的  $3 \times 3 \times 3$  跨步卷积层,第 2 层包含 1 个卷积层数为 2 的残差单元和 1 个步长为 2 的  $3 \times 3 \times 3$  跨步卷积层。两个跨步卷积层使图像的分辨率分别为原始输入图像的  $1/2, 1/4$ 。接着,使用具有两个卷积层的残差单元连接编解码底端,构成瓶颈模块。在解码

器端,使用上采样操作来恢复空间分辨率,与编码阶段类似,上采样过程层重复两次,第 1 个上采样过程层包括 1 个 Upsampling 上采样层和 1 个卷积层数为 2 的残差单元,第 2 个上采样过程层包括 1 个 Upsampling 上采样层和 1 个标准的  $3 \times 3 \times 3$  卷积层。第 1 个 V-Net 网络的输出为粗略配准后的图像  $M(\varphi^1)$ ,  $M(\varphi^1)$  与固定图像  $F$  级联共同构成第 2 个 V-Net 的输入,第 2 个 V-Net 的网络结构与第 1 个 V-Net 的网络结构相同。特别的,从第 1 个 V-Net 的瓶颈模块进行延伸提取更深层次的特征,形成第 3 个 V-Net。第 3 个 V-Net 的网络结构与前两个 V-Net 的网络结构相同,两个步长为 2 的跨步卷积分别使图像的分辨率为原始输入图像的  $1/8, 1/16$ ,在解码阶段使用两个上采样恢复空间分辨率至原始输入图像的  $1/4$ 。将提取的复杂特征与第 2 个 V-Net 编码器提取的特征进行融合,一起送入第 2 个 V-Net 的瓶颈模块。第 2 个 V-Net 输出的位移向量场  $\varphi$  为模型最终预测的位移向量场。

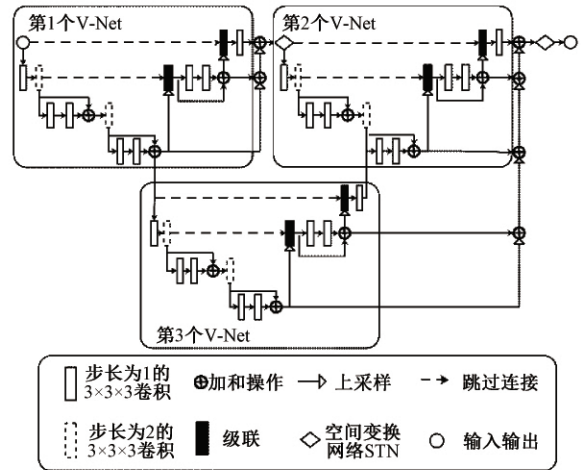


图 2 可变形配准网络 VV-Net

为了更好地训练配准网络,在 VV-Net 中添加了深度监督辅助分支,从第 1 个 V-Net 的底部依次提取不同分辨率层次的阶段性预测结果进行融合,监督第 1 个 V-Net 网络。依次提取第 3 个和第 2 个 V-Net 的阶段性预测成果进行融合作为最终生成层的输入,以此监督整体 VV-Net 模型更好的优化,防止过拟合。

#### 1.2 损失函数

空间变换网络(STN)可以使一个图像重建为另一个图像,所提出的方法使用随机梯度下降最小化期望损失函数  $L$ ,在训练的过程中无需人工标注的真实变形场,测量空间变换网络扭曲运动图像  $M$  得到的扭曲图像  $M(\varphi)$  和固定图像  $F$  的相似性,并引入正则化损失以防止预测变形场不连续或者是过拟合,从而找到最佳参数  $\theta$ 。因此,模型的损失函数由两部分组成:一部分是扭曲图像  $M(\varphi)$  与固定图像  $F$  之间的相似性度量  $L_{sim}$ ,另一部分是施加于位移向量场  $\varphi$  的正则化损失项  $L_{smooth}$ ,相似性度量惩罚外观上的

差异,正则化损失项惩罚位移向量场  $\varphi$  的局部空间变化。损失函数  $L$  计算如式(2)所示。

$$L(F, M, \varphi) = L_{sim}(F, M(\varphi)) + \lambda L_{smooth}(\varphi) \quad (2)$$

其中,  $\lambda$  是正则化参数。

模型使用均方误差 MSE 作为相似性度量,如式(3)所示。

$$MSE(F, M(\varphi)) = \frac{1}{\Omega} \sum_{p \in \Omega} \|F(p) - M(\varphi(p))\|^2 \quad (3)$$

其中,  $\Omega$  表述输入图像的空间域,  $p$  表示空间域内的体素。

最小化  $L_{smooth}$  将鼓励  $M(\varphi)$  近似  $F$ , 但可能产生不连续的  $\varphi$ 。因此在其空间梯度上使用扩散正则化器平滑  $\varphi$ , 如式(4)所示。

$$L_{smooth}(\varphi) = \sum_{p \in \Omega} \|\nabla \varphi(p)\|^2 \quad (4)$$

其中,  $\Omega$  表述输入图像的空间域,  $p$  表示空间域内的体素,  $\nabla \varphi$  是相邻体素之间的差异。

### 1.3 数据集

使用 4 个公共的医学影像数据集 ADNI、ABIDE、ADHD200 以及 OASIS 中的 T1 加权脑 MRI 数据对提出的模型进行评估。每个数据集的采集细节、受试者年龄范围和健康状况都不同。为了规范配准网络的输入以及促进配准网络的优化,对数据集中的数据进行了预处理。首先,将数据集中的 MRI 扫描重采样到  $256 \times 256 \times 256$  的网格,体素间距为  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ 。然后对数据做标准的预处理,使用 FreeSurfer<sup>[14]</sup> 对获取到的数据进行颅骨剥离,将图像尺寸裁剪为  $160 \times 192 \times 224$ ,对裁剪后的数据进行像素值归一化处理,此处像素的归一化通过将所有像素值除以最大像素值来实现,最大像素值为 255,归一化后的像素取值范围从  $[0, 255]$  转化成  $[0, 1]$  区间。本文的工作主要进行的是可变形配准的研究,即非线性变换的研究,在进行可变形变换之前假设所使用的 MRI 数据是线性对齐的,因此,将 Affine 作为 MRI 数据的预处理步骤之一,本文使用工具 ANTs<sup>[15]</sup> 对数据进行 Affine。同时,为了能够更好地训练,使用数据增强对数据集中的数据进行了两种程度的变形以达到扩充数据集的目的。最终 ADNI、ABIDE、ADHD200 和 OASIS 四个数据集中 T1 加权脑 MRI 数据分别为 4 710、1 740、2 193、2 646 个,共 11 289 个训练数据,4 个数据集的测试集分别包含 150、120、120、120 个测试数据。

### 1.4 实验

实验在 Ubuntu16.04 操作系统下进行,使用的编程语言是 python。使用 Keras2.3.1 和 Tensorflow1.15.0 后端实现所提出的模型,硬件设施为显存 16 G 的 NVIDIA Quadro RTX 5000 GPU,内存大小为 128 G。训练过程中使用 Adam 优化器,学习率为  $1 \times 10^{-4}$ ,批量大小为每训练批一对。

## 2 结果

### 2.1 评估方法

使用 FreeSurfer<sup>[14]</sup> 对 MRI 数据进行解剖分割,使用 30 个解剖结构的体积重叠来评估提出的配准模型,表 1 所示为 30 个解剖结构。

表 1 30 个解剖结构索引表

索引	名称	索引	名称
1	左侧大脑白质	16	左侧大脑腹侧
2	左侧大脑皮层	17	左脉络丛
3	左侧脑室	18	右侧大脑白质
4	左侧小脑白质	19	右侧大脑皮层
5	左侧小脑皮层	20	右侧脑室
6	左丘脑	21	右侧小脑白质
7	左尾状核	22	右侧小脑皮层
8	左壳核	23	右丘脑
9	左苍白球	24	右尾状核
10	第三脑室	25	右壳核
11	第四脑室	26	右苍白球
12	脑干	27	右海马体
13	左海马体	28	右杏仁核
14	左杏仁核	29	右侧大脑腹侧
15	脑脊液	30	右脉络丛

解剖结构之间的体积重叠使用戴斯相似性系数 (dice similarity coefficient, DSC) 来量化,当预测的位移向量场  $\varphi$  越准确,则扭曲图像  $M(\varphi)$  和固定图像  $F$  之间相同解剖结构的体积重叠区域就越大。设  $A^i$  和  $B^i$  分别为固定图像  $F$  和扭曲图像  $M(\varphi)$  的第  $i$  个解剖结构体素集,则  $A^i$  和  $B^i$  之间的 DSC 计算如式(5)所示。

$$DSC(A^i, B^i) = 2 \cdot \frac{|A^i \cap B^i|}{|A^i| + |B^i|} \quad (5)$$

固定图像  $F$  和扭曲图像  $M(\varphi)$  重叠的体积越大解剖结构的平均 DSC 的数值越大,固定图像  $F$  和扭曲图像  $M(\varphi)$  完全重叠则两者之间解剖结构的平均 DSC 为 1。

### 2.2 评估结果

为了评估所提出模型的性能,将本文提出的方法与最新的传统配准算法(包括基于 ANTs 的 Affine 和 SyN)以及 VoxelMorph<sup>[8]</sup> 进行了比较。表 2 所示为 Affine、传统的可变形图像配准方法 SyN、VoxelMorph 以及本文提出的模型在 30 个解剖结构上的平均 DSC,由表 2 中的结果可以看出,与前 3 种方法相比,本文提出的模型在 ADNI、ABIDE、OASIS 以及 ADHD200 四个数据集的测试集上的配准精度都达到了最先进的性能。

接下来进一步的展示 ADNI 测试集中 30 个解剖结构体积重叠的 DSC,将表 1 中所有解剖结构对应的 DSC 可视

表 2 平均 DSC

	Affine	SyN	VoxelMorph	本文
ADNI	0.486 (0.202)	0.601 (0.224)	0.720 (0.169)	0.733 (0.163)
ABIDE	0.590 (0.201)	0.592 (0.239)	0.677 (0.185)	0.685 (0.184)
OASIS	0.574 (0.203)	0.601 (0.240)	0.674 (0.188)	0.678 (0.190)
ADHD200	0.581 (0.218)	0.589 (0.247)	0.715 (0.175)	0.727 (0.170)

化为图 3 中的直方图。由图 3 可以看出,在 30 个解剖结构中,本文提出的方法有 26 个解剖结构的 DSC 表现出最佳性能。

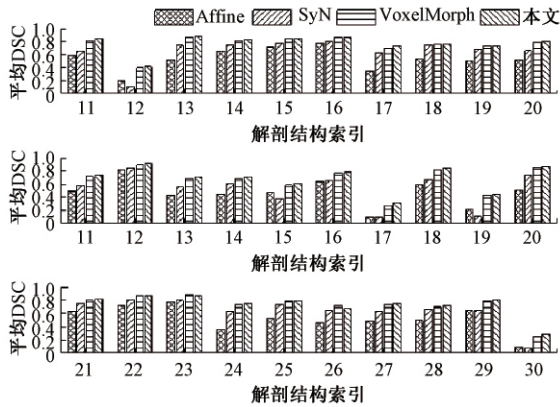


图 3 Affine, SyN, VoxelMorph 以及本文所提出的方法配准图像与固定图像 30 个解剖结构的平均 DSC 直方图

图 4 所示为上述方法在大脑数据集 ADNI 中的一个示例上的可视化结果。本节中描述大脑的图形都显示为二维切片,仅用于可视化目的,所有配准都是在三维中完成的,切片的位置为[80, 96, 112]。

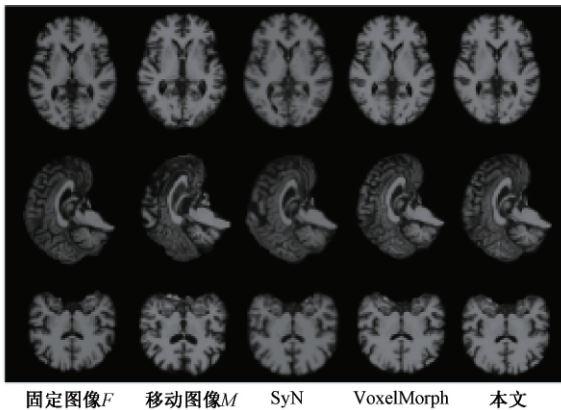


图 4 配准结果可视化

### 3 结 论

本文提出一种用于可变形医学图像配准的 VV-Net 网

络,并利用 ADNI、ABIDE、ADHD200 以及 OASIS 四个数据集对其配准精度进行评估,评估的度量指标为 DSC。VV-Net 网络中第 2 个 V-Net 对第 1 个 V-Net 输出的扭曲图像进行进一步的位移向量场预测,通过两个 V-Net 渐进的预测位移向量场,并通过延申第 1 个 V-Net 的深度为第 2 个 V-Net 提供复杂的特征,这使得每个图像对之间的配准精度更高。实验结果表明,所提出的方法在配准精度方面达到了最佳的性能,以 ADNI 测试集上的结果为例,与最新的传统配准算法(包括 Affine 和 SyN)以及 VoxelMorph 比较分别获得 24.7%、13.2%以及 1.3%的精度提升。由于将图像配准到同一模板图像是大多数医学图像分析方法(如 atlas 比对)的一个重要预处理部分,因此该模型的重点是将目标图像配准到一个固定的模板图像上,未来将致力于将目标图像配准到不同模板图像上的研究。

### 参考文献

- [1] 李冠宇,汪友生.一种改进的 SIFT 血管图像特征匹配算法[J].电子测量技术,2015,38(12):63-66.
- [2] 张年,钱盛友,谭乔来,等.基于 SURF 算法的 HIFU 治疗前后超声图像配准研究[J].电子测量与仪器学报,2016,30(6):917-922.
- [3] ROHÉ M M, DATAR M, HEIMANN T, et al. SVF-Net: Learning deformable image registration using shape matching [C]. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, Berlin: Springer, 2017:266-274.
- [4] CAO X H, YANG J H, ZHANG J, et al. Deformable image registration based on similarity-steered CNN regression [C]. Medical Image Computing and Computer Assisted Intervention-MICCAI 2017, Berlin: Springer, 2017:300-308.
- [5] EPPENHOF K A J, PLUIM J P W. Pulmonary CT registration through supervised learning with convolutional neural networks[J]. IEEE Transactions on Medical Imaging, 2019, 38(5):1097-1105.
- [6] VOS B D, BERENDSEN F F, VIERGEVER M A, et al. End-to-end unsupervised deformable image registration with a convolutional neural network[M].Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, 2017: 204-212.
- [7] FAN J, CAO X, YAP P T, et al. BIRNet: Brain image registration using dual-supervised fully convolutional networks[J]. Medical Image Analysis, 2018, 54(5):193-206.
- [8] BALAKRISHNAN G, ZHAO A, SABUNCU M R, et al. An unsupervised learning model for deformable

- medical image registration[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Silver Spring, MD: IEEE Computer Society, 2018: 9252-9260.
- [9] 程章培. 基于深度学习的 3D 生物医学脑图像配准算法研究[D]. 合肥: 安徽大学, 2020.
- [10] 马英君. 基于无监督学习的三维医学图像配准的研究[D]. 济南: 济南大学, 2020.
- [11] YANG T J, TANG Q, LI L, et al. Non-rigid medical image registration using multi-scale residual deep fully convolutional networks[J]. Journal of Instrumentation, 2021, 16(3): P03005.
- [12] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [13] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Berlin: Springer, 2015: 234-241.
- [14] FISCHL B. FreeSurfer[J]. NeuroImage, 2012, 62(2): 774-781.
- [15] AVANTS B B, TUSTISON N J, SONG G, et al. A reproducible evaluation of ants similarity metric performance in brain image registration [J]. NeuroImage, 2011, 54(3): 2033-2044.

#### 作者简介

李姗姗, 硕士研究生, 主要研究方向为图像处理、医学图像配准、深度学习等。

E-mail: shanshan\_l@163.com

张娜娜, 硕士研究生, 主要研究方向为图像处理。

E-mail: 18838987572@163.com

张媛媛, 硕士研究生, 主要研究方向为图像处理。

E-mail: 1750162458@qq.com

丁维奇, 硕士研究生, 主要研究方向为图像处理、深度学习等。

E-mail: tobias.ding@hotmail.com