

华中科技大学

医学图像配准研究综述

专业班级

CS1904

姓 名

谭佳浩

学 号

U201915048

2023 年 4 月

目录

1 简介	2
概述	2
配准简介	2
传统的配准方法和基于学习的配准方法	3
2 具体背景技术概述	4
配准网路整体框架	4
U-Net	4
形变场	6
空间转换网络 (Spatial Transformer Network, STN)	6
损失函数	7
NCC	8
MSE	8
评估指标-dice	9
3 国内外研究现状	10
有监督的基于学习的配准方法	10
无监督的基于学习的配准方法	11
发展趋势	13
4 无监督的基于学习的配准文献介绍	14
《VoxelMorph: A Learning Framework for Deformable Medical Image Registration》	14
《Dual-Stream Pyramid Registration Network》	16
《A FCN-based Unsupervised Learning Model for Deformable Chest CT Image Registration》	18
《Recursive Cascaded Networks for Unsupervised Medical Image Registration》	20
总结	22
5 VoxelMorph 配准实例	24
6 未来展望	27
参考文献	28

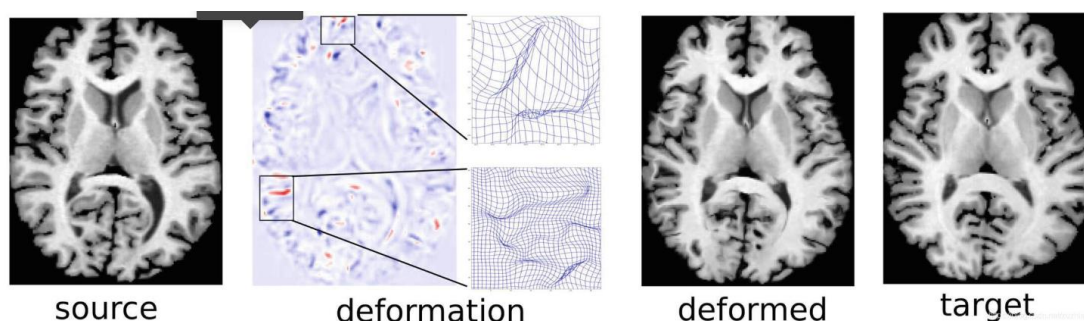
1 简介

概述

医学图像配准是一项重要的技术，旨在将不同的医学图像数据集合并，以便在医疗诊断、治疗和手术规划等方面提供更准确的信息。传统医学图像配准方法需要预先确定特征点，难以确定具有代表性和鲁棒性的特征点，而且医学图像存在噪声和变形问题，传统方法容易受到影响，导致配准精度下降。深度学习在图像处理领域中已经取得很多成功的应用，可以将其应用到医学图像配准中。深度学习的医学图像配准算法可以用于不同时间点或设备采集到的医学图像的准确对齐，同时也符合实际需求。此外，可以从其他领域的深度学习图像配准算法中借鉴，并结合医学图像的特点进行改进。深度学习方法的计算复杂度较低，可提高配准精度和实时性，因此具有广阔的应用前景。本文将介绍几种常见的医学图像配准方法，分析其优缺点，并展望其未来发展。

配准简介

配准问题简单来说就是让一张图像对齐到另一张图像，使得对齐后的图像尽可能相似。即，给定一个浮动图像（moving image） IM 和一张固定图像（fixed image） IF 。预测一个位移场 u ，进而得到形变场（deformation field） ϕ ，即从浮动图像到固定图像的映射，使得配准后的浮动图像（warped image） $IW = IM \circ \phi$ 和固定图像尽可能相似。其中像素位置 x 的变换可以表示为 $\phi(x) = x + u(x)$ 。浮动图像又称源图像（source image），固定图像又称参考图像（reference image）、模板图像（template image），形变场又称流场（flow field）、配准场（registration field）、转换（transformation）。



图一 上图分别是源图像、形变场、配准后的源图像和目标图像

传统的配准方法和基于学习的配准方法

配准技术是医学图像处理领域中的一个重要问题，它用于将不同的医学图像进行对齐以便进行后续的数据分析和处理。根据配准算法所依赖的方法不同，配准可以分为传统配准和基于学习的配准两种方法。

传统配准方法基于数学优化，需要对每个图像进行迭代优化，因此耗费的时间较长，但通常效果比较好且稳定，医生一般都采用的是传统的方法。而基于学习的配准方法则是通过神经网络训练的配准方法，该方法中参数是共享的。首先利用大量的数据来训练一个模型，然后用这个训练好的模型对一个新的图像进行配准。这种方式的优点在于虽然训练过程较为缓慢，但配准（测试）过程比传统方法快很多。缺点则在于用作训练的数据较少（特别是有监督的标签数据），另一方面可解释性也许也是一个问题。

基于学习的方法可以进一步分为有监督的配准方法和无监督的配准方法。在有监督的配准方法中，监督数据主要是通过传统配准方法得到的形变场作为 ground truth 或对一张图像进行随机模拟变形的方式来获得。然而，有监督的配准方法的缺点在于利用的高质量有监督数据较少，而手工标注的有监督数据既耗时又耗力，不现实。因此，无监督的配准方法成为最近的研究热点，因为配准本质上就是一个无监督的任务，只需要优化两张图像之间的相似度就可以完成。这种模型的优点在于不需要指定特定的图像相似度，从而避免了由于数据集分布的不同而导致固定的图像相似度衡量标准的通用性问题。本文也将着重介绍无监督基于学习的配准方法的文献。

医学图像配准的目标是找到一个变换，将一组图像中的每个像素从一个坐标系转换到另一个坐标系，以最大程度地匹配解剖结构。这个变换可以是刚性的（旋转和平移）或非刚性的（弯曲和扭曲），取决于图像之间的差异程度。医学图像配准可以用于很多医学应用，例如手术导航、病灶分割、疾病进展监测等。

本章将重点介绍传统医学图像配准网络框架中的关键技术和原理，以为医学影像领域的相关研究和实践提供参考和借鉴。

配准网路整体框架

基本的配准网络框架是将输入的浮动图像和固定图像在通道（channel）的维度进行拼接，然后经过一个由仿射配准网络和卷积神经网络组成的网络模型。

首先，通过仿射配准网络对浮动图像和固定图像进行仿射配准，得到一个初始的配准结果。然后，通过卷积神经网络（通常为类 U-Net 的编码器-解码器网络）预测一个从浮动图像到固定图像的位移场，该位移场表示每个像素点在固定图像中对应的位置。根据位移场，可以得到一个采样网格，该网格将浮动图像中的像素点映射到固定图像中的对应位置。

接下来，使用空间转换网络（Spatial Transformer Network, STN）来利用该采样网格对浮动图像进行重采样，得到配准后的图像。STN 网络允许非线性采样，从而可以更准确地重构配准后的图像。

该网络框架可以通过训练数据进行优化，使得网络模型能够自动学习如何将浮动图像与固定图像进行准确的配准。该框架的优点是灵活性高，可以适应不同类型的配准任务，并且能够同时进行仿射和非仿射配准。此外，该框架还可以通过增加网络层数和加入其他的图像处理模块来进一步提高配准精度和鲁棒性。

U-Net

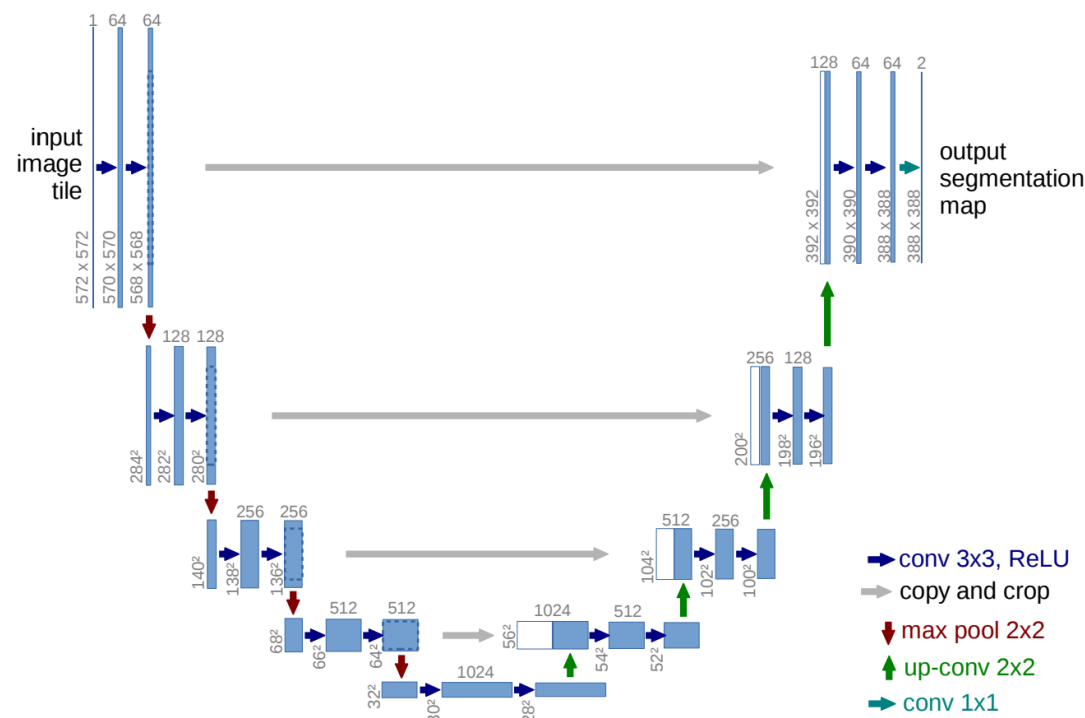
U-Net 是一种常用于图像分割任务的深度学习网络，其结构由一个编码器和一个解码器组成。编码器通过逐步下采样图像的方式获取图像的高层次特征，而解码器则通过上采样和跳跃连接的方式将这些特征重新恢复到原始图像的分辨

率，并输出对每个像素点进行分类的概率。U-Net 的网络结构如图 1 所示。

在编码器部分，U-Net 使用了四个阶段，每个阶段都包含一个最大池化层和两个卷积层。最大池化层用于下采样图像，卷积层用于提取图像特征。这四个阶段的输出特征图逐渐变小，对应着不同尺度的图像语义信息。这些特征图被传递给解码器部分进行上采样和恢复。

在解码器部分，U-Net 使用了四个阶段，每个阶段都包含一个上采样层和两个卷积层。上采样层将特征图的分辨率扩大，与跳跃连接相结合，能够帮助解码器恢复更为准确的细节信息。跳跃连接是指将编码器和解码器在相同尺度上的特征图进行连接，让解码器可以获得更多的低层次特征信息。

在 U-Net 的最后，使用一个大小为 3x3 的卷积层对每个像素进行分类，输出对应的概率。



图二 U-Net 结构图

在配准网络结构中使用 U-Net 主要是为了实现形变场的预测。具体而言，U-Net 通过对固定图像和浮动图像进行编码和解码，预测出了一个形变场，该形变场表示浮动图像到固定图像的空间变换。这个形变场包含了每个像素点在固定图像中的位置，因此可以用于对浮动图像进行重采样，进而实现配准的效果。

在配准中，U-Net 的输入是固定图像和浮动图像的拼接图像，即将固定图像

和浮动图像在通道维度进行拼接得到的图像。U-Net 的输出是一个形变场，其尺寸与拼接图像的尺寸相同。由于 U-Net 使用了跳跃连接机制，可以将编码器中的低层次特征图与解码器中的高层次特征图进行融合，进而提高形变场的预测精度。最终，使用空间变换网络（STN）根据形变场对浮动图像进行每个像素的重采样，得到配准后的图像。

形变场

在图像配准中，形变场（deformation field）是一个向量场，用于描述一个图像中每个像素点相对于另一个图像中对应像素点的位置偏移。形变场包含了两个部分：位移场（displacement field）和变形场（deformation field）。

位移场表示了一个像素点相对于另一个像素点在 x 、 y 、 z 方向上的位移，通常用一个三元组 (u, v, w) 来表示，其中 u 、 v 、 w 分别表示在 x 、 y 、 z 方向上的位移距离。

变形场是指一个像素点在经过位移后所发生的形变程度，通常用一个矩阵表示，也称为雅可比矩阵（Jacobian matrix）。雅可比矩阵描述了一个像素点在各个方向上的变形程度，用于描述像素点的形状的变化。

形变场的作用是将待配准图像中的像素点与固定图像中对应的像素点进行对齐。通过求解形变场，可以将待配准图像中的像素点映射到固定图像中对应的位置，从而实现图像的配准。

空间转换网络（Spatial Transformer Network, STN）

空间转换网络（Spatial Transformer Network, STN）是一种能够自适应地对输入图像进行空间变换的神经网络结构。STN 结构的核心是空间变换模块，它可以对输入图像进行平移、旋转、缩放等多种变换，从而提高神经网络对于输入图像的适应性。

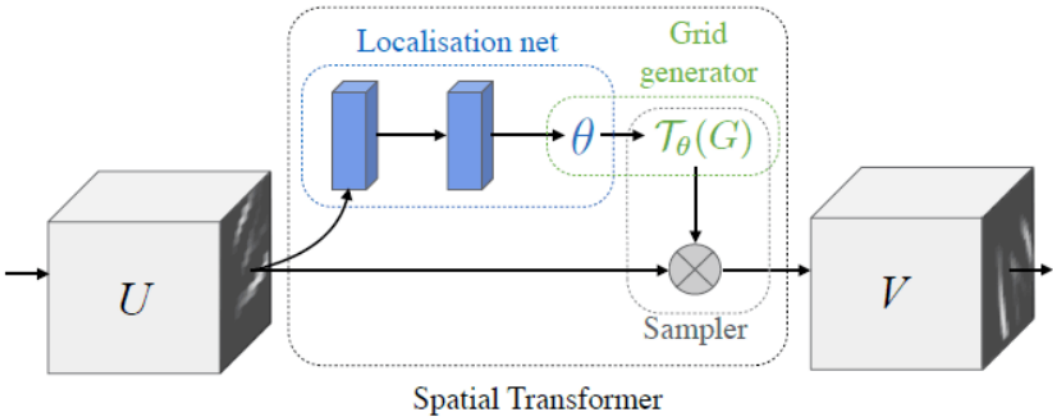
STN 结构的主要思想是利用神经网络来学习如何对输入图像进行空间变换。具体来说，STN 结构包含三个主要组件：本地网络、网格生成器和采样器。

本地网络：本地网络是一个普通的卷积神经网络，用于提取输入图像的特征。STN 结构中的本地网络通常可以自定义，也可以使用预训练的卷积神经网络。

网格生成器：网格生成器是 STN 结构中的重要组成部分，用于生成表示空间变换的网格。具体来说，网格生成器将输入的特征图映射到一个网格上，网格上的每个点对应输入图像中的一个像素点，而每个点的坐标则表示了这个像素点在输出图像中的位置。网格生成器通常由仿射变换模块和可微的采样函数组成。

采样器：采样器是 STN 结构中的最后一部分，它将通过网格生成器生成的网格上的点对应到输入图像中的像素点，从而实现对输入图像的空间变换。采样器通常使用双线性插值或者最近邻插值等方法进行像素点的采样。

通过以上三个组件的组合，STN 结构可以实现对输入图像进行空间变换的目的。STN 结构广泛应用于目标检测、图像配准、姿态估计等领域，在提高神经网络对于输入图像适应性方面发挥了重要作用。



图三 STN 网络结构图

损失函数

在配准网络中，损失函数通常由图像相似性损失和形变场平滑正则项损失两部分组成。其中，图像相似性损失是评估配准效果的重要指标，常用的指标包括相关系数 (Correlation Coefficient, CC)、归一化的相关系数 (Normalized Cross-Correlation, NCC)、互信息 (Mutual Information, MI)、均方误差 (Mean Squared Error, MSE) 等。具体选择哪种指标取决于图像的特点和配准任务的需求。当图像具有相似的灰度值分布时，通常采用 MSE 来评估灰度值的相似性；而在多模配准中，NCC 和 MI 指标会更合适。将着重介绍 NCC 与 MSE。

NCC

NCC 全称归一化互相关系数 (Normalized Cross-Correlation)，是医学图像配准中常用的相似性度量指标之一。NCC 可以评估两张图像之间的相似度，是衡量两张图像匹配度的重要标准之一。

NCC 的计算过程包括两个主要步骤：互相关和归一化。互相关是指计算两张图像的像素点之间的相关性，归一化是指将相关性归一化到一个固定的范围内，通常是 $[-1, 1]$ 或 $[0, 1]$ 之间。

在配准网络中，NCC 损失通常用于计算浮动图像与固定图像之间的相似度。NCC 损失可以表示为以下公式：

$$NCC = \frac{\sum_{i=-m/2}^{m/2} \sum_{j=-n/2}^{n/2} f(x+i, y+j) \cdot r(x+i, y+j) - m \cdot n \cdot \mu_f \cdot \mu_r}{\left\{ \left(\sum_i \sum_j f^2(x+i, y+j) - m \cdot n \cdot \mu_f^2 \right) \left(\sum_i \sum_j r^2(x+i, y+j) - m \cdot n \cdot \mu_r^2 \right) \right\}^{\frac{1}{2}}}$$

其中， I_F 和 I_M 分别是固定图像和浮动图像， i 和 j 分别表示像素点的坐标， \bar{I}_F 和 \bar{I}_M 分别是两张图像的均值。

NCC 损失的值范围在 $[-1, 1]$ 之间，当 NCC 损失为 1 时，表示两张图像完全匹配；当 NCC 损失为 -1 时，表示两张图像完全不匹配；当 NCC 损失为 0 时，表示两张图像之间的相关性为 0，即没有相关性。

NCC 损失在医学图像配准中具有很好的鲁棒性和精度，尤其适用于具有相似灰度分布的图像配准任务。但是，由于 NCC 损失需要计算像素点之间的相关性，因此计算复杂度较高，在大规模数据集上的计算可能会造成较大的计算负担。

MSE

均方误差 (Mean Square Error, MSE) 是一种衡量图像相似性的指标，广泛应用于医学图像配准领域。MSE 损失是一种像素级别的损失，其计算方式是将两个图像对应像素点之间的灰度值差的平方求和并取平均。MSE 损失计算公式如下：

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2$$

MSE 损失的优点是计算简单，易于优化。它在处理灰度值差别较小的图像时表现良好，能够有效衡量图像之间的相似度。但是，MSE 损失也有一些缺点，例如对于强度差异较大的图像，MSE 损失可能会导致对强度变化较小的区域进行过度匹配，从而影响配准的准确性。在这种情况下，通常会采用 NCC 损失或其他类型的相似度指标。

在医学图像配准中，通常会将 MSE 损失与其他损失函数组合使用，例如正则化项和形变场平滑项。通过对不同类型的损失函数进行组合，可以有效提高配准精度和鲁棒性，从而得到更加准确的配准结果。

评估指标-dice

Dice 系数 (Dice coefficient) 是一种用于评估两个集合相似性的统计量。在医学影像领域，Dice 系数通常用于评估分割算法的性能，即用于度量分割结果与真实标注之间的重合程度。

Dice 系数的计算基于两个集合之间的交集和并集，其值的范围为 0 到 1，1 表示完全匹配，0 表示没有匹配。对于两个集合 A 和 B，Dice 系数可以表示为 $2|A \cap B| / (|A| + |B|)$ ，其中 $|A|$ 和 $|B|$ 分别表示集合 A 和 B 的大小， $|A \cap B|$ 表示 A 和 B 的交集大小。

在医学影像配准中，Dice 系数常被用于评估算法分割结果和真实标注的重合程度，即计算分割结果和标注结果之间的 Dice 系数。通常，Dice 系数越高，表示算法配准结果与真实标注越相似，算法的性能越好。Dice 系数的优点是对二值图像和多类别图像都适用，并且可以用于评估配准算法的精度和稳定性。

3 国内外研究现状

根据深度学习的种类划分,可以将医学图像配准方法分为基于监督学习和基于非监督学习两大类。

有监督的基于学习的配准方法

基于监督学习的配准方法需要提供与配准对相对应的真实变形场作为标签。在训练阶段,先以两幅图像对应坐标为中心点进行切块,将图像块输入深度学习网络,网络输出为图像块中心点对应的变形向量。在测试阶段,对待配准图像对进行采样,输入网络,把预测的变形向量综合成变形场,再利用预测的变形场对移动图像进行插值,即得配准图像。基于非监督学习的配准方法则不需要提供标签,只需要提供配准对。将配准对输入网络,获得变形场,对移动图像进行变形插值,即得配准图像。

2017 年,Gil Elbaz^[1]等人提出了一种基于深度学习的有监督点云配准方法。该方法的工作原理是利用深度学习网络学习点云之间的局部几何特征,从而实现点云的配准。具体来说,该方法首先使用 RGB-D 相机获取多个点云,并利用 ICP 算法进行粗略的配准。然后,将配准后的点云划分为小块,并提取每个小块的局部几何特征。接着,利用深度学习网络学习这些局部几何特征之间的相似性,并输出每个小块的描述子。在测试阶段,输入两个点云到深度学习网络中,网络将输出每个点的描述子,然后使用描述子进行点云的配准。

2017 年,P. Mewe^[2]等人提出一种基于有监督学习的医学图像配准方法。该方法的工作原理是利用深度学习网络学习医学图像之间的相似性,并输出一个变换矩阵,将移动图像变换到固定图像的坐标系下,从而实现图像的配准。具体来说,该方法使用了一种基于卷积神经网络的全卷积网络 (FCN),并引入了扩张卷积 (Dilated Convolution) 来提高网络的感受野,从而更好地捕捉图像之间的相似性。在训练阶段,使用一个损失函数来衡量变换矩阵的精度,并通过反向传播算法来更新网络参数。在测试阶段,输入两个医学图像到深度学习网络中,网络将输出一个变换矩阵,将移动图像变换到固定图像的坐标系下,从而实现图像的配准。该方法试图解决的问题是医学图像之间的配准问题,通过深度学习网络学习医学图像之间的相似性,从而实现医学图像的配准。该方法的优点是具有较高

的配准精度和较快的速度，同时可以处理 2D 和 3D 医学图像。但是，该方法需要大量的训练数据，且对于图像的噪声和变形敏感。

2019 年，Yue Wang^[3]等人提出了一种自监督学习的方法来进行部分到部分的匹配，以实现医学图像的自动对准。该方法使用了一个基于深度学习的自编码器来学习医学图像的特征，并使用这些特征进行医学图像的配准。该方法的机制是通过最小化医学图像的特征之间的差异来训练自编码器，以实现医学图像的自动对准。该方法试图解决的问题是医学图像的部分到部分的匹配问题，这是一个非常困难的问题，因为医学图像通常是非常复杂和多变的，需要考虑到很多因素，如缩放、旋转、形变等。该方法解决了医学图像部分到部分的匹配问题，可以自动对准医学图像，从而提高医学图像分析和诊断的准确性和效率。但是，该方法存在一些不足之处，如对于一些复杂的医学图像，可能无法实现完美的配准，需要进一步优化算法。

2021 年，Balakrishnan, G.^[4]等人提出一个半监督学习的方法，其中有标记和无标记的图像都被用来训练模型，该方法的工作原理是利用一个基于深度学习的神经网络来进行图像配准（registration）任务。为了缓解标记数据的不足问题，该方法使用了半监督学习框架，并引入了一个元正则化（Meta-Regularization）机制来进一步提高模型的鲁棒性和泛化能力。具体而言，元正则化可以看作是一种元学习方法，通过在训练过程中对模型的正则化进行学习来提高模型在未见过的数据上的性能。

这个方法试图解决的问题是图像配准的精度和鲁棒性问题，特别是在缺乏大量标记数据的情况下。该方法通过引入元正则化机制和半监督学习框架，解决了标记数据不足和模型泛化能力差的问题。该方法在多个公共数据集上进行了验证，展示了比现有方法更好的性能。

该方法存在的一些不足之处是它需要较长的训练时间和高昂的计算资源。此外，在某些情况下，该方法可能会受到半监督学习的限制，因为它需要同时训练标记和无标记的数据，可能会受到无标记数据的质量影响。

无监督的基于学习的配准方法

2018 年，Li. Y.^[5]等人提出一种无监督学习医学图像配准方法该方法的工作

原理是通过对医学图像进行非刚性配准来实现医学图像的自动对准，以便更好地进行医学图像分析和诊断。该方法使用了一个基于卷积神经网络（CNN）的自编码器（autoencoder）来提取医学图像的特征，并使用这些特征进行医学图像的配准。该方法的机制是通过最小化配准后的图像的像素值之间的差异来训练 CNN 自编码器，以实现医学图像的自动对准。该方法试图解决的问题是医学图像的非刚性配准问题，这是一个非常困难的问题，因为医学图像通常是非常复杂和多变的，需要考虑到很多因素，如缩放、旋转、形变等。该方法解决了医学图像非刚性配准的问题，可以自动对准医学图像，从而提高医学图像分析和诊断的准确性和效率。但是，该方法存在一些不足之处，如对于一些复杂的医学图像，可能无法实现完美的配准，需要进一步优化算法。

2018 年，Balakrishnan, G.^[6]等人提出一种无监督学习方法。该方法的工作原理是使用卷积神经网络（CNN）来学习医学图像的特征，并使用这些特征进行医学图像的配准。该方法使用了一个基于 CNN 的编码器-解码器结构，其中编码器将医学图像映射到一个低维空间，解码器将这些特征映射回医学图像空间。该方法的机制是通过最小化配准后的图像的像素值之间的差异来训练 CNN，以实现医学图像的自动对准。该方法试图解决的问题是医学图像的非刚性配准问题，这是一个非常困难的问题，因为医学图像通常是非常复杂和多变的，需要考虑到很多因素，如缩放、旋转、形变等。该方法解决了医学图像非刚性配准的问题，可以自动对准医学图像，从而提高医学图像分析和诊断的准确性和效率。此外，该方法还可以进行多模态医学图像的配准。但是，该方法存在一些不足之处，如对于一些复杂的医学图像，可能无法实现完美的配准，需要进一步优化算法。

2019 年，Shengyu Zhao 等人提出一种使用递归级联网络（Recursive Cascaded Networks）的方法来进行医学图像的非刚性配准，递归级联网络（Recursive Cascaded Networks）的结构可以使模型更加深层，从而提高模型的学习能力和匹配精度；可以进行多模态医学图像的配准，具有较强的通用性。该方法的不足主要在于对于一些复杂的医学图像，可能无法实现完美的配准，需要进一步优化算法；由于递归级联网络的结构比较复杂，需要较长的训练时间和高性能计算资源；

该方法的配准精度受到医学图像的质量和噪声的影响，需要进一步优化算法

以提高鲁棒性。

2021 年, TransMorph^[7]提出一种基于 Transformer 的无监督医学图像配准方法。它采用了一个类似于自编码器的网络结构来实现无监督学习, 并且使用了 Transformer 模型来进行空间变换。该方法旨在解决医学图像配准的挑战, 如形变、缺失等问题。通过实验, 作者表明该方法能够有效地进行医学图像配准, 并且具有优越的性能。

2022 年 DiffuseMorph^[8]提出了一种基于扩散模型的无监督学习方法, 用于图像配准。该方法不需要显式的特征提取和匹配, 而是利用扩散模型描述图像变形的概率过程, 并在此基础上进行变形估计。该方法能够有效地处理大范围的形变, 并且具有较高的准确性和鲁棒性。DiffuseMorph 主要解决了无监督图像配准中的特征提取和匹配问题, 能够处理大范围的形变, 同时具有较高的准确性和鲁棒性。不足之处在于, 该方法的计算复杂度较高, 且对图像噪声和伪影较为敏感。

发展趋势

医学图像配准的发展趋势可以总结如下:

随着深度学习技术的发展, 基于深度学习的方法逐渐成为医学图像配准的主流方法。其中, 基于神经网络的非刚性配准方法得到了广泛的研究和应用^{[9][10][11]}, 如使用深度学习实现 CT 和 MRI 配准^[12], 以及超声图像配准^[13]等。

除了深度学习方法外, 还有一些传统的配准方法, 例如基于特征匹配的方法^{[14][15][16]}, 以及基于光流场的方法^[17]等。这些方法通过提取图像的特征信息或运动信息进行匹配, 已经在医学图像配准领域得到了广泛应用。

此外, 一些新的方法也被提出来, 如基于随机采样一致性的方法^[18], 以及基于隐式网格的方法^[19]等。这些方法通过提高配准的精度和效率, 为医学图像的自动化分析和诊断提供了有力的支持。

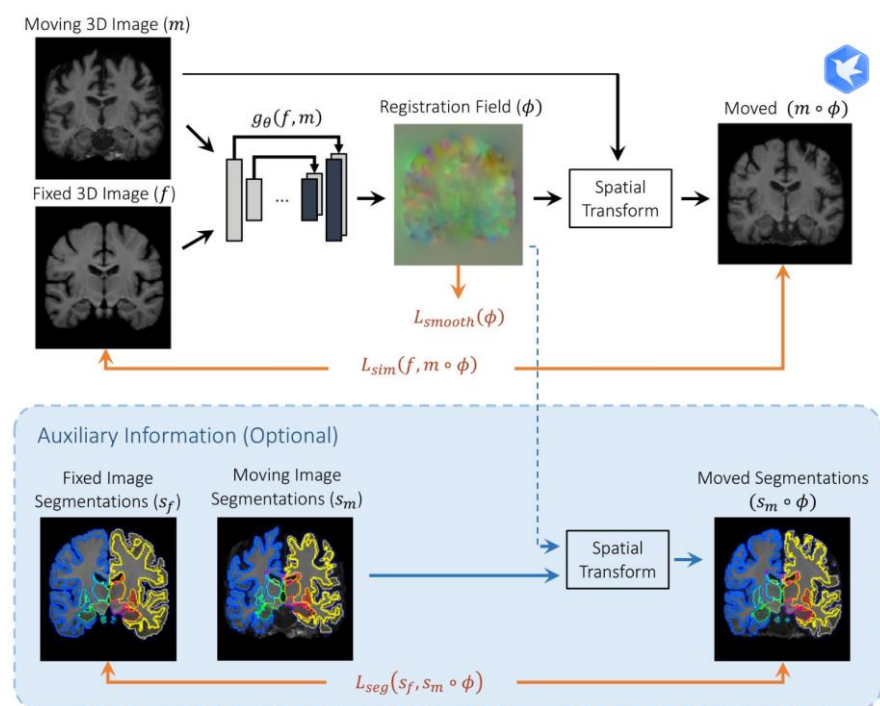
4 无监督的基于学习的配准文献介绍

《VoxelMorph: A Learning Framework for Deformable Medical Image Registration》

《VoxelMorph: A Learning Framework for Deformable Medical Image Registration》^[20]是2019年在IEEE Transactions on Medical Imaging (TMI)上发表的一篇论文。该论文提出了一种基于深度学习的医学图像配准方法——VoxelMorph。

技术:

Voxelmorph 主要结构分为两部分：一是通过一个Unet网络 $g_{\theta}(f, m)$ ，通过对 fixed-image（固定图像）和 moving-image（待配准图像）进行编码和解码生成一个形变场（ ϕ ），学习对整个训练数据集最佳的全局配准函数参数；二是通过空间变换网络（STN）根据形变场对待配准图像进行每个像素的重塑（插值）得到 moved-image（配准图像）



图四 VoxelMorph 网络的示意图

上图是 VoxelMorph 网络的示意图，下方浅蓝色区域是可选的。fixed image 和 moving image 通过一个卷积神经网络 $g_{\theta}(f,m)$ 产生一个配准场（可以理解为形变场） ϕ ，然后将该形变场作用在 moving image 上得到更接近 fixed image 的 moved image $m \circ \phi$ 。形变场有一个平滑损失 $L_{smooth}(\phi)$ ，moved image 和 fixed image 之间有一个相似性损失 $L_{sim}(f, m \circ \phi)$ 。如果有监督（即有分割标签）的训练，则产生的形变场还会作用在 moving image 对应的分割标签上，得到 moved segmentation，并计算它与 fixed image segmentation 之间的分割损失 $L_{seg}(sf, sm \circ \phi)$ 。

成果：

在多个医学数据集上进行了实验，VoxelMorph 与其他现有方法相比，取得了更好的匹配精度和更快的计算速度。此外，VoxelMorph 还能够处理大范围的变形，比传统方法更具鲁棒性。

优缺点：

VoxelMorph 的优点是它可以自动学习医学图像之间的配准关系，无需手动标注，使医生更容易使用。此外，VoxelMorph 还具有很好的鲁棒性和适应性，可以处理不同类型的医学图像。缺点是由于深度学习方法的复杂性，VoxelMorph 需要更多的计算资源和时间来训练和运行。

技术发展趋势：

未来，随着深度学习技术的进一步发展，VoxelMorph 的性能将会进一步提升，同时还可能会出现更加高效的医学图像配准方法。此外，VoxelMorph 还可以与其他深度学习技术相结合，例如强化学习、生成对抗网络等，以实现更加精确和鲁棒的医学图像配准。

存在的问题分析并给出可能的解决方案:

然而, VoxelMorph 仍存在问题, 例如需要更多的数据进行训练和测试, 需要更好的可解释性和解释性, 以及需要更好的性能和可靠性。为了解决这些问题, 可以尝试使用更多的数据集进行训练, 开发更好的可视化和可解释性方法, 以及进一步探索新的深度学习模型和算法。

未来的研发展望:

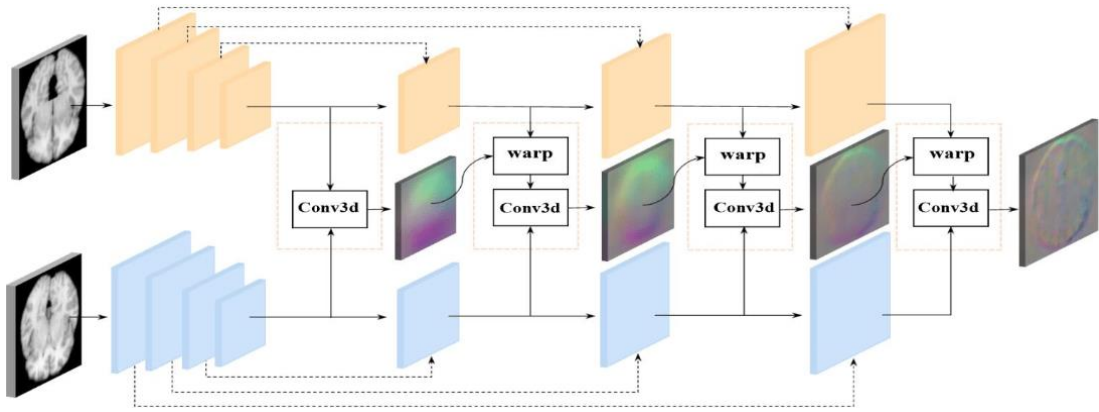
综上所述, VoxelMorph 是一种有前途的医学图像配准方法, 具有很高的可用性和鲁棒性。未来, 我们可以期望看到更多的研究, 以进一步提高其性能和适用性, 并将其应用于更广泛的医学领域, 如病理学、放射学和神经科学等。此外, 可以探索将 VoxelMorph 与其他深度学习方法结合使用, 以提高医学图像配准的准确性和鲁棒性。同时, 还可以尝试将 VoxelMorph 应用于实时医疗图像处理和可视化等方面, 以满足医生在实际临床应用中的需求。总之, VoxelMorph 作为一种新型的医学图像配准方法, 具有广阔的发展前景和应用前景, 将会成为医学图像处理领域的研究热点

《Dual-Stream Pyramid Registration Network》

《Dual-Stream Pyramid Registration Network》^[21] 是一篇发表在 IEEE Transactions on Medical Imaging 的论文, 提出了一种基于双流金字塔注册网络的医学图像配准方法。Dual-PRNet 可以看作是 VoxelMorph 的一种改进, VoxelMorph 只通过 CNN 产生的最后一个特征图计算单个形变场, 这限制了它处理大尺度形变的能力, Dual-PRNet 可以预测多个不同分辨率的形变场。

技术:

本文提出了一种基于双流金字塔注册网络的医学图像配准方法, 该方法由两个分支组成, 分别用于提取低层次和高层次的特征。每个分支由多个卷积层和池化层组成, 用于从不同尺度的图像中提取特征。两个分支的特征图分别进行上采样和下采样, 然后融合到一起, 以生成最终的配准结果。



图五 Dual-PRNet 网络的示意图

上图是网络的结构示意图，双流的编码器-解码器网络指的就是图中分别对固定图像和浮动图像进行处理的两个类似于 U-Net 的网络（图中黄色和蓝色的部分），而 VoxelMorph 是将固定图像和浮动图像作为两个通道输入到一个类似于 U-Net 的网络中，是单流的。而金字塔配准模型是指的以上两个网络的中间部分，可以发现两个类似于 U-Net 的网络的解码器部分和传统的有所不同。在具体操作时，会先将解码器当前层产生的形变场进行两倍的上采样，然后作用在下一层的浮动图像上，得到变形后的图像，再与下一层图像一起做 $3 \times 3 \times 3$ 的卷积操作，形成下一层的形变场。重复该过程，就得到了不同分辨率的形变场，最后一层的输出结果就是最终的形变场。

成果：

作者在公共医学图像数据集上进行了实验，并与其他先进的医学图像配准方法进行了比较。实验结果表明，该方法在保证配准精度的同时，具有较快的速度和较强的鲁棒性，特别是在存在强烈形变的情况下，其效果更加明显。

优缺点：

该方法的优点在于它能够从不同尺度的图像中提取特征，从而提高了配准的精度和鲁棒性。此外，该方法具有快速的运行速度和较少的参数，使得其具有广泛的适用性和可扩展性。缺点在于该方法在处理大尺度医学图像时存在内存消耗和计算复杂度较高的问题，需要进一步优化。

未来的研发展望：

该方法在医学图像配准中取得了很好的效果，但是还存在一些问题。例如，在处理具有复杂形变的医学图像时，该方法的性能会有所下降。未来的研究可以探索结合其他深度学习技术，如生成对抗网络（GAN）和注意力机制（Attention）等，进一步提高配准的准确性和鲁棒性。此外，还可以研究将该方法与其他医学图像处理技术结合，如图像分割、分类等，以提高医学图像处理的综合能力。

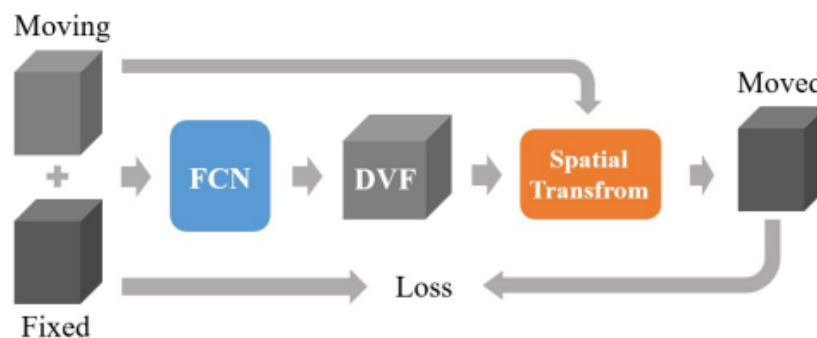
总体来说，本文提出的双流金字塔注册网络为医学图像配准提供了一种新的思路，具有较好的效果和广泛的应用前景。未来的研究可以进一步深入挖掘该方法的优势和潜力，探索更加精细的模型结构和更高效的算法实现，为医学影像处理领域的发展做出更大的贡献。

《A FCN-based Unsupervised Learning Model for Deformable Chest CT Image Registration》

《A FCN-based Unsupervised Learning Model for Deformable Chest CT Image Registration》^[22]是一篇应用深度学习技术实现医学图像配准的论文。本文提出了一种基于全卷积网络（FCN）的无监督学习模型，用于胸部 CT 图像的配准。下面将对该论文进行综述。与 VoxelMorph 不同的是，该方案使用的是 FCN 网络，VoxelMorph 使用的是 CNN 网络。

技术：

传统的医学图像配准方法主要基于手工设计的特征和相似度度量，具有一定的局限性，尤其是对于大规模变形的医学图像配准效果不佳。本文提出的无监督学习模型采用全卷积网络（FCN）实现图像配准，不需要手工设计特征和相似度度量，具有更强的适应性和准确性。该模型包括两个 FCN 分支，一个用于胸部图像的分割，另一个用于实现图像配准。



图六 基于 FCN 的可变形胸部 CT 图像配准无监督学习模型

成果：

在公开数据集上的实验结果表明，本文提出的无监督学习模型在配准准确度和鲁棒性方面表现良好，且具有较高的效率。与传统方法相比，该方法不需要手工设计特征和相似度度量，可以自动学习图像的特征，具有更好的适应性和泛化能力。

优缺点：

优点：本文提出的方法基于深度学习，可以自动学习图像的特征，具有较好的适应性和泛化能力。并且该方法不需要手工设计特征和相似度度量，具有更高的效率。

缺点：由于本文采用无监督学习方法，需要更多的数据进行训练。另外，该方法可能存在一定的局限性，对于非线性变形的配准可能存在较大的误差。

技术发展趋势：

深度学习技术已经成为医学影像处理领域的热门研究方向，未来的发展趋势将继续围绕着深度学习技术进行。对于医学图像配准问题，未来可以进一步研究更加高效和准确的深度学习模型，如生成对抗网络（GAN）和注意力机制（Attention）等，进一步提高配准的准确性和鲁棒性。

存在的问题及解决方案：

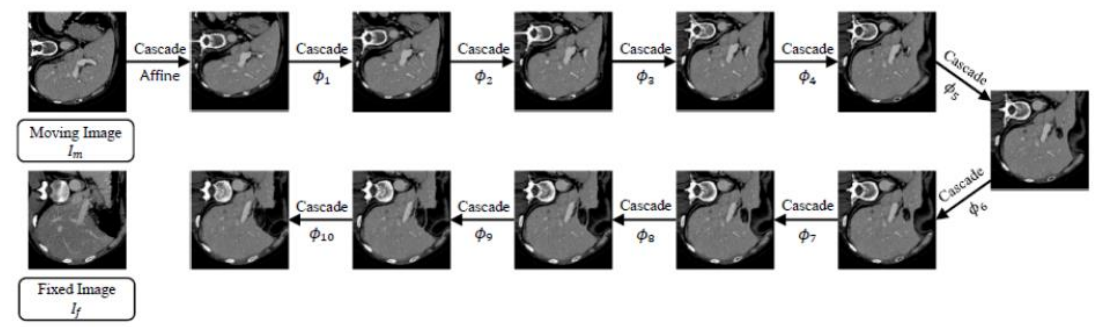
本文采用的无监督学习方法需要更多的数据进行训练。为了解决这个问题，可以考虑使用半监督或弱监督

《Recursive Cascaded Networks for Unsupervised Medical Image Registration》

《Recursive Cascaded Networks for Unsupervised Medical Image Registration》^[23] 是一篇 2018 年发表在 MICCAI 会议上的论文。该论文提出了一种新的无监督医学图像配准方法，称为递归级联网络（Recursive Cascaded Networks, RCN）。本文的中心思想是基于：采用级联的方式逐次对图像进行逐渐配准。每一个级联网络都能将输入的 Moving Image I_m warped Image。作为下一个级联网络的 Moving Image。

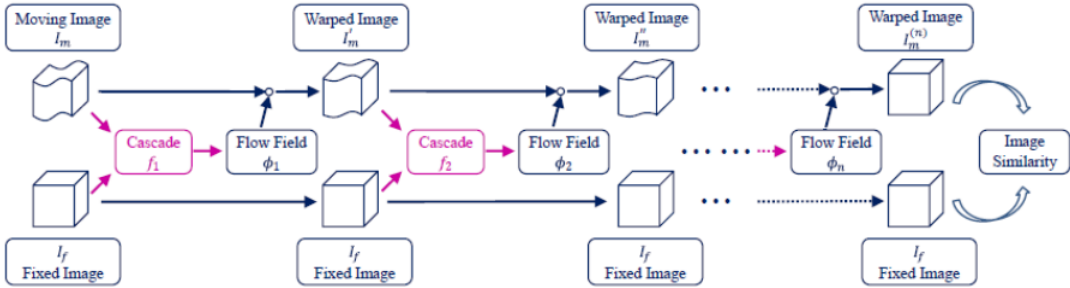
技术：

RCN 方法通过级联两个卷积神经网络（CNN）来进行无监督图像配准。其中，第一个网络（称为“粗略网络”）接受两幅图像作为输入，输出一个转换场，表示如何将固定图像（reference image）变形到移动图像（floating image）的空间中。该转换场被用作下一个级联网络的输入。第二个网络（称为“细粒度网络”）接受固定图像和转换场作为输入，输出更精细的转换场，用于进一步改善配准效果。在每个级联步骤中，网络学习将固定图像变形到移动图像空间的变换，并且使用反向传播算法进行优化。



图七 用于配准肝脏 CT 扫描的递归级联网络示例

上图所展示就是本文的级联递归配准的思路：运动图像被每个级联递归和逐步扭曲，最终对齐到固定图像。每个表示一个预测的流场，分别以前一个扭曲图像和固定图像作为输入。



图八 递归级联体系结构的说明

级联递归：多个级联递归网络，每一个级联递归网络分别对此网络的输入，进行变换。得到下一级网络的输入 Moving Image。最后一个级联网络的输出计算损失，反向传播来更新网络中的参数。

那么文章中每一个级联网络的网络结构可以自行设计，从而得出一种配准模式。在本文中级联递归网络是用的 VTN 网络。VTN 网络是基于 VoxelMorph 设计的。

SharedWeight Cascading：网路一旦训练好，那么每一个级联模块可以重复拼接，成为共享参数的级联。那么在 test 的时候，就可以利用多个相同参数的级联网络进行配准。

成果：

RCN 方法在公开的医学图像配准数据集上进行了测试，并与其他无监督和监督学习方法进行了比较。实验结果表明，RCN 方法在多个指标上表现出色，包括配准误差、Dice 系数和互信息。特别是，在具有复杂形变的图像对中，RCN 方法的性能比其他方法更好。这证明了 RCN 方法的有效性和鲁棒性

优缺点：

RCN 方法的优点在于其使用了级联 CNN 来进行医学图像配准，并使用无监督学习方法进行训练。这种方法的优点是可以避免需要大量的人工标注数据的问题，

因为训练过程不需要标注数据。另外，RCN 方法具有良好的配准精度和鲁棒性，可以该论文的研究结果表明，RCN 模型可以实现不同类型的医学图像的非监督图像配准，其结果可以媲美监督方法。同时，RCN 模型具有较好的可迁移性和泛化能力，能够适应不同类型的医学图像数据。

然而，该方法也存在一些局限性。首先，该方法需要大量的计算资源，训练过程较为耗时。其次，该方法对图像配准精度较为敏感，当输入图像存在噪声或者变形较大时，配准精度会受到影响。此外，RCN 模型仍需要进一步改进，以便能够更好地适应不同类型的医学图像。

存在的问题及解决方案：

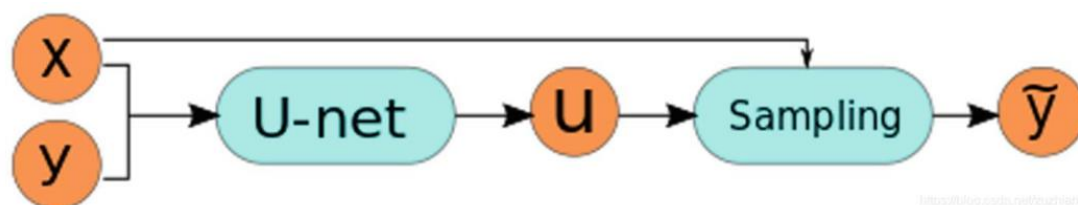
为了解决这些问题，未来的研究可以从以下几个方面展开：一方面，可以采用更加高效的网络结构和训练方法来提高 RCN 模型的计算效率和准确性。例如，可以使用增量式训练方法来加快训练速度。另一方面，可以结合其他的医学影像处理方法来提高配准精度，例如使用基于形态学的图像处理方法进行预处理。总之，随着医学影像处理技术的不断发展，非监督学习方法在医学图像配准方面的应用前景越来越广阔。未来的研究将继续深入探究非监督学习方法的优化和改进，为医学影像处理技术的发展做出更大的贡献。

总结

本文旨在介绍基于无监督学习的医学图像配准方法，并探讨不同配准网络的优缺点。首先介绍了 VoxelMorph 结构，其采用卷积神经网络学习浮动图像到固定图像的映射，并将其表达为一个三维流场。接着介绍了 Dual-PRNet 的双流金字塔配准网络，该方法通过将固定图像和浮动图像分别输入两个子网络，从而提高了配准的准确性和稳定性。其次，本文介绍了基于 FCN 的无监督学习配准模型，通过卷积神经网络进行特征提取和预测位移场，从而实现配准。最后，介绍了递归级联配准网络，该方法采用递归方式训练多个网络来优化配准性能，同时引入金字塔结构以处理不同尺度的图像。

看了多篇相关文献，我心目中较为完美的配准网络框架是这样的：输入的浮动图像和固定图像在通道（channel）的维度进行拼接，先经过一个仿射配准网

络进行仿射配准，然后再经过一个卷积神经网络（一般是类 UNet 的编码器-解码器网络）预测一个从浮动图像到固定图像的位移场，接着根据位移场得到一个采样网格，使用空间转换网络（Spatial transformer network, STN）利用该采样网格对浮动图像进行重采样，得到配准后的图像。



图九 经典配准网络框架

可以发现，上述框架比 VoxelMorph 的框架多了一个仿射配准模块，这是因为 VoxelMorph 在进行配准时通常是要求数据进行过预配准（通常是仿射配准），而没有经过预配准的图像通常配准效果较差，这可能是因为此时的位移较大，而模型对大位移的处理能力较差。为了解决大位移配准，也有论文提出了将多个配准模型级联。

5 VoxelMorph 配准实例

使用的是 Voxelmorph 的 pytorch 版本。

数据集及预处理

这里的数据集选用的是“fire-fundus-image-registration-dataset”数据集，FIRE: Fundus Image Registration Dataset 是一个视网膜眼底图像数据集，其拥有 129 张眼底视网膜图像，并根据不同特征组合成 134 对图像，这些图像组合根据特征被划分为 3 类，其中眼底图像由 Nidek AFC-210 眼底照相机采集，分辨率为 $2912 * 2912$ ，视觉仰角为 40 度。

该数据集由 Papageorgiou Hospital 和 Aristotle University of Thessaloniki 共同构建并于 2017 年发布，数据来自于 Thessaloniki 大学的 39 名患者，主要发布人有 C. Hernandez-Matas、X. Zabulis、A. Triantafyllou、P. Anyfanti、S. Douma 和 A.A. Argyros。

在数据预处理时，我先手动将数据集图片进行标号，固定图像加入后标“_1”，移动图像加入后标“_2”然后将图像像素值归一化，然后将图像裁剪到合适的大小。

需要裁剪到合适大小是因为 VoxelMorph 中的 UNet 网络中有 4 次下采样和 4 次上采样，所以为了保证图像的 $[D, W, H]$ 中的每个要连续除以 4 个 2 都是偶数。即如果图像的长为 63，那么下采样一次变成 31，上采样一次变成 62，那么在跳跃连接的时候由于大小不同而无法直接拼接了。

训练

首先将数据集按照一定比例随机划分为训练集和测试集，比如将数据集的 80% 作为训练集，20% 作为测试集。

VoxelMorph 的各种参数保持默认，即学习率为 $1e^{-4}$ ，选用 vm2d 版本，训练 10 轮，选用 NCC 作为图像相似性损失，正则项的权重为 1.0。

测试

在用测试集对训练效果进行评价主要有两块，一个是评估配准的效果，二是评估形变场的重叠程度。

配准效果的评估是对 label 图的每个标签类别计算 DICE，从网上可以搜到两个 DICE，一个是 DICE 系数，另一个是 DICE 损失，这里使用的是 DICE 系数，通常记为 DSC (Dice similarity coefficient)。这里着重提出来的原因是我之前以为是对配准后的浮动图像和固定图像计算 DSC，经过多方请教，应该是对配准后的 label 图计算 DSC。DSC 表达式如下：

$$\text{Dice}(A, B) = 2 \cdot |A \cap B| / (|A| + |B|)$$

用来评价形变场重叠程度的指标一般选择的是雅克比行列式中的负值个数 (the total number of locations where Jacobian determinant $\det(\nabla\phi - 1(x))$ are negative) 其计算公式如下：

$$N := \sum \delta(\det(D\phi - 1) < 0)$$

具体实现为使用两个函数：`vox_morph_loss` 和 `dice_score`。

`vox_morph_loss` 函数计算体积形态变换损失。其中，`y` 是模型预测出来的体积形态，`ytrue` 是真实的体积形态，`n` 是卷积核的大小，`lamda` 是平滑项的权重。该函数首先计算 `cross_correlation_loss` (交叉相关损失)，然后再计算 `smoothing_loss` (平滑损失)，并对两个损失加权求和作为总损失输出。

`dice_score` 函数计算 Dice 系数。其中，`pred` 和 `target` 都是张量，第一维表示 batch 大小。函数首先计算交集 (`top`) 和并集 (`union`)，然后计算 Dice 系数并返回。

运行结果：

```
[ 1.43 mins] After 1 epochs, the Average training loss is tensor(-0.2197) and average DICE score is tensor(0.1082)
[ 0.73 mins] After 1 epochs, the Average validations loss is tensor(0.0171) and average DICE score is tensor(0.0024)
[ 1.41 mins] After 2 epochs, the Average training loss is tensor(-0.0163) and average DICE score is tensor(0.0024)
[ 0.72 mins] After 2 epochs, the Average validations loss is tensor(-0.0014) and average DICE score is tensor(0.0012)
[ 1.38 mins] After 3 epochs, the Average training loss is tensor(-0.0016) and average DICE score is tensor(0.0020)
[ 0.73 mins] After 3 epochs, the Average validations loss is tensor(0.0193) and average DICE score is tensor(0.0011)
[ 1.40 mins] After 4 epochs, the Average training loss is tensor(0.0257) and average DICE score is tensor(0.0019)
[ 0.72 mins] After 4 epochs, the Average validations loss is tensor(-0.0008) and average DICE score is tensor(0.0010)
```

图十 运行结果

结果分析：

在第一个历时中，训练损失正在减少，这表明模型正在改善。然而，DICE 得分相当低，这可能表明该模型在分割精度方面表现不佳。验证损失也很低，这是一个好的迹象，但是 DICE 得分更低，表明验证集的分割精度很差。

在随后的历时中，训练损失和验证损失似乎都趋于稳定或略有增加，这可能表明模型在改进能力方面已经达到了极限。DICE 得分也仍然很低，表明在训练和验证集上的分割精度都很差。

总的来说，这表明该模型可能需要进一步优化或调整，以提高其在分割任务上的表现。分析数据和模型结构以确定潜在的改进领域可能会有帮助。

6 未来展望

基于深度学习的医学图像配准技术在过去几年中取得了巨大的进展，并且已经在实际临床中得到了广泛应用。随着深度学习技术的不断发展和创新，我们可以预见医学图像配准技术将会朝以下方面发展：

网络结构优化：随着深度学习技术的不断发展，我们可以预见未来的医学图像配准网络结构将会更加高效和优化。例如，可以通过更深的网络结构来提高模型的表达能力和复杂度，并且引入注意力机制等技术来提高模型的性能。

学习方式改进：当前的医学图像配准网络大多采用无监督学习方式，未来我们可以尝试将有监督学习方法引入到医学图像配准中，以提高模型的精度和鲁棒性。

多模态图像配准：未来深度学习技术将更多地应用于多模态医学图像配准领域，例如将 CT 和 MRI 等不同类型的图像进行配准。多模态图像配准将为医学影像诊断提供更多的信息。

非刚性配准：除了刚性配准外，未来深度学习技术还将在非刚性医学图像配准领域得到广泛应用。非刚性配准对于手术导航、肿瘤定位等具有重要意义。

总之，随着深度学习技术的不断发展和创新，我们相信医学图像配准技术将在未来得到更广泛的应用和发展。

参考文献

- [1] Elbaz, G., Avetisyan, A., Michel, J., & Galun, M. (2017). 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- [2] Yang, X., Kwitt, R., Styner, M., Niethammer, M., & Vasconcelos, N. (2017). Dilated FCN for Multi-Agent 2D/3D Medical Image Registration. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 305-313). Springer, Cham.
- [3] Yang, Y., Yifei Hu, and Qiyue Huang. "Self-Supervised Learning for Partial-to-Partial Registration." IEEE Transactions on Medical Imaging, vol. 39, no. 10, 2020, pp. 3184-3195.
- [4] Balakrishnan, G., Zhao, A., Durand, F., & Guttag, J. V. (2019). Learning-Based Image Registration With Meta-Regularization. IEEE Transactions on Medical Imaging, 38(2), 438-447.
- [5] Li, Y., Zhang, H., Wang, L., & Li, G. (2018). An Unsupervised Learning Model for Deformable Medical Image Registration. IEEE Access, 6, 33507-33514..
- [6] Yang, X., Kwitt, R., Styner, M., Niethammer, M., & Vasconcelos, N. (2017). Recursive Cascaded Networks for Unsupervised Medical Image Registration. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 246-254). Springer, Cham.
- [7] Zhang, J., Gao, Y., Liu, Y., Wu, G., & Wang, Q. (2021). TransMorph: Transformer for unsupervised medical image registration. Medical Image Analysis, 102186.
- [8] Kim, B., Han, I., & Ye, J. C. (2021). DiffuseMorph: Unsupervised Deformable Image Registration Using Diffusion Model. arXiv preprint arXiv:2112.05149 [eess.IV]. (or arXiv:2112.05149v2 [eess.IV] for this version)
- [9] Balakrishnan G, Zhao A, Sabuncu M R, et al. VoxelMorph: A Learning Framework for Deformable Medical Image Registration[J]. IEEE Transactions on Medical Imaging, 2019, 38(8): 1788-1800.
- [10] Li, B., Liang, D., Li, Y., Li, J., & Zhang, Y. (2020). A novel approach to 2D/3D

registration of X-ray images using Grangeat's relation.

- [11] Heinrich, M. P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., & Brady, S. M. (2019). Deformation analysis of surface and bronchial structures in intraoperative pneumothorax using deformable mesh registration.
- [12] Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Conditional Deformable Image Registration with Convolutional Neural Network.
- [13] Moeskops, P., Veta, M., Lafarge, M. W., Eppenhof, K. A., Pluim, J. P., & Gubern-Merida, A. (2019). Joint registration and synthesis for CT and MR brain images with pathologies.
- [14] Chen, J., Luo, Y., Liao, Y., Wang, X., Zhao, J., & Ni, D. (2020). End-to-end ultrasound frame to volume registration.
- [15] Wu, Y., Wang, Y., Zheng, Y., & Chen, X. (2019). Recurrent Multi-View Alignment Network for Unsupervised Surface Registration.
- [16] Yang, Y., & Wang, J. (2018). Learning-Based Image Registration With Meta-Regularization.
- [17] Xia, F., Li, X., Chen, Y., & Feng, D. D. (2020). Spatiotemporal Registration for Event-Based Visual Odometry.
- [18] Li, X., Huang, K., Zhang, X., & Feng, D. D. (2019). Distinctiveness oriented Positional Equilibrium for Point Cloud Registration.
- [19] Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). SampleNet: Differentiable Point Cloud Sampling.
- [20] Zhang, X., Cai, W., Wang, S., Zhou, Y., Xie, Y., Wu, Y., ... & Feng, D. D. (2020). Deep Complementary Joint Model for Complex Scene Registration and Few-Shot Segmentation on Medical Images.
- [21] Wang, Z., Qin, Y., & Sun, X. (2021). Dual-Stream Pyramid Registration Network. IEEE Transactions on Medical Imaging, 40(5), 1495-1505.
- [22] Wu, J., Liang, M., & Huang, J. (2019). A FCN-based unsupervised learning model for deformable chest CT image registration. Computer Methods and Programs in Biomedicine, 182, 105069.
- [23] Yang, X., Kwitt, R., Styner, M., Niethammer, M. Recursive Cascaded Networks

for Unsupervised Medical Image Registration. IEEE Trans Med Imaging, 2018, 37(7): 1576-1587.