

卷积神经网络的变分推断图像配准方法研究

摘要

可形变图像配准是一种在几何上对齐两张图像的技术。在医学图像处理中，医学图像配准可以应用于医学图像融合、解剖结构变化检测和器官图谱生成等应用。随着深度学习和 GPU 硬件的发展，可形变图像配准可以通过卷积神经网络实现快速、准确的配准。由于无监督图像配准不需要昂贵稀有的专家标注作为监督信息，无监督深度学习图像配准方法，已经成为了主流。但大多数基于深度学习的方法采用的是非参数化的形变模型，使之难以控制形变场的不平滑和异常。因此，本文采用基于控制点的径向基函数形变模型，在保证精度的前提下，生成较为平滑、形变异常情况较少的形变场。本文首次将径向基函数形变模型与变分生成模型结合，并且创新性地提出了自适应控制点分布方法和改进损失函数。研究内容主要包含以下部分：

通过引入基于变分推断的概率模型，将图像配准视为带有潜变量的生成过程。鉴于后验分布的难以计算，我们采用条件变分自动编码器来逼近后验分布，并通过重构目标函数来推导变分下界。同时，我们引入了紧支撑径向基函数形变模型，其中包含非均匀控制点分布。通过将潜变量的先验分布与控制点位置相关联，我们成功地推导出变分下界中与控制平滑的扭曲能量相关的项。对于非均匀控制点分布，本文首次提出了可偏移控制点方法，并设计了相应的网络结构。在不确定性估计方面，我们描述了如何利用该概率模型进行测量，以评估模型对配准结果的置信度。这种不确定性估计在临床医疗中具有实际应用，可避免基于没有根据和错配的结果做出错误决策。实验证实了本模型在配准精度和形变场质量上的卓越性，优于其他方法。

关键词：可形变图像配准；变分推断；卷积神经网络；径向基函数

1 引言

图像配准是一种寻找两张图像之间的空间变换关系的技术，这种技术使得源图像和目标图像在某些特征上达到空间上一致。在可形变图像配准中，这种一致就是源图像经过空间变换后在图像强度上与目标图像的尽可能相似。图像配准在计算机视觉、医学图像处理、遥感图像以及材料力学等领域中都有广泛的应用。在医学图像处理中，医学图像配准是一个重要而且极具挑战性的课题，它可以用在医学图像融合、解剖结构变化检测和器官图谱生成等应用。本文主要研究图像配准在心脏运动估计中的应用。医学图像配准在临床应用中有着高实时性和高精度的要求，因此研究一种快速准确的图像配准技术是非常有意义的。

目前来说，心血管疾病已经成为全球一大死亡原因之一，而心脏是人体的重要器官，因此研究心脏运动对于心血管疾病的诊断和治疗有着重要意义，而心脏运动估计实际上是从心脏影像序列中估计心脏一些感兴趣的区域的位移或形变。图像配准是应用心脏运动估计的技

术之一，其目的是获得心脏在两个相位（主要是舒张末期和收缩末期）之间的对应关系并估计出形变参数。将卷积神经网络应用到可形变配准中可以直接通过卷积神经网络来估计形变参数，不需要传统图像配准算法那样要迭代优化形变参数，因此基于卷积神经网络的可形变配准在配准时间上远远优于传统配准算法，而且其配准精度与传统算法的精度相当。

目前心脏运动估计的挑战如下：一是心脏在一个心跳周期中存在剧烈且复杂的运动，这导致了不同相位间的图像差异巨大，估计的形变很容易存在不平滑和异常的情况；二是左心室乳头肌的成像强度和心肌的成像强度高度同质，而且随心脏心跳变化极大，不同病例心脏乳头肌运动特征也不同，如图 1.1 所示，三是心脏成像在长轴方向的分辨率不高，片间影像信息不连续，如图 1.2 所示。

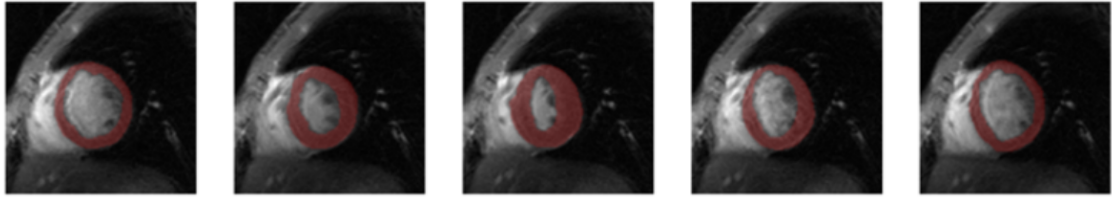


图 1.1. 心脏影响序列的示例

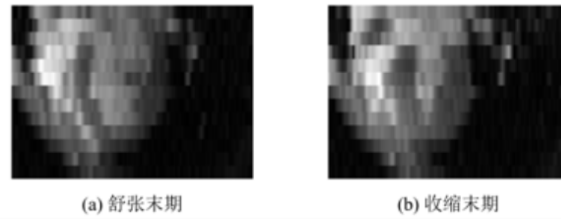


图 1.2. 切片方向（长轴方向）的 Cine-MRI 截面

2 相关工作

2.1 可形变图像配准研究现状

可形变图像配准是寻找两张图像（源图像和目标图像）之间稠密的非线性的空间变换，这种空间变换使得这两张图像尽可能相似。在传统的机器学习算法中，可形变图像配准实际上是一个优化问题：通过迭代优化算法去寻找一个最优的空间变换。这个空间变换一般可以表示成一个稠密偏移向量场（DVF）。令 M 和 F 分别为源图像和目标图像， T 为空间变换，那么图像配准可以表示成以下公式：

$$D = \operatorname{argmin} C(M, F, T) \quad (1)$$

其中， D 为最优的空间变换， C 是一个关于源图像 M 、目标图像 F 和空间变换 T 的目标函数。这个 C 一般包含两个部分：一个用来度量形变后的源图像 $M(T)$ 和目标图像 F 相似性的相似性度量和一个用来约束空间变换 T 某些性质的正则化项。

Sotiras 等人^[1] 总结归纳了深度学习被广泛应用之前的可形变图像配准方法，将可形变配准分为形变模型、目标函数和优化方法三个部分。在传统机器学习中，目前用在心脏运动估

计的常用的形变模型为自由变形形变模型 (FFD) [2], 多仿射模型[3, 4], 弹性体模型[5]。基于传统机器学习算法的可形变图像配准最大的挑战在于优化方法。每一对图像的配准都需要迭代优化, 这需要大量的算力和时间, 并不适合医学图像配准要求的实时性。基于深度学习的可形变图像配准在配准过程中则不需要重新优化, 它是通过已经训练好的神经网络分析图像对的对应关系, 直接生成 DVF。在同等 CPU 算力下, 配准速度已经远胜传统机器学习, 而且配准精度可以与当前最好的传统算法相当, 甚至更好。

2.2 基于深度学习的图像配准方法

基于深度学习的可形变图像配准同样是一个优化问题, 只不过优化的对象不是空间变换 T 的参数而是神经网络的参数。

对于基于深度学习的可形变图像配准方法, 已有学者[6,7] 进行了总结和归纳。早期基于深度学习的配准方法[8,9] 是传统的机器学习算法的扩张, 利用卷积神经网络 (CNN) 去提取图像特征然后进行配准, 配准依然需要迭代优化。基于强化学习的方法[10,11] 同样需要在配准中进行迭代优化, 与常规的可形变图像配准方法不同, 强化学习的迭代优化在于智能体 (agent) 与环境 (environment) 在环境反馈状态和奖励与智能体执行行动的过程中的不断交互。基于强化学习的方法大部分都是应用于只包含低维参数的简单形变, 比如仿射形变和刚性形变, 因为大量的形变参数会导致一个超大的行动空间, 使得这种互动变得非常复杂。与以上这两种方法不同, 端对端的图像配准方法则直接使用训练好的神经网络去学习图像对的对应关系和生成 DVF, 配准过程中不需要优化的过程, 极大地提升配准速度。根据不同的训练网络的策略, 端对端的图像配准方法可以分为两种: 有监督学习和非监督学习。

3 本文方法

3.1 本文方法概述

本文提出了一个基于变分推断的概率模型去实现这种生成过程。在这个过程中, 后验分布是十分难计算的, 因此我们使用了条件变分自动编码器来逼近这种后验分布并重构目标函数, 然后推导出变分下界。我们还采用了带有非均匀控制点分布的紧支撑径向基函数形变模型, 并且与变分推断概率模型结合, 通过将潜变量的先验分布假设为一个与控制点位置相关的矩阵, 将控制平滑的扭曲能量从变分下界中推导出来。对于非均匀控制点分布, 我们首次提出了可偏移控制点方法并且设计了对应该分布的网络结构。

3.2 基于变分推断的图像配准

生成模型一般使用无监督方法去从训练样本中学习数据特征, 然后从这个分布中生成符合输入数据特征的带有变化的新数据。相较于确定性模型, 生成模型在小数据上会有更好的结果, 因为生成模型学习的是潜在的数据结构并且会先定义一些结构假设约束模型以防止过拟合。生成模型可以很轻易地从不同的潜在数据结构特征中生成对应的数据, 即便这个特征是训练样本没有的。在医学图像配准中, 由于昂贵的获取过程和标注过程, 训练数据数量相较于其他非医疗任务比较少。生成过程可以使用较小的训练数据集学习出一个概率形变模型。

在某种意义上，图像配准可以表示成一种源图像 M 经过带有形变参数 z 的空间变换 ϕ_z 形成目标图像 F 的生成过程。形变参数 z 可以作为生成过程的潜变量。目标图像 F 可以通过以形变参数 z 和源图像 M 作为条件的随机分布采样得来。目标图像 F 的生成过程如下图所示，主要分为两步：

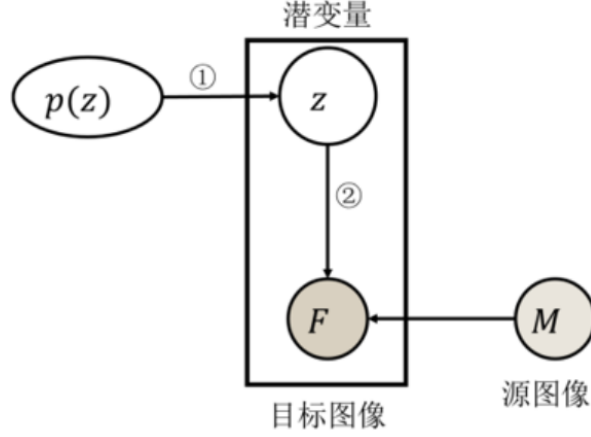


图 3.1. 图像配准生成过程

- 1) 从先验分布 $p(z)$ 中采样得到潜变量 z ；
- 2) 然后从带有生成参数 θ 的生成分布 $P_\theta(F|z, M)$ 采样得到目标图像 F 。

先验分布 $p(z)$ 通常假设成高斯分布。在深度生成模型中，生成分布 $P_\theta(F|z, M)$ 一般是包含了神经网络的函数，而生成参数 θ 是网络的参数。

在这个生成过程中，根据贝叶斯定理，只要知道后验分布 $P_\theta(z|F, M)$ ，那么就可以通过生成分布 $P_\theta(F|z, M)$ 重构出目标图像 F ，但是直接求后验分布几乎是不可能的，所以在这里使用近似分布去逼近这些复杂的后验分布。

变分推断可以用于解决后验分布 $P_\theta(z|F, M)$ 难以计算的问题，它通过使用近似分布去逼近这些复杂的后验分布，从而将目标图像 F 的边缘分布转化成以下公式，最后得到两个结果，一个是变分下界，一个是 KL 散度，这里的 kl 散度代表的是变分分布和后验分布的相似性，变分推断引入了带有变分参数 β 的变分分布 $q_\beta(z|F, M)$ 去逼近后验分布 $P_\theta(z|F, M)$ ，因此变分推断的目的是最大化似然分布 $P_\theta(F|M)$ 和最小化变分分布 $q_\beta(z|F, M)$ 和后验分布 $P_\theta(z|F, M)$ 之间的 KL 散度。最大化似然分布 $P_\theta(F|M)$ 的目的是使得生成网络能够生成一个质量好的目标图像；最小化 KL 散度的目的则是使变分分布 $q_\beta(z|F, M)$ 逼近于后验分布 $q_\beta(z|F, M)$ ，而变分推断最终将这两个目的转化成最大化变分下界 ELBO。其推到过程如下所示：

$$\begin{aligned}
 \log p_\theta(F|M) &= \mathbb{E}_{q_\beta(z|F, M)}[\log p_\theta(F|M)] \\
 &= \mathbb{E}_{q_\beta(z|F, M)}[\log p_\theta(F, z|M) - \log p_\theta(z|F, M) \\
 &\quad + \log q_\beta(z|F, M) - \log q_\beta(z|F, M)] \\
 &= ELBO + KL[q_\beta(z|F, M)||p_\theta(z|F, M)]
 \end{aligned}$$

图 3.2. 推导过程

3.3 基于 VAE 架构的图像配准概率模型

根据变分推断就引入了变分自编码器，它是一种可以使用随机梯度下降去学习深度潜变量模型和其相关推断模型的一类方法，提供了一个带有参数 β 的推断模型去近似后验分布，和前面所讲的一样，VAE 也是通过优化变分下界 ELBO 去逼近后验分布的，最大化 ELBO 实际上就是同时最大化似然分布 $\log P_\theta(x)$ 和最小化 KL 散度，使得生成模型可以更好的生成数据，并且变分分布更接近后验分布。

我们使用 VAE 去构建图像配准概率模型：它是通过一个概率编码器去生成生成过程中的潜变量 z 的后验分布，然后通过概率解码器重构输入图像。其架构图如下所示，VAE 包含两个部分，分别是编码器和解码器，编码器负责近似潜变量的后验分布，解码器负责生成目标图像 F 的分布，而在深度学习中，变分参数 β 和生成参数 θ 是神经网络的参数。VAE 架构图如下所示：

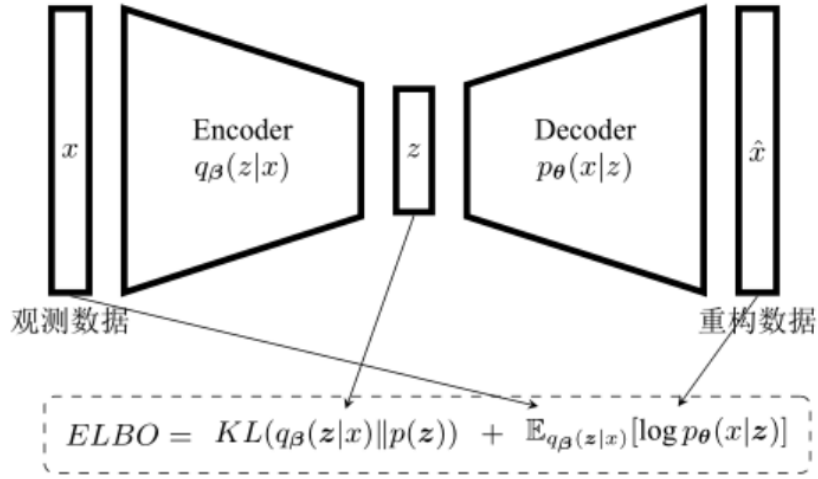


图 3.3. VAE 架构图

VAE 就是通过最大化变分下界 ELBO 去寻找最优的网络参数 β^* 和 θ^* ，其推导过程如下所示，最终的结果为两项，第一项 KL 散度迫使编码器生成的变分分布去逼近先验分布，第二项生成分布 $P_\theta(F|z, M)$ 代表形变后的源图像为目标图像 F 的可能性，它迫使解码器重构的图像要与目标图像相似。

$$\begin{aligned}
 & \max ELBO \\
 \Rightarrow & \max \mathbb{E}_{q_\beta(z|F, M)} [\log p_\theta(F, z|M) - \log q_\beta(z|F, M)] \\
 \Rightarrow & \min \mathbb{E}_{q_\beta(z|F, M)} [\log q_\beta(z|F, M) - \log p_\theta(F, z|M)] \\
 \Rightarrow & \min \mathbb{E}_{q_\beta(z|F, M)} [\log q_\beta(z|F, M) - \log p_\theta(F|z, M) - \log p(z|M)] \\
 \Rightarrow & \min \mathbb{E}_{q_\beta(z|F, M)} [\log q_\beta(z|F, M) - \log p_\theta(F|z, M) - \log p(z)] \\
 \Rightarrow & \min KL[q_\beta(z|F, M)||p(z)] - \mathbb{E}_{q_\beta(z|F, M)} [\log p_\theta(F|z, M)]
 \end{aligned}$$

图 3.4. 推导过程

3.4 紧支撑径向基函数形变模型概述

基于 CSRBF 的形变函数实际上是使用 CSRBF，根据给定控制点的位置和系数进行插值，得到一个平滑的 DVF。形变函数 ϕ_z 实际上由 x 方向上的形变函数 $\phi_{z,x}$ 和 y 方向上的形变函数 $\phi_{z,y}$ 组合而成。以 x 方向上的形变函数 $\phi_{z,x}$ 为例，给定包含 n 个控制点的点集 $P = \{P_i\}_{i=1}^n$ ， $\phi_{z,x}$ 的公式如下所示：

$$\phi_{z,x}(v) = v_x + \sum_{i=0}^n z_{i,x} \psi\left(\frac{\|v - p_i\|}{r}\right)$$

图 3.5. 推导过程

其中， $v = v_x, v_y \in R^2$ 为像素， $z_{i,x}$ 表示控制点 p_i 在 x 方向上的系数。对于 y 方向上的形变函数， $\phi_{z,y}$ 的计算方式与 $\phi_{z,x}$ 相同，唯一不同的是 v_y 和系数 $z_{i,y}$ 。 $\|v - p_i\|$ 表示像素 v 和控制点 p_i 之间的欧式距离，而 $\psi(\cdot)$ 则表示紧支撑半径为 r 的 CSRBF。CSRBF 形变函数将合适的参数去插值成一个稠密 DVF 并且使用控制点控制局部形变函数，是一种关于多元函数插值的成熟技术^[12]。

3.5 参数形变模型的图像配准概率模型

在图像配准中，潜变量的选择有两种：低维向量和形变参数。这种使用低维向量作为潜变量的方法更接近典型的 VAE，使用概率编码器将目标图像 F 和源图像 M 编码成低维向量，再用概率解码器在给定源图像 M 作为条件下根据这个低维向量重构目标图像 F。直接使用形变参数作为潜变量的方法与前者的区别在于基于形变参数的方法使用的概率解码器不存在网络参数，因为形变参数本身就是构成稠密 DVF 的所需元素。两种潜变量都可以被认为包含了形变相关的信息，只不过低维向量会与输入的图像对关联性比较小，可以用于形变转移，而形变参数作为潜变量，则可以通过假设特定的先验分布去约束稠密 DVF。

将形变参数当作生成过程中的潜变量，并且使用变分推断去逼近后验分布 $P_\theta(z|F, M)$ 。图像配准的过程可以被看作寻找每个控制点 p_i 的系数 $z_i = z_{i,x}, z_{i,y}$ 的分布。我们假设先验分布 $p(z)$ 为一个均值为 0 和协方差为 Σ_P 的多元正态分布： $p(z) = \mathcal{N}(z; 0, \Sigma_P)$ ，其中协方差 Σ 为一个与控制点位置相关的矩阵：

$$\Sigma_P^{-1} = \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$$

$$B = \begin{bmatrix} \psi\left(\frac{\|p_1-p_1\|}{r}\right) & \psi\left(\frac{\|p_1-p_2\|}{r}\right) & \dots & \psi\left(\frac{\|p_1-p_n\|}{r}\right) \\ \psi\left(\frac{\|p_2-p_1\|}{r}\right) & \psi\left(\frac{\|p_2-p_2\|}{r}\right) & \dots & \psi\left(\frac{\|p_2-p_n\|}{r}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \psi\left(\frac{\|p_n-p_1\|}{r}\right) & \psi\left(\frac{\|p_n-p_2\|}{r}\right) & \dots & \psi\left(\frac{\|p_n-p_n\|}{r}\right) \end{bmatrix}$$

图 3.6. 相关矩阵

其中 0 为零矩阵。这个先验分布 $p(z)$ 对于不同的控制点分布是不同的。另外，我们还假设变分分布 $q_\beta(z|F, M)$ 为多维高斯分布： $q_\beta(z|F, M) = N(z; \mu(F, M), \Sigma(F, M))$ 。这里均值 $\mu(F, M)$ 和协方差 $\Sigma(F, M)$ 是与源图像 M 和目标图像 F 相关的。为了简化分析和减少算力，这里协方差 $\Sigma(F, M)$ 被假设为对角矩阵。编码器使用带有网络参数 β 的 CNN 去预估变分分布 $q_\beta(z|F, M)$ 的均值 $\mu(F, M)$ 和协方差 $\Sigma(F, M)$ ，得到变分分布 $q_\beta(z|F, M)$ ，然后从变分分布 $q_\beta(z|F, M)$ 采样得到潜变量 z 。为了使得采样的过程可导，潜变量 z 是通过重采样技巧采样而来： $z = \mu(F, M) + \Sigma(F, M) * \epsilon$ where $\epsilon \sim N(0, I)$ 。

由于我们将形变参数假设为潜变量 z ，因此当潜变量 z 从变分分布 $q_\beta(z|F, M)$ 采样后，解码器就可以根据前面所述的形变函数直接生成稠密 DVF，然后对源图像 M 进行空间变换。这样，解码器 $P_\theta(F|z, M)$ 就不需要任何网络参数，解码器实际上是一个虚假的解码器，只包含了生成 DVF 的形变函数和对源图像 M 进行变换的 STN。因此，生成分布 $P_\theta(F|z, M)$ 、似然分布 $P_\theta(F|M)$ 和后验分布 $P_\theta(z|F, M)$ 这三个分布都不需要生成参数 β ，可以分别重新表示成 $p(F|z, M)$ 、 $p(F|M)$ 和 $p(z|F, M)$ ； $p(F|M) = \int f(x) dx$ $p(F|z, M)p(z)$ 则变成了一个和变分分布 $q_\beta(z|F, M)$ 无关的常数。那么最大化变分下界 ELBO 就只相当于最小化变分分布 $q_\beta(z|F, M)$ ，我们的概率模型的目的就是通过最大化变分下界 ELBO 来寻找最优参数 β^* 。

将该先验分布代入变分下界 ELBO 中的 KL 项后，KL 项就变成如下公式：

$$\begin{aligned} & KL[q_\beta(z|F, M) \| p(z)] \\ &= \frac{1}{2}(-\log |\Sigma(F, M)| + \log |\Sigma_P| \\ &\quad + tr(\Sigma_P^{-1} \Sigma(F, M)) + \mu(F, M)^T \Sigma_P^{-1} \mu(F, M) - n) \\ &= \frac{1}{2}(-\log |\Sigma(F, M)| - \log |\Sigma_P^{-1}| \\ &\quad + tr(\Sigma_P^{-1} \Sigma(F, M)) + \mu(F, M)^T \Sigma_P^{-1} \mu(F, M) - n) \end{aligned}$$

图 3.7. KL 项

最终的损失函数为：

$$\begin{aligned} & \frac{1}{2}(-\log |\Sigma(F, M)| - \log |\Sigma_P^{-1}| + tr(\Sigma_P^{-1} \Sigma(F, M)) \\ & \quad + \mu(F, M)^T \Sigma_P^{-1} \mu(F, M) - \frac{\lambda}{K} \sum_{k=1}^K LCC(F, M(\phi_{z^k})) \end{aligned}$$

图 3.8. 损失函数

3.6 固定控制点紧支撑径向基函数形变模型

接下来设计了一个网络去实现这个非均匀控制点 CSRBF 形变模型，叫做 NetGI。根据这

个非均匀控制点的特点，NetGI 需要估计潜变量分布的均值 $\mu(F, M) = \mu_g, \mu_c$ 和协方差 $\Sigma(F, M) = \Sigma_g, \Sigma_c$ ，其中 μ_g 和 Σ_g 是全局控制点的分布， μ_c 和 Σ_c 是局部控制点的分布。NetGI 包括了两个模块 A 和 B，分别估计全局控制点和局部控制点的参数分布。其网络结构如下所示：

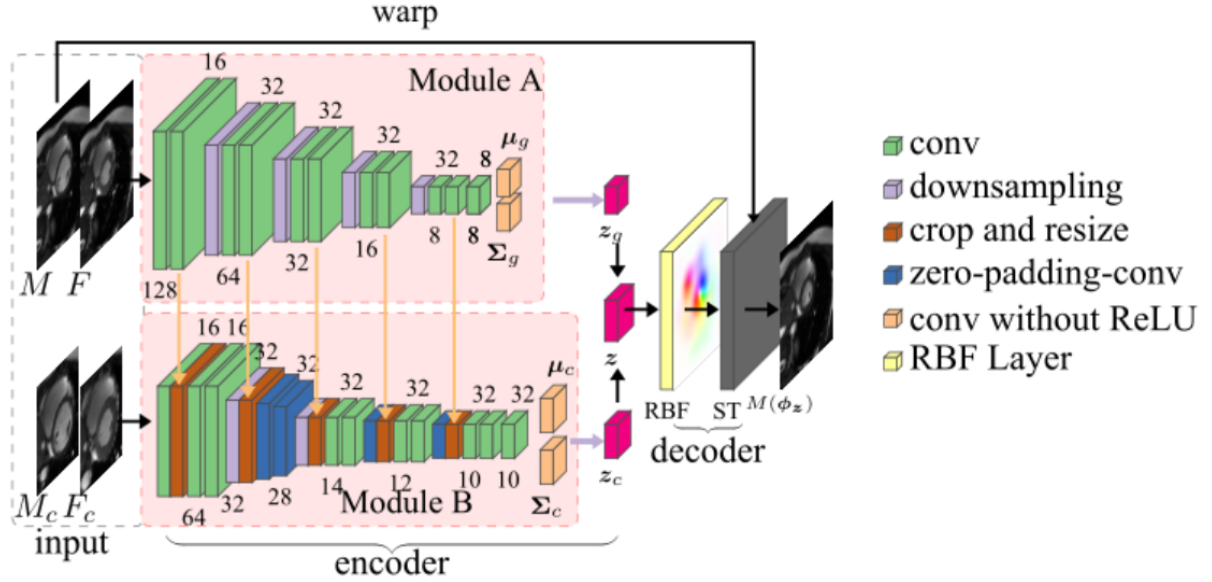


图 3.9. NetGI 网络架构

模块 A 将由源图像 M 和目标图像 F 组成的图像对当作输入，紧接着是 5 层结构。第 1 层结构是两层 16 个卷积核的卷积层，然后剩下 4 层都是一层 32 个卷积核的卷积下采样层，紧接着两层 32 个卷积核的卷积层。5 层结构后是一层 8 个卷积核的卷积层，然后两层卷积层分别输出全局控制点分布的均值 μ_g 和协方差 Σ_g 。

模块 B 则是学习局部控制点的概率分布参数 μ_c 和 Σ_c 。我们把输入的图像对的大小为 64×64 的中心区域裁剪出来当作模块 B 的输入。与模块 A 相似，模块 B 同样有着 5 层网络结构。第一层先对输入进行卷积，然后卷积结果与模块 A 第一层裁剪并且重新调整大小的特征图级联，然后紧接着两层卷积核。其他层的结构除了卷积核数目和下采样卷积外与第一层的结构类似。第二层的卷积核是没有边缘填充的，而且最后两层的结构的下采样卷积也是没有边缘填充的。第一层所有卷积层的卷积核的数目为 16，其余的均为 32。与模块 A 相似，模块 B 最后接着又一个卷积核数目为 8 的卷积层和两个分别输出局部控制点分布参数 μ_c 和 Σ_c 的卷积层。

3.7 可偏移控制点紧支撑径向基函数形变模型

对于可偏移控制点，设计了一个包含了两个分支的网络 NetDC，去估计初始控制点的偏移和对应控制点的系数分布参数。可偏移控制点分布的一个重要问题是如何将偏移后的控制点与网络中的关于参数的特征对应起来。我们采取的办法是使用 STN 层去将概率参数特征图与偏移后的控制点对齐。其网络结构如下苏轼，NetDC 同样由编码器和解码器组成。编码器包括了两种模块：1) 模块 F 负责计算概率参数相关特征和偏移控制点相关特征；2) 模块 O 负责输出初始控制点的偏移和对应的概率系数分布参数。

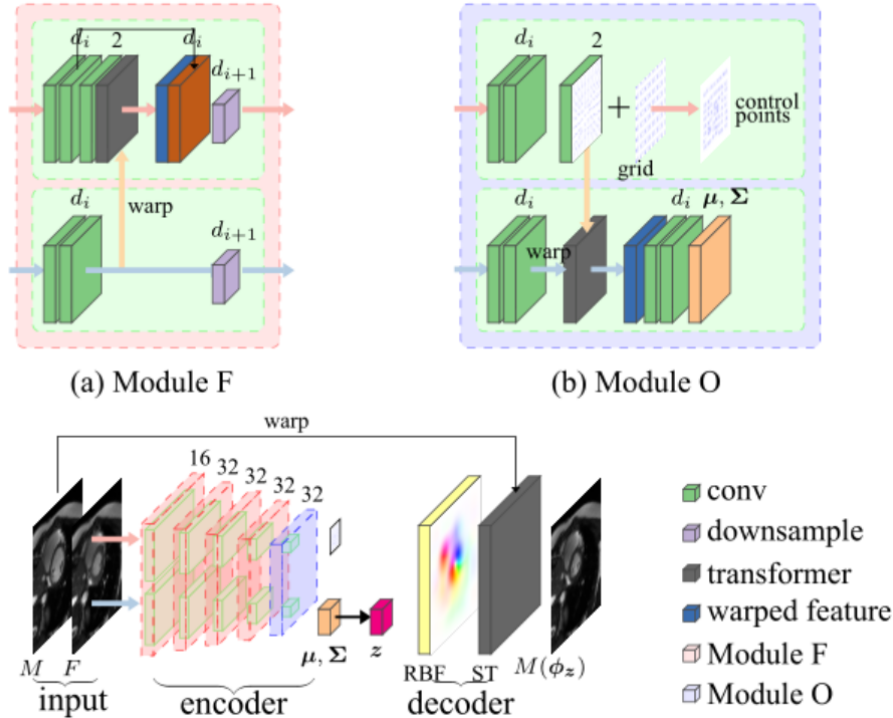


图 3.10. NetDC 网络架构

模块 F 的网络结构如图 3.10a 所示，它包含了上下两个分支。两个分支都是将图像对（如果作为第一个模块时）或者是该模块对应分支的输出作为输入。下分支负责提取概率分布参数的特征，它包含了两个卷积核数目为 d_i 的卷积层和一个卷积核数目为 d_{i+1} 的下采样卷积层。上分支负责分析初始控制点的偏移特征。该分支包含了两个核数目为 d_i 的卷积层和一个用于输出 DVF 的核数目为 2 的卷积层。接着，一个 STN 层会使用这个 DVF 将之前下分支计算好的特征图进行形变，然后与上分支的特征图级联，然后接着一个卷积核数目为 d_{i+1} 的下采样卷积层。两个分支的输出也是下一个模块的输入。

模块 O 的网络结构如图 3.10b，它同样包括上下两个分支。上分支包括了两个核数目为 d_i 的卷积层和一个用来输出初始控制点偏移的卷积层。初始控制点经过该偏移后就会得到图像对对应的控制点分布。下分支包括了两个核数目为 d_i 的卷积层，生成了与概率分布参数相关的特征图。这个特征图目前并不是和偏移后的控制点对应的上的，因此 NetDC 通过 STN 根据上分支输出的偏移将这个特征图与偏移后的控制点对齐。最后模块 O 紧接着两个核数目为 d_i 的卷积层和两个用于输出偏移后控制点的系数分布参数的卷积层。

4 复现细节

4.1 与已有开源代码对比

将紧支撑径向基函数形变模型从固定单一的支撑半径拓展到多支撑径向基半径，并且与变分推断模型结合，在损失函数中加入稀疏约束以从多个不同的支撑半径中选择最优的支撑半径以提升精度。我们还改进了损失函数，结合了互相关和均方差两种常用于深度学习的相似性度量，提出了加权局部互相关相似性度量方法。另外，雅可比行列式可以检测配准偏移

向量场的形变异常，我们利用雅可比行列式这个特性，基于此提出了雅可比行列式形变约束去约束偏移向量场的形变异常。实验验证了我们方法同样优于其他目前先进的方法。

4.2 实验环境搭建

表 1. 实验环境配置

操作系统	Windows10
开发环境	Python
虚拟环境	Anaconda
深度学习框架	PyTorch

4.3 数据集

4.3.1 York 数据集

该数据集收集了从 33 个 18 岁以下的病人获取的心脏短轴 MR 图像序列，这些序列有加拿大多伦多疾病儿童医院诊断影像系提供，共 7980 张 2 维图像。大多数病人显示有多种如心肌病、主动脉瓣反流和心肌病等心脏异常情况。每个心脏的影像序列包含了 20 帧的心脏图像，且每个图像有 8 到 15 片切片（同一个心脏的切片数量不变）。切片厚度在 6 13mm 之间。切片图像大小为 256×256 像素，且像素间距为 0.93 1.64mm。该数据集还提供了每幅图像的有专家标记的左心室心内膜和心外膜的手动分割图。

4.3.2 MICCAI2009 数据集

MICCAI2009 数据集是用于 2009 年心脏左心室分割挑战赛的数据，包含了 45 个从混合了不同病人和病理特征（包括健康，肥大，伴有梗塞的心力衰竭和无梗塞的心力衰竭）的心脏 MRI 图像。每个心脏 MR 序列包含了 20 帧的心脏图像，而且每个图像都有 6 12 片切片。切片厚度在 6 13mm 之间，图像像素分辨率为 256×256 ，且像素间距为 1.21 1.56mm。该数据集将 45 个心脏数据分成了三组，分别为 15 个训练数据，15 个测试数据和 15 个在线测试数据。同时，该数据集在每个心脏数据的 ED 相位提供了心内膜和心外膜的分割信息，在 ES 相位仅提供了心内膜的分割信息。

4.3.3 ACDC 数据集

ACDC 数据集采集的数据都是从第戎大学医院获得的真实临床检测数据，总共包含了 150 组来自不同病人的心脏数据。这些数据可以均分成五类：正常心脏，具有心肌梗塞病史的心脏，扩张型心肌病，肥厚型心肌病以及异常右心室。每个心脏 MR 序列包含了 28 40 帧的心脏图像，每个图像有 8 15 片切片，切片厚度为 5 或者 10mm。切片图像大小不固定，且像素间距为 0.833 1.95mm。该数据集将 150 个心脏数据分成了包含 100 个带有手动分割解剖结构（仅在 ED 相位和 ES 相位分割心内外膜和右心室）的心脏信息的训练集和包含 50 个没有心脏信息的测试集。

这三个数据集一共有 228 个短轴 cineMR 心脏序列图像。我们使用同一个心动周期的 ED 相位和 ES 相位中的同一个切片位置的 2 维图像切片对进行配准。数据集分为训练集、验证集和测试集。训练集包含了 136 个随机挑选的心脏序列图像中的 1257 个图像切片对；验证集包含了 16 个随机挑选的心脏序列图像中的 130 个图像切片对；测试集包含了剩余 76 个心脏序列图像中的 698 个图像切片对。我们以心脏为中心，将图像裁剪成 128×128 的大小。在训练阶段，我们对图像对使用了随机旋转、随机缩放、随机翻转和随机平移等数据增强手段。

4.4 创新点

将紧支撑径向基函数形变模型从固定单一的支撑半径拓展到多支撑径向基半径，并且与变分推断模型结合，在损失函数中加入稀疏约束以从多个不同的支撑半径中选择最优的支撑半径以提升精度。

5 实验结果分析

5.1 超参数

超参数 λ 可以用于控制 DVF 平滑度，越低，那么平滑约束就越强。但由于超参数 λ 是直接作用到损失函数的相似性度量上的，因此过低的 λ 会导致配准精度上的损失。为了探讨 λ 对配准精度的影响，我们使用了不同的 λ 训练 4 个不同的网络并且在验证集测试了 Dice、HD、APD 和 BE。下图展示了 NetGE64、NetGE256、NetGI 和 NetDC 的测量结果。可以看到，随着超参数 λ 的增加，配准精度也会上升，但是 DVF 的平滑度则会逐渐变差。当 λ 达到一定程度的时候，则配准精度就不会显著上升。

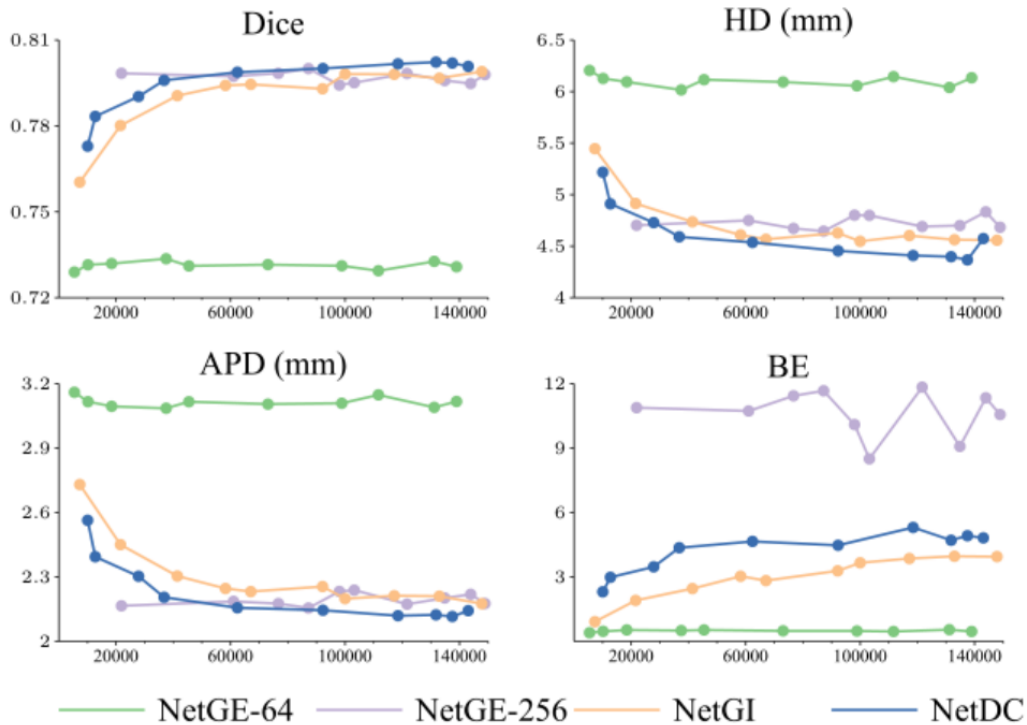


图 5.1. 不同网络不同超参数结果

另外, 对比 NetGE-256 和 NetGE-64 可以发现使用更多的控制点可以达到更好的配准精度。尽管 NetDC 和 NetGE-64 的控制点数目是一样的, 但很明显 NetDC 的配准精度要比 NetGE-64 好很多, 这说明了合理的控制点位置分布会影响配准精度, 这也验证了我们的可偏移控制点分布的有效性。NetGE-256 在 Dice 和 APD 精度上略高于 NetGI 的原因是 NetGE-256 使用了更多的控制点, 但 NetGI 的 HD 和 BE 要优于 NetGE-256。尽管以上的结论是基于验证集数据的, 但这个在测试集的结论是一致的。我们根据验证集的实验结果选择了最优的一个超参数。对于 NetGI, λ 为 147000; 对于 NetDC, λ 为 130000。

5.2 配准结果

5.2.1 创新前

为了对比目前最好的深度学习方法, 我们采用并实现了三个不同的网络模型 KrebsDiff^[13]、DalcaDiff^[14] 和 VoxelMorph^[15]。KrebsDiff 和 DalcaDiff 是两个使用了变分推断实现概率同胚微分配准; 而 VoxelMorph 则是使用无监督的图像对图像的配准网络。对于这些网络, 我们采取了最优参数去训练网络: 对于 KrebsDiff, 超参数 λ 和潜变量维度分别为 60000 和 64; 对于 DalcaDiff, 超参数 λ 和 σ 分别为 50 和 0.03; 对于 VoxelMorph, 超参数 λ 为 1。下图列出了不同方法在不同数据上的结果。在 MICCAI 和 York 数据集上, NetDC 是最好的; 而 NetGI 仅次于 NetDC。尽管 DalcaDiff 在 ACDC 数据集上获得了最好的 Dice 结果, NetDC 在 HD 和 APD 的结果上稍微好过 DalcaDiff。对于 DVF 的平滑度, NetDC 和 NetGI 在三个数据集上都获得了非常好的性能。由于 KrebsDiff 和 DalcaDiff 使用了同胚微分配准模型, 因此这两种方法都能生成形变异常比较小的 DVF, 同样的, NetDC 和 NetGI 在没有同胚微分约束的情况下, 也能生成形变异常与这两种方法相当的 DVF。可以看到 NetDC 总体上的配准精度是最好的, 而且可以生成近乎平滑的几乎一一对应的 DVF。结果如下所示:

数据集	方法	BE [†]	$ J_{\phi_x} \leq 0$	Dice	HD	APD
MICCAI	KrebsDiff	26.20(13.80)	0.33(0.67)	0.857(0.074)	5.57(2.16)	2.60(1.09)
	DalcaDiff [†]	62.40(21.76)	0.01(0.02)	0.856(0.000)	5.755(0.00)	2.66(0.00)
	VoxelMorph [†]	154.58(24.63)	22.58(12.87)	0.829(0.105)	6.55(2.58)	3.18(1.57)
	NetGI	5.16 (1.72)	0	0.858(0.081)	5.48(2.07)	2.54(1.13)
	NetDC	6.92(2.85)	1.08(3.37)	0.861 (0.076)	5.34 (2.03)	2.48 (1.09)
York	KrebsDiff	22.65(6.89)	0	0.793(0.093)	6.64(2.40)	2.98(1.14)
	DalcaDiff [†]	91.56(55.12)	0.01(0.03)	0.785(0.094)	7.67(1.57)	3.25(1.03)
	VoxelMorph [†]	194.51(36.04)	38.06(17.89)	0.795(0.083)	7.20(2.36)	3.05(1.13)
	NetGI	4.89 (0.92)	0	0.798(0.078)	6.59(2.17)	2.92(1.07)
	NetDC	6.94(1.50)	0	0.812 (0.078)	6.25 (2.18)	2.73 (1.01)
ACDC	KrebsDiff	23.96(16.91)	1.50(10.47)	0.800(0.086)	5.49(1.63)	2.24(0.67)
	DalcaDiff [†]	143.73(77.66)	0.20(0.50)	0.807(0.092)	5.64(1.15)	2.19(0.32)
	VoxelMorph [†]	166.72(37.50)	27.84(26.08)	0.804(0.096)	5.81(2.22)	2.29(0.98)
	NetGI	4.14 (1.51)	0.12 (0.75)	0.798(0.093)	5.45(1.99)	2.27(0.87)
	NetDC	5.57(2.52)	0.16(0.72)	0.805(0.087)	5.18 (1.71)	2.14 (0.72)
所有	KrebsDiff ^[55]	24.21(15.29)	1.05(8.52)	0.802(0.132)	5.65(2.25)	2.36(0.99)
	DalcaDiff ^[54]	120.13(75.12)	0.13(0.41)	0.807(0.133)	5.92(2.45)	2.36(1.06)
	VoxelMorph ^[49]	168.35(37.03)	28.28(23.39)	0.804(0.139)	6.04(2.81)	2.45(1.31)
	NetGI	4.45 (1.55)	0.08 (0.61)	0.802(0.137)	5.60(2.42)	2.37(1.17)
	NetDC	6.03(2.55)	0.32(1.65)	0.810 (0.131)	5.33 (2.17)	2.24 (0.99)

图 5.2. 模型对比结果

5.2.2 创新后

我们同样使用了 KrebsDiff, DalcaDiff 和 VM 与我们的网络进行对比, 同时我们使用超参数优化工具 Hyperopt 在验证集上寻找最优的超参数。

下图列出了我们的模型与其他网络在 MICCAI2019、York 和 ACDC 数据集上的配准结果。“Ours-a” 代表了我们的模型使用了 a 个支撑的配准结果。可以看到, 我们的模型在配准精度上要比其他网络的要高, 而且无论是否使用了多支撑, 从扭曲能量看, 我们的模型生成的 DVF 都要比其他网络的要平滑。由于 DalcaDiff[54] 和 KrebsDiff[55] 采用的是同胚微分配准模型, 因此生成的形变场的拓扑保持性很好, 尤其是 DalcaDiff[54]。尽管如此, 我们的网络在没有使用同胚微分模型的情况下, 生成的型变成拓扑保持情况与 DalcaDiff 接近, 甚至比 KrebsDiff 更好。另外, 与比使用单个支撑的模型相比, 使用多个支撑的模型在配准精度上更好, 这说明了多支撑 RBF 模型是有用的。从形变场平滑度上看, 使用多支撑的模型生成的形变场要比使用单支撑的要稍微不平滑, 这可能是多支撑的模型在部分控制点选择了半径更小的支撑集, 导致生成了更为复杂更精确的形变场。

网络	BE	$ J_f \leq 0$	Dice	HD	APD
VM	572.16(150.15)	56.21(34.20)	0.841(0.070)	6.10(2.08)	2.21(0.97)
DalcaDiff	113.92(82.06)	0.03(0.09)	0.850(0.052)	5.58(1.63)	2.04(0.56)
KrebsDiff	25.99(14.85)	1.59(12.51)	0.847(0.056)	5.46(1.70)	2.10(0.64)
Ours-1	5.61(2.44)	0.07(0.57)	0.854(0.053)	5.23(1.62)	2.01(0.62)
Ours-3	9.15(3.79)	0.13(0.68)	0.865(0.047)	5.03(1.56)	1.85(0.53)

图 5.3. 不同网络在 MICCAI2009、York 和 ACDC 数据集的配准结果

M&Ms 数据集提供了 4 个设备供应商的心脏 MRI 数据, 但训练集中并没有设备供应商 D 的 MRI 数据。下图列出了我们的模型与其他方法在 MMs 数据集上不同设备供应商不同生理结构的配准结果以及平均配准精度。可以看到, 我们的模型在平均的配准精度上和大部分分类上都取得最优的精度。而且从网络在设备供应商 D 的配准精度看, 我们的网络是具有一定的泛化能力的。

方法	A			B			C			D			平均值
	LV	Myo	RV	LV	Myo	RV	LV	Myo	RV	LV	Myo	RV	
VM	0.944	0.834	0.842	0.934	0.783	0.782	0.946	0.813	0.834	0.935	0.713	0.79	0.845
DalcaDiff	0.947	0.841	0.842	0.927	0.827	0.783	0.938	0.832	0.836	0.93	0.81	0.787	0.858
KrebsDiff	0.949	0.841	0.844	0.932	0.817	0.777	0.938	0.816	0.821	0.935	0.769	0.785	0.852
Ours-1	0.949	0.835	0.853	0.929	0.817	0.778	0.937	0.826	0.826	0.937	0.809	0.778	0.856
Ours-3	0.944	0.829	0.855	0.934	0.835	0.806	0.944	0.844	0.847	0.938	0.816	0.807	0.867

图 5.4. 不同网络在 MMs 数据集四个数据供应商的 Dice 对比结果

6 总结与展望

在医学图像处理中, 医学图像配准在许多医疗临床应用中是必不可少的步骤。如何更准确快速地实现两张图像之间的对齐一直是研究的重点。深度学习由于其可以实现更加精确而

且较传统算法难以匹敌的速度使其成为了可形变图像配准（其实在许多其他领域也是）的热门。但在基于传统机器学习的图像配准中已有许多借鉴之处。本文针对目前卷积神经网络在图像配准应用中的一些问题，将用在传统机器学习算法的紧支撑径向基函数和深度生成模型结合，并且对此做出改进：在基于变分推断的概率生成模型中结合紧支撑径向基函数的特点在变分下界中推导出扭曲能量；在紧支撑径向基函数的非均匀控制点的基础上拓展出自适应的可偏移控制点方法以及拓展支持多支撑半径；根据局部互相关和均方差的特性提出加权局部互相关和根据雅可比行列式的特性提出新的形变约束。我们通过实验验证了以上的改进都是有效的，在配准精度上和形变场的平滑度和拓扑保持上均有提高。

6.1 本文主要工作和创新点

(1) 图像配准可以看作是一个源图像经过涉及某种潜变量的空间变换生成目标图像的生成过程。我们构建了一个基于变分推断的图像配准概率生成模型，使用条件变分编码器生成潜变量难以计算的后验分布并且根据源图像和采样的潜变量重构目标图像。非参数化形变模型易导致 DVF 的平滑性难以控制，因此我们采用了 CSRBF 来实现基于深度学习的可形变图像配准。我们使用了非均匀控制点来实现 CSRBF，并且根据控制点分布方式的不同设计了两种不同的网络。在本文中，我们将潜变量的先验分布假设成一个关于控制点分布的矩阵，使得可以在变分下界中推导出约束平滑的扭曲能量。对于非均匀控制点分布方式，我们提出了可偏移控制点分布去使用网络自适应地根据图像对来对初始控制点进行偏移，同时得到控制点系数分布和位置分布。我们在该概率模型中测量并分析了配准结果的不确定性。

(2) 我们还采用了多支撑 CSRBF 来在一个控制点中使用多个不同支撑半径的 CSRBF 来提高配准精度。同时，基于目前常用于深度学习图像配准的相似性度量的优缺点，我们结合了 LCC 和 MSE 的特点，利用 MSE 计算的强度差作为 LCC 计算每个像素附近局部区域的互相关值的权重，提出了加权 LCC 相似性度量。为了约束 DVF 的形变异常，我们根据雅可比行列式的特性，提出了基于雅可比行列式的形变约束。

6.2 未来的工作展望

(1) 可偏移控制点分布与多支撑径向基函数的结合：本文提出了可偏移控制点分布方式和基于多支撑径向基函数的网络，这两者虽然都是对径向基函数在基于深度学习的图像配准中的创新和改进，但在实现中是分开实现的。接下来我们将尝试将这两者结合，有望在保持 DVF 一定平滑度的前提下，提高配准精度。

(2) 加权 LCC 相似性度量的鲁棒性研究：尽管加权 LCC 相似性度量在心脏数据中拥有很不错的性能，但我们对该相似性度量的适应性尚未研究。我们将会使用不同的方法和不同的医疗配准任务（比如脑部、肺部图像的配准）进行实验，去继续探索加权 LCC 相似性度量的性能以及不足。

参考文献

- [1] SOTIRAS A, DA V A TZIKOS C, PARAGIOS N. Deformable Medical Image Registration: A Survey[J]. IEEE Transactions on Medical Imaging, 2013, 32(7) : 1153 –1190.

- [2] PERPERIDIS D, MOHIADDIN R, RUECKERT D. SpatioTemporal FreeForm Registration of Cardiac MR Image Sequences[C] // BARILLOT C, HAYNOR D R, HELLIER P. Medical Image Computing and Computer Assisted Intervention – MICCAI 2004. Berlin, Heidelberg : Springer Berlin Heidelberg, 2004 : 911 –919.
- [3] MCLEOD K, SERMESANT M, BEERBAUM P , et al. SpatioTemporal Tensor Decomposition of a Polyaffine Motion Model for a Better Analysis of Pathological Left V entricular Dynamics[J]. IEEE Transactions on Medical Imaging, 2015, 34(7) : 1562 –1575.
- [4] VERESS A I, GULLBERG G T, WEISS J A. Measurement of strain in the left ventricle during diastole with cineMRI and deformable image registration.[J]. Journal of biomechanical engineering, 2005, 127 : 1195 –207.
- [5] PHA TAK N S, MAAS S A, VERESS A I, et al. Strain Measurement in the Left V entricle During Systole with Deformable Image Registration[C] // SACHSE F B, SEEMANN G. Functional Imaging and Modeling of the Heart. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007 : 32 –40.
- [6] HASKINS G, KRUGER U, YAN P. Deep learning in medical image registration: a survey[J]. Machine Vision and Applications, 2020, 31 : 1 –18.
- [7] FU Y, LEI Y, WANG T, et al. Deep Learning in Medical Image Registration: A Review[J]. Physics in medicine and biology, 2020.
- [8] WU G, KIM M, WANG Q, et al. Scalable HighPerformance Image Registration Framework by Un supervised Deep Feature Representations Learning[J]. IEEE Transactions on Biomedical Engineering, 2016, 63(7) : 1505 –1516.
- [9] EPPENHOF K A J, PLUIM J P W. Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks[J]. Journal of Medical Imaging, 2018, 5(02) : 1.
- [10] KREBS J, MANSI T, DELINGETTE H, et al. Robust Nonrigid Registration Through AgentBased Action Learning[C] // DESCOTEAUX M, MAIERHEIN L, FRANZ A, et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. Cham : Springer International Publishing, 2017 : 344 –352.
- [11] MIAO S, PIA T S, FISCHER P W, et al. Dilated FCN for MultiAgent 2D/3D Medical Image Registration[C] // AAAI. 2018.
- [12] BUHMANN M. A new class of radial basis functions with compact support[J]. Mathematics of Computation, 2000, 70(233) : 307 –318.