

基于训练-推理解耦架构的 2D-3D 医学图像配准

李文举¹, 孔德卿¹, 曹国刚^{1*}, 李思诚¹, 戴翠霞²¹上海应用技术大学计算机科学与信息工程学院, 上海 201418;²上海应用技术大学理学院, 上海 201418

摘要 2D-3D 医学图像配准在解决放射治疗的摆位验证问题上发挥着至关重要的作用。针对现有配准方法精度不高、耗时较长等问题,提出一种基于训练-推理解耦架构的 2D-3D 医学图像配准方法。该方法在训练阶段采用多分支结构,利用多尺度卷积增强特征多样性,提高配准精度;在推理阶段,利用重参数化将多分支结构转换为单路结构,加快配准速度;另外,引入自适应激活函数 Meta-ACON,提高网络的非线性表达能力。在胸部和骨盆两个数据集上进行训练和测试,实验结果表明,所提方法预测结果的位移误差约为 0.08 mm,角度误差约为 0.05°,配准时间仅需 26 ms。所提方法有效解决了摆位验证中的医学图像配准问题,不仅提高了精度,还满足了临床实时性要求。

关键词 图像处理; 2D-3D 配准; 卷积神经网络; 重参数化; 激活函数

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1610015

2D-3D Medical Image Registration Based on Training-Inference Decoupling Architecture

Li Wenjü¹, Kong Deqing¹, Cao Guogang^{1*}, Li Sicheng¹, Dai Cuixia²¹School of Computer Science & Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China;²School of Sciences, Shanghai Institute of Technology, Shanghai 201418, China

Abstract The registration of 2D-3D medical images is crucial in solving radiotherapy positioning verification. A 2D-3D medical image registration approach based on training-inference decoupling architecture is proposed to address the issues of low accuracy and time-consuming processes. The multibranch structure and multiscale convolution were employed in the training phase to enhance feature diversity and improve registration accuracy. During the inference phase, the multibranch structure was reparameterized into a single-channel structure to speed up the registration speed. Additionally, an adaptive activation function, Meta-ACON, was used to increase the network's nonlinear expression. Two datasets of the chest and pelvis were used for training and testing. The experimental results show that the mean translation error of the proposed method is approximately 0.08 mm, the mean angular error is approximately 0.05°, and the registration time reaches 26 ms. The proposed method significantly improves the accuracy of medical image registration in positioning verification while meeting the real-time requirements of clinical applications.

Key words image processing; 2D-3D registration; convolution neural network; re-parameterization; activate function

1 引言

放射治疗是治疗恶性肿瘤的重要手段之一,原理是利用放射性射线照射勾画靶区内的肿瘤。在放疗过程中,病人摆位、器官形变等因素会使靶区位置发生偏移,导致肿瘤周围的正常组织受到损害。随着生物影

像技术和放疗设备的发展,图像引导放射治疗(IGRT)通过对病人摆位验证和靶区位置实时跟踪,提高肿瘤细胞的杀灭率和病人的存活率^[1]。其中,2D-3D 医学图像配准在解决病人摆位验证的问题上发挥着关键作用,通过减小病人摆位误差,有效地避免试剂照射到正常组织。

收稿日期: 2021-11-16; 修回日期: 2022-01-26; 录用日期: 2022-02-25

基金项目: 国家自然科学基金(62175156, 61675134, 81827807)、上海市科委科技创新行动计划(19441905800)、温州医科大学重点实验室开放项目(K181002)

通信作者: *guogangcao@163.com

医学图像配准是在两幅图像间寻找最优空间变换参数来对齐图像的过程,其中 2D-3D 配准是指体数据(如 CT、MRI 等)和二维数据(如 X-Ray、超声等)之间的配准^[2]。传统 2D-3D 配准方法常利用图像的灰度值来实现配准,无需特征提取等人工干预,具有较好的配准精度和鲁棒性^[3]。然而,传统配准方法需要进行循环迭代生成数字重建放射影像(DRR),这其中涉及大量运算,占用过多配准时间。为了提高配准效率,刘坤等^[4]提出一种基于 Bresenham 直线改进的光线投射法,该方法在配准实验中的速度提升 3~6 倍。另外,配准中常用的优化算法容易陷入局部极值。差异进化算法、粒子群算法等全局优化算法的提出^[5],有效地缓解了局部最优现象,但该类算法仍存在收敛慢等问题。

近年来,一些学者将深度学习应用于医学图像配准中,这些方法利用卷积神经网络(CNN)自动提取图像特征,预测空间变换参数,如 VoxelMorph 框架^[6]等。按照不同训练方式,基于深度学习的配准方法可分为全监督、弱监督和无监督三类^[7]。由于 DRR 图像具有真实标签的特点,在 2D-3D 医学图像配准领域大多采用监督学习的方法。Miao 等^[8]将 2D-3D 配准视为回归问题,通过构建 5 层网络结构,并引入局部图像残差、参数空间划分等算法策略,首次成功地将 CNN 应用到 2D-3D 图像配准中。陈向前等^[9]利用 GoogLeNet 构建配准网络,分别训练 6 个分类模型,输出变换参数的预测概率。由于该方法构建的分类模型预测变换参数为整数,配准精度有待提高,更适用于粗配准。贾瑞明等^[10]提出一种具有分支解码结构的深度神经网络,该网络使用叠加的数字重建影像库进行训练,预测术中 X 射线图像的位置姿态参数,从而为医生手术提供导航辅助。Gao 等^[11]提出投影空间变换(ProST)模块生成 DRR 图像,并将其应用到 2D-3D 医学图像配准,实现端到端的配准模型。但该方法的位移误差较大,达到 7 mm 左右,难以满足配准的精度要求,需进一步优化网络结构。

综上所述,传统 2D-3D 配准方法存在配准时间长、

局部最优等问题;深度学习方法利用 CNN 自动提取特征,有效提升了配准效率,但配准精度仍需进一步提高。为了达到高精度和实时性的需求,本文采用训练-推理解耦的网络结构。在训练阶段,构建多分支的深层卷积网络,利用不同尺寸的卷积核提取更丰富的图像特征,从而提高配准精度;在推理阶段,利用重参数化方法将多分支结构转换为单路结构,从而缩短配准时间。另外,常用的激活函数存在强制稀疏处理操作,减少了有效特征的提取步骤,本文引入自适应激活函数,利用可学习参数激活神经元,提高网络的非线性表达能力。

2 配准方法

2.1 配准定义

设 2D-3D 配准的输入为二维 X-Ray 图像 X 和三维 CT 图像 V ,配准目标是在统一的空间坐标系下,寻找目标变换参数 T_{out} :

$$T_{out} = \text{MaxSim}(X, V_t), \quad (1)$$

式中: V_t 表示三维 CT 图像在空间变换参数 t 下生成的二维 DRR 图像;MaxSim(\cdot)表示最大化 X-Ray 图像和二维 DRR 图像的相似性。利用 CNN 自动提取图像特征,预测空间变换参数,将 2D-3D 配准转换为一个回归问题,即把 T_{out} 转换为

$$T_{out} = f(X, V_t), \quad (2)$$

式中: $f(\cdot)$ 表示 CNN 回归器。由于 X-Ray 图像的标签制作依赖于传统方法和医生人工标定,数量较少且难以保证数据真实性,使用已知真实变换参数 T_g 的模拟 X-Ray 图像 X_s (即 DRR 图像)代替,式(2)转换为

$$T_{out} = f(X_s, T_g). \quad (3)$$

至此,将配准目标转换为基于监督学习的回归模型。不同于以往训练和推理同一模型的方式,本文对两阶段模型进行解耦处理,即根据需求设计不同的架构。考虑到训练模型和推理模型的要求,将多分支结构应用于训练阶段,将单路结构应用于推理阶段。方法的整体配准框架如图 1 所示。首先,术前获取患者的三维 CT,使用 Siddon 光线追踪法生成大量模拟 X-

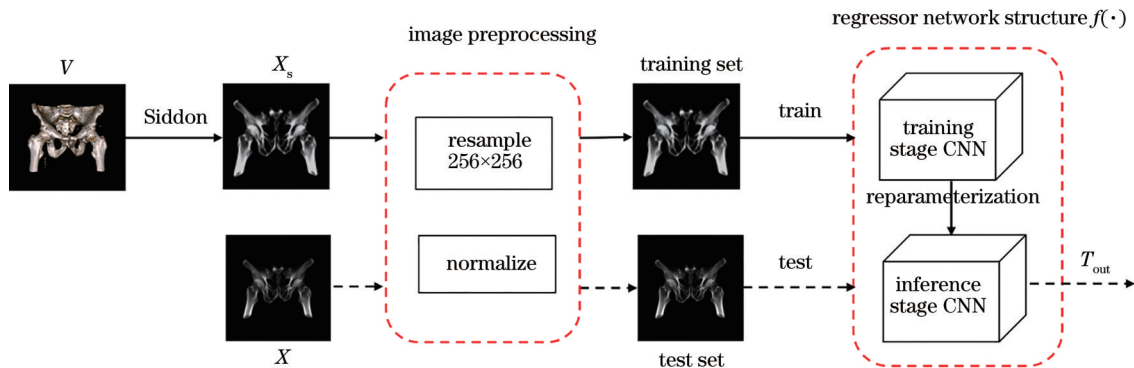


图 1 配准框架

Fig. 1 Registration framework

Ray 图像;接着,进行图像预处理,将预处理结果作为训练集输入训练阶段 CNN;然后,待网络模型训练完毕,利用重参数化方法将训练模型转换为推理模型;最后,术中将 X-Ray 图像预处理后作为测试集,输入推理阶段 CNN,经过一次前向传播,即可输出配准变换参数。将预测的空间变换参数用来调整病人位置,从而

减小病人摆位误差,有效地避免试剂照射到正常组织。

所提方法适用于 2D-3D 医学图像刚体配准,包括 3 个位移参数 t_x, t_y, t_z 和 3 个角度参数 $t_\theta, t_\alpha, t_\beta$, 变换参数引起的效果如图 2 所示。其中位移参数 t_x, t_y 和角度参数 t_θ 的效果近似 2D 刚体变换,位移参数 t_z 使图像细微缩放,角度参数 t_α, t_β 引起图像的复杂形状变化。

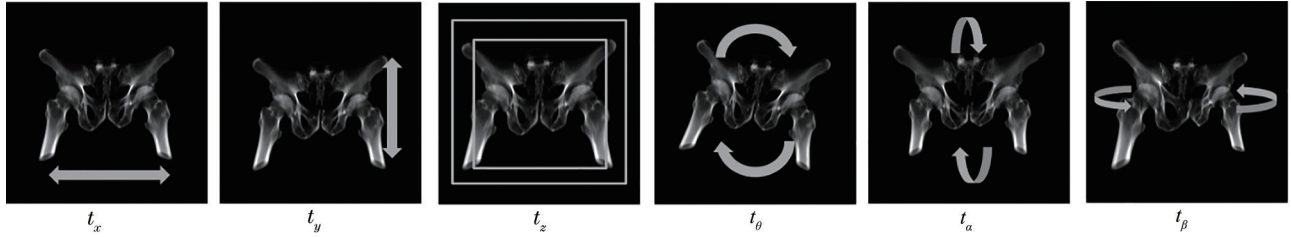


图 2 变换参数效果图

Fig. 2 Transform parameter renderings

2.2 网络结构

回归器的网络结构基于 Ding 等^[12] 提出的 RepVGG 网络,具有训练-推理解耦的架构体系,并引入自适应激活函数 Meta-ACON^[13]。在训练阶段使用多分支结构,在推理阶段利用重参数化将其转换为单路结构。其中训练阶段的网络结构如图 3 所示,采用 28 层的卷积神经网络。首先,输入 256×256 的灰度图像,图像经过 5 个阶段的卷积,每个阶段卷积层数分别为 [1, 4, 6, 16, 1], 通道数分别为 [64, 128, 256, 512, 2048];接着,在每阶段利用 F-Block (First Block) 中步长为 2 的卷积层进行下采样,并将

Block 中的卷积层步长设为 1,从而保留更多空间细节信息;最后,通过全局平均池化层 (GAP) 和全连接层 (FC) 输出 6 个配准变换参数。另外, F-Block 和 Block 的结构如图 3 虚线框所示,在主干网络采用 3×3 的卷积层,同时加入 1×1 的卷积层和恒等映射 (Identity) 分支。由于 Identity 要求特征输入和输出形状一致, F-Block 只加入 1×1 的卷积分支,且各分支都经过批归一化层 (BN),使数据分布一致,加快了收敛速度。对各分支的结果执行 Add 操作后,通过 Meta-ACON 自适应激活函数,增强网络的非线性拟合能力。

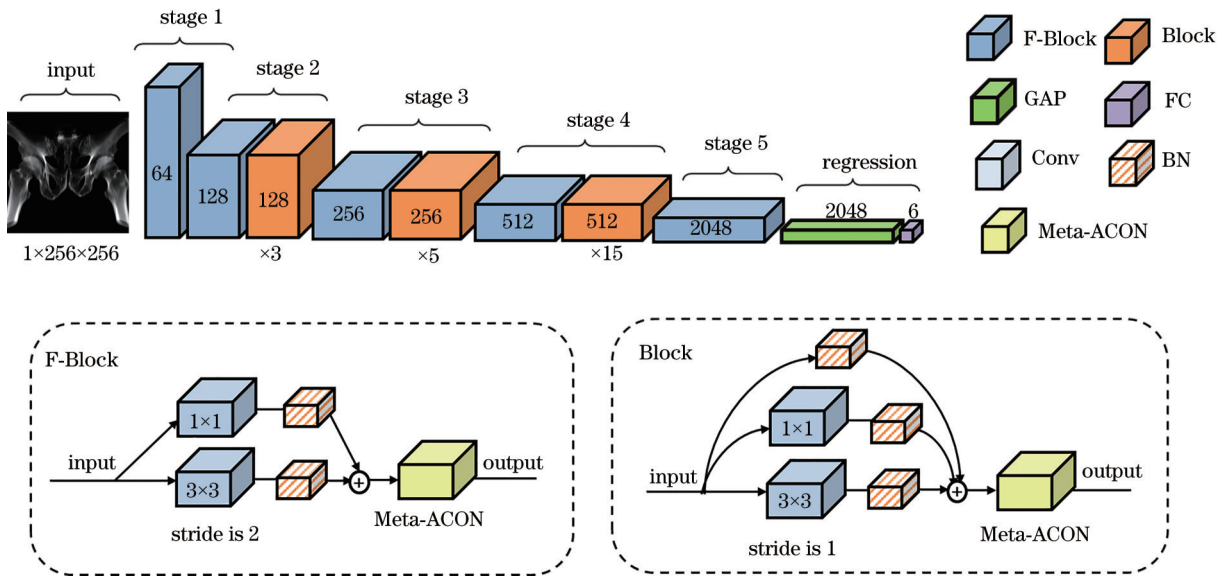


图 3 回归器网络结构

Fig. 3 Network structure of regressor

2.3 重参数化

为了加快推理速度,使用重参数化方法将训练模型转换为推理模型,即重新整合多分支网络的参数,使之等价于单路网络。具体来说,即把每个多分支 Block

等价为一个 3×3 卷积层,重参数化的转换过程如图 4 所示。首先将卷积层和 BN 层融合,定义为

$$\text{Conv}(x) = \omega(x) + b, \quad (4)$$

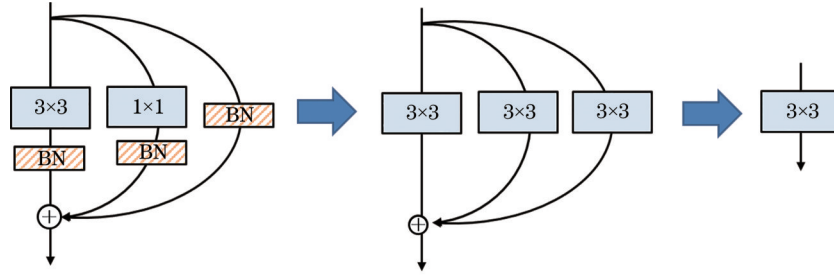


图 4 重参数化过程

Fig. 4 Re-parameterization process

$$\text{BN}(x) = \gamma \cdot \frac{(x - m)}{\sqrt{v_{\text{ar}}}} + \beta_{\text{bias}}, \quad (5)$$

式中: x 是输入数据; ω 和 γ 分别为卷积层和 BN 层的权重; b 和 β_{bias} 分别为两者的偏置项, 此处 b 设为 0; m 为均值; $\sqrt{v_{\text{ar}}}$ 为标准差。将卷积层的结果代入 BN 层, 公式为

$$\text{BN}[\text{Conv}(x)] = \gamma \cdot \frac{[\omega(x) - m]}{\sqrt{v_{\text{ar}}}} + \beta_{\text{bias}}. \quad (6)$$

融合结果等价为一个加入 BN 层权重和偏置项的卷积层:

$$\text{BN}[\text{Conv}(x)] = \frac{\gamma \cdot \omega(x)}{\sqrt{v_{\text{ar}}}} + \left[\frac{\gamma \cdot (-m)}{\sqrt{v_{\text{ar}}}} + \beta_{\text{bias}} \right]. \quad (7)$$

接着, 将融合 BN 层的 Identity、 1×1 卷积分别转换为 3×3 的卷积层。由于 Identity 操作前后值不变, 考虑将其转换为 1×1 卷积, 并以单位矩阵作为卷积核。具体地, 将 Identity 等价于分通道卷积, 将当前通道的权重设为 1, 其他通道权重设为 0; 将 Identity 转换为 1×1 卷积后, 再将所有的 1×1 卷积转换为 3×3 的卷积层。通过观察发现, 在同一特征图上进行卷积操作, 1×1 卷积的路径等同于 3×3 卷积的中间核, 因此在 1×1 卷积核周围用 0 填充为 3×3 的卷积核大小, 即可转换为 3×3 卷积。

最后, 将各分支的卷积核相加, 对应的偏置项也相加, 即可等价推理模型。

2.4 激活函数

为了提高模型的非线性表达能力, 神经网络中往往需要加入激活函数。目前常用的是线性整流单元 (ReLU), 作为分段线性函数, 其正值不变, 负值置为 0。ReLU 函数非饱和性的特点有效地避免梯度消失问题, 加快了训练速度, 但由于其对负梯度完全不激活, 强制的稀疏处理会屏蔽一些有效特征。随着神经网络架构搜索 (NAS) 的发展, 具有平滑、非单调特性的 Swish 激活函数被广泛使用, 公式为

$$\text{Swish}(x) = x \cdot \sigma(\beta_c x), \quad (8)$$

式中: $\sigma(\cdot)$ 表示 Sigmoid 函数; β_c 表示常数。

为了探究激活函数之间的关系, 将 ReLU 函数归类为 MAX 函数, 将 MAX 函数的平滑称为 Smooth

Maximum, 公式为

$$S_{\beta}(x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}}, \quad (9)$$

式中: β 是平滑因子, $\beta = 0$ 时, 该平滑相当于线性函数。进而考虑 $n = 2$ 的具体情况, 公式为

$$S_{\beta}(x_1, x_2) = \frac{x_1 e^{\beta x_1}}{e^{\beta x_1} + e^{\beta x_2}} + \frac{x_2 e^{\beta x_2}}{e^{\beta x_1} + e^{\beta x_2}} = \frac{x_1}{1 + e^{-\beta(x_1 - x_2)}} + \frac{x_2}{1 + e^{-\beta(x_2 - x_1)}} = x_1 \cdot \sigma[\beta(x_1 - x_2)] + x_2 \cdot \sigma[\beta(x_2 - x_1)]. \quad (10)$$

由于 $\sigma(-x) = 1 - \sigma(x)$, 式 (10) 可转换为

$$S_{\beta}(x_1, x_2) = (x_1 - x_2) \cdot \sigma[\beta(x_1 - x_2)] + x_2. \quad (11)$$

将 $S_{\beta}(x, 0)$ 代入式 (11), 发现 Swish 函数其实是 ReLU 函数的一种平滑形式, 两个函数在 $(-5, 5)$ 范围内的曲线如图 5(a) 所示。进一步推广该平滑形式, 得到一系列 ACON (Activate Or Not) 激活函数, 其中最广泛的平滑形式 ACON-C 定义为

$$\text{ACON-C}(x) = (p_1 - p_2) x \cdot \sigma[\beta(p_1 - p_2)x] + p_2 x, \quad (12)$$

式中: p_1 和 p_2 为可学习参数, 控制函数上下界的值。

图 5(b) 为不同参数下的 ACON-C 函数图像, $p_1 = 1$ 和 $p_2 = 0$ 时等价于 Swish 函数。 β 参数控制神经元激活与否 ($\beta = 0$ 时视为不激活), 因此利用卷积等操作计算 β 值, 在 ACON-C 的基础上提出 Meta-ACON, 实现自适应激活。具体操作: 设输入特征大小为 $C \times H \times W$, 首先在 H, W 维度上求均值, 接着经两个 1×1 的卷积层, 最后使用 Sigmoid 函数将结果映射到 $(0, 1)$ 范围内, 得到计算的 β 值, 用于判断是否激活。

3 实验与分析

3.1 实验设置

实验配置为 Intel Core i5-9300H CPU 和 NVIDIA GeForce GTX 1660Ti GPU, 操作系统是 Windows 10, 开发软件包括 PyCharm Professional 2019.2 和 Microsoft Visual Studio 2019, 实验环境为 Python3.6

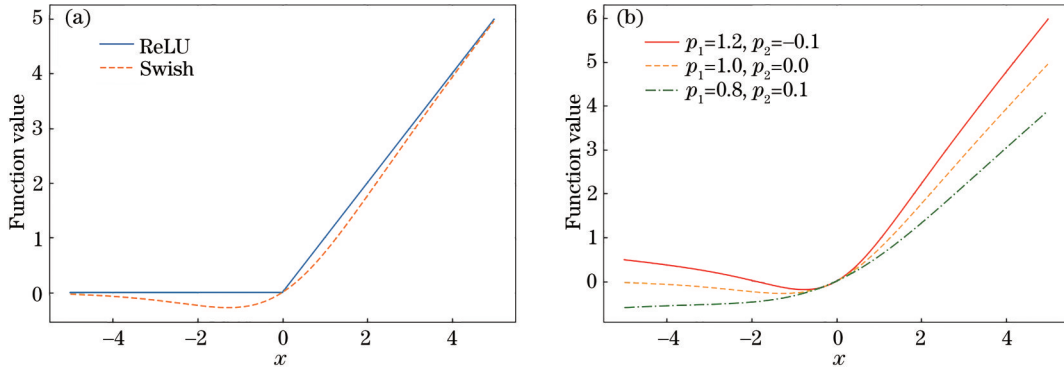


图5 激活函数。(a)ReLU/Swish;(b)不同参数下的ACON-C

Fig. 5 Activate function. (a) ReLU/Swish; (b) ACON-C under different parameters

和深度学习框架Pytorch 1.6。

训练时采用的损失函数是HuberLoss,公式为

$$L = \begin{cases} \frac{1}{2} \sum_{i=1}^n [y_i - f(x_i; \omega)]^2, & |y_i - f(x_i; \omega)| \leq \delta \\ \delta \sum_{i=1}^n |y_i - f(x_i; \omega)| - \frac{1}{2} \delta^2, & |y_i - f(x_i; \omega)| > \delta \end{cases}, \quad (13)$$

式中: n 是训练样本的数量; y_i 是第 i 个训练样本的真实标签; ω 是要学习的权重; $f(x_i; \omega)$ 是第 i 个训练样本的预测值; δ 用来调节误差的阈值,设为1。HuberLoss使用参数调节误差范围,避免了L2损失函数对局外点敏感的问题,且相对于L1损失函数,寻找最优解的过程更为高效。

对于权重 ω ,使用Adam优化器学习,初始学习率设为 3×10^{-4} ,使用固定步长衰减,每隔10个epoch,学习率降为原来的1/10,公式为

$$r_{\text{new_lr}} = r_{\text{initial_lr}} \cdot \gamma^{\lfloor \frac{E_{\text{epoch}}}{s_{\text{step_size}}} \rfloor}, \quad (14)$$

式中: $r_{\text{new_lr}}$ 是更新后的学习率; $r_{\text{initial_lr}}$ 是初始学习率; γ 是衰减系数,设为0.1; $s_{\text{step_size}}$ 是固定步长,设为10; E_{epoch} 为当前训练的时期,且 $\frac{E_{\text{epoch}}}{s_{\text{step_size}}}$ 的结果需要向下取整。训练时的其余设置包括批量大小batchsize设为8,最大迭代次数设为25。各参数设置可根据经验选择,且需要结合实验的硬件环境。

3.2 数据集

所提方法使用两个数据集进行训练和测试,以证明广泛的适用性。

1) Male-Pelvis^[14]:该数据集由瑞典国家放射治疗项目收集,包括大量男性骨盆的CT图像,每组CT的切片数量为80~117,切片大小为 512×512 。

2) LIDC-IDRI^[15]:该数据集由美国国家癌症研究所(National Cancer Institute)收集,用于早期癌症检测,包括大量胸部CT图像,每组CT的切片数量为130~279,切片大小为 512×512 。由于所提方法预测刚体配准参数,因此采用阈值法限制影像的窗宽窗位

(窗宽为400 Hu~450 Hu,窗位为200 Hu~220 Hu),以保留结构清晰、不易形变的胸部骨骼结构,避免肺结节等软组织的干扰。

由于同患者同部位的X-Ray图像难以获取,使用上述收集的三维CT图像,利用Siddon光线追踪法生成模拟X-Ray图像(即DRR图像)进行训练和测试。由于DRR图像具有真实变换参数的特点,保证了数据标签的真实性,且避免了繁琐的手工标注。理论上,通过设定不同的空间变换参数,可以获取无数张DRR图像。考虑到临床应用,设定变换参数的范围如表1所示,所有参数在范围内满足均匀分布。深度学习模型需要大量形态各异的DRR图像,避免训练过程发生欠拟合,因此在参数范围内生成 4×10^4 张骨盆DRR和 4×10^4 张胸部骨骼DRR,划分75%作为训练集,25%作为测试集,数据集的数量分布如表2所示。另外,将DRR图像重采样到 256×256 的大小,并进行归一化处理,便于模型更高效地计算。

表1 变换参数分布

Table 1 Transformation parameter distribution

Parameter	Distribution range
t_x /mm	U(-20, 20)
t_y /mm	U(-20, 20)
t_z /mm	U(-10, 10)
t_θ /($^\circ$)	U(-10, 10)
t_α /($^\circ$)	U(-10, 10)
t_β /($^\circ$)	U(-5, 5)

表2 数据集的数量分布

Table 2 Number distribution of the datasets

Structure	Number of images / 10^4		
	Training set	Test set	Data set
Pelvis	3	1	4
Chest	3	1	4

3.3 评价指标

基于DRR图像具有真实标签的特点,使用预测变换参数和真实变换参数的平均绝对误差(MAE)、均方

根误差 (RMSE) 来评价配准算法的拟合参数能力, 误差越小, 则代表拟合能力越强, 公式分别为

$$E_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i; \omega)|, \quad (15)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n [y_i - f(x_i; \omega)]^2}. \quad (16)$$

在 2D-3D 医学图像配准中, 相似性测度函数常用来比较图像的配准效果, 因此选取常用的测度函数归一化互信息 (NMI) 和归一化互相关 (NCC)^[16] 作为评价指标。其中 NMI 定义为

$$C_{\text{NMI}}(x, y) = \frac{2 \cdot [H(x) + H(y) - H(x, y)]}{H(x) + H(y)}, \quad (17)$$

$$\begin{cases} H(x) = -\sum p(x) \log p(x) \\ H(y) = -\sum p(y) \log p(y) \\ H(x, y) = -\sum p(x, y) \log p(x, y) \end{cases}, \quad (18)$$

式中: $H(x)$ 和 $H(y)$ 分别表示预测图像和真实图像的信息熵; $H(x, y)$ 则为两幅图像的联合熵。

另外, NCC 的定义为

$$C_{\text{NCC}}(x, y) = \frac{\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y})^2}}, \quad (19)$$

式中: x_{ij} 和 y_{ij} 分别表示预测图像和真实图像在 (i, j) 处的像素值; \bar{x} 和 \bar{y} 分别表示像素均值。两种测度函数的值越大, 代表配准效果越佳。

另外, 为了综合考量图像的配准效果, 使用结构相似性 (SSIM) 从亮度、对比度和结构三个方面来进行评估, 公式为

$$C_{\text{SSIM}}(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (20)$$

式中: μ_x 和 μ_y 分别为预测图像和真实图像的像素均值; σ_x 和 σ_y 为标准差; σ_{xy} 则表示两者的协方差; C_1 和 C_2 常数项分别取值为 $(0.01 \times 255)^2$ 和 $(0.03 \times 255)^2$ 。SSIM 指标值越高, 代表两张图像结构相似性越大, 从而证明算法的配准能力越高。此外需要注意的是, 真实图像是网络模型的输入图像, 即带有真实变换参数的 DRR 图像或 X-Ray 图像。预测图像是使用模型预测变换参数生成的 DRR 图像, 生成过程: 将真实图像输入推理模型, 得到 6 个预测变换参数, 并利用 Siddon 光线追踪法生成对应的 DRR 图像, 即为预测图像。

3.4 实验结果与分析

实验设计包括性能分析和方法对比两部分。首先, 进行性能分析, 比较各种分支结构的配准效果, 重参数化前后的配准时间, 以及自适应激活函数对配准精度的影响; 进而, 选取配准中常用的网络结构, 比较

各方法的拟合参数能力; 最后, 与传统配准算法和现有深度学习方法作对比, 分析所提方法的配准效果。

3.4.1 性能分析

首先比较网络超参数, batchsize 分别选取 8, 16, 32, 其余各参数设置参照 3.1 节。使用 3×10^4 张骨盆 DRR 作为训练集, 将数据分批次输入训练网络, 直至损失函数收敛 (损失函数的差小于 1×10^{-4}) 或达到最大迭代 epoch 时, 则结束模型训练。执行重参数化操作, 将训练模型转换为推理模型, 并使用 1×10^4 张骨盆 DRR 进行测试。使用位移误差 (T-MAE)、角度误差 (R-MAE) 和均方根误差 (RMSE) 评估网络性能, 实验结果如表 3 所示, 其中最小误差用粗体标注。通过比较实验结果, 发现当 batchsize 设为 8 时, 更利于网络的训练。

表 3 网络超参数对比

Table 3 Comparison of network hyperparameters

Batchsize	T-MAE /mm	R-MAE /($^{\circ}$)	RMSE
8	0.08	0.05	0.08
16	0.10	0.06	0.09
32	0.15	0.12	0.12

为了探究多分支结构对配准精度的影响, 调整训练阶段的网络结构, 设计消融实验。

1) 无分支: 每个 Block 只保留 3×3 卷积的主干网络。

2) Identity: 在 3×3 卷积的主干网络上, 加入 Identity 分支。

3) 1×1 卷积: 在 3×3 卷积的主干网络上, 加入 1×1 卷积分支。

4) 所提方法: 在 3×3 卷积的主干网络上, 加入 Identity 和 1×1 卷积分支。

训练集分别使用 3×10^4 张骨盆 DRR 和 3×10^4 张胸部 DRR, 测试集分别使用 1×10^4 张骨盆 DRR 和 1×10^4 张胸部 DRR, 使用 T-MAE、R-MAE 和 RMSE 来评估配准性能, 实验结果如表 4 所示, 其中最小误差用粗体标注。首先, Identity 通过直接结合前一层的特征, 推动网络内的信息流动, 在不增加参数量的基础上, 误差值均小于无分支网络。接着, 由于 1×1 卷积层实现跨通道的信息交互和整合, 有助于提高配准精度, 在所有评价指标上均优于无分支网络。最后, 与无分支网络相比, 所提方法在骨盆数据上减少 27.2% 的位移误差, 30.1% 的角度误差; 在胸部数据上减少 25.8% 的位移误差, 31.3% 的角度误差。这归因于所提方法的多分支结构具有更丰富的特征表达性, 从而显著地提升配准精度。另外, 模型的预测时间如表 5 所示, 其中 T-time 表示训练模型的预测时间 (重参数化前), I-time 表示推理模型的预测时间 (重参数化后), 较少的预测时间使用粗体进行标注。通过对比分析, 发现推理模型的预测时间显著减少, 从而证明重参数化的重要性。

为了比较激活函数对配准模型的影响, 分别将网

表 4 分支结构消融实验的误差对比

Table 4 Error comparison of branch structure ablation experiments

Dataset	Network structure	T-MAE /mm	R-MAE /($^{\circ}$)	RMSE
Pelvis	No branch	0.103	0.073	0.115
	Identity	0.087	0.062	0.098
	1×1 Conv	0.083	0.057	0.093
	Proposed method	0.075	0.051	0.084
Chest	No branch	0.155	0.099	0.169
	Identity	0.147	0.088	0.156
	1×1 Conv	0.129	0.082	0.143
	Proposed method	0.115	0.068	0.125

表 5 分支结构消融实验的时间对比

Table 5 Time comparison of branch structure ablation experiments

Network structure	Parameter / 10^5	T-time /ms	I-time /ms
No branch	50.98	29.8	25.7
Identity	50.98	30.2	25.8
1×1 Conv	56.53	33.2	25.8
Proposed method	56.53	34.6	26.0

络结构中的激活函数设为 ReLU、Swish、ACON-C 和 Meta-ACON, 实验结果如表 6 所示, 其中最小误差用粗体标注。首先, 对比 Swish 和 ReLU, 发现它们在两个数据集上的效果相当, 这归因于 Swish 函数只是 ReLU 函数的简单平滑形式, 在提高非线性表达能力上有所限制。接着, 对比 ACON-C 和 ReLU, ACON-C 在两个数据集上的效果要略优于 ReLU, 这是由于 ACON-C 函数利用可学习参数来获取 ReLU 函数的广泛平滑形式, 从而增加模型的非线性因素。最后, 对比 Meta-ACON 和 ReLU, Meta-ACON 在骨盆数据集上效果相当, 在胸部数据上减少 19.6% 的位移误差,

表 7 不同网络的配准效果对比

Table 7 Comparison of registration effects of different networks

Dataset	Network structure	T-MAE	R-MAE	RMSE	Time /ms
Pelvis	GoogLeNet	0.36±0.19	0.23±0.13	0.39±0.24	16.0
	VGG16	0.32±0.27	0.21±0.18	0.35±0.21	19.2
	ResNet50	0.09±0.08	0.06±0.07	0.10±0.08	18.4
	DenseNet121	0.11±0.08	0.07±0.06	0.12±0.09	34.5
	Net+GradNCC ^[11]	7.83±19.8	4.94±8.78		
	Proposed method	0.08±0.04	0.05±0.03	0.08±0.06	26.0
Chest	GoogLeNet	0.41±0.18	0.16±0.09	0.39±0.20	21.5
	VGG16	0.56±0.45	0.31±0.24	0.60±0.41	19.1
	ResNet50	0.33±0.09	0.20±0.05	0.35±0.12	24.6
	DenseNet121	0.15±0.08	0.09±0.05	0.16±0.08	37.0
	Inception+branch ^[10]	0.61	1.93		2.3
	Proposed method	0.12±0.07	0.07±0.04	0.13±0.08	26.4

以胸部数据为例, 图 6 为位移参数和角度参数的误差箱线图, 其中圆点表示离散点, × 号表示平均值。

表 6 激活函数对比

Table 6 Activation function comparison

Dataset	Activation function	T-MAE /mm	R-MAE /($^{\circ}$)	RMSE
Pelvis	ReLU	0.078	0.052	0.087
	Swish	0.079	0.053	0.088
	ACON-C	0.078	0.052	0.087
	Meta-ACON	0.075	0.051	0.084
Chest	ReLU	0.143	0.084	0.156
	Swish	0.151	0.082	0.162
	ACON-C	0.129	0.077	0.129
	Meta-ACON	0.115	0.068	0.125

19.0% 的角度误差, 这是因为 Meta-ACON 在 ACON-C 的基础上加入 β 参数来学习神经元激活与否, 有利于提取有效特征, 从而提高模型精度。

3.4.2 方法对比

首先, 选取 2D-3D 配准中常用的网络结构来对比拟合参数能力, 如 GoogLeNet^[17]、VGGNet^[18]、ResNet^[19] 和 DenseNet^[20], 各参数设置同所提方法保持一致。使用 1×10^4 张骨盆 DRR 图像和胸部 DRR 图像进行测试, 实验结果如表 7 所示, 其中 VGG16 的全连接层参数设为 2048, 便于模型训练。由于采用大量分支结构, 所提方法在参数量和浮点运算数方面略显劣势, 但位移误差和角度误差均超过其他方法。在配准时间上, 利用重参数化转换后, 所提方法与其他方法的配准效率相当。实验结果表明, 与常用配准网络相比, 所提方法在配准时间相当的情况下, 显著地提高了配准精度。表 7 中 Net+GradNCC 方法的实验结果来自文献[11], Inception+branch 方法的实验结果来自文献[10], 原文缺失部分评价指标。另外, 最小误差和最短时间使用粗体进行标注。

通过比较分析发现, 所提方法在位移误差和角度误差上的平均值、中位数、最大值均小于其他方法, 且离散

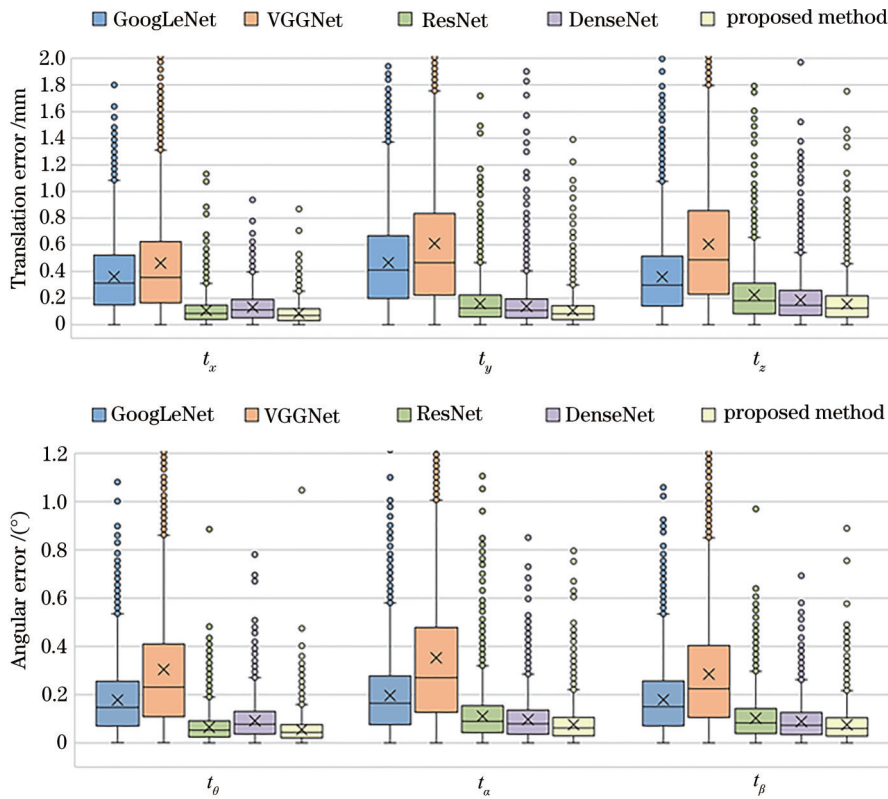


图 6 参数误差箱线图。(a)位移误差;(b)角度误差

Fig. 6 Parameter error box diagrams. (a) Translation error; (b) angular error

点分布也较为合理。

进而,选取常用的 2D-3D 传统配准方法和现有深度学习方法作为对比实验。其中,传统方法包括 NCC_Powell 和 NMI_Powell,使用性能优良的 NCC 和 NMI 作为测度函数,使用 Powell 法作为优化算法,当两次测度值的差小于 1×10^{-3} 时停止迭代。深度学习方法包括陈向前等^[9]应用在骨盆的配准方法,贾瑞明等^[10]应用在胸部的配准方法。使用 100 张骨盆 DRR 和 100 张胸部 DRR 进行测试,使用 NCC、NMI、SSIM 作为评价指标,实验结果如表 8 所示,值越大代表配准

效果越佳。在配准效果上,所提方法在 3 个评价指标上都优于传统方法,满足配准的高精度要求;在配准时间上,所提方法仅需 26 ms,远低于传统方法的配准时间。另外,对比现有的深度学习方法,由于所提方法采用训练-推理解耦架构网络和模拟 X-Ray 图像,在配准效果和时间上均优于其他方法。表 8 中 Deep learning method 的实验结果来自文献[9],Inception+branch 方法的实验结果来自文献[10],原文缺失部分评价指标。另外,各评价指标的最大值和最短时间使用粗体进行标注。

表 8 配准方法对比

Table 8 Comparison of registration methods

Dataset	Method	NCC	NMI	SSIM	Time /ms
Pelvis	NCC_Powell	0.56±0.07	0.16±0.02	0.57±0.09	96.4×10 ³
	NMI_Powell	0.61±0.09	0.15±0.01	0.57±0.09	28.6×10 ³
	Deep learning method ^[9]	0.82±0.07	0.32±0.03		30.0
	Proposed method	0.99±0.05	0.72±0.03	0.99±0.08	26.0
Chest	NCC_Powell	0.53±0.11	0.11±0.02	0.55±0.09	101.5×10 ³
	NMI_Powell	0.57±0.09	0.11±0.02	0.58±0.08	29.2×10 ³
	Inception+branch ^[10]			0.42	2.3×10 ³
	Proposed method	0.99±0.06	0.56±0.02	0.99±0.09	26.4

综上所述,由实验结果可知,所提方法不仅满足了 2D-3D 医学图像配准的高精度需求,而且做到了实时配准,图 7 为胸部的配准效果图,骨骼结构均对齐。在

骨盆、胸部两个数据集上测试的结果证明所提方法具有一定的泛化性能,有利于临床应用。

另外,所提方法只使用 DRR 图像进行训练和测

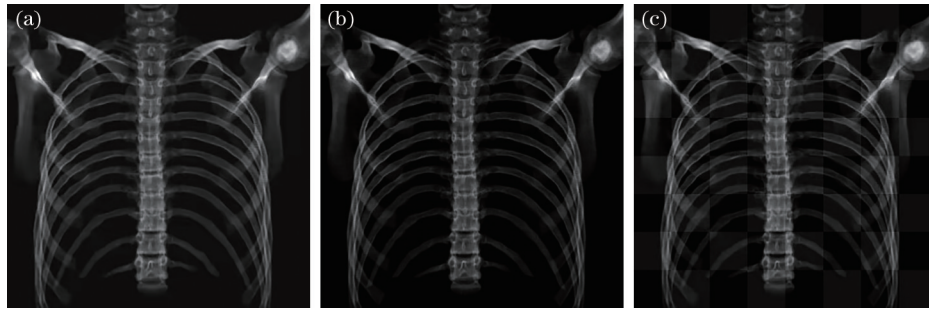


图 7 配准效果图。(a)参考图像;(b)浮动图像;(c)配准棋盘格效果图

Fig. 7 Registration rendering. (a) Reference image; (b) moving image; (c) registered checkerboard rendering

试,陈向前等^[9]提到 DRR 图像和 X-Ray 图像之间存在图像风格、数据分布等差异,因此用 DRR 图像训练的网络模型可能不完全适用于 X 光图像的预测。针对该问题,上述文献在测试时对 X-Ray 图像进行直方图匹配等图像预处理,然而文中没有 X-Ray 图像风格转换前后的对比实验,因此未证实图像预处理的有效性。所提方法考虑到 CNN 学习类似位姿参数的信息,对不同的图像风格并不敏感,认为对拟合参数的能力影响不大,Xie 等^[21]在训练和测试时也均采用 DRR 图像。最后,受实验条件限制,未使用真实 X-Ray 图像测试所提方法的配准效果,是本文的不足之处,有待进一步完善和探讨。

4 结 论

为了达到高精度和实时性的需求,提出一种基于训练-推理解耦架构的 2D-3D 医学图像配准方法,用于预测刚体变换参数。该方法在训练阶段采用多分支结构,显著地提高配准精度;在推理阶段,使用重参数化方法将多分支结构转换为单路结构,大大加快配准速度。另外,引入自适应激活函数,增加网络的非线性因素,提升拟合参数的能力。所提方法可应用于放射治疗的摆位验证系统中,配准精度满足临床需求,且实现了实时配准。同时,所提方法仍存在一些不足之处,如未使用真实 X-Ray 图像验证配准结果,未考虑软组织器官形变的情况,下一步计划收集成对 CT 图像和 X-Ray 图像,并研究非刚性配准方法,预测图像的形变问题。

参 考 文 献

- [1] 张一戈,邱小平,陈维军,等. 基于图像引导技术对头颈部肿瘤临床摆位误差的剂量学影响[J]. 中国医学物理学杂志, 2015, 32(1): 25-30.
Zhang Y G, Qiu X P, Chen W J, et al. Dosimetry influence study on head & neck cancer clinical placement error based on image guided radiotherapy technique[J]. Chinese Journal of Medical Physics, 2015, 32(1): 25-30.
- [2] 张丹,黄欢,尚振宏. 基于互信息的 Active Demons 非刚性图像配准算法[J]. 激光与光电子学进展, 2020, 57(16): 161009.
Zhang D, Huang H, Shang Z H. Nonrigid image registration algorithm based on mutual information Active Demons[J]. Laser & Optoelectronics Progress, 2020, 57(16): 161009.
- [3] 张冉,王雷,夏威,等. 2D/3D 图像配准中的相似性测度和优化算法[J]. 激光与红外, 2014, 44(1): 98-102.
Zhang R, Wang L, Xia W, et al. Comparison of similarity measurement and optimization methods in 2D/3D image registration[J]. Laser & Infrared, 2014, 44(1): 98-102.
- [4] 刘坤,吕晓琪,谷宇,等. 快速数字影像重建的 2 维/3 维医学图像配准[J]. 中国图象图形学报, 2016, 21(1): 69-77.
Liu K, Lü X Q, Gu Y, et al. The 2D/3D medical image registration algorithm based on rapid digital image reconstruction[J]. Journal of Image and Graphics, 2016, 21(1): 69-77.
- [5] 宋涛,曹利波,赵明富,等. 三维点云中关键点的配准与优化算法[J]. 激光与光电子学进展, 2021, 58(4): 0415008.
Song T, Cao L B, Zhao M F, et al. Registration and optimization algorithm of key points in three-dimensional point cloud[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0415008.
- [6] 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: 1788-1800.
- [7] Fu Y B, Lei Y, Wang T H, et al. Deep learning in medical image registration: a review[J]. Physics in Medicine and Biology, 2020, 65(20): 20TR01.
- [8] Miao S, Wang Z J, Liao R. A CNN regression approach for real-time 2D/3D registration[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1352-1363.
- [9] 陈向前,郭小青,周钢,等. 基于深度学习的 2D/3D 医学图像配准研究[J]. 中国生物医学工程学报, 2020, 39(4): 394-403.
Chen X Q, Guo X Q, Zhou G, et al. 2D/3D medical image registration using convolutional neural network[J]. Chinese Journal of Biomedical Engineering, 2020, 39(4): 394-403.
- [10] 贾瑞明,李浩轩,陈彧,等. 胸主动脉腔内修复术前 CT 血管造影与术中 X 射线图像的配准算法[J]. 中国组织工程研究, 2019, 23(35): 5658-5663.
Jia R M, Li H X, Chen Y, et al. Preoperative CT

- angiography and intraoperative X-ray image registration algorithm for thoracic aortic endovascular repair[J]. Chinese Journal of Tissue Engineering Research, 2019, 23(35): 5658-5663.
- [11] Gao C, Liu X T, Gu W H, et al. Generalizing spatial transformers to projective geometry with applications to 2D/3D registration[M]//Martel A L, Abolmaesumi P, Stoyanov D, et al. Medical image computing and computer-assisted intervention-MICCAI 2020. Lecture notes in computer science. Cham: Springer, 2020, 12263: 329-339.
- [12] Ding X H, Zhang X Y, Ma N N, et al. RepVGG: making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13728-13737.
- [13] Ma N N, Zhang X Y, Liu M, et al. Activate or not: learning customized activation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 8028-8038.
- [14] Nyholm T, Svensson S, Andersson S, et al. MR and CT data with multiobserver delineations of organs in the pelvic area: part of the gold atlas project[J]. Medical Physics, 2018, 45(3): 1295-1300.
- [15] Armato S G I, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans[J]. Medical Physics, 2011, 38(2): 915-931.
- [16] Malinsky M, Peter R, Hodneland E, et al. Registration of FA and T1-weighted MRI data of healthy human brain based on template matching and normalized cross-correlation[J]. Journal of Digital Imaging, 2013, 26(4): 774-785.
- [17] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 1552-1560.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2021-02-03]. <https://arxiv.org/abs/1409.1556>.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [20] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [21] Xie Y, Meng C, Guan S Y, et al. Single shot 2D3D image registration[C]//2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), October 14-16, 2017, Shanghai, China. New York: IEEE Press, 2017: 1761-1768.