

融合多尺度特征及注意力机制的医学图像检索

周林鹏, 姚剑敏, 严群, 林志贤

引用本文:

周林鹏, 姚剑敏, 严群, 等. 融合多尺度特征及注意力机制的医学图像检索[J]. 液晶与显示, 2021, 36(8): 1174–1185.

ZHOU Lin-peng, YAO Jian-min, YAN Qun, et al. Medical image retrieval with multiscale features and attention mechanisms[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(8): 1174–1185.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2020-0248>

您可能感兴趣的其他文章

Articles you may be interested in

融合注意力机制的移动端人像分割网络

Mobile-based portrait segmentation network with attention mechanism

液晶与显示. 2020, 35(6): 547–554 <https://doi.org/10.3788/YJYXS20203506.0547>

改进的多尺度火焰检测方法

Improved multi-scale flame detection method

液晶与显示. 2021, 36(5): 751–759 <https://doi.org/10.37188/CJLCD.2020-0221>

基于注意力掩模融合的目标检测算法

Object detection algorithm based on attention mask fusion

液晶与显示. 2019, 34(8): 825–833 <https://doi.org/10.3788/YJYXS20193408.0825>

基于改进Faster R-CNN的交通标志检测算法

Traffic sign detection algorithm based on improved Faster R-CNN

液晶与显示. 2021, 36(3): 484–492 <https://doi.org/10.37188/CJLCD.2020-0195>

基于多尺度残差网络的CT图像超分辨率重建

CT image super-resolution reconstruction based on multi-scale residual network

液晶与显示. 2019, 34(10): 1006–1012 <https://doi.org/10.3788/YJYXS20193410.1006>

文章编号:1007-2780(2021)08-1174-12

融合多尺度特征及注意力机制的医学图像检索

周林鹏¹, 姚剑敏^{1,2*}, 严 群^{1,2}, 林志贤¹

(1. 福州大学 物理与信息工程学院, 福建 福州 350108;

2. 晋江市博感电子科技有限公司, 福建 晋江 362200)

摘要:针对目前医学图像普遍存在病理区域尺寸分布较分散、细节特征不明显以及同类组织影像间的视觉差异较大等问题,本文在 CBMIR 系统的基础上,提出了一种融合多尺度特征及注意力机制的医学图像检索方法。该方法通过融合多尺度特征并设置可学习权重系数来自适应平衡浅层图像纹理特征和深层图像语义特征的关系,提高网络对不同尺度上的病理特征提取能力。同时,引入注意力模块,对网络输出的特征图进行通道加权求和,提高关键特征通道的特征表达能力,使网络更能关注到图像中的具有辨识性的病理特征区域。最后,在损失函数设计时,使用多重损失进一步优化样本特征在特征空间的分布。最终在 Mura 数据集上的 mAP@100、mAP@20 两个指标上分别达到了 0.95、0.98 的检索精度,基本符合实际场景对模型的检索精度要求。

关 键 词:医学影像处理; CBMIR; 多尺度特征融合; 注意力机制; 多重损失函数

中图分类号:TP391.4; TH691.9 **文献标识码:**A **doi:**10.37188/CJLCD.2020-0248

Medical image retrieval with multiscale features and attention mechanisms

ZHOU Lin-peng¹, YAO Jian-min^{1,2*}, YAN Qun^{1,2}, LIN Zhi-xian¹

(1. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China;

2. Jinjiang RichSense Electronic Technology Co., Ltd., Jinjiang 362200, China)

Abstract: In order to solve the common problems of current medical images, such as relatively scattered size distribution of pathological areas, the ambiguous detail features, and the big visual difference of similar tissue images, this paper proposes a medical retrieval method integrating multi-scale features and attention mechanism based on the CBMIR system. This method adaptively balances the relationship between shallow image texture features and deep image semantic features by fusing multi-

收稿日期:2020-09-24; **修订日期:**2021-01-18.

基金项目:国家重点研发计划课题(No.2016YFB0401503); 广东省科技重大专项(No.2016B090906001); 福建省科技重大专项(No.2014HZ0003-1); 广东省光信息材料与技术重点实验室开放基金(No.2017B030301007)

Supported by National Key Research and Development Program of China (No.2016YFB0401503); Science and Technology Major Program of Guangdong Province (No.2016B090906001); Science and Technology Major Program of Fujian Province (No.2014HZ0003-1); Guangdong Provincial Key Laboratory of Optical Information Materials and Technology (No.2017B030301007).

* 通信联系人, E-mail: yaojm@fzu.edu.cn

scale features and setting learnable weight coefficients, thereby the network's ability to extract pathological features at different scales is improved. At the same time, this method introduces the attention module and perform channel weighted summation on the feature maps output by the network to improve the feature expression ability of key feature channels, so that the network can pay more attention to the recognizable pathological feature areas in the image accurately. Finally, the multiple losses are used to further optimize the distribution of sample features in the feature space when the loss function is designed. As a result, the retrieval accuracy of 0.95 and 0.98 is achieved on the mAP@100 and mAP@20 indicators on the Mura dataset, which basically meets the retrieval accuracy requirements of the actual scene on the model.

Key words: medical image processing; CBMIR; multi-scale feature fusion; attention mechanism; multiple loss function

1 引言

医学影像技术日益成熟,如何有效地利用已有的医学影像数据辅助医生进行分析和诊断是目前相对有挑战性的任务。医学图像数据主要包括磁共振成像(MRI)、电子计算机断层扫描图像(CT)、数字减影血管造影图像(DSA)以及正电子发射断层扫描图像(PET)。海量的数据、各种各样的归类标准给医学影像的有效组织和管理带来了巨大的挑战,与之而来的是专业影像医生的极度紧缺。在中国,医学影像的年增长率约为30%,但放射科医生的年增长率仅有4.1%。据统计,以肺结节检测为例,三甲医院平均每天需要接待200例左右的肺结节筛查患者,每个患者在检查环节中会产生200~300张左右的CT影像,如何利用现有技术及相关影像数据来辅助医生进行临床诊断成为了现在亟待解决的问题^[1-3]。

目前,医学影像信息系统(PACS)可以通过各种接口将临床收集的医学影像以数字化的形式保存起来,初步解决了影像数据的存储问题^[4]。为了利用这些医学数据库来辅助医生进行病情分析及诊断,需要设计有效的医学图像检索系统(CBMIR)。通过检索相似的图像和病历,医生可以综合参考多个维度的信息来给出更加全面精准的病情定位及诊疗方案。因此,围绕CBMIR系统设计及优化的相关研究也越来越多。Jiji等提出了一种基于内容的皮肤病变图像检索方法^[5], Mizotin等提出了一种基于SIFT特征的视觉词

袋的方法,用于脑磁共振图像的检索,以诊断阿尔茨海默氏病^[6]。Rahman提出了一种基于类别信息作为监督信号的生物医学图像检索方法^[7]。陈等人提出基于多参数Gabor的消化道超声图像的处理方法,强化了超声图像边缘信息的特征提取^[8]。近年来,深度学习在图像处理领域取得了巨大的成功,为图像特征提取提供了新思路。Qayyum等人提出了通过迁移学习的方法在自然图像上预先训练的CNN模型上使用医学图像进行微调,并将模型所学习的特征和分类结果用于医学图像检索^[9]。吕等人提出基于三维卷积的肺结节图像处理办法^[10]。熊等人提出基于vgg16及哈希编码的医学图像检索模型DHCNN^[11],彭晏飞等人提出引入注意力机制进行图像特征提取^[12],周国华等人提出使用多幅不同角度图像进行CT图像检索^[13]。

医学图像具有不同于通用数据集的固有特征:异质性、模糊性、高分辨率、多模态等^[14],而文献[6]中的模型无法很好地提取图像语义特征,文献[9]中模型只用了网络最后几层特征描述图像,忽略了底层纹理特征,因此都未取得较满意的检索精度。本文提出了一种融合多尺度特征及注意力机制的医学图像检索方法,该方法通过抽取不同尺度的特征进行学习,有效融合了浅层视觉特征及深层语义特征,并引入注意力机制来提高网络对关键区域的关注度,抑制无关背景区域对检索结果的干扰。最后在损失函数设计上,结合了交叉熵损失及中心损失的优点,有效缓解了检索过程中误检索及漏检索的现象。

2 系统概述

根据图 1 所示,一个完整的医学图像检索系统一般包括以下 3 个流程:首先是数据集线下特

征抽取并组建特征矩阵库的阶段,其次是线上输入图像特征提取阶段,最后是将输入图像的特征与特征矩阵库中的特征进行相似度计算,并返回相似度排名靠前的 top- k 图像。

上述流程中主要包括图像预处理、特征提取

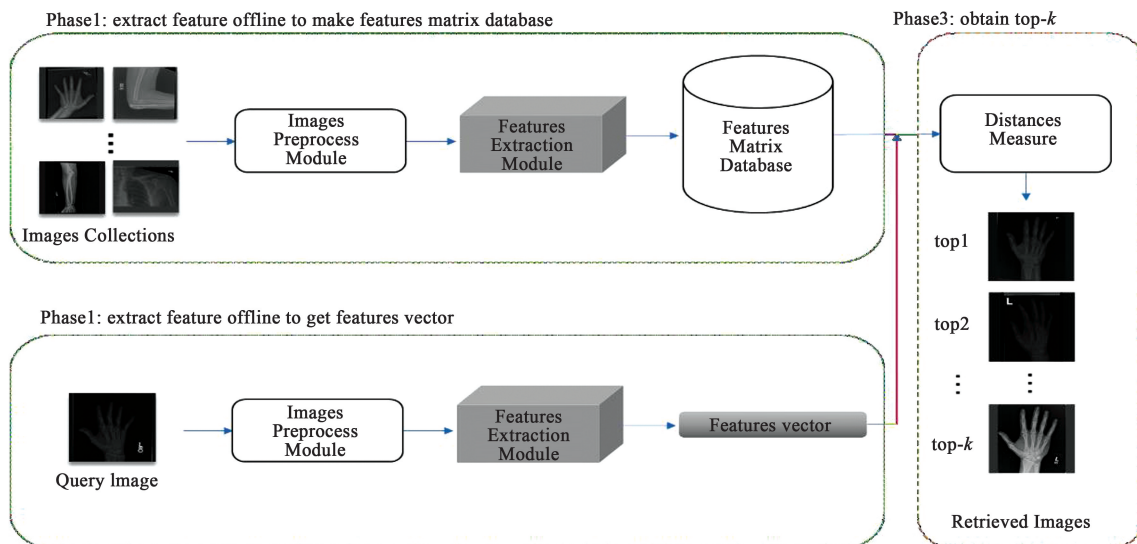


图 1 医学图像检索系统示意图

Fig.1 Schematic diagram of medical image retrieval system

以及距离度量 3 个功能模块,本节将就这 3 个功能模块的具体实现展开介绍,并重点介绍本文在特征提取模块的设计及优化上所做的相关工作。

2.1 图像预处理

在进行医学图像检索时,通常需要对不同成像设备采集到的图像采取不同的预处理措施,比如常见的 CT 图像中,像素值分布较广,直接归一化到 0~255 会损失较多的信息,因此需要根据不同组织的 Hu 值来选择合适的窗宽窗位做特定区间的像素延展,使图像的细节信息得以凸显。对于 X-ray 图像,通常会由于不同采集设备以及不同放射剂量使数据库中 X-ray 样本的亮度、对比度等分布不均匀,需要对图像数据进行直方图平衡预处理,以减轻外界因素对模型特征学习的干扰。

2.2 特征提取

本文的主要工作主要集中在本模块的设计及优化上,首先是设计了一个多尺度特征提取网络,其次是引入自注意力模块,最后是结合多重损失对模型进一步优化。本模块的主体结构如图 2 所示。

2.2.1 多尺度特征提取网络

本文的特征提取模块选用了经典的 Resnet^[15] 结构,我们希望通过一个深层网络来获取医学图像中深层次的语义特征。然而由于网络层数变深,同时也带来了梯度爆炸或梯度弥散的问题,并且梯度在从深层向浅层传递的过程中逐步减弱,使得浅层网络无法得到有效的训练。由于梯度的不稳定及反传的低效性,导致网络很难收敛。针对这些问题,Resnet 网络进行了相应的结构改进。

梯度在传播过程中的不稳定性主要由以下几点导致:首先,在权重随机初始化过程中权值被赋予较大的值,导致反传的梯度与权值相乘大于 1,并在后续传播过程中逐层放大导致梯度爆炸,Resnet 网络通过对权重进行高斯初始化可以较好避免梯度爆炸的问题;其次,sigmoid 激活函数的特性决定了它对较大或较小的输入值表现出梯度低敏感性,导致梯度无法有效地经过 sigmoid 激活函数反向传播。基于此,Resnet 网络通过对激活函数的输入进行批归一化(BatchNorm)操

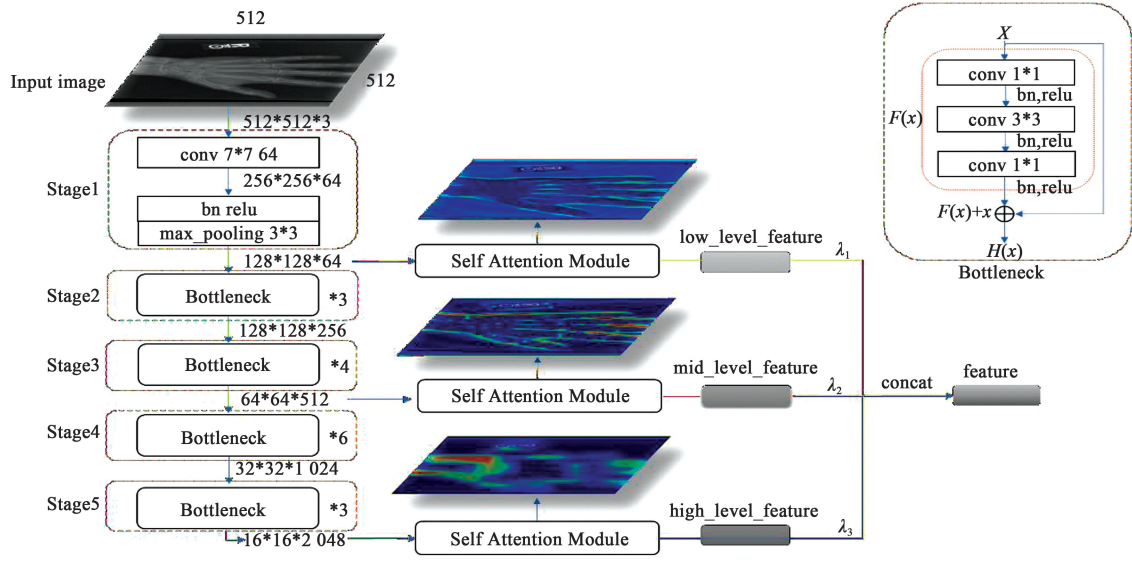


图 2 特征提取模块结构图

Fig.2 Structure diagram of feature extraction module

作,将输入限制在激活函数的梯度敏感区间,并引入计算更为简单且对梯度反向传播更高效的 relu 激活函数来缓解梯度经过激活函数损耗较多的问题。

尽管采取 BatchNorm 操作及选用 relu 激活函数缓解了梯度经过激活函数时的损耗,但还是未彻底解决由网络加深带来的浅层网络学习不充分的问题。为此,Resnet 网络提出了经典的残差块结构,即图 2 中的 Bottleneck 结构。在原始顺序堆叠的 3 个卷积层的基础上,通过一个跳跃连接将输入叠加到输出上。由于跳跃连接的存在,为靠近输出端得到的梯度向靠近输入端的浅层网络传递提供了可能性,避免了梯度只能经过深层网络回传引起的梯度弥散问题。同时,图 2 中的残差块为优化之后的结构,原始残差块由两个 3×3 卷积组成,新结构通过使用 1×1 卷积来对特征图通道进行压缩和扩张,保证网络精度的同时又减少了模型的参数量,加快了网络前向推理的速度。

最后,针对本数据集特征尺度差异较大的问题,为了使网络能充分学习到不同尺度的特征,提高特征的有效性,本文在 Resnet 网络的基础上分别抽取 Stage1、Stage3、Stage5 输出的特征图,对于 512×512 尺寸的输入,输出的特征图尺寸分别为 $128 \times 128 \times 64$ 、 $64 \times 64 \times 512$ 、 $16 \times 16 \times 2048$,分别对应图像的浅层纹理特征、中间层过渡特征

以及深层语义特征,并输入到后续的自注意力模块中对逐层特征进行进一步通道筛选。

2.2.2 自注意力模块

对于残差网络输出的不同尺度的特征图,本文通过设计一个自注意力模块来对特征图中不同通道特征进行进一步筛选,来提高关键通道的特征表达能力,进一步引导网络将注意力聚焦到包含关键信息的区域。该模块的设计思路来源于非局部均值(NLM)降噪算法。NLM 算法最初在 2005 年由 Buades 等人提出^[16],并被广泛用于图像复原及视频降噪领域,NLM 的滤波过程可以用下面公式表示:

$$\hat{X}_p^{(\text{NL-means})} = \frac{1}{C(p)} \sum_{q \in B(p,r)} x_q w(p,q), \quad (1)$$

$$C(p) = \sum_{q \in B(p,r)} w(p,q), \quad (2)$$

$$w(p,q) = \exp\left(-\frac{\max(d^2(p,q) - 2\sigma^2, 0)}{h^2}\right), \quad (3)$$

其中, $\hat{X}_p^{(\text{NL-means})}$ 表示经过非局部均值滤波得到的像素值, $C(p)$ 为权重归一化值, $B(p,r)$ 表示中心为 p , 大小为 $(2r+1) \times (2r+1)$ 的图像块, $w(p,q)$ 表示 q 点对 p 点像素值的权重值, 权重值大小由两个点的相似程度来决定, 相似度可由两个点极小邻域间像素值的欧氏距离来衡量。

NLM 算法的核心思想是为了克服双线性滤波、中值滤波等滤波算法仅考虑图像局部信息的

局限性,从而提出以图像块为单位,在全局范围内根据不同图像块之间的相似度进行像素值加权平均,更好地实现图像高斯噪声的滤除,并且不损失图像细节。虽然本文的目的并不是做图像降噪,但是 NLM 算法实现降噪的思路其实就是在抑制图像中的无关信息,进而使有效信息得到充分表达,基于此,我们可以将这一思想用于高维特征图的特征通道筛选任务中,从而达到抑制无关通道特征、强化关键通道特征的目的。事实上,后面的 Non Local Neural Network^[17] 以及 Attention GAN^[18-19] 中的注意力模块正是借鉴了 NLM 算法的思想,通过计算任意两个特征通道之间的交互来直接捕捉远程依赖,得到更多的全局辅助信息以弥补小卷积核信息获取不足的缺陷,进而对所有特征通道进行更加合理权重分配。自注意力模块结构图如图 3 所示。

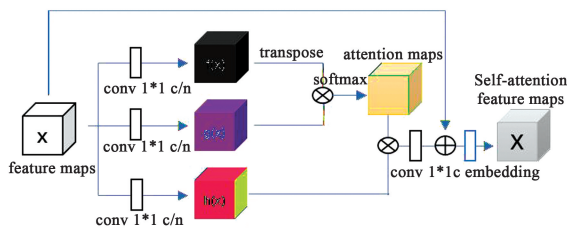


图 3 自注意力模块结构图

Fig.3 Structure diagram of self-attention-module

下面从自注意力模块结构图对其中原理做进一步阐述,首先,对于输入的特征图,经过 3 条不同的分支 $f(x)$ 、 $g(x)$ 及 $h(x)$,通过 3 组数量相同的 1×1 卷积进行通道压缩,并保留通道维度将宽高展平成一维,这里主要是为了减少输入特征图的信息冗余,同时降低后面相似度计算的复杂度。其次,对分支 $f(x)$ 的特征图进行转置操作再和分支 $g(x)$ 的特征图进行矩阵相乘,然后将结果经过 softmax 进行归一化,从向量积数学表达式可以看出矩阵的乘积其实表征了向量间的余弦相似度。实际上,这里进行的就是 NLM 算法中的不同通道之间特征图的相似度计算。最后,将归一化输出后的注意力矩阵和分支 $h(x)$ 得到的特征图进行相乘,这里其实就是根据相似度对不同通道进行权重重分配,再次经过 softmax 得到以及 1×1 卷积对通道扩张至输入特征图的通道数,此时输出的特征图中的关键细节特征相对于原特征图得到了更充分的表达,从而实现注意

力重分配。

在上述模块中对输入不同尺度的特征图进行特征压缩时,对于通道数较少的浅层特征,容易因压缩率过高而损失掉有效信息,而对于通道数较多的深层特征,则会因为压缩率过低导致存在较高的信息冗余并增加后续的相似度计算过程的耗时。因此,本文考虑到不同尺度特征图的通道维度上信息冗余的差异性,从浅层至深层分别采用 4 倍、8 倍、16 倍的压缩率,有效地平衡各尺度的信息利用率及计算效率。同时,借鉴上文残差块结构,将输入特征图通过跳跃连接直接叠加到自注意力模块的输出,进一步优化梯度反向传播的效率。

最后,为了优化系统在高维特征在检索时的时间开销及存储上的空间开销,本文采用主成分分析法将不同层的高维输出嵌入到低维空间中,最终输出能够高效表征图像特征的 128 维特征向量。

2.2.3 损失函数

考虑到医学图像背景大多相似度较高,而同一类别的数据会因不同采集对象而呈现较大的视觉差异,从而导致不同类之间的样本特征因高度相似的背景区域而相互混杂,同类之间数据由于存在较大视觉差异使得在特征空间中距离被拉大,因此,本文在损失函数设计上,采用了交叉熵损失和中心损失相结合的思路,来改善上述问题。本文损失函数公式如下:

$$L = \lambda_1 L_{CE} + \lambda_2 L_C, \quad (4)$$

$$L_s = - \sum_{i=1}^m \lg \left(e^{w_{yi}^T x_i + b_{yi}} / \sum_{j=1}^n e^{w_j^T x_i + b_j} \right), \quad (5)$$

$$L_C = \frac{1}{2} \sum_{i=1}^m \|x_i - C_{yi}\|^2, \quad (6)$$

对于交叉熵损失,从计算公式(5)可以看出,交叉熵损失反映的是预测值的概率分布和真实标签的概率分布之间的差异程度。在网络不断训练迭代的过程中,网络能够学习到类别间的区分特征,使得预测的概率分布能够逐步拟合真实标签的概率分布,然而对于医学影像数据检索模型,仅让不同类别的数据在特征空间实现类间可分还不够,我们还希望同类特征分布能够更紧凑,这样检索得到的结果才能和输入样本表现出强相关性,才能够为临床诊断提供更多有价值的参考信息。

进一步的,为了解决类内特征不紧凑问题,本文引入了中心损失,公式(6)中 x_i 表示网络提取到的样本特征, C_{y_i} 表示第 y_i 个类别的特征中心,中心损失统计的是每个批次中的样本特征与对应类别的特征中心的距离,并在训练过程中,将计算得到的损失值通过梯度反传来优化网络参数,从而缩短同类样本在特征空间中的距离^[20]。

2.3 距离度量

在设计好特征提取模块的基础上,可以离线抽取数据集中的图像特征,并将所有的特征向量拼接成特征矩阵进行存储,同时将数据库中的图片路径与矩阵中对应的特征向量建立索引。在检索过程中,通过计算输入图像的特征向量与特征矩阵中所有向量间的距离,并按距离从小到大排序来检索数据库中的相关样本。常用的距离评估函数有以下几种:

$$\text{欧式距离: } d = \sqrt{\sum_{i=1}^N (x_{i1} - x_{i2})^2}, \quad (7)$$

$$\text{曼哈顿距离: } d = \sum_{i=1}^N |x_{i1} - x_{i2}|, \quad (8)$$

切比雪夫距离:

$$d = \sum_{i=1}^N \max(|x_{i1} - x_{i2}|, |y_{i1} - y_{i2}|), \quad (9)$$

$$\text{余弦距离: } d = \frac{\sum_{i=1}^N (x_{i1} x_{i2} + y_{i1} y_{i2})}{\sqrt{(x_{i1} + y_{i1})^2 (x_{i2} + y_{i2})^2}}, \quad (10)$$

上述距离度量函数中,欧式距离、曼哈顿距离以及切比雪夫侧重描述特征空间中向量间的数值关系,余弦距离则表示特征向量中不同维度间相对层面的差异。由于医学图像固有的异质性,同类样本可能在数值上存在较大区别,因此本文采用余弦距离来衡量输入图像与数据库中图像特征间的相似度。

3 实验及结果分析

3.1 实验准备

3.1.1 数据集

本文使用的是斯坦福吴恩达老师团队收集的MURA数据集,包含来自14 892位不同年龄段患者的40 895张骨骼X光片,分别采集自患者的肩部、肱骨、手肘、前臂、手腕、手掌和手指7个不

同的部位。首先,为了保证模型的检索性能,需要将数据集按最具有区分度的特征进行组织再送入网络进行特征学习,这里选择按不同采集部位进行数据归类。数据集中各类样本的数量分布如图4所示。

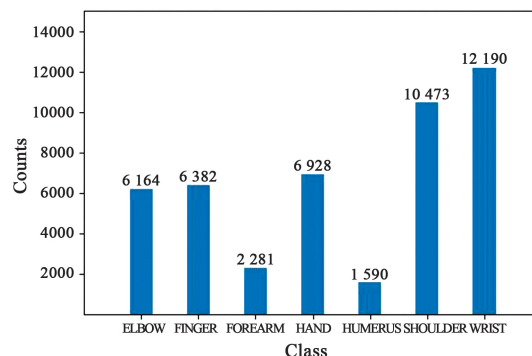


图4 样本数量分布图

Fig.4 Distribution diagram of different classes

观察到数据集中前臂、肱骨这两个类别数量不足,而肩部、手腕两个类别数量偏多,本文在预处理阶段针对数量较少的前臂、肱骨类别做了图像旋转、剪裁等数据增强操作,并适当减少数量较多的肩部、手腕两个类别的训练样本数来平衡各类样本数量。其次,注意到数据集中图片长宽比分布不均,且长边均为512,短边长度在80~512区间呈随机分布,短边长度分布如图5所示。

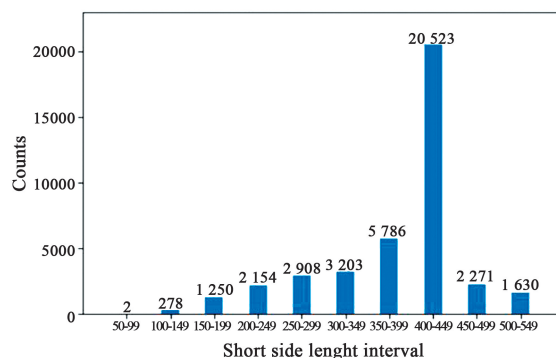


图5 短边长度区间分布图

Fig.5 Distribution diagram of short side length interval

为了避免送入特征提取网络时图像被直接resize而导致特征失真,在预处理阶段将短边沿图像两侧以图像均值像素填充至与长边一致,从而保证图像中包含有效信息区域的纵横比不受破坏。图像预处理前后的图片如图6所示。

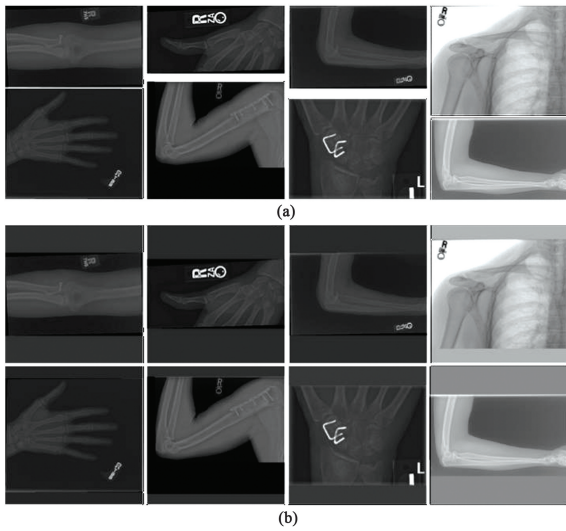


图 6 (a)数据集原图;(b)预处理后图。

Fig.6 (a) Original images of dataset; (b) Preprocessed images.

3.1.2 评估指标

一般而言,图像检索系统的性能可以分别从查准率(Precision)、查全率(Recall)、 F_1 度量(F_1 -score)、平均检索精度(mAP)以及检索时间几个指标来评估。不同的应用场景各个指标的关注度不同,对查询准确率较高的场景,比如在医学图像检索中,需要得到最相关的检索信息,且不相关样本误检索会带来较大的负面作用,所以更关注查准率。

查准率和查全率的公式为:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (11)$$

$$\text{Recall} = \frac{TN}{TP + FN}, \quad (12)$$

其中:TP为检索结果中相关样本的数量,FP是检索结果中不相关样本的数量,FN是数据库中未检索到的相关样本数量。

事实上,查准率和查全率是相互影响的。一般情况下,当查准率高时,容易漏检索,导致查全率低;而查全率高时,容易检索到错误样本,导致查准率低。因此,通过计算查全率和查准率的加权调和平均值 F_1 -score 可以综合考虑这两个指标。 F_1 -score 的计算公式为:

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (13)$$

在一些图像检索比赛中,通常还会参考检索结果

中 top- k 的平均检索精度($mAP@k$),如 2020 年的华为 DIGIX 数码设备检索比赛中,以 top1 的检索精度以及 top- k 的平均检索精度加权得到最终的成绩。一般情况下,用户只会选择性浏览排名靠前的 10~20 条检索结果,因此, top- k 平均检索精度更能反映用户在实际检索场景中的直观感受。top- k 平均检索精度的公式为:

$$mAP@k = \frac{1}{s} \sum_{i=1}^s \sum_{j=1}^k \frac{j}{\text{Position}(j)}, \quad (14)$$

其中 s 为查询次数、 $\text{Position}(j)$ 指搜索到的第 j 个相关样本在检索结果中的位置。

3.1.3 训练

本实验在开源 linux 操作系统 ubuntu18.04 下进行,相关硬件设备为 NVIDIA-1080 显卡、32 G 内存主机。并使用通用的深度学习框架 pytorch 进行网络设计,在 pycharm 编辑器中进行代码调试。

在训练前,为了更好地衡量模型在查准率和查全率两个指标上的评估,本文在测试集构造时统一了各个类别的数量。分别从每个类别中抽取 1 100 张图片,其中 1 000 张作为图像库,100 张作为待检索的输入图片。这样可以避免在召回率计算时,数量多的类别召回率表现很低的情况。在此基础上,对数据集中剩余样本按类别进行 5:5 的训练集、验证集划分。

在数据加载时,为了尽可能保留数据集原始信息,图像以每批次 4 张,尺寸为 512×512 输入网络。为了进一步平衡样本数量差异带来的少数样本特征学习不充分的问题,采用类别平衡采样法来保证每次采样中少数样本类别的被采样概率。其次,为了使模型对实际检索场景中输入图像的光照、角度、尺寸变换有更强的适应能力,对每个批次的数据进行在线数据增强。相比于离线增强,在线数据增强能够节省大量的数据存储空间,并且由于每个批次增强方式的随机性,能得到更丰富的输出,提高模型的鲁棒性。

最后,为了加快网络收敛,采用初始学习率为 0.001,权重衰减因子为 $1e-4$ 的 adam 优化器对模型参数沿负梯度方向更新,并在 20,50,90 训练轮数时对学习率进行衰减,使网络在训练初期保持较高的学习率,加快损失值下降的速度,在训练后期通过降低学习率来抑制损失振荡现象,使网

络逐步收敛。

3.2 对比实验及结果分析

3.2.1 定量分析

为了验证本文方法的有效性,分别对比了 SIFT-BoVWs、DHCNN、RAN 在 Mura 数据集上

的各个指标上的表现,其中查准率及查全率采用相似度 0.8 为阈值,即只取相似度大于 0.8 的作为最终检索结果,并统计了各个模型在 Mura 数据集上每个类别的 mAP@100、mAP@20 指标,表 1 是实验具体数据。

表 1 对比试验模型性能比较

Tab.1 Performance comparison of comparative test models

	Precision	Recall	F ₁ -score	mAP@100	mAP@20	Search time/s
SIFT-BoVWs	0.742	0.691	0.725	0.724	0.833	1.52
DHCNN	0.774	0.752	0.764	0.788	0.885	0.86
RAN	0.832	0.807	0.815	0.819	0.902	0.92
Ours	0.945	0.882	0.922	0.954	0.980	0.83

从对比实验可以看出,基于视觉词袋表征图像特征的 SIFT-BoVWs 模型在本数据集上精度比较低并且检索时间较长,主要是因为模型更关注图像的纹理及形状信息,而无法提取并利用图像的深层语义信息来进行图像检索,检索时间较长主要是图像 SIFT 特征提取阶段耗时过多。DHCNN 模型则利用了 vgg16 特征提取网络来代替 SIFT 特征提取并对高维的特征进行哈希值编码,在 GPU 设备的加速下,加快了特征提取的速度,并且由于训练过程中学习到了每个类的抽象特征,使得模型精度有了 6.2% 的 mAP@20 精度指标的提升。RAN 模型同样是采用深度学习的方法来提取图像特征,并在此基础上引入了自注意力模块,使得模型精度有了大幅度提升,但是 RAN 的特征提取网络采用了结构较复杂的 Resnet101 网络,检索耗时相比于 DHCNN 网络有所增加。本文设计的模型,在特征提取网络上参考了相较于前两者更轻量的 resnet50 主干网络,并在此结构上进行了一定改进,通过抽取不同层的特征,并利用注意力模块对其进行权重重分配,最后在训练阶段通过交叉熵损失和中心损失融合进一步让每个类别的特征在特征空间分布更加合理,最终在 Mura 数据集上 mAP@20 取得了 0.98 的检索精度。

3.2.2 定性分析

为了使检索效果得到更直观的体现,对同一

张输入图片分别用 4 个模型进行检索,并得到 top10 检索结果,如图 7 所示。

从检索结果 top10 可以直观地看出, SIFT-BoVWs 模型更关注样本的颜色、形状等特征,而对于输入的肘部测试图像,由于检索结果中第四幅肱骨图像和输入图像在视觉上的相似性导致误检索。模型 DHCNN 和 RAN 则在误检索上有所改善,但仍存在个别误检。综合来看,本文的模型在 top10 的检索精度表现较好,也比较符合实际场景对模型的检索精度要求。

3.3 消融实验及结果分析

3.3.1 定量分析

本节对本文第二部分中提到的主要改进点进行消融实验分析,并以此对各个模块引入的目的及取得的效果做更直观地叙述。

在实验过程中尝试过的且对精度提升有较大帮助的主要 3 点:(1)融合多个尺度特征对样本进行更全面的描述,优化模型对于输入图像中不同尺度范围的检索能力;(2)加入自注意力模块,强化图像中关键细节特征的表达能力;(3)结合多重损失优化,在加大类间距离的同时,缩短类内距离,使样本特征在特征空间的分布更加合理。

为了方便地描述上述改进点在数据集中每个类别上的提升效果,统计了实验中模型在 Mura 数据集 7 个类别的 mAP@100 指标,表 2 是消融实验的具体数据。

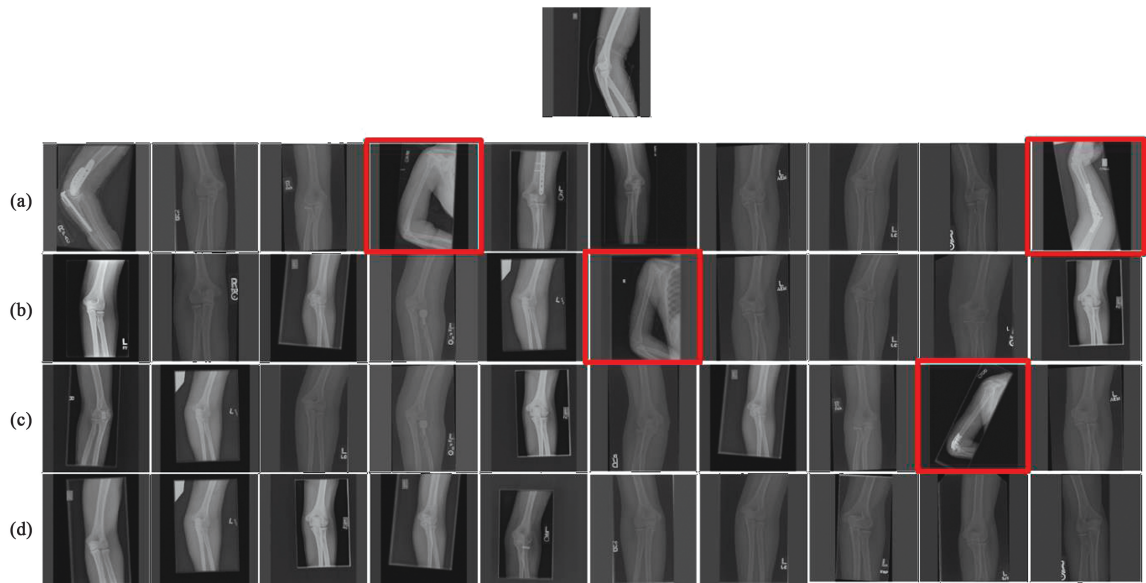


图 7 不同模型的检索效果图。(a) SIFT-BoVWs; (b)DHCNN; (c)RAN; (d)本文模型。红框中为误检索图片。

Fig.7 Effect diagram of different models.(a) SIFT-BoVWs; (b)DHCNN; (c)RAN; (d)Ours. The picture in the red box is the wrong picture.

表 2 消融实验模型性能比较

Tab.2 Performance comparison of ablation experimental models

	Elbow	Finger	Forearm	Hand	Humerus	Shoulder	Wrist
Res50	0.729	0.807	0.837	0.811	0.857	0.752	0.843
Res50+多层特征	0.834	0.856	0.872	0.848	0.828	0.854	0.826
Res50+注意力	0.776	0.892	0.856	0.883	0.863	0.796	0.850
Res50+多重损失	0.854	0.832	0.898	0.886	0.861	0.810	0.872
Res50+多层特征+注意力	0.904	0.931	0.915	0.953	0.941	0.933	0.924
Res50+多层特征+注意力+多重损失	0.933	0.964	0.973	0.971	0.943	0.945	0.957

在 Resnet50 的基础上,结合上文提到的优化措施,设计了 6 组实验。通过对模型在各类的检索精度分析可知,模型在手肘、肩部这两类的精度较低。而通过观察这两类易检索出错的样本发现,模型对于肘关节的局部图像以及包含前臂和肱骨的肘部图像容易检索成其他类,而肩膀这类样本也是如此,由此猜想模型对于尺度变化大的样本的特征辨别能力还不够,因此有了引入多尺度特征的尝试,通过对不同尺度特征的组合尝试,这两类的检索精度得到了平均 10 个点的提高。同时对比了注意力及多重损失单独作用的模型精

度提升,在单一尺度的注意力作用下,手肘、前臂、肩膀 3 个类的提升并不如多尺度的明显,而多重损失的加入则能够在前臂、手掌这两个易混淆的类上有十分明显的提升。

结合上述实验可以发现,在引入多尺度特征的基础上,虽然模型的整体精度提高了,但是手腕、前臂这两类的精度有所下降。对这两类的特征图可视化之后发现,前文引入的浅层纹理会对手腕的特征造成一定程度的影响,使得模型的注意力被边缘纹理特征破坏,导致手腕与前臂这两类混淆的几率加大。基于此,引入自注意力机制,

使网络能够关注到重要的特征并抑制无关的干扰性特征,实验表明,引入注意力机制后,网络的注意力能够关注到不同类别的关键特征区域,从而较好地解决了类间易混淆问题。

在引入注意力模块之后,每个类别的精度都得到了平衡。为了进一步提高模型精度,引入中心损失和交叉熵损失结合来优化各类样本在特征空间的分布,减少位于边界区域的样本混淆概率。

3.3.2 定性分析

为了使每个模块的改进更加直观,本文随机抽取了部分样本的特征进行了可视化处理,并以热力图的形式叠加到原图进行展示,图8是具体效果。

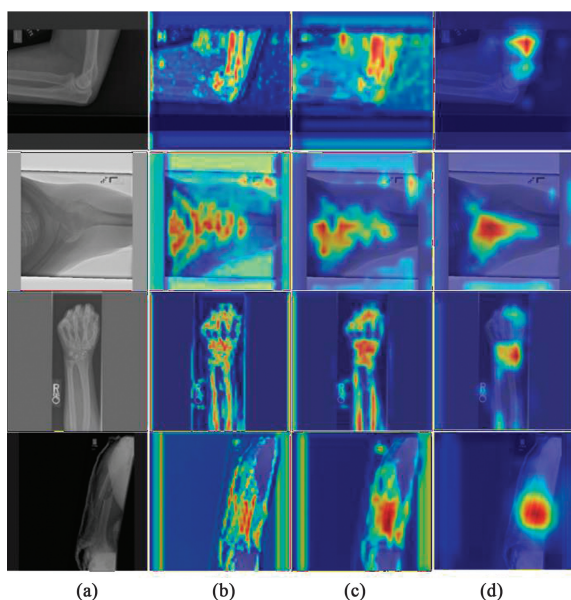


图8 消融实验效果图。(a)原图;(b)RvesNet50;(c)ResNet50+多层特征;(d)ResNet50+多层特征+注意力。

Fig.8 Effect diagram of ablation experiment. (a) Original image; (b) ResNet 50; (c) ResNet 50 + Multiple feature; (d) ResNet 50 + Mutiple feature+Self-attention.

可以看出,相对于 ResNet50 基础模型,多尺度特征的引入可以从全局角度对不同尺度特征进行更合理的组合。引入自注意力模块后,网络的关注度进一步集中到了关键区域。

引入多重损失前后在注意力图中无明显变化,这里将样本特征进行降维处理,降维到二维后,在平面图中进行展示,图9是使用多重损失前后的每类样本特征分布图。

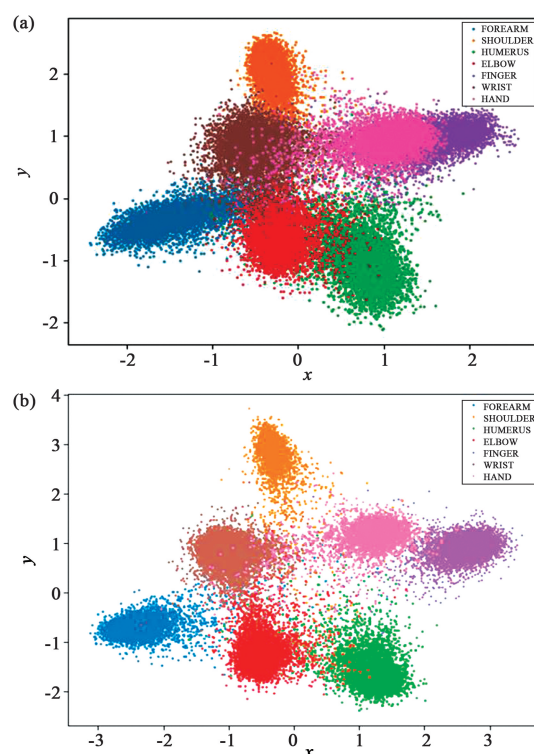


图9 样本特征分布图。(a)原分布图;(b)优化后分布图。

Fig.9 Distribution map of sample features. (a) Original distribution map; (b) Optimized distribution map.

4 结 论

针对医学图像的一些固有特征造成现有的一些图像检索方案偏低的问题,本文提出了一种融合多尺度特征及注意力机制的医学图像检索系统优化思路。在特征提取阶段,借鉴了深度残差网络的结构设计,并融合不同层次、不同尺度的特征图,充分利用了图像的浅层纹理特征及深层语义特征,较好地缓解了不同尺度目标的特征提取问题。同时,设计了一个改进的注意力模块以适应不同尺度的特征图输出,并对所有通道特征进行权重重分配,提高了关键通道的特征表达能力,使图像中的重要细节特征更加突出。最后,在模型训练阶段,采用交叉熵损失和中心损失相结合的思路,使得各个类的样本特征在样本空间的分布更加合理,进一步提高了模型的检索精度。实验证明,本文的方案相较于其他医学图像检索模型在 Mura 数据集上 mAP@20 能够获得 0.98 的精度,基本符合实际场景对模型的检索精度要求。

参 考 文 献:

- [1] MÜLLER H, MICHOUX N, BANDON D, *et al.* A review of content-based image retrieval systems in medical applications—clinical benefits and future directions [J]. *International Journal of Medical Informatics*, 2004, 73(1): 1-23.
- [2] 周进凡. 基于大数据的肺部 X 光图像的分析与研究方法[D]. 贵阳: 贵州大学, 2019.
ZHOU J F. Analysis and research methods of lung X-ray images based on big data [D]. Guiyang: Guizhou University, 2019. (in Chinese)
- [3] 陈静. 基于深度学习的肺结节检测系统[D]. 上海: 上海交通大学, 2018.
CHEN J. Lung nodule detection with deep learning [D]. Shanghai: Shanghai Jiao Tong University, 2018. (in Chinese)
- [4] 彭朋, 蒋涛, 李敏. PACS 系统与临床信息系统互联对医学影像教学的意义[J]. 中国病案, 2016, 17(1): 82-84.
PENG P, JIANG T, LI M. The significance of PACS system and clinical information system interconnection to medical imaging teaching [J]. *Chinese Medical Record*, 2016, 17(1): 82-84. (in Chinese)
- [5] JIJI G W, RAJ P S J D. Content-based image retrieval in dermatology using intelligent technique [J]. *IET Image Processing*, 2015, 9(4): 306-317.
- [6] MIZOTIN M, BENOIS-PINEAU J, ALLARD M, *et al.* Feature-based brain MRI retrieval for Alzheimer disease diagnosis [C]//*Proceedings of the 2012 19th IEEE International Conference on Image Processing*. Orlando: IEEE, 2012: 1241-1244.
- [7] RAHMAN M, ANTANIS K, THOMA G R. A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback [J]. *IEEE Transactions on Information Technology in Biomedicine*, 2011, 15(4): 640-646.
- [8] 陈晓冬, 盛婧, 杨晋, 等. 多参数 Gabor 预处理融合多尺度局部水平集的超声图像分割[J]. 中国光学, 2020, 13(5): 1075-1084, doi:10.37188/CO.2020-0025.
CHEN X D, SHENG J, YANG J, *et al.* Ultrasound image segmentation based on a multi-parameter Gabor filter and multiscale local level set method [J]. *Chinese Optics*, 2020, 13(5): 1075-1084. doi: 10.37188/CO.2020-0025. (in Chinese)
- [9] QAYYUM A, ANWAR S M, AWAI S M, *et al.* Medical image retrieval using deep convolutional neural network [J]. *Neurocomputing*, 2017, 266: 8-20.
- [10] 吕晓琪, 吴凉, 谷宇, 等. 基于三维卷积神经网络的低剂量 CT 肺结节检测[J]. 光学 精密工程, 2018, 26(5): 1211-1218.
LÜ X Q, WU L, GU Y, *et al.* Detection of low dose CT pulmonary nodules based on 3D convolution neural network [J]. *Optics and Precision Engineering*, 2018, 26(5): 1211-1218. (in Chinese)
- [11] 熊舒羽. 基于深度学习技术的医学图像检索方法研究[D]. 重庆: 重庆理工大学, 2019.
XIONG S Y. Research on medical image retrieval using deep learning technology [D]. Chongqing: Chongqing University of Technology, 2019. (in Chinese)
- [12] 彭晏飞, 梅金业, 王恺欣, 等. 基于区域注意力机制的遥感图像检索[J]. 激光与光电子学进展, 2020, 57(10): 101017.
PENG Y F, MEI J Y, WANG K X, *et al.* Remote sensing image retrieval based on regional attention mechanism [J]. *Laser & Optoelectronics Progress*, 2020, 57(10): 101017. (in Chinese)
- [13] 周国华, 蒋晖, 顾晓清, 等. 多视角判别度量学习的乳腺影像检索方法[J]. 液晶与显示, 2020, 35(6): 619-630.
ZHOU G H, JIANG H, GU X Q, *et al.* Multi-view metric learning with Fisher discriminant analysis and its applications for breast image retrieval [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6): 619-630. (in Chinese)
- [14] GU Y, YANG J. Multi-level magnification correlation hashing for scalable histopathological image retrieval [J]. *Neurocomputing*, 2019, 351: 134-145.
- [15] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016*

- IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016.
- [16] BUADES A, COLL B, MOREL J M. Non-local means denoising [J]. *Image Processing on Line*, 2011, 1: 208-212.
- [17] WANG X L, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks [C]//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018.
- [18] TANG H, XU D, SEBE N, *et al.* Attention-guided generative adversarial networks for unsupervised image-to-image translation [C]//*Proceedings of 2019 International Joint Conference on Neural Networks*. Budapest: IEEE, 2019: 1-8.
- [19] TANG H, LIU H, XU D, *et al.* AttentionGAN: unpaired image-to-image translation using attention-guided generative adversarial networks [J]. *arXiv*: 1911.11897, 2019.
- [20] WEN Y D, ZHANG K P, LI Z F, *et al.* A discriminative feature learning approach for deep face recognition [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016.

作者简介:



周林鹏(1996—),男,江西赣州人,硕士研究生,2018年于江西理工大学获得学士学位,主要从事深度学习、图像处理方面的研究。E-mail: 961031645@qq.com



姚剑敏(1978—),男,福建莆田人,博士,副研究员,2005年于中科院长春光学精密机械与物理研究所获得博士学位,主要从事人工智能、图像处理、信息显示技术等方面的研究。E-mail: yaojm@fzu.edu.cn