

基于小样本学习和多尺度残差网络的特纳综合征预测研究

刘璐

(北京工业大学信息学部软件学院 北京 100124)

摘要 为了提高特纳综合征(Turner Syndrome, TS)的诊断效率,提出一种基于小样本学习和多尺度残差网络的TS预测模型。对TS人脸图像进行预处理获取人脸主要区域,提出具有多级注意力机制的多尺度残差模块,其中,多尺度残差模块以集成多尺寸卷积核的残差结构实现,多级注意力机制用来学习特征通道关系和不同卷积核的重要性,利用该模块构建多尺度残差网络。使用小样本学习进行模型训练。实验结果表明,该模型能够提升TS的诊断准确率。

关键词 特纳综合征 注意力机制 残差网络 小样本学习

中图分类号 TP183

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.09.026

A PREDICTION MODEL FOR TURNER SYNDROME BASED ON FEW-SHOT LEARNING AND MULTISCALE RESIDUAL NETWORK

Liu Lu

(School of Software Engineering, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract A prediction model is proposed for improving the diagnosis efficiency of Turner syndrome (TS) based on a multiscale residual network (MRN) and few-shot learning. TS facial images were pre-processed to obtain the main facial areas. A multiscale residual block (MRB) with multilevel attention mechanisms (MAM) was designed. The MRB was implemented by integrating the residual structure of multi-scale convolution kernels, and the MAM was used to learn feature channel relationships and the importance of different convolution kernels. The MRN was built using the MRB. The few-shot learning was utilized to train the MRN. The experimental results demonstrate that the prediction model can improve the diagnostic accuracy of TS.

Keywords Turner syndrome Attention mechanisms Residual network Few-shot learning

0 引言

特纳综合征是一种严重危害女性生理健康的染色体异常疾病。TS是由细胞中部分或全部X染色体缺失或者结构异常引起的,主要的临床症状包括身材矮小、青春期无性征发育和卵巢衰竭等。该病的发病率约为0.25%~0.5%,且无法完全治愈。目前,临床常用激素补充治疗方法缓解TS造成的各种症状,因此及时的诊断对治疗策略的制定和患者的恢复都具有非常重要的意义。

常用的TS诊断方法是染色体核型检查,但是这种方法比较昂贵且检测时间较长。因此,相关学者研究使用计算机辅助诊断技术预测TS,进而提高医生的诊断效率。

有些研究人员基于TS的临床症状指标数据构建立贝叶斯网络模型^[1];也有学者使用TS患者的复杂颅面相关特征数据构建逻辑回归模型^[2];还有学者结合孕妇血清筛选数据、超声波检查数据,以及TS患者统计学数据构建神经网络分类模型^[3]。但是,上述方法所用的实验数据获取难度较大且需要研究人员具备较高的医学专业知识。最近,基于人脸图像的TS分类

研究开始受到学术界的关注。相比上述医疗数据,人脸数据更容易获取。根据相关医学研究,人脸图像也可以作为一种诊断 TS 的重要依据。TS 人脸图像包括一些 TS 患者的特有面部特征,比如宽额头和高鼻梁等,这是基于人脸图像构建 TS 分类模型的医学理论基础。现有的基于 TS 人脸图像的分类模型的构建过程是首先利用统计学知识抽取人脸图像的特征表示,比如离散小波特征和纹理特征等,然后使用经典的机器学习分类算法,如支持向量机、构建 TS 分类模型^[4-5,7]。此外,Liu 等^[6]提出使用一种无监督特征学习方法,即卷积递归神经网络,实现对图像的自动特征抽取,然后使用支持向量机分类 TS。综上所述,现有 TS 分类模型主要是将手动设计的图像特征和机器学习算法结合构建 TS 预测模型。这些模型的缺点是手动设计的图像特征可能存在不完整或者冗余等问题,会影响模型的预测性能。文献[6]方法的分类准确率也有待提升。

近年来,深度学习技术在医疗图像处理领域有非常成功的应用。比如,Jun 等^[8]提出一种基于卷积神经网络(Convolutional Neural Networks, CNNs)的白内障分级模型。此外,也有许多基于 CNNs 的医疗图像分割模型被提出^[9-11]。这些深度学习模型在这些医疗图像处理任务中都取得了较好的效果。

因此,本文将深度学习应用于 TS 自动诊断任务中,提出一种基于小样本学习(Few-shot Learning)和多尺度残差网络(Multiscale Residual Network, MRN)的 TS 预测模型^[12]。首先,设计一种具有多级注意力机制的多尺度残差模块(Multiscale Residual Block, MRB),其中,多尺度残差模块将多个尺寸的卷积核融入残差结构中进而获取图像的多尺度特征,多级注意力机制包括通道级注意力模块和尺度级注意力模块,分别对源于多个卷积核的特征图的通道关系和不同卷积核的重要性进行建模;其次,利用该 MRB 构建面向 TS 预测任务的 MRN;最后,为了解决标注数据不足的问题,使用一种小样本学习技术,即模型无关的元学习算法(Model-agnostic Meta-learning Algorithm, MAML)进行模型训练^[13]。实验结果表明,所提出的 TS 预测模型能够基于有限的标注数据获得理想的分类性能。

1 TS 预测模型

1.1 多尺度残差模块

所提出的 MRN 主要是以堆叠 MRB 的方式实现网络模型的构建。因此,MRB 是实现人脸特征学习的核

心组件,其具体计算流程如图 1 所示。

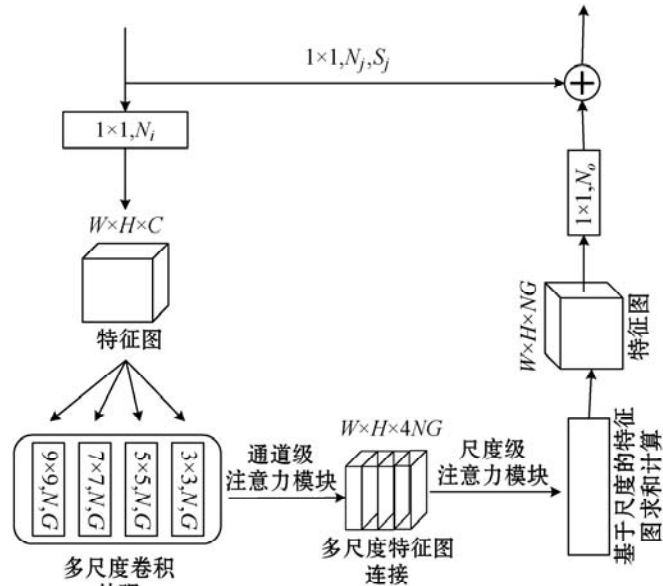


图1 多尺度残差模块

可以看出,所提出的 MRB 涉及的主要技术内容有残差结构设计、多尺度分组卷积、通道级注意力模块和尺度级注意力模块。

(1) 残差结构设计。残差结构是卷积神经网络中非常经典的模块设计技巧^[12]。在残差结构中,以结合卷积计算和跳跃连接的方式进行图像特征学习。因此,可以认为残差结构是以增加跳跃连接的方式改进传统基于卷积计算的特征学习过程。跳跃连接将浅层的特征图与深层特征图进行对应位置相加,进而实现特征的跨层传递。该计算过程的理论基础是恒等映射理论,其优点是在模型优化时可以有效避免因模型过深而引起的梯度消失。所提出的 MRB 也使用了残差结构设计。对于原始设置中的卷积模块以多尺度分组卷积结合多级注意力机制替换。对于跳跃连接维度不匹配问题,以 1×1 的卷积对输入特征图进行处理,其步长和卷积核数量分别为 S_j 和 N_j 。设 MRB 的特征学习过程为函数 $f(x)$, 输入为 x , 模块目标函数为 $D(x)$, 则 MRB 可以表示为:

$$D(x) = f(x) + x \quad (1)$$

其中,当 $f(x) = 0$ 时, $D(x) = x$ 为恒等映射。这种以最小化 $f(x)$ 逼近恒等映射的改进,比直接优化 $D(x) = x$ 更为容易。

(2) 多尺度分组卷积。在卷积神经网络中,特征学习是通过卷积核实现的。卷积核可以视为一种滤波器,通过对图像进行滤波处理,获取图像的有效特征表示。常规卷积的计算过程如式(2)所示。

$$O = b + \sum_{r=1}^c W_r X_r \quad (2)$$

式中: O 、 b 和 C 分别表示输出、偏置项和输入通道数; X_r 和 W_r 分别表示第 r 通道的输入和卷积核权重。可以发现,常规卷积的计算过程涉及所有的输入通道。随着输入通道数量的增加,常规卷积计算的参数也会快速增长。因此,为了缓解这种问题,研究人员提出使用分组卷积减少卷积核参数。分组卷积首先将输入通道进行平均分组,然后在各个子通道组中执行式(2)。因此,分组卷积计算过程可描述为:

$$O = \bigg|_{k=1}^G \left(b + \sum_{r=1}^{C/G} W_r^k X_r^k \right) \quad (3)$$

式中: G 表示通道分组数; $\big|_{k=1}^G (\cdot)$ 表示将 G 组卷积输出特征图进行连接处理。当生成的特征图数量相同时,常规卷积的参数数量为 $K \times K \times C \times M$; 对于分组卷积,参数数量为 $K \times K \times C/G \times M$; 在生成 M 个特征图的情况下,分组卷积的参数数量为传统卷积的 $1/G$ 。

多尺度分组卷积是指在分组卷积中设置多个尺寸的卷积核^[14-15]。根据图 1,MRB 可以设置 4 个尺寸的卷积核,针对每个卷积核,分组数设置为 G ,卷积核数量设置为 N 。多尺度卷积的目的是获得图像的多尺度特征表示,但是常规的多尺度卷积会产生过多的参数,因此本文将多尺度分组卷积设置引入 MRB 设计中。

(3) 通道级注意力模块。注意力机制是一种能让网络在学习过程中更关注有价值特征信息的模型设计方式。在计算机视觉领域,注意力机制可以分为空间域注意力机制和通道域注意力机制^[16]。空间域注意力机制的目标是让模型更加关注特征图的局部特征信息。本文所提出的 MRB 采用了多尺度分组卷积的设计技巧,能够充分捕获输入的局部特征信息。因此,本文不考虑空间域注意力机制设置。

对于输入图像,一次卷积计算可以生成该输入的多个特征图表示。这些特征图对于输入的重要性是不同的。因此,独立地对 MRB 中每个尺寸的卷积核所生成的各个特征通道进行建模是很有必要的。MRB 所用的通道级注意力模块如图 2 所示。

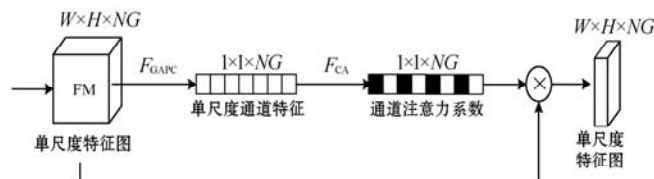


图 2 通道级注意力模块

该模块借鉴了方法^[17]的学习技巧。根据图 2,将源于某一尺寸卷积核的特征图称为单尺度特征图。根据图 1 所示,当输入的维度为 $H \times W \times C$ 时,那么单尺

度分组卷积的输出维度为 $H \times W \times NG$,即为图 2 中的单尺度特征图。针对该单尺度特征图计算通道级注意力系数包括三个计算步骤,即单尺度特征图降维、通道注意力系数学习以及通道注意力系数应用。对于特征图降维,使用沿通道维度的全局均值池化计算 $F_{\text{GAP}}(\cdot)$ 。设单尺度特征图 FM 经过 $F_{\text{GAP}}(\cdot)$ 生成的特征向量为 $s \in \mathbf{R}^{1 \times 1 \times NG}$ 。根据全局均值池化计算, s 中第 l 个元素的计算过程为:

$$s_l = \frac{1}{W \times H} \left(\sum_{i=1}^W \sum_{j=1}^H FM_l(i, j) \right) \quad (4)$$

利用生成的 s ,通过两层全连接神经网络 $F_{\text{CA}}(\cdot)$ 计算通道注意力系数 CA :

$$CA = F_{\text{CA}}(s) = \sigma(W_1 \cdot \delta(W_0 \cdot s)) \quad (5)$$

式中: σ 和 δ 分别表示激活函数 Sigmoid 和 ReLU; $W_0 \in \mathbf{R}^{NG/r \times NG}$ 和 $W_1 \in \mathbf{R}^{NG \times NG/r}$ 为神经网络的权重参数; r 为缩放系数,设置为 16。

通道级注意力系数的应用方式以注意力系数与对应通道特征图相乘的方式实现:

$$FM'_l = CA_l \times FM_l \quad l \in [1, NG] \quad (6)$$

式中: CA_l 表示通道级注意力系数 CA 中第 l 个元素; FM_l 为 FM 第 l 维特征图; FM'_l 为经过通道级注意力模块生成的单尺度特征图 FM' 的第 l 维特征图。

(4) 尺度级注意力模块。因为 MRB 设置了多尺度分组卷积,所以仅对单尺度特征图的通道重要性建模是不够的。本文认为不同尺度的特征图对于输入的重要性也是不完全相同的。因此,本文将通道域注意力机制扩展到多尺度卷积核重要性分析任务。所提出的尺度级注意力模块如图 3 所示。尺度级注意力模块通过对多尺度特征图的重要性进行分析,进而实现对不同尺度卷积核重要性的建模。尺度级注意力模块与通道级注意力模块最显著的区别是尺度级注意力系数是针对单尺度特征图所有通道的,而通道级注意力系数只对单尺度特征图对应通道特征图起作用。根据图 3 所示,尺度级注意力模块包括四个计算步骤,即多尺度特征图降维、尺度级注意力系数学习和应用以及多尺度特征融合。

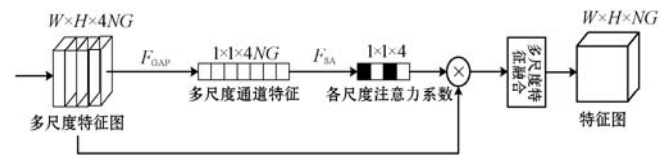


图 3 尺度级注意力模块

多尺度特征图降维与图 2 中单尺度特征图降维的计算过程完全一致。设 MRB 中使用了四个不同尺度的卷积核,通过通道级注意力模块和多尺度特征图连

接,生成的多尺度特征图 $MF\mathcal{M}$ 的维度为 $H \times W \times 4NG$ 。对 $MF\mathcal{M}$ 执行全局均值池化 $F_{CAP}(\cdot)$,生成 $MF\mathcal{M}$ 的特征向量 $ms \in \mathbf{R}^{1 \times 1 \times 4NG}$ 。基于 ms ,使用两层全连接神经网络 $F_{SA}(\cdot)$ 计算尺度级注意力系数 SA :

$$SA = F_{SA}(ms) = \sigma(W_1 \cdot \delta(W_0 \cdot ms)) \quad (7)$$

式中: $W_0 \in \mathbf{R}^{4NG/r \times 4NG}$ 和 $W_1 \in \mathbf{R}^{4 \times 4NG/r}$ 为神经网络的权重参数; r 、 σ 和 δ 设置同式(5)。

尺度级注意力系数的应用方式以注意力系数与对应单尺度特征图相乘的方式实现:

$$MF\mathcal{M}'_l = SA_l \times MF\mathcal{M}_l \quad l \in [1, 4] \quad (8)$$

式中: SA_l 表示尺度级注意力系数 SA 中第 l 个元素; $MF\mathcal{M}_l \in \mathbf{R}^{W \times H \times NG}$ 为 $MF\mathcal{M}$ 中第 l 个单尺度特征图分量; $MF\mathcal{M}'_l$ 为经过尺度级注意力模块生成的多尺度特征图 $MF\mathcal{M}'$ 中第 l 个单尺度特征图分量。

为了减少特征维度和计算量,将生成的 $MF\mathcal{M}'$ 中的四个不同的单尺度特征图进行对应位置求和来实现多尺度特征融合,具体的计算过程如下:

$$SFM = \sum_{l=1}^4 MF\mathcal{M}'_l \quad (9)$$

式中: $SFM \in \mathbf{R}^{W \times H \times NG}$ 为通道级注意力模块最终生成的特征图。

(5) MRB 特征学习过程。所提出的 MRB 是将多尺度分组卷积、通道级注意力模块和尺度级注意力模块以残差结构的设计技巧进行融合的特征学习过程。根据图 1 所示,MRB 首先将输入的特征图以数量为 N_i 的 1×1 常规卷积进行预处理;其次,对于处理输出的特征图进行多尺度分组卷积处理,需要注意的是,在多尺度分组卷积中,卷积核的尺寸设置并不固定,可以随着网络层数的加深逐渐减少卷积核的种类;然后使用通道级注意力模块对生成的多个不同尺度的特征图进行处理,并对处理结果进行特征图连接并;最后,以通道级注意力模块处理连接生成的多尺度特征图并以 N_o 个 1×1 常规卷积进行特征图输出处理。

1.2 MRN 构建

所提出面向 TS 预测的 MRN 的网络架构如图 4 所示。根据图 4,首先使用 64 个 7×7 的常规卷积核对待 TS 人脸图像进行特征处理,其次,对于输出的特征图,使用三组不同的 MRB 进行特征学习,然后,使用三个依次连接的 3×3 常规卷积对 MRB 的输出做进一步特征处理。最后使用全局均值池化对特征图进行降维计算,并将生成的特征向量传入全连接层实现对 TS 的预测。

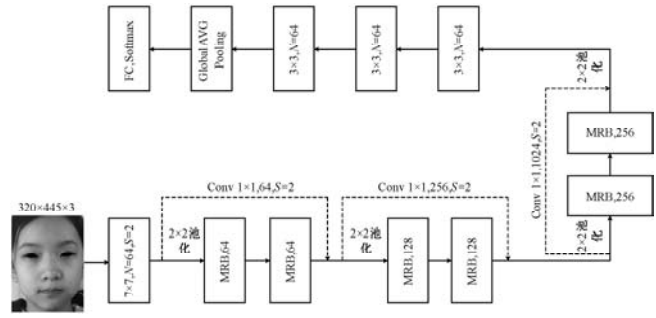


图 4 多尺度残差网络

表 1 网络模型参数

模块	卷积核尺寸	注意力模块	各层输出
预处理	$7 \times 7, 64, /2$	—	$160 \times 223 \times 64$
MRB, 64	$1 \times 1, N_i = 64$ $\begin{bmatrix} 9 \times 9, \\ 7 \times 7, N = 16 \\ 5 \times 5, G = 4 \\ 3 \times 3, \\ 1 \times 1, N_o = 64 \end{bmatrix} \times 2$	$Sca = 4,$ $Cha = 64$	$80 \times 112 \times 64$
MRB, 128	$1 \times 1, N_i = 128$ $\begin{bmatrix} 7 \times 7, N = 32 \\ 5 \times 5, G = 8 \\ 3 \times 3, \\ 1 \times 1, N_o = 256 \end{bmatrix} \times 2$	$Sca = 3,$ $Cha = 256$	$40 \times 56 \times 256$
MRB, 256	$1 \times 1, N_i = 256$ $\begin{bmatrix} 5 \times 5, N = 64 \\ 3 \times 3, G = 16 \\ 1 \times 1, N_o = 1024 \end{bmatrix} \times 2$	$Sca = 2,$ $Cha = 1024$	$20 \times 28 \times 1024$
Conv	$1 \times 1, N_i = 512$ $\begin{bmatrix} 3 \times 3, \\ N = 512 \\ G = 1 \end{bmatrix} \times 3$ $1 \times 1, N_o = 512$	—	$10 \times 14 \times 512$
Global AVG Pooling			512
FC(512, 2), Softmax			2

MRN 的网络参数设置如表 1 所示。在表 1 中,模块“MRB, 64/126/256”表示 MRB 模块中输入端的 1×1 卷积核的数量 N_i 为 64/126/256;Conv 模块表示图 4 中三个连续的 3×3 常规卷积;在注意力模块中, Sca 表示尺度级注意力系数维度, Cha 表示通道级注意力系数维度;“—”表示模块无该设置;MRN 中填充模式设置为 $Padding = SAME$,MRB 模块和 Conv 模块中的卷积的步长设置为 1。根据图 4 和表 1,输入图像经过卷积预处理后输出的特征图维度为 $160 \times 223 \times 64$ 。接下来,使用 2×2 的最大值池化处理预处理输出,生成 $80 \times 112 \times 64$ 的特征图;然后,依次使用两组“MRB, 64”、两组“MRB, 128”和两组“MRB, 256”对特征图进行处理,不同组之间使用 2×2 的最大值池化进行特征图降维处理。在两组“MRB, 64”中,首先用 64 个 1×1

卷积进行处理,然后以四个不同尺度的分组卷积对处理结果进行多尺度特征学习,卷积核设置为 3×3 、 5×5 、 7×7 和 9×9 ,分组数 $G=4$,卷积核数量 $N=16$ 。经过通道级注意力模块和尺度级注意力模块以及 64 个 1×1 常规卷积处理,最后生成表达能力更强的 $80 \times 112 \times 64$ 特征图。在 MRN 中,在两个“MRB,64”之上也设置了更高级别的跳跃连接,可以视为残差结构的残差结构。对于“MRB,128”和“MRB,256”分组,相关的设置和特征学习过程与“MRB,64”类似。随着网络的增加,特征图的空间尺寸会变小,因此在“MRB,128”和“MRB,256”分组中依次减少卷积核的种类。在“MRB,128”分组中,设置三种卷积核,即 3×3 、 5×5 和 7×7 ;在“MRB,256”分组中,设置两种卷积核,即 3×3 和 5×5 。为了保证特征的多样性,也增加了两个分组的分组数和卷积核数量。在“MRB,128/256”中, $G=8/16$, $N=32/64$ 。“MRB,128/256”最后的输出使用 1×1 卷积处理,卷积核数量为 $N_o=NG$ 。

经过“MRB,64/126/256”特征处理模块处理,生成的特征图维度为 $20 \times 28 \times 256$ 。由于预处理中特征图的空间尺寸为 7×7 ,因此在 Conv 模块设置三个连续的 3×3 常规卷积,其中 $N=512$,对该特征图进行进一步处理。这种设置的理由是三个连续的 3×3 卷积的感受野与一个 7×7 卷积的感受野是相同的。特征图经过处理维度变为 $10 \times 14 \times 512$,然后使用全局均值池化处理,生成 TS 图像的 512 维特征向量表示,最后通过全连接层实现对 TS 图像的预测。

1.3 基于小样本学习的模型训练

深度学习模型的良好泛化性往往需要依赖大量的标注数据。当使用少量训练数据训练深度学习模型时,极有可能造成模型的过度拟合。最近,小样本学习技术被证明能够有效地完成基于少量标记数据的深度学习模型训练任务。

基于元学习(Meta Learning)的小样本学习方法能以一种简单且有效的方式建立具有良好泛化性的深度学习图像分类模型。元学习是指利用模型过去学习到的知识训练新的任务,这个过程可以视为“学习如何学习”。元学习包括元训练阶段和元测试阶段。在元训练阶段,通常包括多个 N-way K-shot 任务,代表每个任务具有 N 个类别,支撑集(训练集)的样本数量为 NK 。此外,每个任务有一个查询集(测试集),样本数量为 $Q \times N$ 。元学习基于这些任务完成模型的训练。生成模型即为学习器,能够扩展到新的相关任务中。

对于 TS 预测任务,本文应用 MAML 算法完成 MRN 的训练。MAML 是一种不限制网络结构与损失

函数的元学习算法。因此,基于 MAML 的 TS 预测模型的训练主要涉及 MRN 网络,损失函数设置和生成 2-way K-shot 任务的任务分布 $P(T)$ 。针对每个任务,使用交叉熵损失函数:

$$LT_i(f_\theta) = \sum_{x_i, y_i \sim T_i} y_i \log f_\theta(x_i) + (1 - y_i) \log(1 - f_\theta(x_i)) \quad (10)$$

式中: x_i 和 y_i 表示源于任务 T_i 的训练样本对; f_θ 是MRN的函数表示; θ 为MRN的参数。

TS 预测模型的训练在元测试阶段完成,元训练阶段主要是训练学习器。在元测试阶段,采用两步微调(Fine-tuning)的方式训练 TS 预测模型,即首先使用 TS 人脸数据对整个学习器进行微调,然后冻结全连接层之前的网络参数,以 TS 人脸数据微调最后的全连接层,进而获取最终的 TS 预测模型。基于 MAML 的 TS 预测模型具体训练过程如算法 1 所示,其中 iterations 是指迭代次数 epoch。

算法 1 基于 MAML 的 TS 预测模型训练

输入:人脸图像 X 和标签 Y ,学习率 α 和 β 。

输出:TS 预测模型 Model。

1. 随机初始化 f_θ
2. for ite in iterations do
3. 抽取 (Batch Size) M 个图像二分类任务 $T_i \sim P(T)$
4. for all T_i do
5. 从 T_i 中抽取 K 个样本 $D_i = \{x_i, y_i\}$
6. 基于 D_i 和 LT_i 计算梯度 $\nabla_\theta LT_i(f_\theta)$
7. 执行参数更新 $\theta'_i = \theta - \alpha \nabla_\theta LT_i(f_\theta)$
8. 从 T_i 中抽取 K 个样本 $D_m = \{x_i, y_i\}$ 用于学习器参数更新
9. end for
10. 更新学习器参数 $\theta = \theta - \beta \nabla_\theta \sum_{T_i \sim P(T)} LT_i(f_{\theta'_i})$
11. end for
12. 基于人脸数据对整个模型 $f_\theta(X, Y)$ 进行微调
13. 对 MRN 的全连接层进行微调,获取 TS 预测模型 Model = $f_\theta(X, Y)$

2 实验结果

2.1 数据集

实验所用的 TS 人脸数据集源自北京协和医院的相关数据库。TS 人脸数据集包含 108 例 TS 患者以及 622 例正常实验对象(Normal Control, NC)。参与采集的实验对象的年龄在 9 ~ 14 岁之间。本文使用 mini-ImageNet 训练基于 MAML 的学习器。miniImageNet 是源于 ILSVRC-2012 的一个子集,包括 100 个类别的图像,每个类别有 600 个样本^[18]。

2.2 实验设置

在模型微调阶段,使用 70% 的 TS 人脸图像和正常人脸图像作为训练集,剩余 30% 作为测试集。对于 miniImageNet,以随机重复抽样的方式抽取两类图像构成图像分类任务。在元训练阶段,设置 Batch Size = 64,即每次训练有 10 个 2-way K-shot 图像分类任务,迭代次数 epoch = 60。对于测试阶段,设置 $Q = K$ 。MAML 中的两个学习率设置为 0.4。图像预处理主要是以 MTCNN 获取人脸主要面部区域^[19]。

性能指标包括准确率 Accuracy,灵敏度 Sensitivity 以及特异度 Specificity。具体的数学定义如下所示:

$$A_{accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (11)$$

$$S_{sensitivity} = \frac{T_p}{T_p + F_n} \quad (12)$$

$$S_{specificity} = \frac{T_n}{T_n + F_p} \quad (13)$$

式中: T_p 和 T_n 分别表示被正确分类的 T_s 和 N_c 的数量; F_p 和 F_n 分别表示被错误分类的 TS 和 NC 的数量。

2.3 训练样本数量设置影响

该实验研究不同训练集大小对学习器的性能影响。学习器的性能以元测试阶段生成的 TS 预测模型的分​​类准确率评估。在学习器的训练阶段,训练样本是 2-way K-shot 任务,通过设置不同的 K 来改变训练集的大小。针对学习器的训练,设置 K 分别为 1、5、10 和 15,即训练集包含 2、10、20 和 30 个样本。基于不同学习器的 TS 预测模型的准确率如图 5 所示。

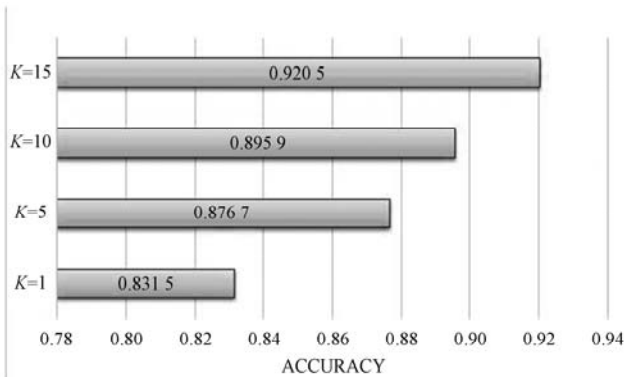


图 5 不同学习器性能

根据实验结果,可以发现 K 的数量超过 10 时学习器的性能有显著提升,因此在接下来的实验中,均采用 2-way 15-shot 设置进行学习器训练。

2.4 模型训练方法讨论

针对 TS 人脸标注数据不足的问题,该实验研究了三种模型训练方法:方法一,直接使用 TS 人脸数据训练 MRN;方法二,基于 ImageNet2012^[20] 和微调技术的

模型训练方法;方法三,基于小样本学习的模型训练方法(2-way 15-shot)。其中微调技术使用两步微调设置和全连接层微调设置。此外,还对元测试阶段中不同训练样本数量对 TS 预测模型的性能影响进行了研究。实验结果如图 6 所示。

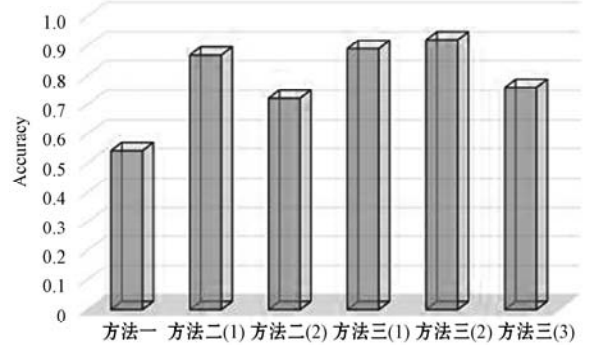


图 6 不同模型训练方法的性能比较

实验所用的 Batch Size 和 epoch 设置与 2.2 相同。

图 6 中(1)表示使用全连接层微调,(2)表示使用两步微调方法。方法一取得了最差的实验性能,主要原因是 TS 人脸标注数据较少不足以充分训练 MRN;方法二使用迁移学习中常用 ImageNet2012 数据集进行模型预训练,即从该数据集中随机抽取 2 个类别的样本,将每种类别的样本数量设置为 1 000 进行 MRN 预训练。该设置能保证方法二与方法三中每个 epoch 所用的样本数量保持相近,执行该过程五次以平均准确率作为方法二的性能指标。方法三(1)、(2)的性能要好于对应的方法二(1)、(2),说明基于 MAML 的模型训练方法能获得性能更好的 TS 预测模型,产生这种结果的原因可能是方法二的训练目标在于更好地预测当前目标任务,而 MAML 则更注重模型在新任务上的泛化性。在方法三中,微调策略(2)的性能好于微调策略(1)说明基于 MAML 的学习器在经过适当地微调之后能取得较好的实验性能。在方法二中,微调策略(1)好于微调策略(2)说明在数据不足的情况下对模型进行充分微调不能提升模型性能。

方法三(3)在元测试阶段使用 15-shot 的 TS 人脸图像进行模型训练。此实验的主要目的是研究经过不同数量的 TS 图像微调所获得的预测模型的性能。根据方法三(3)的实验结果,当使用较少数据进行训练时,TS 预测模型并不能取得理想的分类性能,主要原因是元训练阶段和元测试阶段所处理的任务分布不同。但是,根据方法三(1)、(2)和(3)的实验结果可知,基于 miniImageNet 的 MAML 训练方法可以提供较好的模型初始化参数。因此,对于 TS 预测模型的训练,方法三(2)的训练策略是一种更好的选择。

为了研究不同数据对 TS 预测模型的初始化影响,

在另一个常用的小样本学习数据集上进行了实验,即手写字符数据集 Omniglot^[21]。基于该数据集的预训练设置与方法三(2)保持一致,微调之后的模型准确率为0.850 7,显著低于基于 miniImageNet 数据集的 TS 预测模型,主要原因是该数据集的图像数据分布与 TS 图像数据分布差别较大,导致学习器的参数不够理想,即使借助充分微调也不能获得有效的 TS 预测模型。虽然 miniImageNet 和 TS 人脸数据也有一定的分布差异,但是借助它和 MAML 算法可以为 TS 预测模型提供一个较好的参数初始化。因此,基于 miniImageNet 的小样本学习方法将被用来训练 TS 预测模型。

2.5 注意力模块顺序影响

MRB 包含通道级注意力模块和尺度级注意力模块,但是这两个注意力模块的应用顺序并不是固定的,可以优先应用通道级注意力模块再应用尺度级注意力模块(通道+尺度),也可以是尺度级注意力模块结合通道级注意力模块的设置(尺度+通道)。需要注意的是,尺度+通道设置需要将最后的多尺度特征融合集成到通道级注意力模块尾部。基于不同设置的 MRB 构建的 TS 预测模型的性能如表 2 所示。

表 2 注意力模块设置性能分析

模块设置	Accuracy	Sensitivity	Specificity
通道+尺度	0.920 5	0.833 3	0.935 7
尺度+通道	0.904 1	0.824 1	0.918 0

根据表 3,基于通道+尺度设置的 TS 预测模型取得了更好的 TS 分类性能。因此,在所提出的 MRB 中使用通道级注意力模块结合尺度级注意力模块的设置方式。

2.6 模型比较

该实验将本文所提出的 TS 预测模型与最近发表的基于人脸图像数据的 TS 分类方法进行比较,对比结果如表 3 所示。

表 3 模型性能比较

方法	Accuracy	Sensitivity	Specificity
TS 模型	0.920 5	0.833 3	0.935 7
AlexNet ^[20]	0.902 7	0.814 8	0.918 0
ResNet-18 ^[12]	0.865 8	0.796 3	0.877 8
Duta 等 ^[15]	0.910 9	0.824 1	0.926 0
Song 等 ^[4]	0.846 0	0.803 0	0.889 0
Yao 等 ^[5]	0.832 0	0.831 0	0.824 0
Liu 等 ^[6]	0.849 5	0.823 5	0.892 4
Li 等 ^[7]	0.912 7	0.500	0.990 5

根据表 4,所提出的 TS 预测模型取得了最高的 Accuracy 和 Sensitivity, Specificity 略低于方法[7]。实验结果证明,结合小样本学习技术和深度学习算法能够提升 TS 的预测性能。此外,本文还将 MRN 与现有的经典 CNNs 和相关多尺度 CNNs 进行比较,即 AlexNet^[20]、ResNet-18^[12]以及基于金字塔卷积的 CNNs^[15],其中金字塔卷积采用多级多尺度卷积设置,即卷积核尺寸越小对应的分组数(级别)越小,且不同等级的卷积核数量设置也不同。本文以去掉注意力模块的 MRN 轻量级版本模拟基于金字塔卷积的 CNNs。实验结果显示,在相同的实验设置条件下,所提出的 TS 预测模型的性能要优于现有的经典 CNNs 和相关多尺度 CNNs。

3 结 语

为了提高 TS 的诊断准确率,基于 TS 人脸图像数据,提出一种基于小样本学习和多尺度残差网络的 TS 预测模型。首先,提出具有多级注意力机制的多尺度残差模块,其中,多尺度残差模块以应用多尺度分组卷积的残差结构实现,多级注意力机制包括通道级注意力模块和尺度级注意力模块,分别对单尺度特征图的通道关系和不同尺寸卷积核的重要性进行自动学习。然后基于该改进模块构建面向 TS 预测任务的多尺度残差网络。为了解决训练数据不足的问题,小样本学习技术被用来完成模型的训练。实验结果表明,所提出的模型优于现有 TS 分类方法。

参 考 文 献

- [1] Filho H P L. Applicability of data mining technique using Bayesians network in diagnosis of genetic diseases[J]. International Journal of Advanced Computer Science and Applications, 2013, 4(1): 47-50.
- [2] Dumancic J, Kaic Z, Varga M L, et al. Characteristics of the craniofacial complex in Turner syndrome[J]. Archives of Oral Biology, 2010, 55(1): 81-88.
- [3] Catic A, Gurbeta L, Kurtovic-Kozaric A, et al. Application of neural networks for classification of Patau, Edwards, down, turner and Klinefelter syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics [J]. BMC Medical Genomics, 2018, 11(1): 1-12.
- [4] Song W A, Lei Y, Chen S, et al. Multiple facial image features-based recognition for the automatic diagnosis of turner syndrome[J]. Computers in Industry, 2018, 100: 85-95.
- [5] Yao G H, Li J Q, Pei Y, et al. An automatic turner syndrome identification system with facial images [C]//2019

- International Conference on Frontier Computing. Springer, 2019:105 – 112.
- [6] Liu L, Sun J, Li J, et al. Automatic classification of Turner syndrome using unsupervised feature learning [C]//2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2020: 1578 – 1583.
- [7] Li J Q, Liu L, Sun J C, et al. Diagnosis and knowledge discovery of Turner syndrome based on facial images using machine learning methods [J]. IEEE Access, 2020, 8: 214866 – 214881.
- [8] Kim D, Jun T J, Eom Y, et al. Tournament based ranking CNN for the cataract grading [C]//41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019: 1630 – 1636.
- [9] Kayalibay B, Jensen G W, Smagt P V D. CNN-based segmentation of medical imaging data [EB]. arXiv: 1701.03056, 2017.
- [10] Moeskops P, Wolterink J M, Velden B H M, et al. Deep learning for multi-task medical image segmentation in multiple modalities [C]//International Conference on Medical Image Computing and Computer-assisted Intervention, 2016: 478 – 486.
- [11] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation [C]//4th International Conference on 3D Vision, 2016: 565 – 571.
- [12] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [13] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks [C]//International Conference on Machine Learning, 2017: 1126 – 1135.
- [14] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 – 9.
- [15] Duta I C, Liu L, Zhu F, et al. Pyramidal convolution: Re-thinking convolutional neural networks for visual recognition [EB]. arXiv: 2006.11538, 2020.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
- [17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.
- [18] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning [J]. Advances in Neural Information Processing Systems, 2016, 29: 3630 – 3638.
- [19] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23 (10): 1499 – 1503.
- [20] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097 – 1105.
- [21] Lake B M, Salakhutdinov R, Tenenbaum J B. The Omniglot challenge: A 3-year progress report [J]. Current Opinion in Behavioral Sciences, 2019, 29: 97 – 104.
- ~~~~~
- (上接第 181 页)**
- [18] 石跃祥, 曾智超. 基于特征传播的时域分割网络行为识别 [J]. 计算机辅助设计与图形学学报, 2020, 32 (4): 582 – 589.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB]. arXiv: 1409.1556, 2014.
- [20] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 – 9.
- [21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [22] Zolfaghari M, Singh K, Brox T. ECO: Efficient convolutional network for online video understanding [C]//European Conference on Computer Vision, 2018: 713 – 730.
- [23] Lin J, Gan C, Han S. TSM: Temporal shift module for efficient video understanding [C]//2019 IEEE International Conference on Computer Vision, 2019: 7082 – 7092.
- [24] Shahroudy A, Liu J, Ng T, et al. NTU RGB + D: A large scale dataset for 3D human activity analysis [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1010 – 1019.
- [25] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [EB]. arXiv: 1502.03167, 2015.
- [26] Farnebeck G. Two-frame motion estimation based on polynomial expansion [C]//13th Scandinavian Conference on Image Analysis, 2003: 363 – 370.
- [27] Zolfaghari M, Oliveira G, Sedaghat N, et al. Chained multi-stream networks exploiting pose, motion, and appearance for action classification and detection [C]//2017 IEEE International Conference on Computer Vision, 2017: 2923 – 2932.
- [28] Wang P, Li W, Wan J, et al. Cooperative training of deep aggregation networks for RGB-D action recognition [C]//32nd AAAI Conference on Artificial Intelligence, 2018: 7404 – 7411.
- [29] Wang X, Gupta A. Videos as space-time region graphs [C]//European Conference on Computer Vision, 2018: 413 – 431.