

基于 BCNN 的胎儿颅脑超声横切面识别算法

束鑫¹ 常锋¹ 张歆² 杜睿² 余转²

1 江苏科技大学计算机学院 江苏 镇江 212003

2 江苏大学附属医院超声医学科 江苏 镇江 212003

摘要 孕期超声检查是评估胎儿大脑发育、检测生长异常的重要步骤,开展对胎儿早期检查准确高效的诊断研究具有重要的临床价值。文中使用双线性卷积神经网络 BCNN 进行胎儿颅脑横切面识别,提出了 BCNN-R 和 BCNN-S 两种算法。BCNN 算法首先对输入的胎儿颅脑超声影像数据进行预处理,去除个人信息等敏感信息;其次,利用两路并行的子网络从影像数据中提取辨识度高、鲁棒性强的横切面特征,并将其融合得到有助于识别的细微特征;最后使用线性连接层进行识别和分类。为了验证算法的有效性,在自建胎儿超声数据集 JFU19 上进行了对比实验,实验结果表明,所提算法相比常见的深度网络(GoogleNet, DenseNet, SeNet 等)在分类性能上有明显的提升,其中 BCNN-S 算法的总体准确率达到 88.95%,BCNN-R 在水平横切面的识别上达到了 97.22% 的精确度和 88.61% 的召回率。此外,在公开数据集 HC18 上进行了实验,BCNN 算法的准确率、精确度、召回率分别达到了 89.48%,87.66% 和 87.71%,进一步验证了算法的有效性。

关键词: 胎儿超声; 颅脑图像; 横切面识别; 深度学习; BCNN

中图分类号 TP391

Transverse Section Recognition Algorithm Based on BCNN for Fetal Craniocerebral Ultrasound

SHU Xin¹, CHANG Feng¹, ZHANG Xin², DU Rui² and YU Zhuan²

1 School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China

2 Department of Medical Ultrasound, Affiliated Hospital of Jiangsu University, Zhenjiang, Jiangsu 212003, China

Abstract Ultrasound examination during pregnancy is an important step to discover whether the fetus is abnormal, so it is of great value to carry out accurate and efficient examination and diagnosis on fetus. In this paper, bilinear convolutional neural network(BCNN) is used to identify the transverse section of the fetal head. On this basis, we propose BCNN-R and BCNN-S. The BCNN model takes the fetal craniocerebral ultrasound image as input, and firstly preprocesses the input data. Secondly, the two parallel sub-networks can extract the cross-sectional features with high identification and strong robustness from the input, after that the model integrates the extracted features, which is helpful to extract the fine features for recognition. Finally, the linear layer outputs the result of classification. In order to verify the effectiveness of the proposed algorithm, this paper makes contrast experiment on self-built fetal ultrasound dataset JFU19. The experimental results show that the proposed algorithms have obvious improvement on classification performance compared with the basic network(GoogleNet, DenseNet, SeNet, etc.), the overall accuracy of BCNN-S reaches 88.95%, and the precision and recall of BCNN-R in horizontal cross-sections achieves 97.22% and 88.61%. In addition, we also use the public dataset HC18 to conduct classification experiments. The accuracy, precision and recall of BCNN reach 89.48%, 87.66% and 87.71% respectively, which further verifies the effectiveness of the proposed algorithms.

Keywords Fetal ultrasound, Craniocerebrum image, Identification of transverse section, Deep learning, BCNN

1 引言

超声是一种低成本、实时、无创的成像方法,被广泛应用于产前胎儿结构畸形的检查。在胎儿超声的常规检查中,最重要的就是在胎儿标准切面对相关组织结构进行测量^[1]。然而,超声图像容易受到噪声、阴影和混响的影响,很难对其进行准确识别,需要有丰富经验的超声医师来获取和解释图

像^[2]。为了减轻超声医师的工作负担,使用人工智能技术辅助诊断显得尤为重要。

机器学习是从人工智能中建立发展起来的一种方法,是一种跨学科领域,旨在构建能够从数据中学习和预测结果的算法^[3]。随着高分辨率超声成像设备越来越多地应用于临床,积累了大量的胎儿超声影像数据,使得机器学习在胎儿超声上的应用变得可行^[4]。其中,最常见的应用是利用低层特

收稿日期:2020-05-13 返修日期:2020-09-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61876072);江苏省妇幼健康科研面上项目(F201822);镇江市重点研发计划(社会发展)项目(SH2019038)

This work was supported by the National Natural Science Foundation of China(61876072), Maternal and Child Health Research Project of Jiangsu Province(F201822) and Key Research and Development Project of Zhenjiang(Social Development)(SH2019038).

通信作者:束鑫(shuxin@just.edu.cn)

征^[5](如 SIFT, Haar 和 HoG 等特征)作为图像特征描述符来表达图像。Ojala 等^[6]使用局部二值模式(Local Binary Patterns, LBP)算子和直方图统计算法对医学图像进行识别分类。Song 等^[7]应用 HOG-LBP 特征和最大最小后验概率算法(Max-Min posterior Pseudo-probabilities, MMP)学习识别肺部病变的影像学特征。随后,研究人员对低层特征进行进一步编码来改善识别性能,常见的编码方法有视觉词袋(Bag of Visual Word, BoVW)、局部特征聚合描述符(Vector of Locally Aggregated Descriptors, VLAD),以及 Fisher 向量(Fisher Vector, FV)等。Lei 等^[8-9]提出利用 Fisher 向量识别胎儿颜面部标准平面,随后又提出利用多层 Fisher 向量进行胎盘成熟自动分期的判别学习。种种研究表明,计算机辅助诊断作用明显。然而,随着医疗数据的日积月累,手动选择特征的传统机器学习算法逐渐暴露出其局限性;特征选择的优劣直接影响着分类器的分类精度。

与此同时,大规模数据集 ImageNet 的出现和具有极强表达能力的深度卷积网络(Deep Convolution Neural Network, DCNN)的发展,使得深度学习在图像处理领域取得了巨大成功,并迅速推广至医学图像处理领域。基于深度学习的图像处理技术通过叠加多个卷积层来提取图像特征,并使用多层感知器对其进行分类。在没有明确表达式的情况下,通过损失函数和反向传播算法对深度学习模型进行训练^[10]。其端到端的学习模式允许网络自动提取特征,使研究人员专注于数据的预处理和网络模型的设计。在网络的初始化方面,Chen 等^[11]首先提出使用基于迁移学习的训练策略来定位胎儿标准面,这为后续研究深度学习在胎儿超声上的应用奠定了基础。Yu 等^[12]提出使用深度学习识别胎儿颜面部超声标准切面,实验结果表明在胎儿超声应用上深度网络模型比传统机器学习算法有更强的表现力。Yaqub 等^[2]使用临床标准约束深度网络实现自动检测胎儿神经超声平面,依据临床检查诊断的顺序,首先定位胎儿的颅脑,然后检查其是否对称,最后定位透明隔腔,进行相关指标的测量。然而,该研究所使用的数据样本均为胎儿颅脑水平横切面,需要医师在检查过程中花费大量精力来寻找水平横切面,因此拥有一种能够准确识别胎儿颅脑横切面是否为标准水平方位的算法显得尤为重要。

针对上述问题,本文提出了两种基于双线性卷积神经网络(Bilinear Convolution Neural Network, BCNN)^[13]的胎儿颅脑超声横切面识别算法,即 BCNN-S 和 BCNN-R。首先,对输入影像进行预处理,包括清除隐私信息、裁剪感兴趣的区域以及数据增强操作;然后利用两路并行的子网络(SqueezeNet, ResNet)提取鲁棒性强、辨识度高的横切面特征,并进行特征融合,得到有利于识别的细微特征;最后,使用线性连接层对融合后的特征进行识别和分类。通过交叉熵损失函数训练学习,两路子网络可以相互监督,在提取细微特征的同时避免过拟合,最终无需手动选择特征,实现胎儿颅脑超声横切面水平方向的精准识别。

2 BCNN

本文的主要任务是对胎儿颅脑超声横切面进行识别。本文提出的基于双线性卷积神经网络(Bilinear Convolution Neural Network, BCNN)的胎儿颅脑超声横切面识别算法(以

下称简 BCNN)的结构如图 1 所示。其创新点在于:1)利用两路并行的子网络同时提取输入影像中辨识度高的横切面特征;2)将提取到的特征融合,得到有助于识别的细微特征;3)通过优化算法最小化损失,两路子网络可以相互监督,在不花费大量时间调节参数的情况下就能提高网络的性能,得到准确的识别结果。下面将从网络初始化、数据的获取及预处理、BCNN 网络设计等方面详细介绍本文方法。

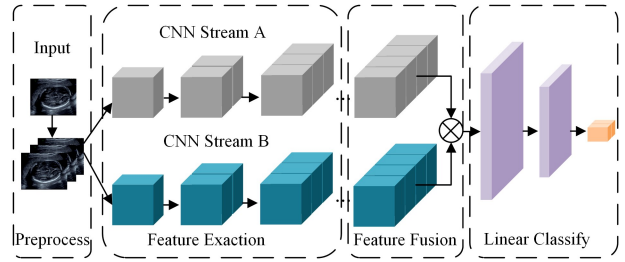


图 1 BCNN 网络结构

Fig. 1 BCNN network structure

2.1 网络初始化

网络初始化对后期的训练有很大影响,糟糕的初始化容易让神经网络陷入梯度消失的陷阱中。Yosinski 等^[14]指出低层 CNN 学到的特征类似于 Gabor 核滤波得到的掩膜边缘部分,随着深度的增长,CNN 逐渐学到物体的鲜明特征,越是靠近输入端的特征越是通用,靠近输出端的高层特征才包含指定任务的具体信息^[15]。因此,很多研究者尝试使用基于迁移学习的训练策略微调模型分类层的参数来应对不同的任务场景。针对本研究中有标签的胎儿颅脑超声横切面数据样本较少、训练容易出现过拟合的情况,采取迁移学习的策略,使用 ImageNet 预训练模型初始化 BCNN 模型的子网络,其余参数则使用 Kaiming^[16]初始化的方式进行初始化。Kaiming 初始化使得训练中的输出更接近高斯分布,有利于网络的收敛。在训练过程中分别对迁移学习参数层和 Kaiming 初始化参数层设置不同的学习率:迁移学习参数层的学习率设置为 0.1;Kaiming 初始化参数层的学习率设置为 1,并随着模型迭代次数的增多而逐渐减小,以起到微调的作用。该策略能让网络模型最大程度地学习目标数据的分布情况。

2.2 数据的获取及预处理

深度学习是数据驱动的算法,需要大量标记过的数据来训练网络模型,然而受到数据隐私、样本采集和标注困难等诸多因素的影响,大规模医学图像数据集的构建困难重重。因此,对训练样本进行增强是必不可少的操作,通过数据增强扩充数据集,可以有效地缓解因数据不足而无法训练出有效模型的问题。

本文使用的实验数据来自江苏大学附属医院超声科自主搭建的数据集(Jiangbin Fetal Ultrasound database),即 JFU19 数据集,共包含 2113 张胎儿颅脑横切面超声影像,由 GE Voluson E10 全身彩色多普勒超声诊断仪在临床中采集得到,实验以 8:2 的比例将 1690 张影像作为训练集、423 张影像作为测试集。本研究获得了医院伦理委员会的同意,根据《赫尔辛基宣言》的原则,患者标识符被移除,所有数据都是匿名的^[17]。数据标签由两位 10 年以上临床工作经验的医师标注,由一位超过 20 年工龄的主任医师进行复核,数据集包含 3 类胎儿颅脑超声横切面图像,分别是水平横切面(超声声束方向垂直于脑中线)、垂直横切面(超声声束方向平行于脑

中线)和斜向横切面(超声声束方向成角于脑中线)。

数据预处理及增强的步骤如下:

(1)为了保持训练样本中胎儿水平横切面与其他横切面数量上的均衡,实验前手动调整了各个横切面的样本数量。首先随机选择部分水平横切面数据样本,将其副本旋转 90° , 添加到垂直横切面的训练数据集;然后随机选择部分垂直和斜向横切面的数据样本,将其副本旋转 180° 后放回原处,共增加 689 份样本。

(2)ImageNet 预训练模型要求输入数据是 3 通道的彩色图,而超声影像是单通道的灰度图,因此实验中把单张灰度图复制 2 份,添加到另外 2 个通道组成 3 通道图像。

(3)为了降低服务器计算时的 GPU 显存消耗,实验中所有输入影像均缩放为 448×448 像素,并在网络加载数据时进行以下数据增强操作:首先进行了包括水平翻转、垂直翻转等增加样本量且不改变数据标签的操作;其次,任意的缩放图片会丢失很多高频信息,而高频信息对于目标识别非常重要,考虑到影像数据中胎儿不一定处于图像的中心位置,实验中单独对每张影像进行五角裁剪,裁剪出影像的左上、左下、右上、右下及中心块部分,其标签与原图保持一致。图 2 给出了部分示例图。

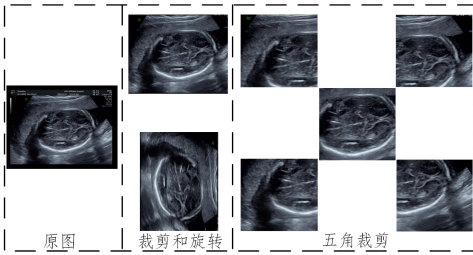


图 2 数据预处理示例图

Fig. 2 Examples of data preprocess

2.3 BCNN 网络

在本文构建的 BCNN 胎儿颅脑超声横切面识别算法中,特征提取部分由两路并行的 CNN 子网络组成,分别使用 ResNet 和 SqueezeNet 作为子网络的骨干网,在下文中称为 BCNN-R 和 BCNN-S。用四项式 $\beta=(\mathbf{F}(A), \mathbf{F}(B), P, C)$ 表示整个模型,其中 F 表示网络的特征提取层, $\mathbf{F}(A)$ 和 $\mathbf{F}(B)$ 分别表示从两路子网络提取到的特征, P 表示外积池化函数, C 是分类函数。特征提取层 F 的功能是把输入图片 \mathcal{L} 上的一个点 l 映射为一个 $J \times M$ 维的特征,经过 P 函数池化后的特征由式(1)给出,分类函数 C 使用式(2)中的 Softmax 函数输出识别出的类型,其中 O_i 是对第 i 类的输出。

$$(l, \mathcal{L}, \mathbf{F}(A), \mathbf{F}(B)) = \mathbf{F}(A)(l, \mathcal{L})^T \mathbf{F}(B)(l, \mathcal{L}) \quad (1)$$

$$C_i = \frac{\exp(O_i)}{\sum_i \exp(O_i)} \quad (2)$$

2.3.1 特征提取模块

BCNN-R 以 ResNet^[18] 为子网络的骨干网,主要使用残差结构,其结构如图 3 所示。令 x 表示输入特征, $F(x)$ 为卷积后得到的特征,反向传播优化算法需要计算出其梯度值 $F'(x)$,然而当深层网络模型任意一层特征的梯度 $F'(x)$ 为 0 时,根据链式求导法则,整个网络的梯度都变为 0,导致模型无法继续训练,而残差结构输出特征 $F(x) + x$ 的梯度为 $F'(x) + 1$,避免了 $F'(x)$ 为 0 导致的梯度消失问题。BCNN-

R 还使用 Leaky ReLU^[19] 作为激活函数,相比 ReLU 激活函数,当输入特征小于 0 时其可以保证梯度不为 0,从而避免了梯度消失。因此,BCNN-R 可以比其他网络更深,学到更丰富的特征。

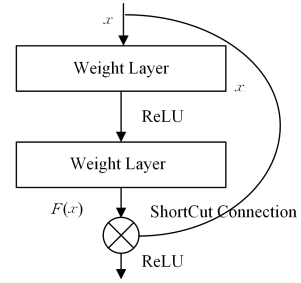


图 3 残差模块

Fig. 3 Residual module

BCNN-S 选择 SqueezeNet^[20] 作为骨干网,主要使用了 Fire 模块,由压缩和扩展两部分构成。压缩模块将多个维度的特征融合到更少的维度,首先使用 3×3 的卷积核对输入图像进行卷积,得到通道数较多的特征图;然后使用较少的 1×1 卷积核对上一层的特征图进行卷积,跨通道整合特征,达到减少特征维数的目的。扩展模块采用类似 Inception 网络^[21] 的设计,从网络宽度上进行扩展,分别使用 1×1 和 3×3 的卷积核卷积,将得到的特征图级联,增加模型的线性表示,提升泛化能力。Fire 模块的结构如图 4 所示。

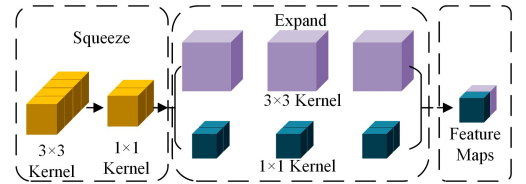


图 4 Fire 模块结构图

Fig. 4 Structure of Fire module

2.3.2 特征融合模块

在目标识别任务中,常用的分类模型只有单支网络,由卷积层、池化层和全连接层组成。池化层主要通过下采样将输入数据降维,概括全局特征以便进行全连接层分类,这样一方面可以节省训练时的存储开销,另一方面可以避免全连接层的过拟合。但是,下采样会丢失重要的特征,增加算法出错的几率,因此本文算法中的特征融合模块将两路分支特征 $\mathbf{F}(A)$ 和 $\mathbf{F}(B)$ 通过外积的方式进行融合来代替池化层。其中, $\mathbf{F}(A) = (a_{k1} \cdots a_{kj})^T$, $\mathbf{F}(B) = (b_{k1} \cdots b_{kj})^T$, $k \in [1, m]$, J 和 M 代表特征图的维数。 $\mathbf{F}(A)$ 和 $\mathbf{F}(B)$ 融合后得到双线性特征 $\mathbf{X} = \mathbf{F}(A)^T \mathbf{F}(B)$, 大小为 $M \times M$, 将 \mathbf{X} 转化成 $M^2 \times 1$ 的双线性向量,以便使用线性层进行分类。直观地说,双线性形式允许特征 $\mathbf{F}(A)$ 和 $\mathbf{F}(B)$ 相互制约,外积可以计算出它们组合的所有形式,类似完全平方公式的展开。特征融合公式如式(3)所示:

$$\begin{aligned} \mathbf{F}(A)^T \mathbf{F}(B) &= \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} (b_1 \cdots b_m) \\ &= \begin{pmatrix} a_{11} & \cdots & a_{j1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{jm} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{j1} & \cdots & b_{jm} \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \mathbf{a}_1^T \mathbf{b}_1 & \cdots & \mathbf{a}_1^T \mathbf{b}_m \\ \vdots & \ddots & \vdots \\ \mathbf{a}_m^T \mathbf{b}_1 & \cdots & \mathbf{a}_m^T \mathbf{b}_m \end{pmatrix} \quad (3)$$

由于卷积神经网络是端到端的,即输入的是原始数据,输出的是预测结果,可以使用反向传播算法优化训练。令 $d\ell/dx$ 为损失函数 ℓ 对输入 x 的梯度,则 ℓ 分别对 $\mathbf{F}(A)$ 和 $\mathbf{F}(B)$ 进行链式求导,梯度如式(4)所示:

$$\frac{d\ell}{dA} = \mathbf{F}(B) \left(\frac{d\ell}{dx} \right)^T, \quad \frac{d\ell}{dB} = \mathbf{F}(A) \frac{d\ell}{dx} \quad (4)$$

从式(4)中可以发现 A 和 B 相互影响对方的梯度,即起到了监督的作用,因此模型能够充分学习不同类别间的细微差异特征来避免过拟合,即便没有特意地调整模型超参数,也能得到较好的识别效果。

3 实验结果与分析

3.1 实验数据

实验使用了自建数据集 JFU19 和公开数据集 HC18。JFU19 数据集经过预处理后共包含 2802 张超声影像,按照 8:2 的比例划分为训练集和测试集,训练过程中从训练集中随机抽取 20% 作为验证集来评估模型,取验证集上识别效果最好的模型用于测试。HC18 数据集^[22-23] 包含 1832 张超声影像。数据集的具体情况如表 1 所列。

表 1 训练和测试数据的分布情况

Table 1 Distribution of training set and test set

Dataset		horizontal	vertical	diagonal
JFU19	Train	958	289	994
	Test	236	80	245
HC18	Train	735	152	611
	Test	251	22	61

3.2 硬件配置

本实验所用的硬件平台如下:内存(RAM)为 128 GB,中央处理器(CPU)为 Intel © Xeon(R) Silver 4110 CPU @ 2.10 GHz × 32,图形处理器为(GPU)NVIDIA 1080Ti × 2;软件平台为 Ubuntu 16.04 操作系统,开发语言选用 Python 3.6,深度学习框架为 Pytorch-1.2.0。

3.3 评价指标

本文除了使用准确率(Accuracy)外,还将精确度(Precision)和召回率(Recall)作为性能评价指标。当正负样本比例不平衡时,这种评价更有效。各指标计算公式如下:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (7)$$

其中, TP 代表真阳性,即真值为正例预测也是正例的样本数; TN 代表真阴性,即真值为负例预测也是负例的样本数; FP 为假阳性,即真值为负例预测但为正例的样本数; FN 为假阴性,即真值为正例预测但为负例的样本数。

3.4 参数设置

BCNN-S 共包含 8 个 Fire 模块,BCNN-R 使用了 4 个残差模块,它们统一使用外积运算代替最后一个卷积层的池化层,并对特征进行 L2 范数归一化,最后由一层线性层输出分

类结果。具体参数信息如表 2 所列。训练前使用预训练模型初始化,先冻结特征提取层,复制 SqueezeNet 和 ResNet 除最后 3 层全连接层之外的所有卷积层参数至子网络中的对应层,运行 40 个 epoch 微调模型的分层;然后解冻特征提取层,与分类层共同微调 20 个 epoch。在图像处理过程中,图像的像素信息可以看作一种数据分布,减去均值可以移除共同部分,凸显个体,这极大地影响了超声图像的识别,因此训练时输入数据减去所有训练数据的平均值进行中心式归一化,以加速模型收敛。选择 SGD 随机梯度下降法作为优化算法,将初始学习率设为 1,每经过 20 个 epoch 学习率乘以 0.1,动量衰减为 0.9,权重衰减为 1×10^{-5} 。根据文献[13]复现出 BCNN-V,使用相同的超参数进行配置。

表 2 BCNN 网络的参数信息

Table 2 Parameters of BCNN network

BCNN-S		BCNN-R	
Layer name	Kernel	Layer name	Kernel
Conv1	$7 \times 7, 64, \text{stride } 2$	Conv1	$7 \times 7, 64, \text{stride } 2$
	$2 \times 2 \text{ max pool, stride } 2$		$3 \times 3 \text{ max pool, stride } 2$
Fire 1-3	$\begin{bmatrix} 1 \times 1, 3 \times 3, 128 \\ 1 \times 1, 3 \times 3, 128 \\ 1 \times 1, 3 \times 3, 256 \end{bmatrix}$	Block 1	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
	$2 \times 2 \text{ max pool, stride } 2$		$3 \times 3 \text{ max pool, stride } 2$
Fire 4-7	$\begin{bmatrix} 1 \times 1, 3 \times 3, 256 \\ 1 \times 1, 3 \times 3, 384 \\ 1 \times 1, 3 \times 3, 384 \\ 1 \times 1, 3 \times 3, 512 \end{bmatrix}$	Block 2	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
	$2 \times 2 \text{ max pool, stride } 2$		$3 \times 3 \text{ max pool, stride } 2$
Fire 8	$\begin{bmatrix} 1 \times 1, 3 \times 3, 512 \end{bmatrix}$	Block 3	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
	$2 \times 2 \text{ max pool, stride } 2$		$3 \times 3 \text{ max pool, stride } 2$
	$\begin{bmatrix} 1 \times 1, 3 \times 3, 512 \end{bmatrix}$	Block 4	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 6$
Pool		Outer Product	
Classify		—	

本文还使用常用的深度学习网络 GoogleNet^[24], SeNet^[25], DenseNet^[26] 作为对比,其同样使用预训练模型初始化,然后运行 20 个 epoch 整体微调模型,使用 Adam^[27] 算法作为模型学习算法,该算法在更新梯度时保证了梯度平滑、稳定地过渡,因此默认参数即可将模型训练至稳定水平。

3.5 实验结果

根据式(5)一式(7)在 JFU19 数据集上统计出各算法的评价指标,实验结果如表 3 所列。由表 3 可知,文献[24-26]中的基础深度网络性能一般,反映出目标类间差异小、类内差异大对深度网络的性能影响非常大,未经过特殊设计的基础网络只能学习到有限的特征,继续训练将面临过拟合的风险,而基于 BCNN 的模型在各个指标上都有明显的优势。

表 3 JFU19 数据集中各算法的性能对比

Table 3 Comparison of different algorithms on JFU19 database

(单位: %)

Model	Accuracy	Precision	Recall
GoogleNet ^[24]	80.64	77.06	82.11
SeNet ^[25]	80.21	92.64	73.79
DenseNet ^[26]	81.70	75.32	85.71
BCNN-V ^[13]	82.99	90.08	87.64
BCNN-R	85.74	97.22	88.61
BCNN-S	88.95	90.25	87.83

此外,本文分别提取了 BCNN-V, BCNN-R 和 BCNN-S 实验过程中的特征图以便分析算法性能好坏的原因,特征图如图 5 所示。其中,BCNN-V 提取到的特征相对平稳,而 BC-

NN-R 和 BCNN-S 算法得到的特征图有明显的激活区域,激活区域越多,说明提取到有利于网络识别分类的特征就越多。BCNN-R 在水平横切面上的精确度达到了 97.22%,相比 BCNN-S,是因为激活区域更多,所以得到了更多有辨识能力的特征;同时,由于超声影像多噪声、回响,BCNN-R 更复杂的结构不可避免地会获取到更多噪声,因此准确率低于 BCNN-S。

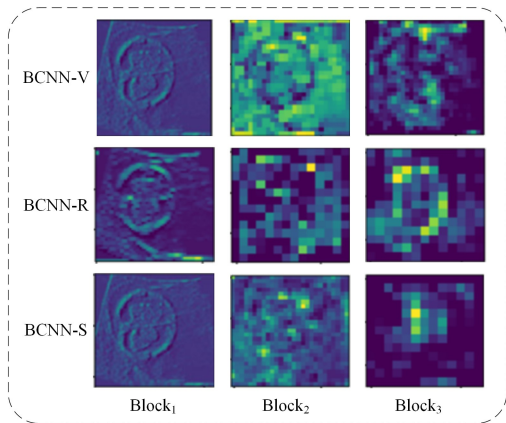


图5 BCNN-V, BCNN-R 和 BCNN-S 3种算法在 JFU19 上三阶段的部分特征图

Fig. 5 Part three-stage feature maps from BCNN-V, BCNN-R and BCNN-S on JFU19

本文还使用 HC18 数据集进行了实验,其结果如表 4 所列。表 4 中,BCNN-S 的准确率最高,为 89.48%;BCNN-R 的精确度最高,为 87.66%,而其召回率仅有 73.31%,与 BCNN-S 相差 15%,相比在 JFU19 数据集上两种算法的召回率相差 0.8%,这意味着图像质量对算法的影响非常大,召回率更低代表因噪声导致的识别错误更多,符合我们分析特征图得到的结果,更加验证了本文算法、实验的有效性。

表 4 HC18 数据集中各算法的性能对比

Table 4 Comparison of different algorithms on HC18 database

Model	Accuracy	Precision	Recall
GoogleNet ^[24]	70.96	76.37	72.67
SeNet ^[25]	71.05	54.37	73.66
DenseNet ^[26]	79.61	55.35	64.58
BCNN-V ^[13]	81.82	73.93	87.29
BCNN-R	83.24	87.66	73.31
BCNN-S	89.48	83.98	87.71

3.6 消融实验

(1)数据增强。本实验首先评估数据增强的情况。由于本实验的数据量相对常规图像数据集较小,在不使用数据增强以及迁移学习的情况下,庞大的深度模型会呈现极端不平衡的状况,即全 0 或者全 1,这样的结果没有对比的意义,故不详细说明。

(2)迁移学习。为了验证迁移学习在胎儿颅脑超声横切面识别中的作用,本文使用预训练模型对所有的实验网络进行了对比实验,结果如图 6 所示。根据 He^[28]的研究,预训练模型并不会使模型性能提高,但其研究的前提是有足够多的数据支撑,而本实验受制于数据集的规模,从头训练并不能将模型训练至最优,因此使用迁移学习加载预训练模型,其准确率可以提高 10%左右。本文算法 BCNN-S 和 BCNN-R 在未使用预训练模型时效果明显低于基础网络,其主要原因是从

头训练的模型在提取低层特征时更容易学习到影像中的噪声,并且双线性网络的外积操作对这一现象起到了放大作用,加剧了噪声对模型的干扰程度。

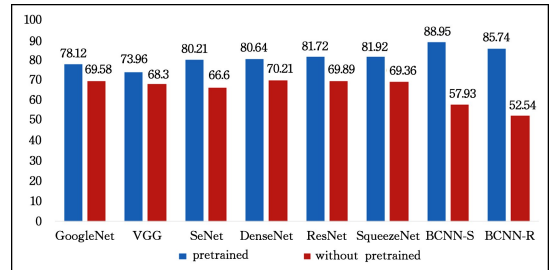


图6 是否使用预训练模型的实验效果对比

Fig. 6 Comparison between models with and without pretrained model

(3)BCNN 模块。在 JFU19 数据集上评估 VGG, ResNet 和 SqueezeNet 与使用了 BCNN 模块改善后 3 种网络的表现。如图 7 所示,在使用了数据增强和迁移学习的前提下,使用 BCNN 模块的网络性能均比未使用 BCNN 模块的网络性能提升 7%左右,进一步说明了本文使用 BCNN 模块构建的算法切实有效的。

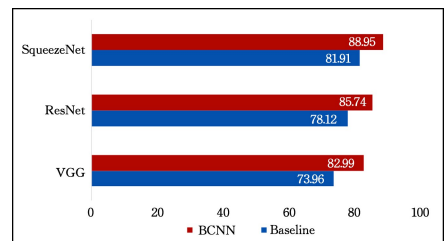


图7 是否使用 BCNN 模块的实验对比

Fig. 7 Comparison between models with and without BCNN module

结束语 本文利用 BCNN 模块和 SqueezeNet 网络、ResNet 网络构建了胎儿颅脑超声横切面识别算法,实现了横切面方向的自动检测。本文提出的胎儿颅脑超声横切面识别算法是基于胎儿超声检查中横切面影像所具有的共同特征构建的。一方面,使用超声影像数据作为算法的输入,利用两路子网络从中提取鲁棒的横切面特征,通过特征融合模块提取出细微的差异特征。另一方面,反向传播算法允许子网络之间互相监督,从而使 BCNN 网络的鲁棒性更强,辨识度更高。此外,在数据输入网络前对其进行了多种数据增强操作,并使用对原影像的五角裁剪来解决降采样导致的高层信息丢失问题。对于胎儿颅脑横切面的检测,本文算法的准确率和召回率分别达到了 88.95% 和 88.61%,在识别水平横切面的精确度上达到了 97.22%。在 HC18 数据集中,该算法也能取得较好的效果,其准确率、精确度和召回率分别达到了 89.48%、87.66% 和 87.71%。因此,深度学习结合临床标准的研究具有实际意义,值得进一步的探索。另一方面,虽然在实验数据集上模型的表现良好,但离投入临床使用还有一定的距离。考虑到临床实践中采集的数据多为动态,需要轻量级的模型来进行动态的胎儿识别检测;并且动态的影像中有更多的噪声干扰,需要更加鲁棒的模型支撑,这将是我们的下一步的研究方向。

参 考 文 献

- [1] VAN DEN HEUVEL T L A, PETROS H, SANTINI S, et al. Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries[J]. *Ultrasound in Medicine & Biology*, 2019, 45(3):773-785.
- [2] YAQUB M, KELLY B, PAPAGEORGHIU A T, et al. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints[J]. *Ultrasound in Medicine & Biology*, 2017, 43(12):2925-2933.
- [3] MURTY M N, DEVI V S. Introduction to pattern recognition and machine learning[M]. Co-Published with Indian Institute of Science(IISc), Bangalore, India, 2015.
- [4] SHAN J, ALAM S K, GARRA B, et al. Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods[J]. *Ultrasound in Medicine & Biology*, 2016, 42(4):980-988.
- [5] LEI B, ZHUO L, CHEN S, et al. Automatic recognition of fetal standard plane in ultrasound image[C]//2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). IEEE, 2014: 85-88.
- [6] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 24(7):971-987.
- [7] SONG L, LIU X, MA L, et al. Using HOG-LBP features and MMP learning to recognize imaging signs of lung lesions[C]//2012 25th IEEE International Symposium on Computer-Based Medical Systems(CBMS). IEEE, 2012:1-4.
- [8] LEI B, TAN E L, CHEN S, et al. Automatic recognition of fetal facial standard plane in ultrasound image via fisher vector[J]. *PloS one*, 2015, 10(5):e0121838.
- [9] LEI B, YAO Y, CHEN S, et al. Discriminative learning for automatic staging of placental maturity via multi-layer fisher vector[J]. *Scientific Reports*, 2015, 5(1):1-11.
- [10] MA Y, WU L, LI Z. A novel face presentation attack detection scheme based on multi-regional convolutional neural networks[J]. *Pattern Recognition Letters*, 2020, 131:261-267.
- [11] CHEN H, NI D, QIN J, et al. Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks[J]. *IEEE Journal of Biomedical and Health Informatics*, 2015, 19(5):1627-1636.
- [12] YU Z, WU L Y, NI D, et al. Automatic Recognition of Fetal Facial Ultrasound Standard Section based on Deep Learning[J]. *Chinese Journal of Biomedical Engineering*, 2017(3):267-275.
- [13] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear cnn models for fine-grained visual recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015:1449-1457.
- [14] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [C]//Advances in Neural Information Processing Systems. 2014:3320-3328.
- [15] MALLAT S. Understanding deep convolutional networks [J]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2016, 374 (2065): 20150203.
- [16] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C] // Proceedings of the IEEE International Conference on Computer Vision. 2015:1026-1034.
- [17] VAN DEN HEUVEL T L A, PETROS H, SANTINI S, et al. Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries[J]. *Ultrasound in Medicine & Biology*, 2019, 45(3):773-785.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [19] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C]//Proc. ICML. 2013.
- [20] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv:1602.07360, 2016.
- [21] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:1-9.
- [22] VAN D H T L A, DAGMAR D B, et al. Automated measurement of fetal head circumference using 2D ultrasound images [J]. *PloS One*, 13, 8(2018):e0200412.
- [23] VAN D H T L A, DAGMAR D B, et al. Automated measurement of fetal head circumference using 2D ultrasound images [DB/OL]. Zenodo. <http://doi.org/10.5281/zenodo.1322001>.
- [24] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:1-9.
- [25] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7132-7141.
- [26] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
- [27] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.
- [28] HE K, GIRSHICK R, DOLLAR P. Rethinking imagenet pre-training[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019:4918-4927.



SHU Xin, born in 1979. Ph.D, associate professor, is a member of China Computer Federation. His main research interests include medical image analysis, face recognition and pattern recognition.