

基于深度残差网络的皮肤镜图像黑色素瘤的识别

李 航 余 镇 倪 东 雷柏英 汪天富*

(深圳大学医学部生物医学工程学院, 医学超声关键技术国家地方联合工程实验室,
广东省生物医学信息检测和超声成像重点实验室, 广东 深圳 518060)

摘 要: 恶性黑色素瘤是最常见和最致命的皮肤癌之一。临床上, 皮肤镜检查是恶性黑色素瘤早期诊断的常规手段。但是人工检查费力、费时, 并且高度依赖于皮肤科医生的临床经验。因此, 研究出自动识别皮肤镜图像中的黑色素瘤算法显得尤为重要。提出一种皮肤镜图像自动评估的新框架, 利用深度学习方法, 使其在有限的训练数据下产生更具区分性的特征。具体来说, 首先在大规模自然图像数据集上预训练一个深度为 152 层的残差神经网络 (Res-152), 用来提取皮肤病变图像的中间卷积层特征, 并对其使用均值池化得到特征向量, 然后利用支持向量机 (SVM) 对提取的黑色素瘤特征进行分类。在公开的皮肤病变图像 ISBI 2016 挑战数据集中, 用所提出的方法对 248 幅黑色素瘤图像和 1 031 幅非黑色素瘤图像进行评估, 达到 86.28% 的准确率及 84.18% 的 AUC 值。同时, 为论证神经网络深度对分类结果的影响, 比较不同深度的模型框架。与现有使用传统手工特征的研究 (如基于密集采样 SIFT 描述符的词袋模型) 相比, 或仅从深层神经网络的全连接层提取特征进行分类的方法相比, 新方法能够产生区分性能更强的特征表达, 可以在有限的训练数据下解决黑色素瘤的类内差异大、黑色素瘤与非黑色素瘤之间的类间差异小的问题。

关键词: 皮肤镜检查图像; 黑色素瘤识别; 残差网络; 深度学习

中图分类号: R318 **文献标志码:** A **文章编号:** 0258-8021(2018) 03-0274-09

Melanoma Recognition in Dermoscopy Images via Deep Residual Network

Li Hang Yu Zhen Ni Dong Lei Baiying Wang Tianfu*

(Department of Biomedical Engineering, School of Medicine, Shenzhen University, National-Regional Key Technology Engineering
Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging,
Shenzhen 518060, Guangdong, China)

Abstract: Malignant melanoma is one of the most common and deadly skin cancers. Clinically, dermoscopy is a routine method for early diagnosis of malignant melanoma. However, human's visual examinations are laborious, time-consuming, and highly dependent on dermatologist's clinical experience. Therefore, it is important to design an algorithm for recognizing melanoma automatically in dermoscopy images. This study proposed a novel framework for the evaluation of dermoscopy images, using deep learning to generate more discriminative features with limited training data. Specifically, we first extracted the intermediate convolutional features of each skin lesion image using a very deep residual neural network including 152 network layers (i. e. Res-152) which was pre-trained on a large natural image dataset, and the final deep representation was obtained by averaging the spatial feature maps into single feature vector, then, the support vector machine (SVM) was used to classify the melanoma. By using the proposed method 248 melanoma images and 1031 non-melanoma images in published ISBI 2016 challenge datasets of skin lesion images were evaluated, obtaining accuracy rate of 84.96% and AUC of 84.18%. In addition, in order to demonstrate the effect of neural network depth on the classification results, we compared the different depth of the model framework. Our approach, which could solve large variations in melanoma and small differences between melanoma and non-

doi: 10.3969/j.issn.0258-8021.2018.03.003

收稿日期: 2017-07-21, 录用日期: 2018-03-22

基金项目: 深圳市基础研究项目 (JCYJ20150525092940986)

* 通信作者 (Corresponding author), E-mail: tfwang@szu.edu.cn

melanoma with the limited training data, can produce more discriminative representations than existing studies using hand-crafted features (i. e. the BoF models based on densely sampled SIFT (DSIFT) descriptors) or only to extract features from the fully connected layers.

Key words: dermoscopy image; melanoma recognition; residual network; deep learning

引言

黑色素瘤皮肤癌是世界上速度增长最快和最致命的癌症之一,占皮肤癌死亡病变的 75%^[1-2]。早期诊断对于其治疗非常重要,因为在早期阶段这种疾病很容易被治愈^[1-3]。为了改善皮肤细胞疾病的诊断效果,临床上引入皮肤镜检查以协助皮肤科医生进行诊断。作为一种无创性的皮肤成像技术,皮肤镜图像可以为临床医生提供高质量的皮肤病变信息。与传统的临床图像相比,皮肤镜检查图像具有更少的表面反射、更深层次的细节和更低的筛选误差,能够获得更好的可见度和识别精度。由于黑色素瘤皮肤癌比非黑色素瘤皮肤癌更致命,因此黑色素瘤与非黑色素瘤皮肤镜图像之间的区分识别已经引起了极大关注^[1-3]。目前,临床上提出了几种启发式的方法,如 Menzies 方法^[4]和 CASH 方法^[5],以增强临床医生辨别黑色素瘤与非黑色素瘤的能力。然而,即使对于经验丰富的专业人士,正确诊断黑色素瘤也是有一定困难的。此外,通过人眼目视检查进行的皮肤镜诊断通常耗时费力,且较为主观。因此,准确性不佳、可重复性差仍然是诊断这种疾病的主要难题。

为了解决这些问题,临床研究提出了许多用于自动分析皮肤镜图像算法。这些计算机辅助分析模型通常包括以下 4 个步骤:

- 1) 图像预处理,如去除毛发^[6-8]和图像增强^[9-10];
- 2) 边界检测或分割^[2,11];
- 3) 特征提取,即颜色、纹理、边界梯度、形状相关描述符^[2,12];
- 4) 分类 (k -最近邻 (KNN)), 支持向量机 (SVM)^[2],神经网络,AdaBoost^[12]。

大多数现有研究主要集中在特征工程和分类,或多或少都假设输入图像包含条件良好的病变对象。然而,皮肤镜图像可能并不总是包含整个病变区域,或者病变对象仅仅占据图像的很小部分,如图 1 所示。

由于黑色素瘤与非黑色素瘤(良性)之间的类内差异大和类间差异小(见图 1),所以这些人工特

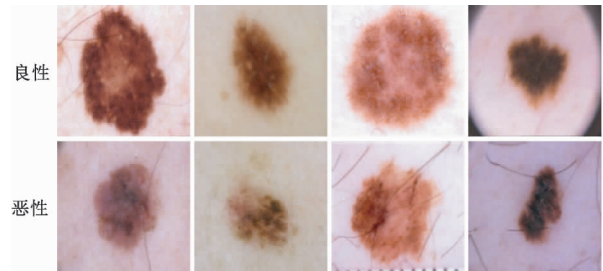


图 1 皮肤病变的皮肤镜检查图像。黑色素瘤与非黑色素瘤(良性)之间类内差异大、类间差异小

Fig. 1 Dermoscopy image of the skin lesions. Small variations between different classes and large variations in the same class between melanoma and non-melanoma (benign)

征所提供的诊断结果并不能令人满意。此外,这些基于手工特征的方法大多数包含复杂而繁琐的步骤,这导致临床实践中其泛化能力差,适用性低。与依靠人工特征的方法不同,深度卷积神经网络(CNN)等学习方法过去几年在图像识别任务中具有显著性的优势^[13-14]。CNN 的主要优点是根据给定的训练数据集,自动地学习相应检测或者分类任务所需的特征表达^[15],它在许多应用中都能达到最佳的表现性能^[16]。然而,CNN 需要大量的训练数据进行特征的学习,在医学图像领域,通常存在样本数据缺乏、采集和标注困难的情况。因此,其应用往往存在局限性。有研究^[17-18]表明,即使没有经过重新训练,迁移的深度卷积特征也可以用作通用视觉表达,比如在大型自然图像 ImageNet 数据集^[19]上预先训练的 CNN 特征,在其他很多图像识别任务中也取得了很好的结果。因此,近年来迁移的 CNN 特征也被应用于皮肤镜图像分类任务。一般情况下,深度表达主要从 CNN 模型的全连接(FC)层进行提取。虽然高层次的 CNN 特征具有良好的图像表达和泛化能力,但是这些深层描述符缺乏对局部图像特征的描述。在临床诊断中,局部纹理结构的信息对于皮肤细胞病变的识别至关重要。例如,黑色素瘤通常包含多种纹理结构(网状结构、条纹状结构、点状结构等),而良性皮肤细胞损伤则只包含一种纹理结构且呈均匀对称分布。

深度网络中间的卷积层特征可以看作是神经

元对于输入图像的局部感知结果,通常包含丰富的局部细节描述信息。基于此,本研究提出了一种基于深度卷积特征的黑素瘤图像识别方法。具体而言,首先将原图像统一调整尺度至 256 像素×256 像素大小,并对其进行预处理,包括图像归一化、数据增强等,然后将其输入到深度为 152 层的残差卷积神经网络(ResNet-152)中进行特征提取,其中最后一组残差块的输出特征被提取作为局部描述符,再将这些局部特征利用均值池化得到特征向量。在本实验中,使用的预训练卷积神经网络模型首先在大型自然图像数据集 ImageNet 上预训练,然后复制预训练网络中的参数作为网络中的初始化参数。最后,将得到的特

征向量用来训练一个线性支持向量机分类器,进行黑色素瘤与非黑色素瘤的分类。同时,对于所提出的框架,对比不同深度的网络架构以及数据增强技术对分类识别性能的影响。本研究基于公开的 ISBI 2016 皮肤数据集进行了大量的实验,实验结果证明了所提出方法的有效性,可以在有限的训练数据下更准确地识别黑色素瘤。

1 方法

本研究是对黑色素瘤与非黑色素瘤的二分类问题,所使用的方法包括残差神经网络、卷积特征提取过程、图像预处理和数据增强方法,整个框架的流程如图 2 所示。

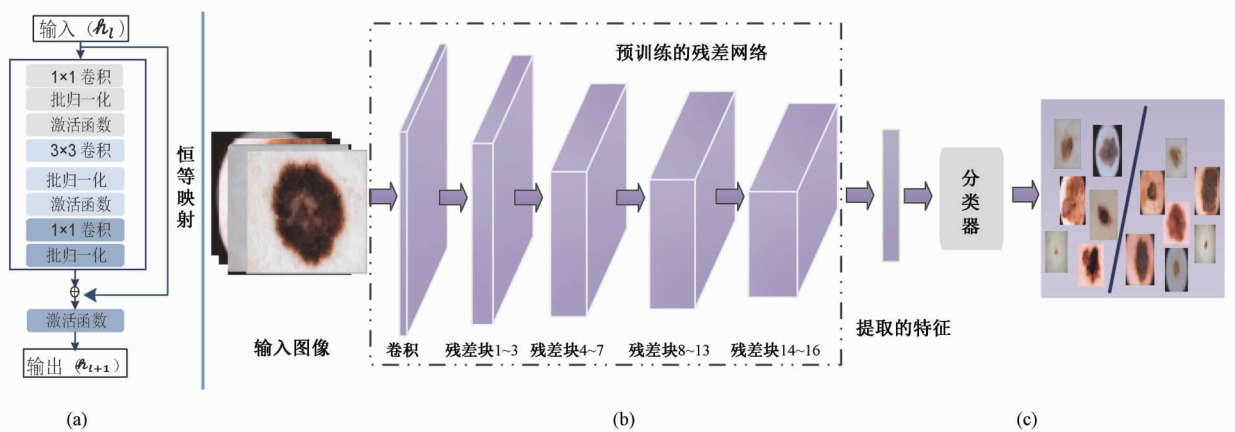


图 2 黑色素瘤识别框架的流程。(a) 残差块;(b) 深度特征提取,提取中间卷积层特征,并对其使用均值池化得到特征表达;(c) 支持向量机分类

Fig. 2 The flowchart of melanoma identification. (a) Illustration of residual block; (b) Extraction of deep representations, the feature vector is given by transforming the intermediate convolutional feature maps using averaging pooling; (c) SVM classification

1.1 深度残差神经网络

CNN 模型的深层结构对其特征学习能力至关重要。本研究采用了 He 等介绍的最新一代卷积神经网络(深度残差神经网络,ResNet-152)^[20],在 2016 年 ImageNet 大规模视觉识别挑战赛(ILSVRC 2016)中,残差网络取得了第一名的成绩。与经典的 CNN 结构相比,ResNet-152 的主要特点在于引入了残差连接,它能够在训练非常深的网络时解决其性能退化问题。已有研究证明,残差连接可以加速深度网络的收敛,并通过大幅度提高网络深度来保持准确率的提高。一般来说,深度残差网络由一组残差块组成,每个残差块包含几个堆叠的卷积层(将修正线性单元(Relu)和批归一化层作为卷积层附属)。具有恒等映射的残差块可以表示为

$$h_{l+1} = \text{Relu}(h_l + F(h_l, w_l)) \quad (1)$$

式中, h_l 和 h_{l+1} 分别是第 l 个残差块的输入和输出, $\text{Relu}(\cdot)$ 是修正线性单元函数, F 表示残差映射函数, w_l 是残差学习单元的参数。

具体来说,当 $F(h_l, w_l)$ 的通道(尺寸)和 h_l 不相等时,通常应用线性投影 φ 来匹配尺寸,因此式(1)可以进一步转换为

$$h_{l+1} = \text{Relu}(\varphi(h_l) + F(h_l, w_l)) \quad (2)$$

图 2(a) 表示了一个基本的残差块,可以参考文献[19]得到更多的细节。

在本研究中,利用了 ResNet-152 模型,并且残差网络在 ImageNet 上进行了预先训练。ResNet-152 和 ResNet-101 模型相比基本的 ResNet-50 结构有更多的残差块堆叠,详细信息如表 1 所示。另外,还比较了几种不同的 CNN 架构(即 VGGNet-16、AlexNet),具体

可参考第 2 部分。

表 1 本研究中使用的 Res-152 架构
Tab.1 Architecture of Res-152 used in this study

层数/残差块		参数
Conv1		7 × 7, 64
最大池化层, 3 × 3, 64		
Conv2_x	ResBlock1 - 3	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$
Conv3_x	ResBlock4 - 7	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 8$
Conv4_x	ResBlock8 - 13	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 36$
Conv5_x	ResBlock14 - 16	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$
平均池化层, 7 × 7, 2048		
全连接层, 分类器输出, 1 × 1, 1000		

1.2 图像预处理和数据增强

考虑到图像特征,图像预处理对深度表示的性能有很大的影响。

1) 图像大小。CNN 通常使用固定和方形大小的图像作为输入(如 227 × 227 或 224 × 224)。在常规情况下,将图像进行尺度大小的调整,使得所有图像具有相同的大小并且裁剪成所需的尺寸,以进行训练或特征提取。例如,在 AlexNet 中,在 CNN 训练期间,所有图像在进入网络之前都会被调整为 227 × 227,但这会导致皮肤病变对象发生形变,而形状轮廓信息是皮肤细胞损伤类型判别的重要依据。在这项研究中,对于皮肤数据集,将这些图像沿着最短边尺寸调整到一个统一的尺度(256),同时保持长宽比,从而保证皮肤病变部位的原有形状结构信息得到更好的保留。然后,将调整好尺度的图像输入到预训练的 CNN 中。

2) 图像归一化和增强。通常,在 CNN 处理之前,通过减去在整个训练数据集上计算出的平均像素值来对图像进行归一化。然而,数据集中的皮肤病变图像的光照、肤色和视角差异很大,减去统一的平均值并不能很好地标准化个体图像的光照,最近的研究也论证了这种影响。为了解决这个问题,通过减去在单个图像上计算出的通道平均强度值(表示为 per-channel-mean)来对每个皮肤图像进行归一化。具体地,给定一张皮肤图像 X , 计算归一化图像 X_{norm} 为

$$\begin{cases} X_{\text{norm},R} = X_R - u(X_R) \\ X_{\text{norm},G} = X_G - u(X_G) \\ X_{\text{norm},B} = X_B - u(X_B) \end{cases} \quad (3)$$

式中, $u(X_R)$ 、 $u(X_G)$ 、 $u(X_B)$ 为分别返回 3 个颜色通道的平均像素值。

虽然卷积神经网络具有极强的特征表达能力,但是训练数据量过少将会使 CNN 训练困难,容易过拟合,抑制 CNN 学习能力。因此,为了避免因有限的训练数据而引起的过拟合问题,采用了基于旋转和平移的实时数据增强技术。在训练期间增加数据,同时网络正在对 GPU 上的一大批数据进行训练,下一批数据将以多进程的方式在 CPU 上生成。具体来说,将每个缩放的图像旋转 4 个角度(0°, 90°, 180° 和 270°),然后在旋转的图像上随机添加像素平移(偏移在 -10 和 10 像素之间),得到一系列增强的子图像,实验结果见表 2。

表 2 CNN 识别结果

Tab.2 Recognition results with different CNN models(%)					
网络	Sen	Spec	mAP	Acc	Auc
AlexNet-noAug	34.67	95.39	53.42	83.38	79.79
VGG-16-noAug	36.00	94.41	53.42	82.25	76.76
Res-50-noAug	30.67	96.05	59.73	83.31	81.49
Res-101-noAug	37.33	95.39	57.13	83.91	80.48
Res-152-noAug	22.67	96.38	52.83	81.79	79.73
AlexNet-Aug	44.00	93.75	58.46	83.91	81.08
VGG-16-Aug	33.33	96.05	60.28	83.64	81.80
Res-50-Aug	40.00	96.05	63.90	84.96	82.25
Res-101-Aug	48.00	94.74	63.94	85.49	83.63
Res-152-Aug	44.00	96.71	65.64	86.28	84.18

1.3 卷积特征的提取

给定预训练的网络模型,首先通过上述操作来处理输入的皮肤病变图像。缩放图像的短边固定为 256(即 $S = 256$),对于第 i 张图像 X_i ,将其输入至预训练的 CNN 模型。在第 l 层卷积层中,得到 $w_{il} \times h_{il} \times d_l$ 空间特征图 M_{il} ,其中 w_{il} 和 h_{il} 分别表示宽度和高度, d_l 是当前特征图的深度或通道。为了方便描述, $N_{il} = w_{il} \times h_{il}$ 。值得注意的是,对于具有不同尺寸的输入图像,所得到的特征图的大小可以不同。在特征图 M_{il} 中,每个位置 $c = (c_x, c_y)$, $1 \leq c_x \leq w_{il}$ 的激活处,得到 d_l 维度向量 $f_{il,c} \in \mathbb{R}_{d_l}$,即本研究中的局部深度特征。因此,得到每个图像 X_i 的局部深度特征表达式为

$$F_{il} = \{f_{il,(1,1)}, \dots, f_{il,c}\} \in \mathbb{R}_{N_{il} \times d_l} \quad (4)$$

CNN 的不同卷积层往往具有不同的性质。一般而言,低层特征往往包含纹理形状等低层信息,

而高层特征则通过不断组合低层级特征得到更为抽象化的语义特征。在特征迁移的过程中,高层级特征往往和预训练的数据集有关,考虑到预训练数据集和皮肤数据之间的差异性以及深层特征的表达能力,在本研究中提取网络的最后卷积层特征。

最后,特征向量通过对提取的卷积特征利用均值池化进行计算,得到

$$\text{Fev}_i = \frac{1}{N_{il}} \sum_{c_x=1}^{w_{il}} \sum_{c_y=1}^{h_{il}} f_{li}(c_x, c_y) \quad (5)$$

本实验利用在大型自然图像 ImageNet 数据集上预训练的深度残差网络(Res-152)提取特征,并对其使用均值池化操作得到特征向量,然后用支持向量机建立分类模型,从而对黑色素瘤与非黑色素瘤进行自动识别分类。在实验中,采用平均精度(mAP)、准确度(Acc)、ROC 曲线下面积(AUC)、灵敏度(Sen)和特异性(Spec)评价分类性能。

1.4 实验

1.4.1 实验设置

1) 数据集。使用公开的皮肤病变图像 ISBI 2016 挑战数据集验证所提出的方法。该数据集基于国际皮肤成像协作(ISIC)档案[<https://isic-archive.com/>],是最大的、公开的皮肤镜图像数据集。2016 皮肤镜图像视觉竞赛发布的数据集包含了 1 279 幅皮肤病变图像,其中相应的类别标签被预分割成 900 幅训练集图像和 379 幅测试集图像。数据集中有两种病变类别:黑色素瘤和非黑色素瘤。其中,248 幅黑色素瘤图像(训练集中的 173 幅图像,测试集中的 75 幅图像)作为正样本,1 031 幅非黑色素瘤图像作为负样本。经过数据增强后,共有 28 800 幅图像作为训练集获得分类器模型。在实验中,SVM 分类器^[2]的超参数通过对训练数据使用 5 折交叉验证获得,其中惩罚因子 C 为 10,终止条件的允差采用默认值 0.001。

2) 评估和实验。对于评价的性能指标,采用平均精度(mAP)、准确度(Acc)、ROC 曲线下面积(AUC)、灵敏度(Sen)和特异性(Spec)。对于 3 种不同的卷积神经网络模型,将利用数据增强技术对数据集扩充所得到的分类结果分别表示为 AlexNet-Aug、VGG-16-Aug 和 Res-50-Aug、Res-101-Aug、Res-152-Aug,而没有利用数据增强技术得到的分类结果分别表示为 AlexNet-noAug、VGG-16-noAug 和 Res-50-noAug、Res-101-noAug、Res-152-noAug。

本实验主要是基于开源库来实现所提出的框架,即用于深度特征提取的 Matlab 工具包

MatConvNet。所有实验都在具有 CPU Inter XeonE5-2680 @ 2.70 GHz、GPU NVIDIA Quadro K4000 和 128GB RAM 的计算机上进行。

1.4.2 网络类型和卷积特征的实验

1) 不同网络架构对分类识别结果的影响。将提出的框架与其他基于卷积神经网络的方法进行比较。在本实验中,采用了 3 种不同的卷积神经网络模型,除了 152 层深度残差网络(Res-152)外,还对不同深度的其他几种 CNN 模型进行性能比较,包括 8 层的 AlexNet、16 层的 VGGNet(VGG-16)、50 层的 Res-50 以及 101 层的 Res-101。所有网络模型都在 ImageNet 上进行了预训练。采用相同的图像预处理以进行公平比较,用随机旋转、翻转和裁剪来增加每幅皮肤图像数据,最终产生了 32 个子图像。

2) 数据增强对分类结果的影响。对数据集进行了有无数据增强的对比实验。

2 结果

2.1 网络类型和卷积特征的实验结果

图 3 列出了不同深度卷积神经网络模型的比较结果,包括各种评价分类结果的指标对比,可以观察到网络架构的深度对识别黑色素瘤与非黑色素瘤之间的差异会产生很大的影响。从图 3 以及表 2 中可以看出,Res-152 相比其他几个浅层网络能提取判别力更强的特征,具有更好的识别黑色素瘤的性能。对于 AlexNet 和 VGG-16,利用第一个全连接层的激活映射来提取数据特征,作为图像表达。对于 Res-50、Res-101 以及 Res-152,利用了倒数第二层(即最后一组)卷积残差块的 2 048 个维度的输出来提取图像特征。由实验结果可以知道,从残差网络 Res-152 提取的深层特征优于从 AlexNet、VGG-16、Res-50 和 Res-101 中提取的深层特征。从表 2 和图 3 可以看出,在 5 种不同深度的卷积神经网络模型中,152 层的深度残差网络 Res-152 都具有最佳的识别分类性能表现。实验结果表明,利用的卷积神经网络架构越深,则提取出的图像特征表达会更具判别力。

对数据集进行有无数据增强的对比实验,其结果如表 2 所示。可以看出,在 3 种不同的网络架构中,没有进行数据增强的分类表现比进行了数据增强的分类结果表现都要差,具体来说,在 AlexNet 中进行了数据增强后,mAP 提升了大约 5%;在 VGG-16 中,mAP 增长了大约 8%;在 Res-50 中,mAP

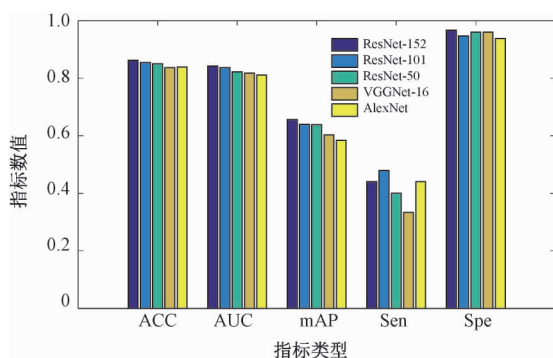


图3 不同网络架构的实验结果

Fig.3 The result with different network architecture

提升了大约 4.2%; 在 Res-101 中, mAP 提升了大约 6.8%; 在 Res-152 中, mAP 提升了大约 12.8%。由此说明, 随着神经网络深度的加深, 数据增强能够有效增加分类结果的准确性, 并且避免过拟合。

2.2 与其他方法的比较

将本研究的方法与对黑色素瘤分类识别的已有文献进行对比, 所采用的数据集同样是基于公开的皮肤病变图像 ISBI 2016 挑战数据集。

2.2.1 与传统手工特征方法的比较

SIFT 描述符是提取手工特征的常用描述符, 已被广泛应用。将基于密集采样 SIFT(DSIFT) 描述符的词袋(Bag of features, BoF) 模型与提出的框架进行比较, 结果如表 3 和图 4 所示。DSIFT-VQ 表示基本直方图量化, 即图像表示由视觉词的直方图给出。另外两种为基于 DSIFT 特征的不同编码方式, 包括局部特征聚合描述符编码(DSIFT-VLAD) 和 FV 向量编码(DSIFT-FV)。可以观察到, 当使用深度卷积特征替代 DSIFT 时, 实现了 mAP 的大约 10% 的改进, 这表明深度卷积特征比低级的手工特征描述符具有更强大的表达能力。

表3 与手工特征方法的对比结果

Tab.3 Comparision results with hand-crafted feature

方法	Sen	Spec	mAP	Acc	AUC
DSIFT-VQ	32.00	93.75	46.94	81.53	76.97
DSIFT-VLAD	32.00	94.41	51.51	82.06	78.53
DSIFT-FV	33.33	95.39	55.63	83.11	78.01
CNN-SVM (our)	44.00	96.71	65.64	86.28	84.18

2.2.2 与基于卷积神经网络方法的比较

表 4 列出了比较结果。对于 CNN-SVM 来说, 利用不同网络架构提取特征, 再通过 SVM 分类。而 CNNaug-SVM 为利用数据增强技术对数据集扩充, 再用 CNN 模型提取特征输入到 SVM 分类。可以看

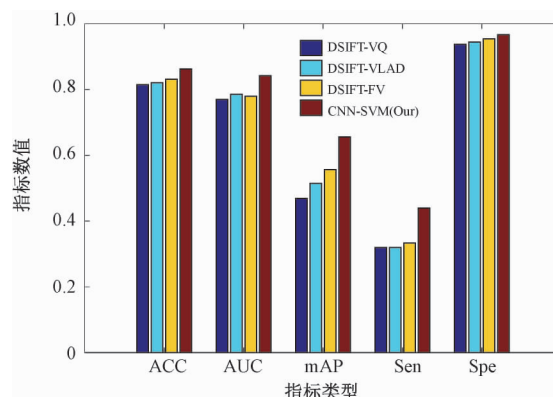


图4 与手工特征方法的对比结果

Fig.4 Comparision results with hand-crafted feature

到, 数据增强可以提高分类结果的准确度。同时, 将本研究的分类结果与最新出版的文献 [20] 和挑战赛中排名第一的方法 [21] 进行对比。Fine-tuned CNN 只对网络进行微调, 没有利用 SVM 分类。CUMED 分类方法虽然灵敏性高于本研究的方法, 但是其过程和计算的复杂度高, 采用的多级方案涉及额外的分段预处理和网络微调。如表 4 和图 5 所示, 本研究提出的框架在 mAP 中优于 Fine-tuned CNN、CUMED 及没有利用数据增强的 CNN-SVM 这 3 种方法, 分别增长约为 2.28%、1.94% 和 5.91%。

表4 与其他卷积神经网络方法的对比结果

Tab.4 Comparision results with other CNN methods

方法	网络	Sen	Spec	mAP	Acc	AUC
CUMED [21]	Res-50	50.70	94.10	63.70	85.50	80.40
Fine-tuned CNN [20]	Res-50	48.00	94.08	63.36	84.96	81.58
	AlexNet	34.67	95.39	53.42	83.38	79.79
	VGG-16	36.00	94.41	52.24	82.85	76.76
CNN-SVM [17]	Res-50	30.67	96.05	59.73	83.31	81.49
	Res-101	37.33	95.39	57.13	83.91	80.48
	Res-152	22.67	96.38	52.83	81.79	79.73
	AlexNet	44.00	93.75	58.46	83.91	81.08
	VGG-16	33.33	96.05	60.28	83.64	81.80
CNNaug-SVM (our)	Res-50	40.00	96.05	63.90	84.96	82.25
	Res-101	48.00	94.74	63.94	85.49	83.63
	Res-152	44.00	96.71	65.64	86.28	84.18

3 讨论

深度神经网络通过组合多层神经元, 提取不同层次的特征, 不断迭代组合成更高层次的抽象特征, 相比传统的手工特征具有更强的特征表达能力和泛化能力。本研究基于深度神经网络, 提出了一种新颖有效的方法, 用于从皮肤镜图像中自动准确地识别黑色素瘤, 并通过广泛的实验来研究其有效

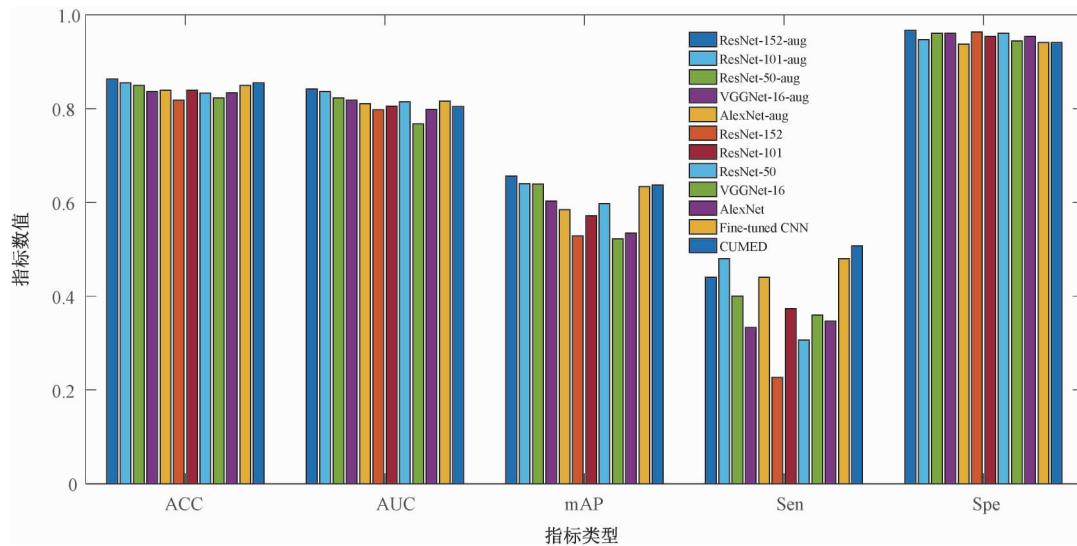


图 5 与其他卷积神经网络方法的对比结果

Fig.5 Comparision results with other CNN methods

性。该方法主要启发于深度网络的卷积层特征,通常包含丰富的区分性特征信息。例如,如图 6 所示,在预训练的残差网络 Res-152 中,从不同的残差块的输出提取 2D 特征图,可以看到提取的特征图清楚地保留了相应局部区域的描述信息。正如在许多文献中所研究的,特征图中的每个激活可以追溯到输入图像的某个局部区域(感受野)。对于卷积层,通常用成百上千的卷积核生成大量的激活图(特征图),以捕获图像的各种特征。因此,可以聚

合这些特征图的局部激活值作为特征向量,以构建比手工提取的描述符和直接从 CNN 全连接层提取的特征更具区分性的表达。与基于端到端训练或微调网络的现有方法相比,本研究的框架仅使用预先训练的 CNN 作为特征提取器,从而避免了训练过程的复杂性,因此在有限的训练样本下更为实用。以下几个因素应该注意:首先,应该考虑图像的预处理和归一化;其次,深度学习对训练数据有严格的要求,训练数据不足、需要学习的权重参数过

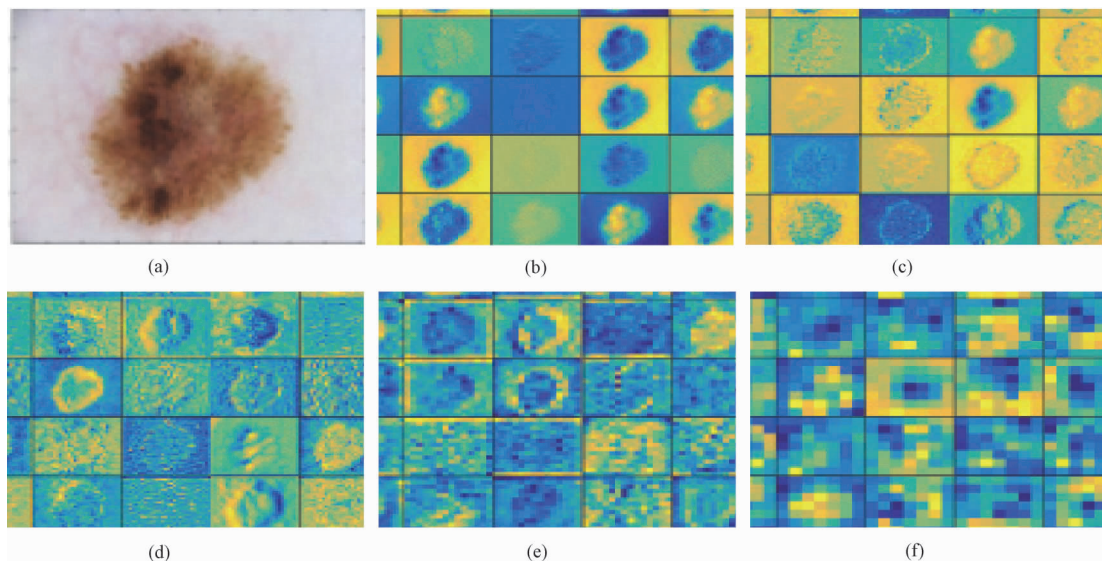


图 6 预训练的残差网络 Res-152 随机捕获一张图片的中间层的特征图。(a) 输入图片;(b)~(f) 残差网络的第 1~5 个残差块随机选取的 16 个通道的特征图

Fig.6 Illustration of Pre-trained residual network Res-152 randomly captures a picture of the intermediate layer of the feature map. (a) is the input image, (b)~(f) Randomly selected 16 channels of the feature map of the 1st~5th residual block

多会导致过拟合等问题。因此,选择适当的 CNN 架构,同时进行数据增强,对于提取深层卷积特征也是非常关键的。值得注意的是,本研究所提出的框架中,网络只对每个输入图像应用一次,然后从激活图中提取局部深层描述符。

本研究所提出的方法虽然取得了一定的成效,但仍然存在一些不足之处。首先,对于迁移问题,原始训练数据对目标任务有重大影响^[22]。例如,与其他数据集上预先训练的网络相比,在 ImageNet 上预先训练的网络可能会产生不同的结果。然而,笔者只调查了在 ImageNet 大型自然图像数据集上预先训练过的网络,没有更多地了解在其他数据集上预训练的网络结构对黑色素瘤的影响。其次,本研究提出的框架只是在 ISBI 皮肤数据集上进行测试,并没有在其他的数据集上进行验证,没有进一步验证所提出网络结构的鲁棒性和泛化能力。此外,实验结果中,灵敏度指标主要计算分类器能正确识别出黑色素瘤的结果,特异性是判断非黑色素瘤的指标,由于数据不均衡,黑色素瘤相比非黑色素瘤数据量差距较大,导致特异性相比灵敏性高出大约 50%。

4 结论

在本研究中,提出了一个用于黑色素瘤与非黑色素瘤皮肤镜图像分类的新框架。它从深度残差网络中提取局部卷积特征,形成更复杂的表达。笔者进行了大量的实验来研究可能影响本方法性能的关键因素,包括图像预处理、数据增强、网络架构。同时,在公开的 ISBI 2016 挑战皮肤病变数据集上显示结果。在未来的研究中,笔者将在更多的数据集上评估这次研究的方法,并提升其在临床实践中的应用。

参考文献

- [1] Kasmi R, Mokrani K. Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule [J]. *Iet Image Processing*, 2016, 10(6): 448-455.
- [2] Celebi M, Kingravi H, Uddin B, et al. A methodological approach to the classification of dermoscopy images [J]. *Computerized Medical Imaging & Graphics*, 2007, 31(6): 362-373.
- [3] Codella N, Cai Junjie, Abedini M, et al. Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images [C] // *Machine Learning in Medical Imaging: 6th International Workshop (MLMI 2015)*. Munich: Springer International Publishing, 2015: 118-126.
- [4] Stolz W, Reimann A, Cognetta A. Abcd rule of dermoscopy: a new practical method for early recognition of malignant melanoma [J]. *European Journal of Dermatology*, 1994, 4(7): 521-527.
- [5] Henning J, Dusza S, Wang S, et al. The cash (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy [J]. *Journal of the American Academy of Dermatology*, 2007, 56(1): 45-52.
- [6] Lee T, Ng V, Gallagher R, et al. Dullrazor: A software approach to hair removal from images [J]. *Computers in Biology & Medicine*, 1997, 27(6): 533-543.
- [7] Sáez A, Serrano C, Acha B. Model-based classification methods of global patterns in dermoscopic images [J]. *IEEE Transactions on Medical Imaging*, 2014, 33(5): 1137-1147.
- [8] Bi Lei, Kim J, Ahn E, et al. Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification [C] // *2016 IEEE 13th International Symposium on Biomedical Imaging*. Prague: IEEE, 2016: 1055-1058.
- [9] Iyatomi H, Celebi M, Schaefer G, et al. Automated color calibration method for dermoscopy images [J]. *Computerized Medical Imaging & Graphics*, 2011, 35(2): 89-98.
- [10] Schaefer G, Rajab M, Celebi M, et al. Colour and contrast enhancement for improved skin lesion segmentation [J]. *Computerized Medical Imaging & Graphics*, 2011, 35(2): 99-104.
- [11] Shimizu K, Iyatomi H, Celebi M, et al. Four-class classification of skin lesions with task decomposition strategy [J]. *IEEE Trans Biomed Eng*, 2015, 62(1): 274-283.
- [12] Capdehourat G, Corez A, Bazzano A, et al. Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions [J]. *Pattern Recognition Letters*, 2011, 32(16): 2187-2196.
- [13] Song Youyi, Zhang Ling, Chen Siping, et al. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning [J]. *IEEE Transactions on Biomedical Engineering*, 2015, 62(10): 2421-2433.
- [14] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [C] // *The 19th International Conference on Neural Information Processing Systems (ICONIP 2012)*. Doha: Curran Associate Inc, 2012: 1097-1105.
- [15] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [16] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C] // *The 18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2015)*. Munich: Springer International Publishing, 2015: 234-241.
- [17] Razavian A, Azizpour H, Sullivan J, et al. Cnn features off-the-shelf: an astounding baseline for recognition [C] // *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Columbus: IEEE Computer Society, 2014: 512-

- 519.
- [18] Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition [J]. 2013, 50(1): 1-647.
- [19] Deng Jia, Dong Wei, Socher R, et al. Imagenet: a large-scale hierarchical image database [C] //IEEE Computer Vision and Pattern Recognition (CVPR 2009) . Florida: IEEE Computer Society, 2009: 248-255.
- [20] He Kaiming, Zhang Xiangyu, Ren Shaoping, et al. Deep residual learning for image recognition [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016) . Las Vegas: IEEE Computer Society, 2016: 770-778.
- [21] Yu Lequan, Chen Hao, Dou Qi, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks [J]. IEEE Transactions on Medical Imaging, 2017, 36(4): 994-1004.
- [22] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [C] //Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014: 3320-3328.