

# 基于多模型特征与精简注意力融合的图像分类<sup>①</sup>



宋东情, 朱定局, 贺 超

(华南师范大学 计算机学院, 广州 510631)

通讯作者: 朱定局, E-mail: zhudingju@m.scnu.edu.cn

**摘 要:** 为了提高图像分类性能, 本文提出一种多模型特征和注意力模块融合的图像分类算法 (image classification algorithm based on Multi-model Feature and Reduced Attention fusion, MFRA). 通过多模型特征融合, 使网络学习输入图像不同层次的特征, 增加特征互补性, 提高特征提取能力; 通过加入注意力模块, 使网络更关注有目标的区域, 降低无关的背景干扰信息. 本文算法在 Cifar-10, Cifar-100, Caltech-101 这 3 个公开数据集上的大量实验对比, 验证了其有效性. 与现有算法对比, 本文算法的分类性能有较为明显的提升.

**关键词:** 多模型; 注意力机制; 图像分类; 特征融合; 深度学习

引用格式: 宋东情, 朱定局, 贺超. 基于多模型特征与精简注意力融合的图像分类. 计算机系统应用, 2021, 30(11): 210-216. <http://www.c-s-a.org.cn/1003-3254/8153.html>

## Image Classification Based on Multi-Model Feature and Reduced Attention Fusion

SONG Dong-Qing, ZHU Ding-Ju, HE Chao

(College of Computer Science, South China Normal University, Guangzhou 510631, China)

**Abstract:** To improve the performance of image classification, this paper proposes an image classification algorithm based on the fusion of Multi-model Feature and Reduced Attention (MFRA). Through multi-model feature fusion, the network can learn the features of different levels of input images, increase the complementarity of features and improve the ability of feature extraction. The introduction of the attention module makes the network pay more attention to the target area and reduces the irrelevant background interference information. In this paper, the effectiveness of the algorithm is verified by a large number of experimental comparisons on three public datasets, Cifar-10, Cifar-100 and Caltech-101. The classification performance of the proposed algorithm is significantly improved as compared with existing algorithms.

**Key words:** multi-model; attentional mechanism; image classification; feature fusion; deep learning

图像分类是计算机视觉领域最基本的任务之一, 其性能取决于算法和特征的选取. 早期图像分类方法主要基于人工提取特征, 如使用 SIFT<sup>[1]</sup> 和 HOG<sup>[2]</sup> 等特

征算子提取特征. 后来学者们开始采用传统机器学习算法, 如随机森林<sup>[3]</sup>, 支持向量机<sup>[4]</sup> 等进行分类, 但分类性能仍有提升空间.

① 基金项目: 中国高等教育学会专项课题 (2020JXD01); 广东高校省级重点平台和重大科研项目 (重大科研项目-特色创新类)(2017KTSCX048); 广东省中医药局科研项目 (20191411); 广东省普通高校“人工智能”重点领域专项 (2019KZDZX1027); 国家自然科学基金重点项目 (U1811263); 广州市大数据智能教育重点实验室 (201905010009); 广东省公益研究与能力建设 (2018B070714018)

Foundation item: Project of China Association of Higher Education (2020JXD01); Provincial Key Platforms and Major Scientific Research Projects of Guangdong Universities (Major Scientific Research Projects-Characteristic Innovation)(2017KTSCX048); Scientific Research Project of Bureau of Traditional Chinese Medicine Guangdong Province (20191411); Key Areas of “Artificial Intelligence” in Guangdong Universities (2019KZDZX1027); Key Program of National Natural Science Foundation of China (U1811263); Guangzhou Key Laboratory of Big Data Intelligent Education (201905010009); Guangdong Public Welfare Research and Capacity Building (2018b070714018)

收稿时间: 2021-01-26; 修改时间: 2021-02-24; 采用时间: 2021-03-03; csa 在线出版时间: 2021-10-22

2012年,随着 AlexNet 卷积神经网络<sup>[5]</sup>在图像分类领域的成功应用,深度卷积神经网络(DCNN)引起了人们高度关注。DCNN 的基本研究理念是开发先进的卷积神经网络架构以及相应的训练算法。随着 DCNN 的蓬勃发展,各种视觉任务的精度得到了极大提升<sup>[6-8]</sup>。其成功主要归功于深度架构和端到端的学习方法,该方法能够自动学习输入图像在不同层次的内部特征,从而大幅度提高分类精度。为了提取更有效的特征进行识别,近年来学者们不断提出先进的深度卷积网络架构和相应训练算法。

最近研究<sup>[9-11]</sup>表明加深网络深度和拓宽网络宽度能够提高卷积神经网络的性能。加深网络深度方面,He 等<sup>[8]</sup>提出了一种 152 层的 ResNet 网络,比 VGGNet<sup>[9]</sup>深 8 倍,在 2015 年的 ILSVRC 多项任务中都取得了最先进的性能。拓宽网络宽度方面,Zagoruyko 等<sup>[12]</sup>提出的 WRN 网络减小了 ResNet 的深度,增加了 ResNet 的宽度,并取得了不错的性能。此外,减小卷积核或池化的步幅<sup>[9,13]</sup>、提出新的非线性激活函数<sup>[14,15]</sup>、增加新的层数<sup>[10,16]</sup>以及有效的规则等都能够提高网络性能。本文在现有的卷积神经网络基础上,通过结合不同的卷积神经网络特征层增加网络宽度,学习互补特征,提取更准确的图像表示进行识别。

原始输入图像往往包含许多无关的背景干扰信息,从而干扰分类决策。图像分类任务中的小物体目标往往只占图像的一小部分,准确识别这小部分特征

对分类决策来说至关重要。本文引入注意力机制,在进行图像特征提取时能够让网络自动选择需要关注的区域,在网络学习过程中,能够更关注图像的关键目标区域,对模型提取的特征进行选择,以获取更有效的信息。

基于以上,本文提出一种多模型特征和注意力模块融合的图像分类算法(image classification algorithm based on Multi-model Feature and Reduced Attention fusion, MFRA),通过多模型特征融合,让网络学习图像的不同特征,增加特征互补性,增强网络特征提取能力,并引入了注意力机制,使网络更关注目标区域。在几大公开数据集上进行了试验对比,验证了本文算法的有效性。

## 1 本文算法

在提高特征表达能力及提取显著性特征方面,目前的主流分类算法仍有上升空间。基于迁移学习技术,本文提出 MFRA 算法。本文网络架构如图 1 所示,不同网络架构能够学习到不同图像特征,为了提高特征互补性,本文使用 Inception-v3<sup>[17]</sup>和 MobileNet<sup>[18]</sup>网络作为特征提取器,输入图像分别经过 Inception-v3 网络的 mixed9 特征层和 MobileNet 网络的 conv\_pw\_12\_relu 特征层,并在提取高层语义特征后嵌入注意力模块,可使网络将注意力放在重要特征上,降低背景信息。并在算法最后添加一个全连接层进行图像分类。

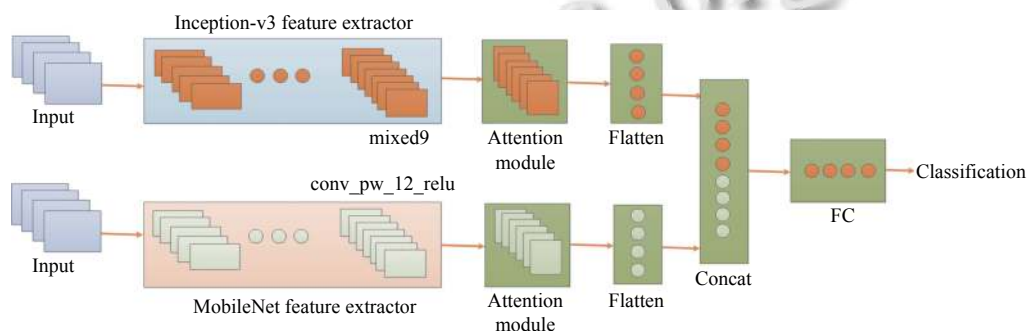


图1 基于多模型特征与注意力模块融合的网络结构

### 1.1 迁移学习与目标模型

为节约训练时间与资源,本文使用在 ImageNet 数据集上预先训练好的 Inception-v3 和 MobileNet 网络,并在目标数据集上进行微调。迁移学习与构建新的网络结构相比,能够取得更好的效果。

Inception-v3 网络在包含 1000 个不同类别的被认为是计算机视觉分类任务基准数据集的 ImageNet 上进行训练,Inception-v3 网络结构如图 2 所示,由图 2 可知,Inception-v3 网络内部集成了多个不同尺度大小的卷积核,能够学习到不同尺度的特征。Inception-v3

还通过将一个大卷积分解为几个小卷积, 以及将空间分解为非对称的卷积等操作减少网络参数, 从而节省计算开销, 加快网络训练速度. 在 ImageNet 数据集上, Inception-v3 达到了目前较高的分类性能, 错误率仅为 17.3%.

MobileNet 网络是针对手机等移动嵌入式设备提出的轻量级深层神经网络模型, 其核心是采用可分解

的深度可分离卷积, MobileNet 网络结构如图 3 所示. 由图 3 可知, MobileNet 网络架构由正常卷积和深度可分离卷积交叉组成, 深度可分离卷积是对常规卷积做出小规模调整, 对于来自上一层的多通道特征图, 将其全部拆分为单通道特征图, 分别对其进行单通道卷积后重新堆叠. 深度可分离卷积可以降低模型计算复杂度, 并降低模型体积.

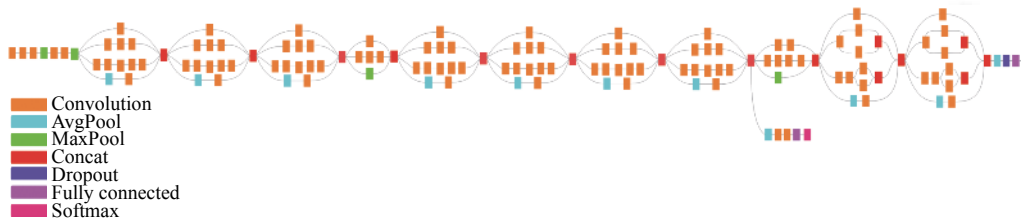


图 2 Inception-v3 网络结构



图 3 MobileNet 网络结构

### 1.2 精简注意力机制

在计算机视觉领域, 研究者们围绕着注意力机制的本质提出了许多注意力机制变体. Hu 等<sup>[19]</sup>提出的通道注意 SE 块采用两个全连接层结构, 从而获取各通道的重要性权重. Woo 等<sup>[20]</sup>提出的卷积注意力模块 (CBAM) 采用通道和空间注意力串行的方式, 对于通道注意模块, 特征向量分别经过最大池化和全局平均池化进入权重共享的网络, 空间子模块同样利用最大池化和全局平均池化的输出, 沿通道轴汇集到卷积层. Park 等<sup>[21]</sup>将通道和空间注意力结果相加形成瓶颈注意模块. 为了在尽可能少增加网络参数的情况下提高网络的特征提取能力, MFRA 算法简化了注意力机制结构, 只保留最本质的注意力结构, 因为本文提出的注意力机制结构简单, 因此称之为精简注意力. 本文提出的精简注意力结构如图 4 所示.

由图 4 可知, 精简注意力由通道注意力和空间注意力串行组成. 特征提取器提取的特征向量先输入通道注意模块, 根据各通道的重要程度得到通道权重矩阵, 然后将通道权重矩阵输入空间注意模块, 根据特征图的重要区域得到空间权重矩阵. 网络通过嵌入精简

注意力模块, 可以使网络更关注有意义的特征. 具体过程可表示为:

$$out = W_a \otimes (W_b \otimes F) = \sigma(F_a) \otimes (\sigma(F_b) \otimes F) \quad (1)$$

其中,  $F$  为输入特征,  $W_a$  和  $W_b$  分别为空间注意特征权重矩阵和通道注意特征权重矩阵,  $F_a$  和  $F_b$  分别为空间注意特征矩阵和通道注意特征矩阵,  $\sigma$  为 Sigmoid 激活函数,  $\otimes$  为矩阵相乘操作,  $out$  为输入特征经过精简注意力后的输出特征.

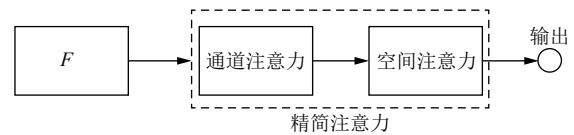


图 4 本文精简注意力机制网络结构

### 1.3 通道注意力

输入图像经过 Inception-v3 和 MobileNet 特征提取器提取特征, 不同通道表示不同类型的视觉特征, 以 Inception-v3 模型的 mixed9 特征层为例, mixed9 特征图中包含 2048 个通道, 每个通道均描述图像部分特征, 而对分类器而言, 不同特征算子提取到的特征重要性不同, 因而每个通道的重要程度不同, 应该给予每个通

道不同权重,让分类器更加关注对网络有意义的通道,加强显著性区域特征,减弱非显著性区域特征。

传统通道注意力网络 SENet<sup>[19]</sup> 在空间维度采用全局平均池化进行压缩,经过两个全连接层和激活函数,得到通道注意力权重矩阵。由于全连接层参数较多,加重网络负担,增加网络过拟合风险,因此本文采用一种精简的通道注意力模块,其结构如图 5 所示。

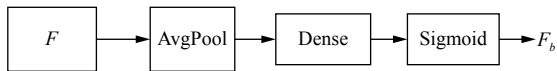


图 5 本文通道注意力机制网络结构

由图 5 可知,  $F$  为特征提取器提取到的特征向量, 维度为  $(B, H, W, C)$ , 经过全局平均池化操作得到维度为  $(B, 1, 1, C)$  的特征矩阵, 输入神经元个数为  $C$  的全连接层, 经过  $\sigma$  激活函数得到通道注意力权重矩阵, 其特征通道权重矩阵和为 1, 其过程表示为:

$$F_b = \sigma(\text{Dense}(\text{AvgPool}(F))) \quad (2)$$

其中,  $F$  为输入特征,  $\text{AvgPool}$  为全局平均池化,  $\text{Dense}$  为全连接层,  $\sigma$  为激活函数,  $F_b$  为通道注意特征权重矩阵。

#### 1.4 空间注意力

在原始输入图像经过特征提取器提取的特征图中, 特征图是原始输入图像的映射。对于图像而言, 目标总是只占据图像部分区域, 而其余区域则是背景信息。如何更好地让分类器更加关注存在目标的区域, 而降低背景信息干扰对图像分类性能起着关键作用。因此, 本文采用一种简单的通道注意力模块, 其结构如图 6 所示。

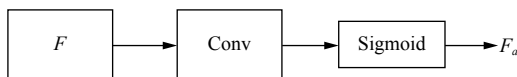


图 6 本文空间注意力机制网络结构

由图 6 可知,  $F$  为特征提取器提取到的特征向量, 维度为  $(B, H, W, C)$ , 经过卷积操作得到维度为  $(B, H, W, 1)$  的特征矩阵, 然后经过激活函数得到空间注意力权重矩阵, 其特征空间权重矩阵和为 1, 其过程表示为:

$$F_a = \sigma(\text{Conv}(F)) \quad (3)$$

其中,  $F$  为输入特征,  $\text{Conv}$  为卷积操作,  $\sigma$  为激活函数,  $F_a$  为空间注意特征权重矩阵。

## 2 实验结果及分析

本文实验环境为 Windows 系统, 显卡型号为 NVIDIA

GeForce RTX 2060, 采用基于 TensorFlow<sup>[22]</sup> 的 Keras 深度学习框架和 Python 编程语言, 分别在 Caltech-101<sup>[23]</sup>, Cifar-10<sup>[24]</sup>, Cifar-100<sup>[24]</sup> 这 3 个数据集上进行实验, 采用网络的错误率来验证本文算法的有效性。

### 2.1 Cifar-10 数据集

Cifar-10 数据集是图像分类常用数据集, 分为 10 类, 每类 6000 张图像, 共 60 000 张图像, 其中 50 000 张图像用于训练, 10 000 张图像用于测试, 图像尺寸为  $32 \times 32$ 。在模型训练过程中, 为了提高模型的泛化性, 使用随机翻转, 亮度变化等数据增强技术。在参数设置方面, 学习率初始化为 0.01, 批处理大小为 32, 使用 SGD 优化器, 并在训练中使用学习率递减策略, 当验证集准确率 3 个 epoch 不再下降时, 学习率递减为原来的一半。

基于 Cifar-10 数据集, 表 1 为本文提出的 MFRA 算法与其他算法的实验结果对比, 从表 1 可知, MFRA 算法的错误率为 3.86%, 相较于其他网络错误率降低 1.11% 至 6.55%, 据作者所知是目前同种类分类网络中错误率较低, 说明 MFRA 具有更佳分类性能。

表 1 不同算法错误率对比 (Cifar-10 数据集)

网络模型	错误率 (%)
NIN+Dropout <sup>[25]</sup>	10.41
DSN <sup>[26]</sup>	7.97
ResNet-CE <sup>[27]</sup>	9.59
SqueezeNet <sup>[28]</sup>	9.74
WideResNet <sup>[12]</sup>	4.97
FractalNet <sup>[29]</sup>	5.22
DSENet (depth=40) <sup>[30]</sup>	7.48
DSENet (depth=100) <sup>[30]</sup>	5.70
MS-FNet <sup>[31]</sup>	6.19
<b>MFRA</b>	<b>3.86</b>

### 2.2 Cifar-100 数据集

Cifar-100 数据集由 100 个类的 60 000 张  $32 \times 32$  彩色图片组成, 每个类有 6000 张图片。分为 50 000 张训练图像和 10 000 张测试图像。与 Cifar-10 不同的是, Cifar-100 数据集的每类图像都带有一个精细标签即它所属的类和一粗糙标签即它所属的超类, 因此 Cifar-100 数据集比 Cifar-10 数据集更具有挑战性。在模型训练过程中, 为了提高模型的泛化性, 使用随机翻转, 亮度变化等数据增强技术。在参数设置方面, 学习率初始化为 0.01, 批处理大小为 32, 使用 SGD 优化器, 并在训练中使用学习率递减策略, 当验证集准确率 3 个 epoch

不再下降时,学习率递减为原来的一半。

基于 Cifar-100 数据集,表 2 为 MFRA 算法与其他算法的实验结果对比,从表 2 可知, MFRA 算法的错误率为 17.51%,与其他网络相比错误率至少降低 8.34%,说明 MFRA 算法具有更佳的分精度。

表 2 不同算法的错误率对比 (Cifar-100 数据集)

网络模型	错误率 (%)
NIN+Dropout <sup>[25]</sup>	35.68
DSN+Dropout <sup>[26]</sup>	34.57
ResNet-CE <sup>[27]</sup>	29.71
SqueezeNet <sup>[28]</sup>	39.80
DSENet (depth=40) <sup>[30]</sup>	30.71
DSENet (depth=100) <sup>[30]</sup>	25.85
DNI <sup>[32]</sup>	30.24
X-CNNs <sup>[33]</sup>	42.82
MS-FNet <sup>[31]</sup>	29.36
<b>MFRA</b>	<b>17.51</b>

### 2.3 Caltech-101 数据集

Caltech-101 是一个非常流行的图像分类数据集,共 9144 张图像,分为 102 个类别。每个类别的图像数量在 31 至 800 图像之间。在实验中,为了比较的结果更具说服力,与比较的其他算法一致,我们使用每个类的随机选择的 30 张图像进行训练,其余的图像用于测试。

在模型训练过程中,为了提高模型的泛化性,使用随机翻转,亮度变化等数据增强技术。在参数设置方面,学习率初始化为 0.001,批处理大小为 16,使用 SGD 优化器,并在训练中使用学习率递减策略,当验证集的准确率 3 个 epoch 不再下降时,学习率递减为原来的一半。

基于 Caltech-101 数据集,表 3 为 MFRA 算法与其他算法的实验结果对比,从表 3 可知,本文算法的错误率为 5.36%,与其他网络模型相比有所降低,说明 MFRA 算法具有更好的分类能力。

### 2.4 消融实验

MFRA 算法的核心在于多模型特征融合及精简注意力模块,在实验参数设置相同的前提下,本节基于 Cifar-100 数据集上进行两组消融实验来分别证明多模型特征融合及精简注意力模块的有效性。其中,第 1 组实验验证精简注意力模块对本文网络结构性能的提升,第 2 组实验验证了多模型特征融合对本文网络结构性能的提升。

表 3 不同算法的错误率对比 (Caltech-101 数据集)

算法	错误率 (%)
文献[34]	18.6
文献[13]	13.5
文献[35]	11.6
文献[16]	6.6
文献[36]	5.3
<b>MFRA</b>	<b>5.36</b>

图 7 为不使用注意力模块的网络和只使用一种注意力模块的网络与使用精简注意力模块的网络在 Cifar-100 数据集的准确率曲线对比,其中, A 曲线为不使用任何注意力模块的网络, B 曲线为使用空间注意力模块的网络, C 曲线为采用通道注意力的网络, D 曲线为采用精简注意力的网络。由图 7 可知,使用精简注意力模块的网络模型与其他网络模型相比,收敛速度更快且分类准确率更高,证明了在网络模型中加入精简注意力可以加快模型的训练速度,使模型更快收敛,从而提高网络性能。

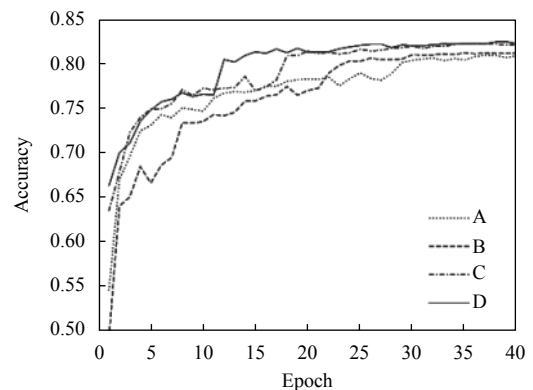


图 7 不同模块的准确率曲线对比

从表 4 可知,在 Cifar-100 数据集上,不使用注意力模块的网络分类错误率为 18.96%,只使用了空间注意力或通道注意力模块网络错误率相比不使用注意力模块的网络性能会有所下降,分类错误率降低了 0.22%至 1.19%,使用精简注意力模块的网络分类性能最佳,达到了 17.51%的分类错误率,这充分验证了精简注意力模块的有效性。

图 8 为使用单一模型和多模型特征网络在 Cifar-100 数据集上的准确率曲线对比。其中, A 曲线为只使用 Inception-v3 模型 mixed9 特征层的准确率曲线, B 曲线为单一 MobileNet 模型 conv\_pw\_12\_relu 特征层的准确率曲线, C 曲线为多模型特征融合的准确率曲线。由图 8 可知,采用多模型特征融合网络比单一模

型网络分类性能更佳. 收敛速度更快且分类准确率较高, 证明了多模型特征融合网络的有效性.

表4 不同模块的错误率对比

算法	错误率 (%)
mixed9+conv_pw_12_relu	18.96
mixed9+conv_pw_12_relu+空间注意力	18.74
mixed9+conv_pw_12_relu+通道注意力	17.77
<b>mixed9+conv_pw_12_relu+精简注意力</b>	<b>17.51</b>

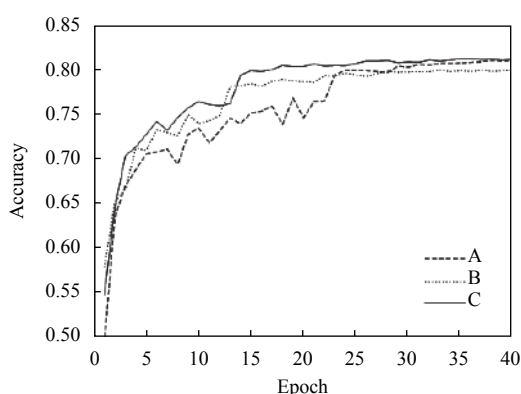


图8 不同模块的融合结果对比

由表5可知, 在Cifar-100数据集上, 使用单一Inception-v3模型特征的网络分类错误率为19.02%, 使用单一MobileNet模型特征的网络分类错误率为19.98%, 而使用多模型特征网络的分类错误率为18.96%. 这充分验证了多模型特征融合的有效性.

### 3 结束语

本文提出一种多模型特征和注意力模块融合的图像分类算法——MFRA算法. 该算法选取当前主流的卷积神经网络进行迁移学习, 节省训练成本. 根据不同的网络架构学习到图像的不同特征, 利用模型的多样性提取互补的视觉特征, 通过融合不同模型的所提取的特征提高网络的特征提取能力, 提高网络分类性能, 并在网络中嵌入精简注意力模块, 使网络将注意力集中于重要特征上, 降低背景干扰. 在几个常用的分类数据集上的实验结果表明, MFRA算法取得了较好的图像分类性能, 相较于其他算法, 其准确率有显著提升.

表5 不同模型的错误率对比

特征层	错误率 (%)
mixed9	19.02
conv_pw_12_relu	19.98
<b>mixed9+conv_pw_12_relu</b>	<b>18.96</b>

### 参考文献

- Lowe DG. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05). San Diego: IEEE, 2005. 886–893. [doi: 10.1109/CVPR.2005.177]
- Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5–32. [doi: 10.1023/A:1010933404324]
- Mao QH, Ma HW, Zhang XH. SVM classification model parameters optimized by improved genetic algorithm. *Advanced Materials Research*, 2014, 889–890: 617–621.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe: Curran Associates, 2012. 1097–1105.
- He KM, Zhang XY, Ren SQ, *et al.* Identity mappings in deep residual networks. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 630–645. [doi: 10.1007/978-3-319-46493-0\_38]
- Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
- He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv: 1409.1556*, 2014.
- Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 1–9. [doi: 10.1109/CVPR.2015.7298594]
- Sermanet P, Eigen D, Zhang X, *et al.* OverFeat: Integrated recognition, localization and detection using convolutional networks. *arXiv: 1312.6229*, 2013.
- Zagoruyko S, Komodakis N. Wide residual networks. *Proceedings of the British Machine Vision Conference (BMVC)*. York: BMVA Press, 2016. 1–87. [doi: 10.5244/C.30.87]
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. 13th European Conference on

- Computer Vision. Zurich: Springer, 2014. 818–833.
- 14 Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. ICML Workshop on Deep Learning for Audio, Speech and Language Processing. Atlanta: ICML, 2013.
  - 15 He KM, Zhang XY, Ren SQ, *et al.* Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015. 1026–1034. [doi: [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123)]
  - 16 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. 13th European Conference on Computer Vision. Zurich: Springer, 2015. 346–361. [doi: [10.1007/978-3-319-10578-9\\_23](https://doi.org/10.1007/978-3-319-10578-9_23)]
  - 17 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 2818–2826.
  - 18 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861, 2017.
  - 19 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
  - 20 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
  - 21 Park J, Woo S, Lee JY, *et al.* BAM: Bottleneck attention module. arXiv: 1807.06514, 2018.
  - 22 Abadi M, Agarwal A, Barham P, *et al.* TensorFlow: Large-scale machine learning on heterogeneous distributed systems. <https://arxiv.org/pdf/1603.04467.pdf>. (2016-03-16) [2021-01-25].
  - 23 Li FF, Fergus R, Perona P. One-shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 594–611. [doi: [10.1109/TPAMI.2006.79](https://doi.org/10.1109/TPAMI.2006.79)]
  - 24 Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Handbook of Systemic Autoimmune Diseases, 2009, 1(4).
  - 25 Lin M, Chen Q, Yan SC. Network in network. arXiv: 1312.4400, 2014.
  - 26 Lee CY, Xie SN, Gallagher P, *et al.* Deeply-supervised nets. arXiv: 1409.5185, 2014.
  - 27 杨萌林, 张文生. 分类激活图增强的图像分类算法. 计算机科学与探索, 2020, 14(1): 149–158. [doi: [10.3778/j.issn.1673-9418.1902025](https://doi.org/10.3778/j.issn.1673-9418.1902025)]
  - 28 Iandola FN, Han S, Moskewicz MW, *et al.* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv: 1602.07360, 2016.
  - 29 Larsson G, Maire M, Shakhnarovich G. FractalNet: Ultra-deep neural networks without residuals. arXiv: 1605.07648, 2016.
  - 30 Wu Y. Deep convolutional neural network based on densely connected squeeze-and-excitation blocks. AIP Advances, 2019, 9(6): 065016. [doi: [10.1063/1.5100577](https://doi.org/10.1063/1.5100577)]
  - 31 陈鑫华, 钱雪忠, 宋威. 基于轻量级特征融合卷积网络的图像分类算法. 计算机工程: 1–10. <https://doi.org/10.19678/j.issn.1000-3428.0059815>. (2020-12-17) [2021-01-25].
  - 32 Hou S, Liu X, Wang Z. DualNet: Learn complementary features for image recognition. Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 502–510. [doi: [10.1109/ICCV.2017.62](https://doi.org/10.1109/ICCV.2017.62)]
  - 33 Velickovic P, Wang D, Laney ND, *et al.* X-CNN: Cross-modal convolutional neural networks for sparse datasets. 2016 IEEE Symposium Series on Computational Intelligence (SSCI). Athens: IEEE, 2016. 1–8.
  - 34 Bo LF, Ren XF, Fox D. Multipath sparse coding using hierarchical matching pursuit. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013. 660–667. [doi: [10.1109/CVPR.2013.91](https://doi.org/10.1109/CVPR.2013.91)]
  - 35 Chatfield K, Simonyan K, Vedaldi A, *et al.* Return of the devil in the details: Delving deep into convolutional nets. Proceedings of the British Machine Vision Conference. Nottingham: BMVA Press, 2014. [doi: [10.5244/C.28.6](https://doi.org/10.5244/C.28.6)]
  - 36 Mahmood A, Bennamoun M, An S, *et al.* ResFeats: Residual network based features for image classification. 2017 IEEE International Conference on Image Processing (ICIP), Beijing. 2017. 1597–1601. [doi: [10.1109/ICIP.2017.8296551](https://doi.org/10.1109/ICIP.2017.8296551)]