文章编号：2095-2163(2019)06-0101-06 中图分类号：TP391.4 文献标志码：A

# 基于迁移学习与模型融合的犬种识别方法

李思瑶，刘宇红，张荣芬
（贵州大学 大数据与信息工程学院，贵阳 550002）

**摘 要**：犬种识别研究属于细粒度图像分类的典型代表，使用传统图像分类方法与普通卷积神经网络进行犬种识别，会出现准确率普遍很低等问题。本文提出了一种将迁移学习与模型融合相结合的方法。通过运用四种常用的卷积神经网络模型分别进行部分图像的特征提取，选取表现最佳的两种模型 Inception_v3 以及 Resnet152_v1 进行双模型融合，将得到的融合网络用于犬种图像进行迁移学习训练。针对 120 类犬种图片，训练得到了验证集精度可达 93.02% 的网络模型。同时考虑将测试集图片经过 YOLO 目标检测算法识别，定位目标区域后再送入网络，实验结果表明该方法在融合模型中能进一步提高犬种识别检测精度。

**关键词**：迁移学习；模型融合；犬种识别；深度学习

# Dog breed identification method based on transfer learning and model fusion

LI Siyao, LIU Yuhong, ZHANG Rongfen
(Institute of Big Data and Information Engineering, Guizhou University, Guiyang 550002, China)

**【Abstract】** Dog breed identification research is a typical representative of fine-grained image classification. The accuracy of recognition is generally low due to the use of traditional image classification and common convolutional neural networks for dog breed identification studies. This paper proposed a method to combine transfer learning with model fusion. Firstly, the method extracted features of partial images by using four convolutional network models separately, the two best performing models--Inception_v3 and Resnet152_v1, are selected for dual model fusion. The fusion network was applied to the dog breed image for transfer learning training . For the 120 kinds of dog breed pictures, the network model obtain the accuracy of 93.02% on the verification dataset. Meanwhile, the testset was targeted by the YOLO object detection algorithm and cropped into the network. The result shows that the method can further improve the dog species classification accuracy through the model.

**【Key words】** transfer learning; model fusion; dog breed classification; deep learning

## 0 引 言

深度学习(Deep Learning)的概念源于人工神经网络的研究，深度学习实际上是深度神经网络 DNN。而神经网络技术的开展可以追溯到 1943 年[1]。近年来，深度学习在目标检测、表情识别、目标跟踪等诸多领域有了巨大的研究进展，特别是在图像分类方面有许多突破。目前大部分的图像分类工作都集中在通用分类，比如对手写体数字的分类等多种不相关类别。因此，子类别的图像的区分也就是细粒度图像分类成为这几年计算机视觉领域的研究热点。其研究目标从不同类别转换为同一类别不同子类之上[2-4]，是一项极具挑战的研究任务。犬类是与人类最密切相连的动物，人们对于犬类普遍比较熟悉，素材照片等更容易获取，一定程度方便了深度学习的开展。

犬种识别的研究是进阶性的过程，由于犬类的类间相似性和类内差异性以及图片的背景、拍摄光线、目标姿态等的影响，其研究具有一定的难度。犬类的区分一般是通过专家鉴定或者基因检测来实现，然而这些方法会消耗极大的人力与时间。此外，还出现了利用人脸识别、图片局部定位和 PCA 技术的由粗到细的犬种分类[5]，基于地标形状的犬种分类方法[6]，以及使用传统的神经网络算法来进行图像分类[7]等方法。但由于犬类的数据集对于神经网络的训练远远不够、对图像局部(关键)区域的细节特征提取不充分等原因，以上这些方法在识别准确率上普遍不高，识别种类也不是很多。

本文提出了一种基于迁移学习和双模型融合的犬种识别算法。使用已在大规模数据集 ImageNet

**基金项目**：贵州省科技计划项目(黔科合基础[2019]1099)。

**作者简介**：李思瑶(1995-)，女，硕士研究生，主要研究方向：电路与系统、图像处理；张荣芬(1977-)，女，博士，教授，主要研究方向：嵌入式系统、机器视觉、智能算法及大数据应用。

**通讯作者**：张荣芬 Email：rfzhang@gzu.edu.cn。

**收稿日期**：2019-09-25

上预训练的四种常用深度神经网络 Vgg16_bn、Densenet161、Inception_v3 和 Resnet152_v1 分别进行部分图像的特征提取。由于不同的卷积神经网络(CNN)架构在提取图像特征时表现的学习过程不同,导致视觉分类有不同结果。为组合多方位的信息表示,达到更优化的性能,选取表现最佳的两个网络 Inception_v3 和 Resnet152_v1 进行双模型融合。然后将该融合网络用于 120 类犬种图像进行迁移训练。同时将测试集图片经过 YOLO 目标检测算法识别目标区域后再送入融合网络进行实验,进一步减少背景干扰。

## 1 相关理论

### 1.1 Inception_v3 模型

卷积神经网络包含多种模型,例如 Alexnet、VGG、Googlenet 等。其中,Google Inception Net 在 2014 年的 ILSVRC 比赛中取得第一名。Inception _v3 模型在 v2 的基础上改进了三种 Inception 模块:使用两个 3×3 的卷积替代每个 5×5 的卷积,将 $n \times n$ 的卷积分解成一维的 $n \times 1$ 和 $1 \times n$ 卷积的串联。压缩特征维度数不仅促进了高尺寸图像的表示,也减轻了过拟合现象[8-9]。全连接层被全局平均池化层所取代,极大地降低了参数数量。该网络包含 47 层,其详细网络结构如图 1 所示。

Inception _v3 网络输入图片大小为 299×299,在减少计算量的同时提升了网络性能。除了在模型中使用分支,也实现了在分支中使用分支。同时增加了一层非线性扩展模型表达能力,可以处理更多的空间信息,增加特征多样性。

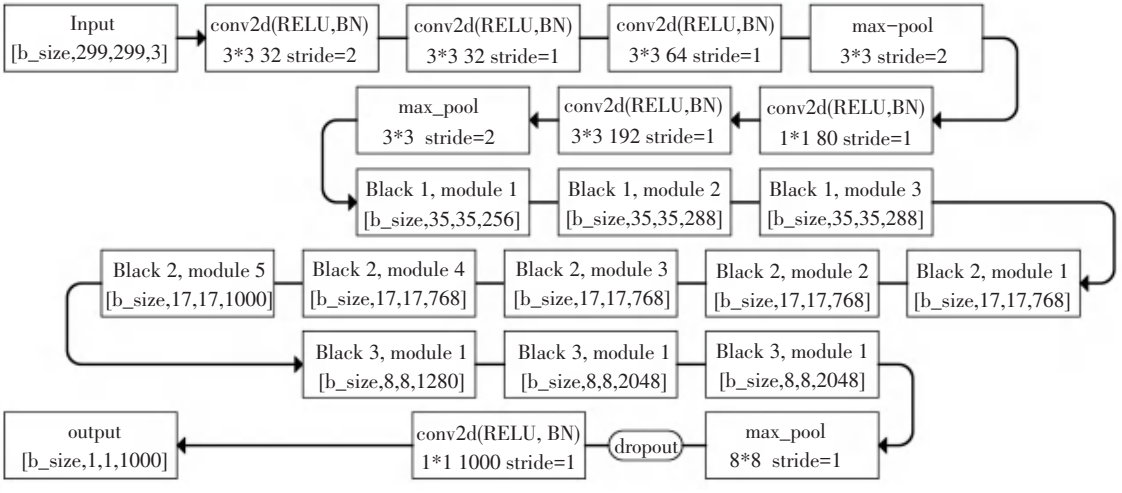



**图 1 Inception _v3 网络结构图**

**Fig. 1 Inception _v3 network structure diagram**

### 1.2 Resnet152_v1 模型

Resnet 在 2015 年的 ILSVRC 比赛中取得第一名。该网络多达 152 层,网络深度和维度的增大使其可以进行更加复杂的特征模式提取[10]。同时该网络的作者还提出了残差学习来解决深度网络的退化问题[11],残差学习单元如图 2 所示。

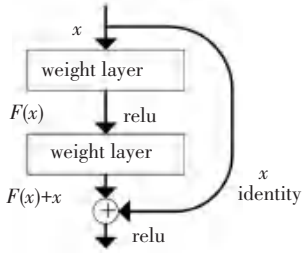



**图 2 残差学习单元**

**Fig. 2 Residual learning unit**

其中,输入为 $x$,学习到的特征记为 $H(x)$,这样残差就表示为 $F(x) = H(x) - x$。残差单元可以表示为:

$$y_j = h(x_j) + F(x_j, W_j), \quad (1)$$

$$x_{j+1} = f(y_j), \quad (2)$$

其中,$x_j$ 和 $x_{j+1}$ 分别表示第 $j$ 个残差单元的输入与输出,$f$ 为激活函数 relu。推导可得从浅层 $j$ 到深层 $J$ 的学习特征为:

$$x_J = x_j + \sum_{i=j}^{J-1} F(x_i, W_i). \quad (3)$$

在实际操作中残差不等于零。因此考虑残差函数会使得堆积层在输入特征基础上学习到新的特征,从而拥有更好性能。

### 1.3 迁移学习

在面对图像分类领域的具体应用场景时,通常

可能无法得到用以构建神经网络模型所需规模已标记的数据。迁移学习(Transfer learning)提出，通过从相关域中的数据中提取有用信息，并将其转移以用于目标任务来解决此类跨域学习问题。即除了目标域中的数据之外，还可以包括不同域中的相关数据，以扩展目标未来数据的先前知识的可用性。本文主要以常见的 120 类犬种数据为基础展开研究。然而犬类的数据集远远不够用来重新训练深层神经网络，因此借助迁移学习的思想，在训练数据集规模较小的情况下可以使用预训练过的模型。通过对预训练网络瓶颈层之后进行截断，保留可重用层的有用神经元，以挖掘出更多的犬种分类特征[12-14]。另外，迁移学习可以使训练数据保持在相同特征空间中或具有与未来数据相同的分布，避免了过度拟合问题，因此可以学习到更出色的底层规则。

### 1.4　YOLO 目标检测网络

YOLO 网络将物体检测任务当做一个回归问题来处理，使用一个神经网络，直接从一整张图像来预测出边界框的坐标、框中包含物体的置信度和物体的可能性。YOLO 将目标区域检测和类别预测整合于一个神经网络中，端到端的训练过程优化了物体检测性能[15-16]。如图 3 所示，YOLO 将输入图像划分为 S * S 的栅格，每个格子负责检测中心落在这个格子中的物体。

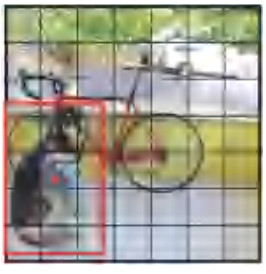


**图 3　YOLO 对图像的划分**

**Fig. 3　YOLO division of images**

其中，每个栅格对其预测的边界框判断条件为：

$$score = \Pr(Object) * IOU_{pred}^{truth}. \qquad (4)$$

通过得到的结果来判断该边界框中是否含有目标。

在本文的研究中，借助 YOLO 网络能够快速检测数据集，并能较为准确地将狗与图片背景分开，在一定程度上减少了背景对品种分类的影响。

## 2　犬种识别

### 2.1　基本思想

犬种分类时，由于不同种类的狗外观具有一定的相似性，同一种类的狗也有毛发颜色、姿势不同等原因，因此需要充分包含目标纹理中的细节信息。为了能在得到全局视觉特征信息的同时，还能够得到图像局部(关键)区域的细节特征共同进行对比分析，本文提出了如图 4 的实验过程以找到表现更好的网络结构，从而进一步提高识别准确率。具体实验步骤如下：

(1)将 4 种基础网络在大规模数据集 ImageNet 上进行训练得到相关预训练的神经网络。

(2)将部分数据集分为训练集与验证集，并分别导入 4 个网络。

(3)选取表现最佳即在验证集上损失函数最小的 2 个网络进行模型融合。

(4)改进网络结构，使其适用于犬种图像的数据集。

(5)再次将全部数据中的训练集与验证集导入融合后的网络，进行迁移学习。

(6)对网络进行微调使其表现出更好的性能，最终得到的网络用于犬种图像的分类预测。

(7)将测试集图像输入网络，查看预测结果。

(8)将测试集图像经过 YOLO 检测算法，分割目标区域后送入网络，比较两次操作的预测结果。

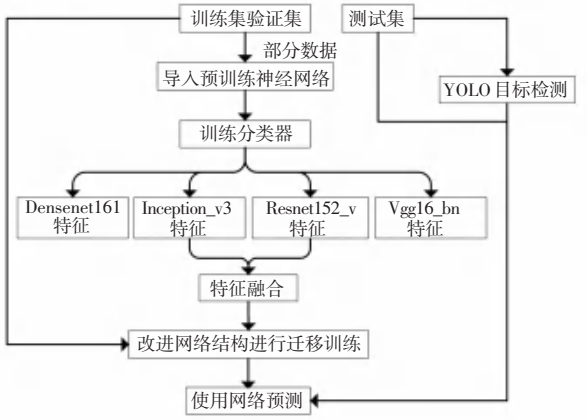



**图 4　实验过程**

**Fig. 4　Experiment procedure**

其中，在网络迁移学习过程中，冻结融合网络的相关权重，训练最后添加的分类头浅层网络。由于训练参数比较少，选择能对梯度的一阶矩估计和二阶矩估计进行综合考虑，计算出更新步长的 Adam 优化器。训练过程中，梯度下降时，每个批次包含 128 个样本，迭代轮数设为 100 轮。

### 2.2　模型融合

由于搜集到的人工标记犬类图片数量远不及训练高广义 CNN 模型所需要的大规模标记数据集合，

尤其是对于复杂且非常深的 CNN 架构。单靠本文准备的数据集无法获得合理的分类模型。因此,本研究利用大规模 ImageNet 数据集,利用预训练模型探索了犬种分类的转移学习策略。通过在预训练模型中用犬类数量的神经元替换最终的分类器层(1 000 个神经元),并保留相同的条件。例如其它层的内核大小数,使用预训练的 CNN 模型的学习参数作为初始值。之后,使用加权表决融合方式将两种网络训练所得特征进行拼接。不同的模型具有不同内核大小和体系结构,这可以学习不同方面的图像表示,例如多尺度属性,并在后层中载入特征提取后的数据以用于犬品种分类学习。本文选择 Inception_v3 与 Resnet152_v1 网络进行实验,具体融合方法如图 5 所示。

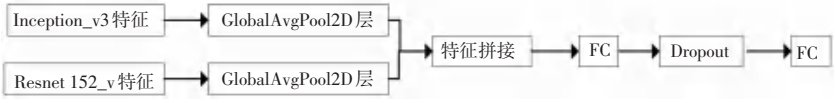



**图 5 模型融合过程**

**Fig. 5 Model fusion process**

其中,添加 GlobalAvgPool2D 层是为了对两个网络的输出尺寸进行调整从而可以合并。GlobalAvgPool2D 层没有数据参数,两个网络的特征输出尺寸均为(28 440, 2 048, 1, 1),是四维矩阵。而全连接层的输入要求是二维矩阵,两个输入尺寸分别进入 GlobalAvgPool2D 层后被调整为(28 440, 2 048)大小,特征拼接后的输出尺寸为(28 440,4 096),可见进行网络融合后可以学习到更多的图像特征。添加的两层完全连接层的神经单元数分别为 256 以及 120。激活函数为 RELU 函数,为了防止训练过拟合,设置 Dropout 为 0.5。

## 3 实验验证与分析

### 3.1 犬种数据集及其预处理

本文所使用的犬种数据集包括斯坦福大学搜集的 120 种犬类数据,以及 Kaggle 竞赛所使用的有标签的训练数据,两个数据集包含的图片有部分重复。同时,用到了自行在网络搜集的部分图片。图片总和为 31 600 张。实验中随机抽取每种类别图片的 80%作为训练集,10%为验证集,剩余 10%为测试集。首先通过对数据集进行随机镜像、增加适当高斯噪声、垂直方向图像随机旋转等处理实行图像增强。数据集增强在一定程度上弥补了数据集样本不足的问题,减少网络的过拟合现象,可得到泛化能力更强的网络,更好地适应应用场景。之后,对图像进行预处理:由于图片将分别进入两个网络,所有图片应缩放为 224×224 以及 299×299 像素,满足两个网络不同的输入要求。图像的均值和方差按照数据集 ImageNet 来设置,根据模型预先训练时的处理方式来处理数据,这样才能保证最好的效果。

### 3.2 实验环境及评价指标

本实验采用的 GPU 显卡为 GTX 1 080 Ti,内存为 64 GB。在 Linux 系统下,犬种分类过程采用基于 mxnet 的 gluon 深度学习框架对图片进行分类。

模型的性能评价指标包括训练集、验证集、测试集的准确率以及训练集和验证集的损失率。其中,训练集的准确率和损失率体现了模型训练时的性能,验证集的损失率用于判断模型在迭代过程中是否出现过拟合等情况,测试集的准确率直接反映了已训练好的模型的预测能力。这两个指标由如下公式定义:

$$Accuracy = \frac{1}{M}\sum_{i=1}^{M} I(y_i = f(x_i)), \tag{5}$$

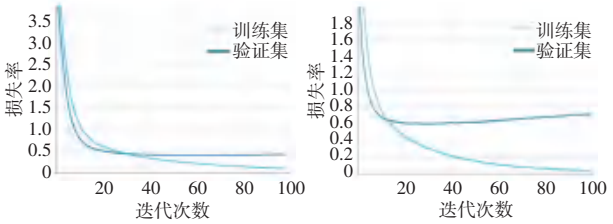
$$loss = -\frac{1}{M}\sum_{x} y\ln a + (1-y)\ln(1-a). \tag{6}$$

其中, $M$ 是样本数量; $y_i$ 是标签; $f(x_i)$ 为模型预测结果; $I$ 是条件判断函数; $a$ 为神经元经过激活函数的非线性输出。

### 3.3 实验验证与分析

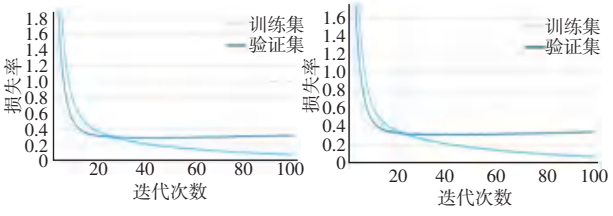
**实验 1** 将数据集中 Kaggle 所包含的 10 222 张图片经过预处理后分别送入 Densenet161、Inception_v3、resnet152_v1 以及 vgg16_bn 的预训练网络。预训练模型是由其它组织使用包含许多类别的大量数据集训练而成,其中有几千万的样本。在这种情况下,还需要训练模型最后的分类头。卷积层从图片中提取特征,完成连接层对图片进行分类,需要调整最后的完全连接层来制作模型以适应犬种的数据。通过比较四种网络的性能,找到在验证集上损失率较小的前两个网络进行双网络融合。网络的批量处理大小为 64,学习速率设置为 0.001,用于微调。四个模型训练后得到的训练集与验证集的损失曲线如

图 6 所示。




（a）Densenet161 损失曲线　（b）vgg16_bn 损失曲线

（a）Denseet161 loss curve　（b）Vgg16_bn 损失曲线




（c）Resnet152_v1 损失曲线　（d）Inception_v3 损失曲线

（c）Resnet152_v1 loss curve　（d）Inception_v3 loss curve

**图 6　四种训练模型损失率曲线**

**Fig. 6　Four training models loss curves**

四种网络训练的性能指标数据见表 1。

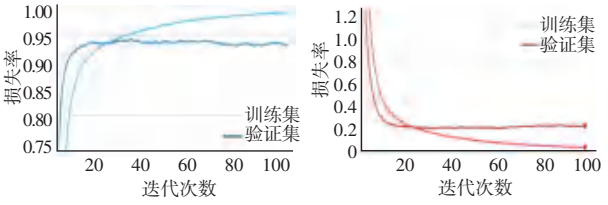
**表 1　四种网络训练模型的实验比较**

**Tab. 1　Experimental comparison of four models**

| model | Train_loss | Val_loss |
|---|---|---|
| Densenet161 | 0.11 | 0.43 |
| vgg16_bn | 0.04 | 0.71 |
| resnet152_v1 | 0.07 | 0.30 |
| Inception_v3 | 0.06 | 0.32 |

由此可见，对于犬种数据学习性能较好的为 resnet152_v1 网络和 Inception_v3 网络。接下来，将这两个网络按照本文 2.2 节所介绍的方法进行模型融合。为了两个网络进行融合，需要自定义一个网络合并层，保证这两个神经网络在合并前的输入尺寸一致。之后，添加分类头，也就是输出层。将上述两个网络拼接起来，逐步构建完整网络结构。然后将总数据集中的 90% 图片作为训练集与验证集送入融合后的网络进行迁移学习，训练得到的准确率与损失率结果如图 7 所示。

由图可见，融合网络进行迁移学习后，验证集的损失率下降到 0.24，验证集的准确率可达到 93.02%。并且，迁移学习能通过较短的训练时间进行特征学习。然而，通过对比数据发现损失率在经过 30 次左右的迭代后出现了下降速度缓慢、不能降到很低等问题。这可能是由于仅训练最后添加的分类头使网络不能更好地匹配犬种数据集所造成的。




（a）融合模型准确率曲线　（b）融合网络损失曲线

（a）Fusion model accuracy curve　（b）Fusion model loss curve

**图 7　融合模型训练结果**

**Fig. 7　Fusion model training results**

**实验 2**　首先将测试集图片送入训练好的网络检测其预测能力。通过观察发现犬种图片仅出现在图片的部分区域，背景在一定程度上分散了网络学习。为验证该设想，随机选取五张图片进行 YOLO 目标检测网络，随后将得到的五组图片送入网络进行犬种分类。实验结果对比见表 2。分析表中结果可知，若是图片中目标面积较大，图片分割前后均能识别出正确的类别名称；若图片中的背景环境较为复杂，分割出目标后再进行检测可以有效提高预测精度；若是拍摄照片时的角度、光线等出现变化，分割前后的结果均容易受到影响。可见，将目标区域定位后再送入网络能够在一定程度上减少识别误差。

**表 2　改进定位后分类结果对比**

**Tab. 2　Comparison of results after improving positioning**

| 犬种类别 | 原始图片 | 目标定位 | 原始识别结果 | 分割后识别结果 |
|---|---|---|---|---|
| 阿富汗猎犬 |  |  | 阿富汗猎犬 | 阿富汗猎犬 |
| 巴辛吉 |  |  | 博得猎狐犬 | 巴辛吉 |
| 巴吉度猎犬 |  |  | 巴吉度猎犬 | 巴吉度猎犬 |
| 伯瑞犬 |  |  | 凯恩犬 | 伯瑞犬 |
| 杜宾犬 |  |  | 巨型雪纳瑞犬 | 比利时黑猫牧羊犬 |

经过将测试集分组实验，对比测试图片经过 YOLO 检测算法分割目标区域前、后的识别效果并取识别率均值，结果得到目标分割前后准确率分别为 67.20% 和 68.45%，提高了 1.05%。表 3 是本文测试结果与其它文献中识别方法的对比。

表 3 各类方法实验结果对比

Tab. 3 Comparison of experimental results of some methods

| 方法来源 | 所用方法 | 识别准确率/% |
|---|---|---|
| 文献[6] | Alexnet 微调 | 63 |
| 文献[8] | 拓展数据+DCNN | 67.31 |
| 文献[9] | Googlenet 迁移学习 | 69.07 |
| 本文(未分割) | 迁移融合网络 | 67.20 |
| 本文(分割后) | 迁移融合分割网络 | 68.45 |

可以看出,文献[6]是基于传统的深度学习网络 Alexnet 而展开的训练,其检测精度明显低于其它方法,这是由于传统的网络参数训练对数据有极大依赖性,图像特征的提取也与先验知识有密切关系。特别是随着网络层数的不断增加,Alexnet 包含 8 个隐藏层,而本文所使用的网络的层数明显更深,单纯利用数据对网络进行训练是不能更好地实现分类任务的。文献[8]使用的数据集多达 163K 张,使用预训练方式对 DCNNs 进行完全初始化训练。该方法虽然增加了数据总量,但训练学习到的特征过于单一。文献[9]的结果虽然比本文方法稍高一点,不过在该实验中使用了比本文更多的图片用于训练,并且在网络训练过程中首先训练了一个狗脸探测器用于之后的宠物识别,这无疑会消耗更多的训练时间,降低工作效率。

最后,值得一提的是,在本文所搜集的数据中,Kaggle 数据集存在少数犬类图片人工标记分类错误等问题。同时总数据集的每个种类包含的图片数量并不均匀,这些都会对网络的训练造成一些影响。

## 4 结束语

本文提出了一套基于迁移学习与融合基本 CNN 模型实现犬种分类的方法。实验证明该方法对于属于细粒度图像分类的犬种识别有一定的性能改进。由于训练样本数量不足等问题,本文研究了迁移学习策略。另外,由于不同网络的内核大小、层数和结构等的不同,深度 CNN 架构可能会提取出图像的不同表示特征,从而导致视觉分类的不同表现。通过将 4 种常用 CNN 模型在大规模 ImageNet 数据集上进行预先训练,学习调整网络的各层参数,以增加不同特征信息的表示。再次将犬种数据集分别输入 4 个网络,选取表现最佳的 2 个网络进行模型融合以结合多方位信息来获取更加准确的分类,这是本文创新之一。

论文在选取 Inception_v3 以及 Resbet152_v1 网络进行双模型融合后,将生成的新网络用于犬种图像的迁移学习。针对 120 类犬种图片,训练得到了验证集精度可达 93.02%的网络模型。之后将测试集图片经过 YOLO 目标检测算法识别犬目标区域后再送入融合网络,结果表明该测试集经过目标定位分割后在模型中提高了 1.05%的检测精度。说明减少背景信息的干扰可以显著提高性能,这是本文另一个创新之处。横向对比实验结果进一步验证了本文方法用于犬种识别的可行性。在之后的工作中,将继续沿着这个方向研究并做出改进,进一步提高犬种识别的性能,探索其应用潜能。

## 参考文献


[1] 娄英欣. 基于深度学习的目标检测[D]. 北京:北京邮电大学, 2018.

[2] NILSBACK M E, ZISSERMAN A. A visual vocabulary for flower classification[C]. In Proc. CVPR, 2006.

[3] KHOSLA A, JAYADEVAPRAKASH N, YAO B, et al. Novel data-set for fine-grained image categorization[C]. In First Workshop on Fine-Grained Visual Categorization, CVPR, 2011.

[4] 王永雄, 张晓兵. 聚焦—识别网络架构的细粒度图像分类[J]. 中国图象图形学报, 2019, 24(4): 493-502.

[5] PRASONG, P, CHAMNONGTHAI, K. Face-recognition-based Dog Breed Classification Using Size and Position of Each Loc-al Part, and PCA[C]. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technol-ogy (ECTI-CON), 2012 9th International Conference on 2012: 1-3.

[6] WANG Xiaolong, Vincint Ly, Scott Sorensen, et al. Dog br-ed classification via landmarks[C].Image Processing (I-CIP), 2014 IEEE International Conference on 2014.

[7] SINNOTT Richard O, WU Fang, CHEN Wenbin. A Mobile Application for Dog Breed Detection and Recognition Based on Deep Learning[C]. Big Data Computing Applications and Technologies (BDCAT), 2018 IEEE/ACM 5th International Conference on 2018.

[8]吴迪, 刘秀磊, 侯凌燕,等. 基于显著性检测和迁移学习的花卉图像分类[J].北京信息科技大学学报(自然科学版), 2019, 34(1): 55-63.

[9] 张泽中, 高敬阳, 吕纲,等. 基于深度学习的胃癌病理图像分类方法[J]. 计算机科学, 2018, 45(S2): 263-268.

[10]LIU Shaopeng, TIAN Guohui, XU Yuan. A novel scene class-ification model combining ResNet based transfer learning and data augmentation with a filter[J].Neurocomputing, 2019, 338.

[11]裴颂文, 杨保国, 顾春华. 网中网残差网络模型的表情图像识别研究[J]. 小型微型计算机系统, 2018, 39(12): 2681-2686.

[12]于奥运. 基于深度学习的犬种识别研究[J]. 现代计算机, 2018(0): 106-109.

[13]TU Xinyuan, LAI Kenneth, Svetlana Yanushkevich. Transferlearning on convolutional neural networks for dog identification[C]. Software Engineering and Service Science (ICSESS), 2018 IEEE 9th International Conference on 2018.

[14]王莉莉, 冯其帅, 陈德运,等. 一种基于正则化判别分析的迁移学习算法[J]. 哈尔滨理工大学报, 2019(2):89-95.

[15]周奇. 基于多特征的轮船运动目标跟踪及轨迹获取方法[D]. 北京:北方工业大学, 2018.

[16]蒋家俊. 基于卷积神经网络的小目标行人检测研究[D].兰州:兰州理工大学, 2018.