基于图像放缩增强网络的图像分类方法

冯继凡[1] 杨 清[1] 石昌鑫[1]
FENG Jifan YANG Qing SHI Changxin

摘 要 随着人工智能的快速发展，基于深度神经网络的图像分类任务性能得到了巨大提升，在图像检索、智能安防和自动驾驶等领域具有重要的应用价值。但目前图像分类网络的性能通常会受到输入图像大小的限制，相比于输入较小的图像，增大输入图像大小可以保留更多的细节信息，保证网络可以提取到更丰富的特征，从而提高分类准确性，但其也会降低网络的推理速度。目前的深度学习算法通常用线性插值的方式调整图像分辨率至固定大小，这种方式往往会限制网络的性能。针对上述图像输入大小和插值方式对分类准确率的影响问题，提出了一种基于图像放缩网络的图像分类方法，以轻量化网络架构 EfficientNet 为基准分类网络，引入混合空洞卷积增大特征的感受野，并设计了一种图像放缩增强网络模块对输入图片进行压缩增强，在少量增加网络参数量的情况下，丰富输入图像包含的信息，提升分类网络的分类性能。多组实验证明，所提出方法对图像分类准确率有着显著的提高。

关键词 图像分类；混合空洞卷积；感受野；EfficientNet；图像放缩增强

doi：10.3969/j.issn.1672-9528.2024.05.041

0 引言

在人类认知世界的过程中，百分之八十的信息 [1] 是通过视觉获得的。与人类视觉相似，计算机视觉同样是计算机和其它机器认知世界的基础。图像分类作为计算机视觉领域的基本任务 [2]，是指提取图像中能够代表其内容的特征信息，并对这些信息进行处理和分析，从而确定图像的类别标签。近年来，伴随着互联网与计算机技术的飞速发展和手机、笔记本电脑等便携式移动网络设备的普及，图像分类技术已成功应用于安防、智慧城市、医疗影像、安全生产等领域 [3]。

虽然图像分类的分类方法因深度学习技术的快速发展取得了质的突破，但是输入图像大小对网络准确性的影响却鲜有人关注。在图像分类研究中，输入的图像大小会直接影响到神经网络的性能。通常调整图像大小的原因有：网络模型通过梯度下降的小批量学习时需要每一批的图像具有相同的分辨率；计算机内存限制以高分辨率的图像训练网络模型；高分辨率图像导致网络模型训练和推理的速度变慢。

为了提高网络的效率，输入图像的大小往往会调整为相对较低的分辨率送入网络训练和推理，但这会降低分类的准确率。同时现有的图像预处理方法采用简单的最近邻线性插值、双线性插值等方法将图像缩放，这些方法速度快且可以灵活集成到训练或者测试框架中，但是也极大减少了图像的有用信息，导致图像分类任务的性能下降。此时，对输入图像增强的方式，可以保留更多的细节信息，在保持推理速率的同时提高分类的准确性。

1 算法框架

本文所提出的基于图像放缩增强网络（image scaling enhancement network，ISEN）的图像分类方法，图像放缩增强网络的结构如图 1 所示。该方法不旨在增强图像的感知质量，而是专注于提高模型的识别性能，同时该模块可以与图像分类模型联合训练，以更好地适应图像分类模型。通过增加图像输入尺寸并进行放缩增强的方式，保留图像中的细节信息，更好的细节信息可以让分类网络提取到鲁棒性更强的图像特征，便能更容易推断出图像所属的类别。模型通过双线性插值的方式和跳转连接的方式，将线性调整的图像与 CNN 特征相结合。其中线性插值的方式可以将图像调整为分类模型所需大小，即可以将原始分辨率的图像调整为任意目标大小和横纵比的图像，且模型中的插值方法可以变换为 Bicubic 或者 Lanczos[4] 等其它采样算法，跳转连接则可以将线性调整的图像与图像特征进行融合，得到保留更多图像信息的增强图像，并且跳转连接使网络参数更容易学习。相较于直接输入图像，输入增强后的图像可以大幅提升分类的准确率。

1. 中国航空工业集团公司西安航空计算技术研究所
陕西西安 710065

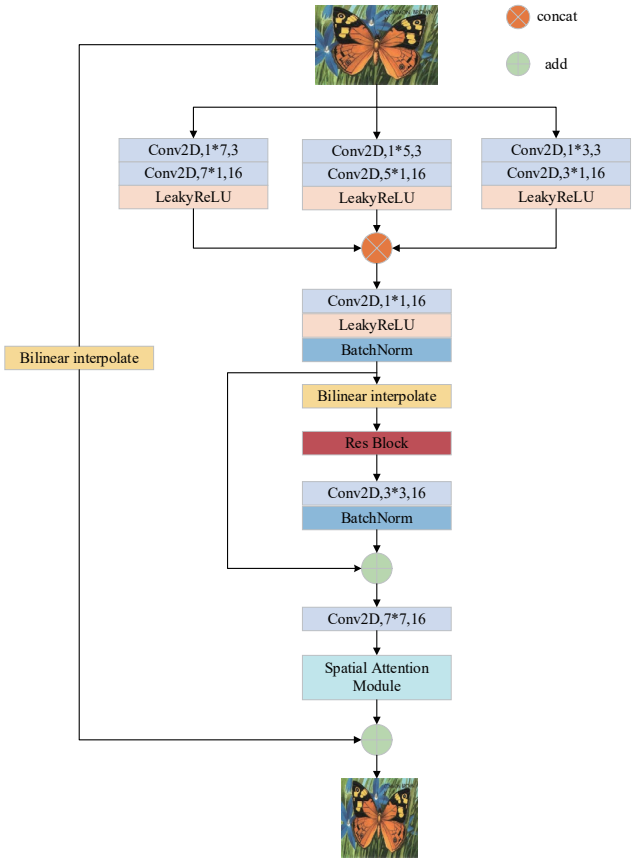

图 1 图像放缩增强网络结构图

2 数据集构建

本文采用 Oxford Flowers102 数据集[5]训练并测试，该数据集是由牛津大学于 2008 年发布的数据集，包含 102 种英国常见的花卉，总计 8189 张图像，每类花卉图像有 40 到 258 张不等。由于该数据集中花卉主体在图像中占比较大，且不同图像之间姿势和光线变化明显，所以常被用于图片分类研究。

在图像生成到最终被处理应用的过程中，存在多种因素会使图像的分辨率降低，例如拍摄设备的性能较差或者受到拍摄参数和拍摄环境的影响。为了尽可能模拟现实中获得的图像，并验证图像放缩增强网络对不同分辨率图像分类性能的提升效果，本文原始清晰图像通过式（1）的方法进行退化，得到不同分辨率图像。

$$I_{LR} = (I_{HR} \otimes K)\downarrow_S + N \tag{1}$$

式中：I_{LR} 表示采样得到的低分辨率图像；I_{HR} 表示对应的清晰图像；$\otimes$ 代表卷积操作；K 表示模糊核；$\downarrow_S$ 表示进行 S 倍的下采样；N 为加性高斯白噪声。然而最近许多低分辨率图像的研究中，通常忽略模糊核和噪声，采用 bicubic 降采样得到对应的低分辨率图像。本文参照文献 [6]，使用 bicubic 下采样高分辨率图像来构建低分辨率图像。图 2 为部分下采样之后图像对比。

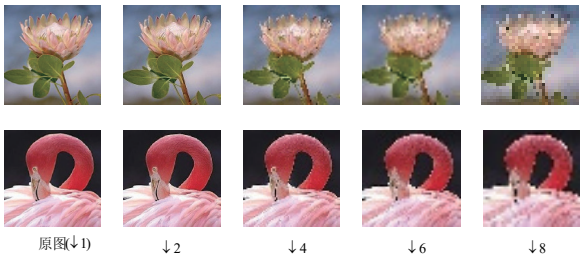

图 2 部分不同倍数下采样图像对比图

3 算法模块分解

3.1 图像放缩网络

图像放缩增强模块首先采用 Inception 网络的思想对输入图像采用多个不同通路卷积，提取高分辨率图像不同尺度感受野的特征信息，卷积使用的卷积核采用 n*1 和 1*n 的非对称卷积来替代 n*n 的对称卷积，在保证相同感受野的同时减少模型的参数量。

在其之后对三个不同感受野的特征信息进行融合，可以得到更好的特征表达，使模型更加稳健，具体融合方式为将三种均为 m 通道的特征信息先进行 Concat 为 3×m 通道，再使用 1*1 的卷积将通道融合压缩至 m 通道，为了保证特征信息与短路连接中缩放的图像尺寸相同，将特征信息线性差值为相同大小。结构中的 Residual Block 结构如图 3 所示，通过堆叠残差结构增加网络深度，提取更深层次的信息，通过实验发现在 1 个 Residual Block 时即可达到较好的效果，因此本文将网络中 Residual Block 的个数设置为 1。

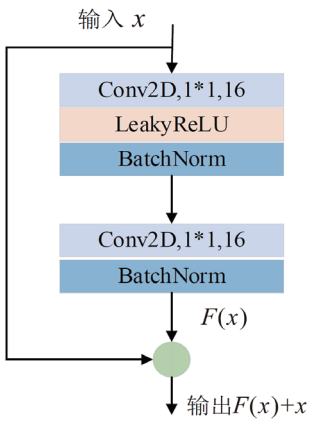

图 3 Residual Block 结构图

残差模块采用“短路连接”将输出特征层和之前的输入层进行融合，残差模块结构可以表示为：

$$\begin{aligned} y &= F(x) + x \\ F(x) &= W_2\sigma(W_1x) \end{aligned} \tag{2}$$

式中：x 为输入，y 为输出，$F(x)$ 为卷积输出，W_1、W_2 为两次卷积的权值，σ 表示 LeakyReLU 函数，输出是将输入与卷积输出结果相加之后融合得到。提取到的图像特征信息通过 SAM 结构，在图像空间中强化关键区域信息，弱化背景等无

用信息，最后与线性插值后的图像进行 add 操作后得到最终预处理之后的图像。

3.2 图像分类网络

EfficientNet 是由谷歌提出的一种高效且准确的卷积神经网络模型系列，是一种极具创新性的网络结构，它不仅具备快速推理的速度，而且在模型精度方面也表现出色。EfficientNet 系列网络中基础的网络模型 EfficientNet-b0 的网络结构主要由 16 个移动倒置瓶颈卷积（mobile inverted bottleneck convolution，MBConv）模块 [7] 构成。MBConv 模块结构如图 4 所示。

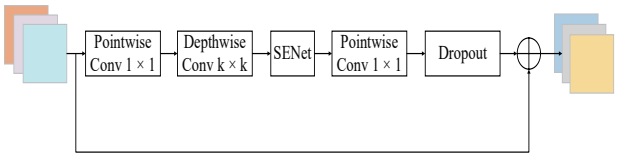

图 4 MBConv 结构图

在卷积神经网络中，特征信息感受野指的是某一层输出的特征图像素在输入图像上对应的感受野大小。特征信息感受野的大小决定了网络能够获取多大范围内的上下文信息，而上下文信息的多少直接影响网络的性能。然而单一空洞率的卷积反复叠加也存在着一些缺陷，当空洞率过大时，会损失信息的连续性，从而导致计算的低效性。

针对空洞卷积叠加引起的问题，本文引入混合空洞卷积 [8]（hybrid dilated convolution，HDC），其采用多种卷积率组合成一种新的卷积操作来扩大特征信息感受野，扩张率较小和较大的卷积分别可以捕获小物体、近距离的信息和大物体、远距离的信息。对于 EfficientNet 系列的图像分类任务而言，高级特征的提取是通过多个结构相似的 MBConv 模块进行提取的，因此本文采用混合空洞卷积对 MBConv 模块进行改进，具体操作为将深度可分离卷积中的深度卷积替换为扩充率为 {1,2,1,2} 循环锯齿状结构的空洞卷积。

3.3 分类损失函数

本文将图像放缩增强模块与图像分类网络联合训练时，采用的方法是监督分类中常用的交叉熵损失 [9]。

$$
\begin{aligned}
L_{CE} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} q_{ij}\log(p_{ij}) \\
q_{ij} &= \begin{cases} 0, if & y_i = j \\ 1, else \end{cases} \\
p_{ij} &= \frac{e^{a_{ij}}}{\sum_{j=1}^{M} e^{a_{ij}}}
\end{aligned}
\tag{3}
$$

对于 $D=\{(x_1,y_1),\ldots,(x_i,y_i),\ldots,(x_N,y_N)\}$ 表示的包含 N 幅图像，可划分为 M 个类别的数据集，其中 x_i 表示第 i 幅图像，y_i 表示第 i 幅图像的真实类别标签，q_{ij} 表示第 i 幅图像的真实类别标签是否为 j，log 表示以 e 为底的对数操作，a_{ij} 表示分类网络第 i 幅图像在第 j 类的输出。

4 实验结果与分析

4.1 实验设置

本文提出的分类方法沿用了图像分类任务中普遍采用的 ImageNet[10] 的预训练模型进行微调。本文在 ImageNet 的预训练 EfficientNet 模型的基础上，将图像放缩增强模块和分类网络模型联合进行训练，设置训练批样本大小（batchsize）为 32，数据集中 80% 的图像作为训练集，20% 的数据作为测试集。在不同下采样之后的图像数据集上进行实验，对比了添加图像放缩增强网络前后不同分辨率图像的分类准确率，同时也将本文分类方法与其他分类算法进行对比。

4.2 实验结果

图像放缩网络的放缩比例选定为 224→112、448→224 两种组合，224→112 表示将输入为 224×224 的图像通过 ISEN 缩放到 112×112 后送入分类网络。表 1 对比了在不同放缩条件下，图像放缩增强网络对于不同分辨率图像的分类实验结果。由实验结果可以看出，将 MBConv 改为混合空洞卷积并添加图像放缩增强网络之后的分类模型，相比使用线性插值的基准网络 EfficientNet-b0 对于不同分辨率数据集有 0.58% ～ 3.15% 不等的提升。在 8 倍的下采样时，准确率由 77.85% 提升至 81.07%，这说明本文提出的方法能够避免图像因线性插值过小而导致的信息丢失问题，有效将大输入图像中的信息通过 ISEN 放缩增强保留进小图像，且 HDC 可以增强分类网络的特征提取能力。对比同一数据集的不同下采样图像数据集的实验结果，可以发现随着图像下采样倍率的增加，本文方法对于分类性能提升更加明显。

表 1 不同分辨率 Oxford Flowers102 数据集实验结果表

模型	448→224		224→112	
	b0(224)	ours	b0(112)	ours
$\downarrow_1$	94.78%	**95.92%**	89.36%	**90.03%**
$\downarrow_2$	93.30%	**94.91%**	88.52%	**89.10%**
$\downarrow_4$	87.93%	**90.45%**	84.65%	**86.18%**
$\downarrow_6$	84.58%	**85.43%**	80.14%	**81.39%**
$\downarrow_8$	77.92%	**81.07%**	76.66%	**77.48%**

注：表中 b0(224) 指 EfficientNet-b0 输入为 224×224，$\downarrow_1$、$\downarrow_2$、$\downarrow_4$、$\downarrow_6$、$\downarrow_8$ 分别表示下采样 1、2、4、6、8 倍之后的图像数据集。

本文同时使用图像放缩增强网络与 EfficientNet-b5 的组合，对比其它先进的分类网络结构，结果如表 2 所示。结果显示，在更少的网络参数量或更少的计算量的情况下，本文方法可以达到相近分类结果，例如与 transformer 相比，在 7/10 参数和 3/5 的计算量的情况下，Oxford Flowers102 分类准确率仅相差 0.03%。实验结果表明，更改 MBConv 为混合空洞卷积，并添加图像放缩增强网络模块可以有效保留图像原始信息，同时可以提取表征性更高的特征，在少量增加模型参数和计算量下可以提升图像分类的精度。

表 2 本文方法与其它图像分类方法对比

模型	Acc	Parameters/MB	Flops/GB
b5	97.53%	28.5	10.6
Inception-v4[11]	98.50%	41	22.9
Transformer[12]	98.89%	86	49.4
GFNet[13]	98.80%	54	8.6
ours	98.86%	28.7	13.1

4.3 增强可视化

为了直观观察到本文方法对于图像的增强效果，本文可视化部分图像增强前后图像，对比图如图 5 所示。这些结果的共同特点是对图像的高频细节信息都有一定程度的增强，而这种增强效果会使分类模型更有效。总的来说，这些增强效果可能并不符合人类视觉的感知标准，但是它们会提升分类任务的准确率。

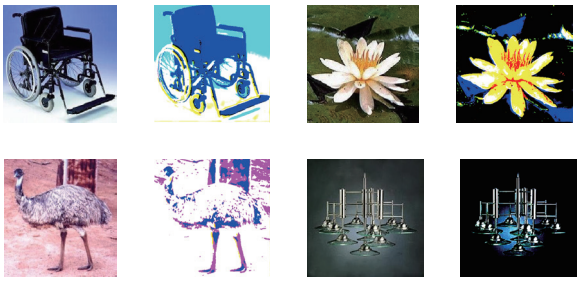

图 5 图像增强可视化效果

5 总结

本文针对图像分类任务，提出了基于图像放缩增强网络的分类方法，针对简单线性插值导致图像信息丢失限制网络的分类性能的问题，从网络结构本身和添加放缩增强模块两方面对图像分类网络进行了改进。首先设计了混合空洞卷积的 EfficientNet 网络结构，增大连续性的特征感受野的同时保证分类网络的推理速度；然后设计了基于注意力机制的图像预处理模块，将高分辨率的输入图像放缩为增强后的较低分辨率的输入图像，从而保证分类网络输入大小不变的情况下，即不增加分类网络计算量的情况下可以提取到更优异的特征；最后通过多组实验证明，本文提出的注意力机制的图像预处理网络对于分类网络的性能有着明显的提升。

参考文献：

[1] QIAO D, LIU G, LV T, et al. Marine vision-based situational awareness using discriminative deep learning: A survey[J]. Journal of marine science and engineering, 2021, 9(4): 1-18.

[2] MASANA M, LIU X, TWARDOWSKI B, et al.Class-incremental learning: survey and performance evaluation on image classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2022,45(5):5513-5533.

[3] 金玮，孟晓旻，武益超．深度学习在图像分类中的应用综述 [J]. 现代信息科技，2022, 6 (16): 29-31+35.

[4] PEHERSTORFER B, WILLCOX K, GUNZBURGER M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization[J]. Siam review, 2018, 60(3): 550-591.

[5] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes[C]//Sixth Indian Conference on Computer Vision, Graphics & Image Processing. Los Alamitos:IEEE Computer Society, 2008: 722-729.

[6] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence,2015,38(2): 295-307.

[7] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2:inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway:IEEE,2018: 4510-4520.

[8] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE winter conference on applications of computer vision (WACV).Piscataway:IEEE,2018: 1451-1460.

[9] SHORE J E, JOHNSON R E. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy[J].IEEE transactions on information theory,1980,26(1):26-37.

[10] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition.Piscataway:IEEE,2009: 248-255.

[11] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence.Palo Alto:AAAI Press,2017:4278-4284.

[12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL].(2020-10-22)[2024-02-05].https://arxiv.org/abs/2010.11929.

[13] WOODWORTH B, PATEL K K, STICH S, et al. Is local SGD better than minibatch SGD[C]//37th International Conference on Machine Learning,Part 14 of 15.New York:Curran Associates, 2020: 10334-10343.

【作者简介】

冯继凡（1999—），男，陕西咸阳人，硕士，助理工程师，研究方向：计算机视觉、图像分类。

杨清（1998—），男，陕西榆林人，硕士，助理工程师，研究方向：计算机视觉。

石昌鑫（1998—），男，陕西渭南人，硕士，助理工程师，研究方向：计算机视觉、图像美学。

（收稿日期：2024-03-04）