

基于多层特征融合机制的细粒度图像分类研究

李焱成

四川旅游学院 四川成都 610100

摘要: 本研究提出三种基于Bilinear CNN的细粒度图像分类改进模型, 采用多层特征融合机制。通过对VGG16基网络结构的改动, 分别使用相加、交叉和混合三种特征融合方式获得更丰富的图像信息, 增强了不同层和同一层中不同子层的信息互通。在此基础上, 引入全局池化、R-S激活函数和自学习权重特征融合等局部优化策略, 进一步提升特征提取和训练效率。实验结果表明, 新模型分类精确度达到86.94%。

关键词: Bilinear CNN; 细粒度分类; 特征融合; 全局池化; R-S激活函数

引言:

细粒度图像分类是计算机视觉领域中一个具有挑战性和研究价值的方向, 它需要对表现相似但存在微小差异的图像进行更精细的分类。为了克服细粒度图像分类任务中的难点, 如微小的局部差异、光照、姿态、无关背景和拍摄角度等因素对图像特征的干扰^[1], Lin^[2]等人提出了双线性卷积神经网络(Bilinear CNN), 该网络能够有效地提取较为丰富的特征信息, 但其缺点是模型参数量过多, 导致训练成本高, 并且识别精度不够理想。针对这些问题, 本研究在Bilinear CNN模型的基础上进行改进, 提出三种引入多层特征融合机制的改进模型, 并结合全局池化、R-S激活函数和自学习权重特征融合等局部优化策略, 实现了更好的细粒度图像分类效果。

一、三种改进模型网络结构设计

1. 多层特征融合机制

多层特征融合机制是指在多个浅层连续的卷积层后加入空间整体加权池化, 再与高层卷积层的输出做双线性池化, 使得网络能够同时获取高层语义和低层语义的分辨率位置信息, 从而提升目标的特征表示能力^[3]。

2. 多层特征融合机制-普通多层特征融合模型

为了提取更加丰富和多样化的特征图, 本节对BCNN中的VGG16网络结构进行了调整, 利用多层特征融合机制。设BCNN两路网络分别为A和B, X为输入图像, $g_{i,j}(X)$ 为第i层卷积层中j子层输出的特征图, $f_k(X)$ 为融合后的特征图。在A、B两路网络中分别选择 $g_{3,1}(X)$ 、 $g_{3,2}(X)$ 、 $g_{3,3}(X)$ 、 $g_{5,1}(X)$ 、 $g_{5,2}(X)$ 和 $g_{5,3}(X)$ 作为待

融合的特征图, 采用相加(\oplus 符号表示)的方式进行 $g_3(X)$ 和 $g_5(X)$ 的内部特征融合, 如下列公式所示:

$$f_3(X) = g_{3,1}(X) \oplus g_{3,2}(X) \oplus g_{3,3}(X) \quad (1)$$

$$f_5(X) = g_{5,1}(X) \oplus g_{5,2}(X) \oplus g_{5,3}(X) \quad (2)$$

相加这种特征融合方式是将对应的特征图逐元素相加, 再把不同的特征向量组成复合向量, 以得到更多的特征信息。如下列公式所示。其中, x是准备融合的特征图, y是融合后的特征图, h、w和c分别表示特征图的长宽和通道数, a、b和d表示卷积层的层数。

$$y_{h,w,c}^{add} = x_{h,w,c}^a + x_{h,w,c}^b + x_{h,w,c}^d \quad (3)$$

经过特征融合操作后, 分别在网络A和B中各做一次双线性池化, 即对相加提取的特征向量进行双线性池化, 得到双线性向量: 设 L_5 为CNN网络中的第五个卷积层, L_3 为第三个卷积层, 它们的通道数分别为 D_5 和 D_3 , 卷积层输出的特征(已通过相加操作获取)分别为 F_{L_5} 和 F_{L_3} , 取特征图 F_{L_5} 的所有通道在位置l的值形成一个 $1 \times D_5$ 的向量 $F_{L_5}(l, I)$, 取特征图 F_{L_3} 的所有通道在位置l的值形成一个 $1 \times D_3$ 的向量 $F_{L_3}(l, I)$, 计算 L_5 和 L_3 两个卷积层融合后在位置l处的融合特征, 其中l代表特征图中的一个位置, I代表输入图像, 该融合特征为一个大小 $D_5 \times D_3$ 的矩阵。同时网络A和B进行正常的双线性池化, 如下列公式所示:

$$G_A(l, l, F_{L_5}, F_{L_3}) = F_{L_5}(l, I)^T F_{L_3}(l, I) \quad (4)$$

$$G_B(l, l, F_{L_5}, F_{L_3}) = F_{L_5}(l, I)^T F_{L_3}(l, I) \quad (5)$$

$$G_{AB}(l, l, f_A, f_B) = f_A(l, I)^T f_B(l, I) \quad (6)$$

选用sum-pooling作为池化函数, 对三个双线性矩阵进行池化操作, 如下列公式所示:

$$P_A(l) = \sum_{i \in L} G_A(l, l, F_{L_5}, F_{L_3}) \quad (7)$$

$$P_B(l) = \sum_{i \in L} G_B(l, l, F_{L_5}, F_{L_3}) \quad (8)$$

$$P_{AB}(l) = \sum_{i \in L} G_{AB}(l, l, f_A, f_B) \quad (9)$$

作者简介: 李焱成(1995—), 男, 汉族, 四川省南充市人, 研究实习员, 农学硕士, 单位: 四川旅游学院, 研究方向: 计算机视觉、高校信息化建设。

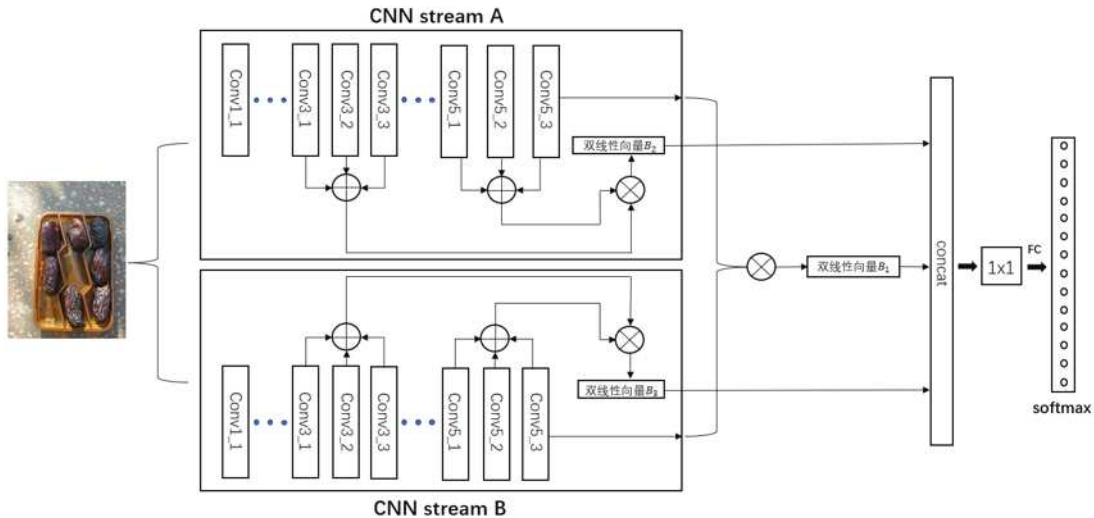


图1 基于多层特征融合策略改进后的BCNN网络结构

经过池化、resize、符号平方根变换和L2正则化后，得到三个双线性特征向量 B_1 、 B_2 和 B_3 。最后通过concat操作得到多层特征融合的双线性输出向量 F ，如下列公式所示：

$$F = \text{concat}(B_1, B_2, B_3) \quad (10)$$

改进后的BCNN模型融合了三组不同维度的特征，这些特征在每个维度中具有不同的语义信息和细节表达，通过与多个特征的交互，丰富了细粒度特征信息的学习，降低了中间层特征信息的丢失，提升了细粒度图像分类

的准确度。

3. 多层特征融合机制-交叉多层特征融合模型

本节提出第二种改进方案：交叉多层特征融合模型^[4]。将三个子层的特征两两交叉按像素比例相乘，然后再相加得到交叉特征融合图。这样可以增强不同层之间特征的关联，提高特征的多样性。另外，通过交叉多层特征融合的方式还可以突出图像中的主要前景目标，消除背景光和噪音的影响，提高定位性能。

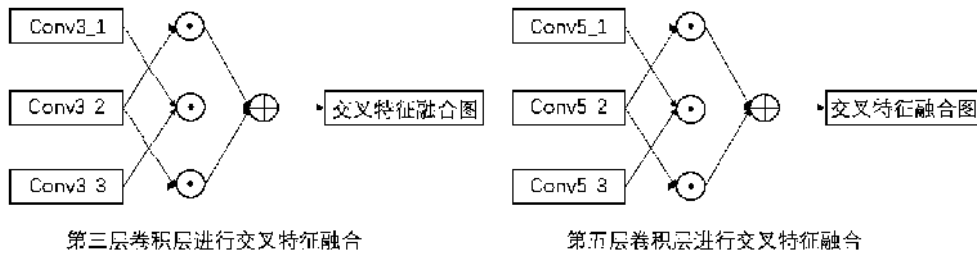


图2 交叉多层特征融合

4. 多层特征融合机制-混合多层特征融合模型

本节提出第三种改进方案：混合多层特征融合模型^[5]。在保留原来网络A和网络B中Conv3层和Conv5层的内部交叉特征融合的基础上，选择将网络A中的Conv1_2层特征图

与网络B中的Conv1_2层特征图、网络A中的Conv3_3层特征图与网络B中的Conv3_3层特征图进行混合融合后输入到下一层卷积中，同理，对于网络B也选择和上面同层次的特征图进行混合融合，然后输入到下一层卷积中。

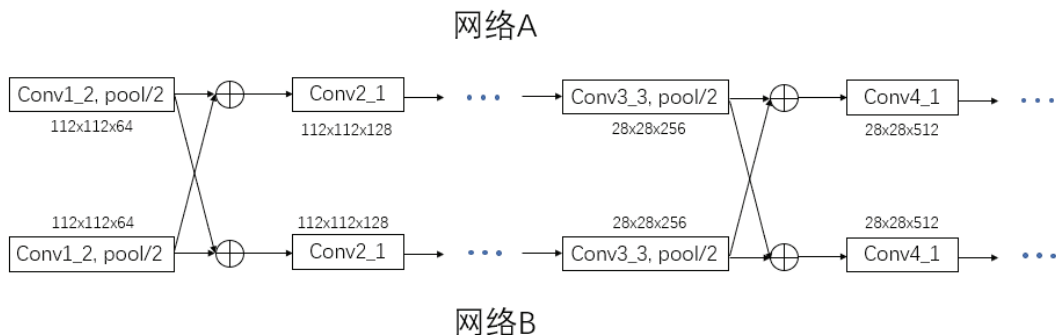


图3 混合多层特征融合

通过不同网络不同层的混合特征融合再配合上一节中同网络中的交叉特征融合, 更加有效地提高了特征的表达能力。

二、局部优化策略

1. 全局池化

为了减少BCNN的参数数量, 提高模型的训练速度和拟合能力, 本节用全局池化层替代原有的全连接层部分, 将整个特征图进行平均或最大池化, 使得输出特征图的通道数等于类别数。这样可以让每个通道都具有类别含义, 提高特征的表达力^[6]。

2. R-S激活函数

为了解决 ReLu 激活函数在 $x < 0$ 时梯度为 0, 导致部分神经元不更新的问题, 本节提出了一种新的激活函数

R-S, 结合了 ReLu 和 Softplus 的特性^[7]。R-S 激活函数表达式如下列公式所示:

$$f(x) = \max [\ln(e^x + 1).x + 0.8] \quad (11)$$

R-S 激活函数 $x > 0$ 时保留了 ReLu 函数的稀疏性和快速收敛的优势, 在 $x \leq 0$ 时采用了 Softplus 函数的平滑性, 避免了梯度消失的问题。同时, R-S 激活函数整体向上平移了 0.8 个单位, 使得输出特征图中的每个通道都具有一定的正值, 增强了特征的表达力。

3. 自学习权重特征融合

本节提出了一种自学习权重特征融合的方法 (SLWFF), 实现了底层特征和高层特征的融合^[8]。该方法通过给不同层级的特征自动分配不同权重, 获取不同特征的影响因子, 让模型自主判断底层特征对分类的贡献。

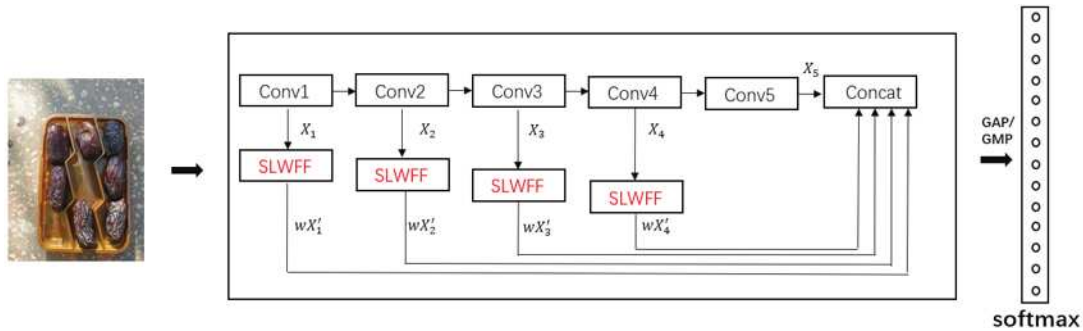


图3 基于自学习权重特征融合的网络结构图

每一层的特征图经过该层的卷积层后都会进入 SLWFF 模块做加权操作, 然后将所有加权后的底层特征 wX'_1 到 wX'_{n-1} 和高层特征 X_n 通过 concat 的方式进行融合, 最后利用 softmax 得到分类结果。

SLWFF 模块的具体步骤如下:

- 对输入的特征 X_1 进行最大池化操作, 得到 X'_1 , 减少计算开销。池化核的大小为 $k_h \times k_w$, 步长为 s , $H'_1 = (H_1 - k_h)/s$, $W'_1 = (W_1 - k_w)/s$;

- 将 X'_1 的尺寸调整为 $H''_1 \times W''_1 \times 1$, 即一张特征图, 这里 $H''_1 \times W''_1 = H'_1 \times W'_1 \times C_1$;

- 采用 1×1 大小的卷积核对调整后的特征进行卷积操作, 得到 $H''_1 \times W''_1 \times 1$ 大小的加权特征 wX'_1 ;

- 调整 wX'_1 的维度和高层特征 X_n 的维度相同, 其中 $C = H''_1 \times W''_1$;

- 使用上一节提出来的 R-S 激活函数, 防止过拟合和保持非线性。

具体的计算过程如下列公式所示:

$$wX'_1 = RS(\text{resize}(\text{conv}(\text{resize}(\text{pool}(X)))))) \quad (4-13)$$

三、数据集准备

本研究在 Fruits-360 数据集上进行测试。该数据集包含 90483 张图片, 分为 131 种水果和蔬菜类别, 其中 67692 张为训练图像, 22791 张为测试图像, 并附有 103 张多水果集合图像。

四、实验结果与分析

本研究比较了 BCNN 及其四种改进模型在分类准确率上的表现, 如表 1 所示。这四种改进模型分别是基于多层特征融合机制的三种模型和基于自学习权重特征融合模型。结果表明, 多层特征融合机制能够提升 BCNN 的分类效果。

表1 准确率对比

网络模型	准确率 (%)
BCNN 模型	84.46
普通多层特征融合模型	85.63
多层交叉特征融合模型	86.16
混合特征融合模型	86.41
自学习权重特征融合模型	86.94

为了验证不同子层交叉特征融合的有效性, 本研究

在同一层卷积层中尝试了所有可能的融合方式，并进行了对比实验，如表2所示。结果表明，将三个子层的特征图按两两元素相乘后再相加的方法具有最高的分类准确率，说明多层交叉特征融合机制能够提高细粒度图像的认识精度。

表2 不同子层的交叉特征融合方式准确率对比

交叉特征融合方式	准确率 (%)
$F_1 \odot F_1 + F_2 \odot F_2 + F_3 \odot F_3$	84.95
$F_1 \odot F_2 + F_2 \odot F_3 + F_3 \odot F_3$	85.84
$F_1 \odot F_2 + F_2 \odot F_2 + F_2 \odot F_3$	85.91
$F_1 \odot F_1 + F_2 \odot F_3 + F_1 \odot F_3$	85.87
$F_1 \odot F_2 + F_1 \odot F_3 + F_2 \odot F_3$	86.31

五、结语

本研究在BCNN模型的基础上，提出了三种多层特征融合模型，即普通、交叉和混合三种方式。实验结果显示，混合方式能够最大程度地利用特征信息，提高细粒度分类的准确率。本研究还用全局池化层替代全连接层，用新的激活函数增强激活效果，并用自学习权重策略融合底层和高层特征，进一步优化了分类的性能。

参考文献：

[1]罗建豪，吴建鑫.基于深度卷积特征的细粒度图像分类研究综述[J].自动化学报，2017，43（08）：1306-1318.

[2]Lin T Y , Roychowdhury A , Maji S .Bilinear CNN Models for Fine-grained Visual Recognition[J]. 2015.DOI: 10.48550/arXiv.1504.07889.

[3]X. Cai et al., “Automatic Extraction of Layover From InSAR Imagery Based on Multilayer Feature Fusion Attention Mechanism,” in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 4017705, doi: 10.1109/LGRS.2021.3105722.

[4]邓亚萍，王新，尹甜甜.基于多层交叉注意力融合网络模型的人脸图像情感分析[J].科学技术与工程，2023，23（3）：1152-1159.

[5]蔡斐.基于混合注意力机制和多层特征融合的服装图像检索研究[D].东华大学，2022.DOI: 10.27012/d.cnki.gdhuu.2022.000703.

[6]孙俊，何小飞，谭文军，等.空洞卷积结合全局池化的卷积神经网络识别作物幼苗与杂草[J].农业工程学报，2018.DOI: CNKI: SUN: NYGU.0.2018-11-020.

[7]陆明远，侯春燕，王劲松.基于Softplus函数的神经网络的Reluplex算法验证研究[J].信息安全研究，2022，8（09）：917-924.

[8]许志龙，施焯，俞鑫春，郭伟.基于自学习特征权重网络的模板匹配研究[J].电子技术与软件工程，2020（14）：139-142.