

# 基于数据增强的秦俑碎片深度分类模型

鱼跃华, 张海波, 李昕, 寇姣姣, 李康, 耿国华, 周明全\*

西北大学信息科学与技术学院, 陕西 西安 710127

**摘要** 在秦俑保护领域, 为了降低秦俑碎片匹配及拼接的工作难度, 更多的计算机辅助技术应用在破碎秦俑复原工作核心环节的碎片分类中。针对传统的秦俑碎片分类方法对碎片特征提取不充分及秦俑碎片数据采集难度较高等导致的分类准确率低下的问题, 提出了一种基于数据增强的秦俑碎片深度分类模型。首先, 通过条件生成式对抗网络对现有秦俑碎片数据集进行数据增强, 实现秦俑数据集的扩充。其次, 通过深度卷积神经网络自动且充分地提取碎片特征信息并实现有效的碎片分类效果。然后, 引入 convolutional block attention module (CBAM) 双通道注意力机制和 CutMix 增强策略来显著提升深度分类模型的性能。最后, 在秦俑实验数据集的对比实验结果表明, 与传统的基于几何特征、尺度不变特征变换特征、形状特征、多特征融合等经典碎片分类方法相比, 所提方法对秦俑碎片的分类取得了更准确的分类结果, 有效降低了后续复原工作中匹配、拼接等工作的复杂度, 进而提高了秦俑文物复原工作的整体效率。

**关键词** 图像处理; 破碎秦俑复原; 碎片分类; 条件生成式对抗网络; 双通道注意力机制; 增强策略

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1810010

## Data Enhanced Depth Classification Model for Terracotta Warriors Fragments

Yu Yuehua, Zhang Haibo, Li Xin, Kou Jiaojiao, Li Kang, Geng Guohua, Zhou Mingquan\*

College of Information Science and Technology, Northwest University, Xi'an 710127, Shaanxi, China

**Abstract** In the field of Terracotta Warriors protection, to reduce the challenge of matching and splicing fragments of the Terracotta Warriors, more computer-aided technology is applied to the core link's debris classification in the restoration of the broken Terracotta Warriors. The classification accuracy is low because of insufficient characteristic extraction of traditional Terracotta Warriors debris classification approaches and increased difficulty associated with data collection. In this paper, a depth classification model of Terracotta Warriors fragments based on data enhancement is presented. First, the existing dataset of Terracotta Warriors fragments was improved using conditional generative adversarial nets to achieve the dataset's expansion of Terracotta Warriors. Second, the deep convolutional neural network was employed to automatically and effectively extract the debris feature information and achieve an effective debris classification effect. Third, the double-channel attention mechanism of the convolutional block attention module (CBAM) and the CutMix enhancement strategy were effectively introduced to significantly improve the deep classification model's performance. Results on the experimental dataset of the Terracotta Warriors reveal that the presented approach is more accurate than the traditional classical debris classification approaches based on geometric, scale-invariant feature transform, and shape features as well as the multifeature fusion. It can effectively reduce the subsequent restoration work's complexity, such as matching and stitching, and therefore improve the overall efficiency of the Terracotta Warriors' restoration work.

**Key words** image processing; restoration of broken Terracotta Warriors; classification of debris; conditional generative adversarial nets; two-channel attention mechanism; enhancement strategy

收稿日期: 2021-06-09; 修回日期: 2021-07-02; 录用日期: 2021-07-28

基金项目: 国家重点研发计划 (2019YFC1521103, 2019YFC1521102)、国家自然科学基金青年基金 (61902317)、国家自然科学基金重点项 (61731015)、陕西省重点产业链项目 (2019ZDLSF07-02)、陕西省自然科学基金青年基金 (2019JQ-166)、青海省重点研发计划 (2020-SF-142)

通信作者: \*mqzhou123@126.com

## 1 引言

经过了千年之久的自然损坏及人为破坏,秦俑被挖掘问世以来,大多数都是以不完整的破损碎片形式呈现的且数量庞大<sup>[1-2]</sup>。而通过人工修复这些秦俑碎片,不仅效率低下,且人工操作很有可能对文物造成二次损坏。因此如何高效率地修复这些数量庞大的碎片,使其拼接成完整的兵马俑,成为了新的研究课题<sup>[3]</sup>。借助计算机辅助技术,研究人员可以更高效地对兵马俑进行数字化修复,实现了手工作业到数字化、智能化的跨越<sup>[4]</sup>。破碎兵马俑碎片的修复工作一般包括:碎片分类、碎片匹配和碎片拼接<sup>[5-6]</sup>。碎片分类作为非常关键的核心环节之一,有效且准确的碎片分类效果可以大幅降低后续修复工作的工程复杂度,提高匹配以及拼接的准确性,提高文物复原过程的整体效率。

近年来,在秦俑碎片分类方面,魏阳<sup>[7]</sup>提出了一种高效的基于支持向量机(SVM)的兵马俑碎片分类方法,该方法将尺度不变特征变换(SIFT)特征和Hu不变矩结合,利用多特征动态加权融合的算法表征秦俑碎片图像,以融合后的碎片特征实现各部位秦俑碎片的分类。康馨月<sup>[8]</sup>提出了一种高效的秦俑碎片分类方法,该方法基于快速拒绝模型匹配和先验知识,利用局部区域内的显著特征构建碎片三维模型的多层描述子,通过迭代方法计算待分类三维碎片模型与待匹配模型的距离,实现各类秦俑碎片的分类。但该方法过度依赖于几何特征,当对平滑碎片分类时,该方法匹配的准确率较低。王娜<sup>[9]</sup>提出了一种流形正则化多核模型的半监督分类方法。该方法针对单一特征表征数据能力低的局限引入多核函数,将多种特征进行融合;同时依据多核学习的高数据映射能力和灵活的特征组合能力,为半监督的多分类算法提供数据依据,从而提高秦俑分类器的分类效果。贺格<sup>[10]</sup>提出了一种基于图的秦俑碎片半监督协同分类方法,该方法将在秦俑碎片多视图特征上学习得到的多个分类器进行协同训练,利用碎片数据多视图特征之间的差异性,合理优化互补,协助提高每个分类器的分类准确率和泛化能力,但是该方法并没有考虑其在多个监督分类器的协同训练的结果是否理想。

卷积神经网络(CNN)可以直接以原始图像作为输入,然后从大量的样本来学习相应的特征,从而避免了复杂的特征提取过程,而且可以提取出更加抽象精细化的特征,取得了更加优越的图像分类性能<sup>[11-12]</sup>。大量的实验结果表明,通过大规模带有标签的数据集来训练卷积神经网络,训练后的网络可以取得高精度的识别率及分类效果。随着深度学习的快速发展,更高性能、更深层次的网络模型相继被提出,从最初的LeNet<sup>[13]</sup>到AlexNet<sup>[14]</sup>、ZFNet<sup>[15]</sup>、VGGNet<sup>[16]</sup>、GoogleNet<sup>[17]</sup>、ResNet<sup>[18]</sup>、SENet<sup>[19]</sup>及DenseNet<sup>[20]</sup>,这些经典的网络

模型在图像分类、物体识别、图像分割等计算机视觉方面的应用领域都取得了革命性的创新成果。近年来,图像分类领域发展迅速,针对深度网络结构复杂、参数量较大的问题,马永杰等<sup>[21]</sup>提出了一种基于稠密连接网络进化的卷积神经网络(D-ECNN)图像分类算法,该算法的模型文件较小、速度更快,分类效果显著。为解决基于深度学习的高光谱图像分类方法对于小样本数据分类精度低的问题,张祥东等<sup>[22]</sup>提出了一种基于多尺度残差网络的分类模型,该模型通过构造基于光谱特征和空间特征的提取模块,实现了空间特征和光谱特征的多尺度融合提取,可充分利用高光谱图像中丰富的空谱信息实现更加高效的分类效果。针对三维卷积网络在训练样本较少时对高光谱图像的分类精度不理想问题,冯凡等<sup>[23]</sup>提出了一种高效的基于多特征融合和混合卷积网络的分类模型,该模型对高光谱图像进行降维处理后用三维卷积层提取深层空谱联合特征,然后引入残差连接并通过特征图连接和逐个像素相加进行多特征融合,实现特征重用、增强信息传递,有效提升了小样本情况下的高光谱图像分类效果。

在提高卷积神经网络性能方面,以往的研究工作主要集中在深度、宽度、基数等三个网络因素方面<sup>[24]</sup>。人类视觉系统的一个重要特性是人们不会试图同时处理整个场景。相反,为了更好地捕捉视觉结构,利用一系列局部瞥见并选择性地聚焦于显著部分区域。因此,根据这样的特性, Park 等<sup>[25]</sup>提出了 bottleneck attention module(BAM)注意力机制, Woo 等<sup>[26]</sup>提出了 convolutional block attention module(CBAM)双通道注意力机制。注意力不仅告诉我们应该关注哪里,还能改善兴趣的表现。利用注意机制可以增加表征权力,关注对象的重要特征,抑制不必要的特征<sup>[27-28]</sup>。随着人们对深度卷积神经网络更高训练效率和质量的追求,更多的数据增强及正则化的训练策略被不断提出,例如 Dropout<sup>[29]</sup>、Mixup<sup>[30]</sup>、Cutout<sup>[31]</sup>、CutMix<sup>[32]</sup>等优化策略可用于进一步提高深度卷积神经网络模型的性能。不论是加深网络的深度,还是更先进的模型优化方法,都在深度网络模型性能优化方面的工作提供了有效的帮助。

针对传统的秦俑碎片分类方法对碎片表面特征提取不充分及秦俑碎片数量庞大、形状不规则、数据采集难度较高、表面特征复杂等导致的分类准确率低问题,本文提出了一种基于数据增强的秦俑碎片深度分类模型。该模型以 ResNet18 作为基础架构,并集成了 CBAM 双通道注意力机制和 CutMix 增强策略两大轻量级优化模块作为分类模型。为了降低训练集样本数量过少对训练结果造成的不良影响,基于条件生成式对抗网络的数据增强方法对秦俑碎片数据集进行了数据集扩展。数据增强不仅可以提升数据集样本的数量及样本的多样性,还可以使得碎片数据集规模满足一般深度卷积神经网络模型训练需要的数据集规模需求。

## 2 基于数据增强的兵马俑碎片深度分类模型

深度卷积神经网络需要在大规模带有标签的数据集下训练,训练后的模型能够取得高的识别率或好的分类效果。由于多方面因素的制约及采集文物数据工作的特殊性,目前累积的秦俑碎片数据集样本远小于一般深度卷积网络模型训练所需要的数据集规模。这就涉及了样本数据的数量问题,小规模的数据集可能会导致训练后的深度卷积神经网络出现过拟合的现象。因此,采用数据增强的方法扩展数据集样本数量是解决网络过拟合等问题的可行方法之一。对比传统分类方法,深度卷积神经网络可以避免复杂的特征提取过程,且能提取出更加抽象精细化的特征,以获得更优的分类效果,CBAM 双通道注意力机制和 CutMix 增强策略两大轻量级的优化模块可以进一步优化以 ResNet18 为基础架构的秦俑碎片分类模型的性能和泛化性能。

### 2.1 条件生成式对抗网络数据增强

Mirza 等<sup>[33]</sup>提出的条件生成式对抗网络(CGAN)模型可以有效进行数据增强,该模型在生成与样本类似的新图像的同时,还能够提升生成样本的多样性。本实验组采用 CGAN 对秦俑碎片图像数据进行数据

增强,实现了让有限的秦俑碎片图像数据样本产生更多的等价图像,从而达到扩展秦俑碎片二维图像数据集的目的,使得秦俑数据样本数量满足一般深度卷积网络模型训练所需要的数据集规模。GAN<sup>[34]</sup>中包括生成器  $G$  和鉴别器  $D$ ,其中生成器  $G$  负责拟合学习训练样本的数据分布,鉴别器  $D$  负责判别数据是来自训练样本还是来自生成器  $G$  模仿训练样本生成的数据,依赖于神经网络的强大的拟合能力,由噪声数据,生成器  $G$  可以生成类似的真实图,实现数据增强的目的。生成器  $G$  和判别器  $D$  进行二元极大极小博弈,损失函数的表达式为

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} \left\{ \log \left\{ 1 - D[G(z)] \right\} \right\}, \quad (1)$$

式中: $x$  为真实数据; $z$  为随机噪声。但是这样  $G$  最后生成的数据是不可控制的,只是和训练样本中的数据类似而已,而不能指定生成什么的数据。本实验组通过指定生成秦俑数据 B1 胸脯、B2 胳膊、B3 手、B4 裙摆、B5 腿、B6 脚及踏板等六大类样本,给原始的 GAN 生成器  $G$  和判别器  $D$  添加额外的条件信息  $y$ ,实现条件生成模型 CGAN。基于该方法的秦俑碎片数据增强模型结构如图 1 所示。

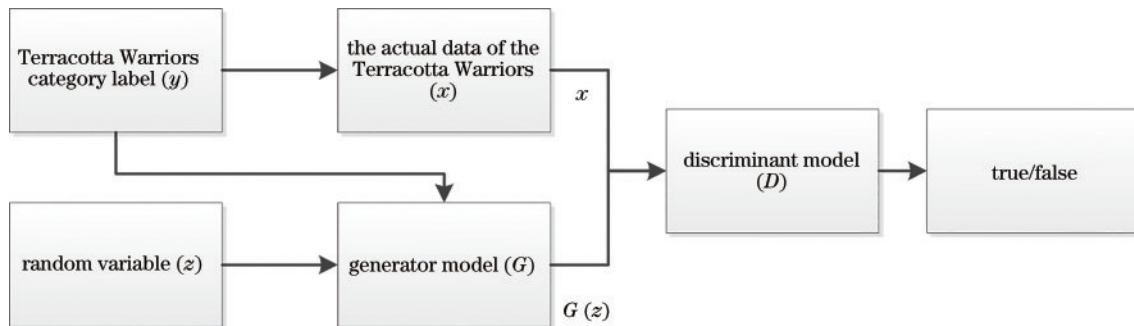


图 1 秦俑碎片数据增强模型结构示意图

Fig. 1 Schematic diagram of Terracotta Warriors fragment data enhancement model structure

使用类别标签信息  $y$  控制  $G$  生成的数据,指定生成不同类别。CGAN 添加的额外信息  $y$  只需要和真实数据  $x$  与随机噪声  $z$  进行合并,作为生成器  $G$  和  $D$  的输入即可,由此 CGAN 的损失函数的表达式为

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} \left\{ \log \left\{ 1 - D[G(z|y)] \right\} \right\}. \quad (2)$$

由式(2)可知,相比 GAN 的目标函数,CGAN 的  $G$  和  $D$ ,其概率表达变成了条件概率表达式,即该目标函数只有在输入为  $y$  标签的时候成立。对于不同的类别标签  $y$ ,有不同的目标函数,通过式(2)更改标签,就可以得到自己需要的生成图像,从而根据秦俑碎片数据集的类别,选择不同的类别标签  $y$  指定类别生成图像,实现对秦俑数据集更高效、高准确的数据增强。基于 CGAN 的新样本的生成过程的具体步骤如下:

- 1) 确定秦俑碎片图像的样本类别标签  $y$ 。
- 2) 基于选定的类别标签  $y$ ,将随机噪声  $z$  和类别标签  $y$  映射到具有校正线性单元(ReLU)激活的隐藏层。再将训练数据  $x$  和类别标签  $y$  映射到隐藏层。
- 3) 将上述两个隐藏层联合映射到最大输出层后馈送到最终的 sigmoid 层作为输出,用于生成秦俑碎片数据的新样本  $G(z)$ 。该算法生成的部分部位新样本与原样本对比如图 2 所示。

本实验组通过该方法将原有二维秦俑碎片数据集按照 1:10 的比例进行数据集增强扩充,最终得到了包含 16000 张的二维秦俑碎片图像数据集。由于数据量较大,需对数据增强后的数据集进行标签处理,使得秦俑碎片数据集可以满足后续深度学习分类模型的数据规模要求。

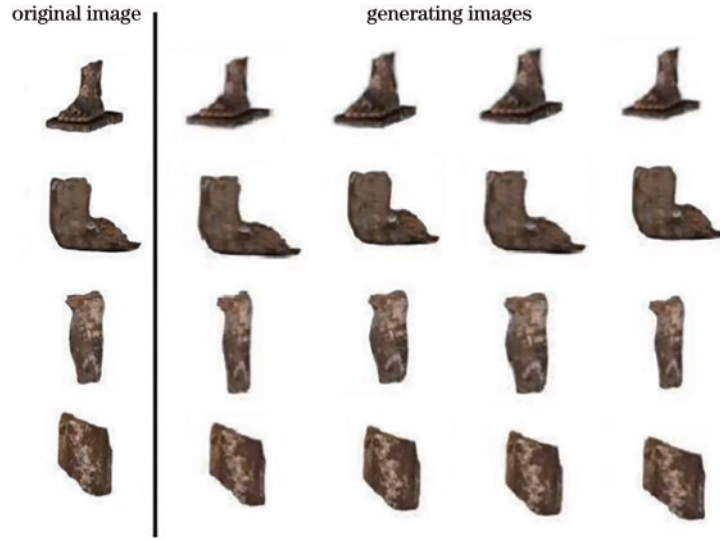


图2 部分部位新样本与原样本对比图

Fig. 2 Comparison between new samples and original samples in some parts

### 2.2 CBAM 双通道注意力机制

将注意力机制引入 CNN 图像分类任务中,可以改善分类任务的性能。所提深度 CNN 分类模型把双通道注意力机制 CBAM 集成到分类模型中,提升分类任务的性能。CBAM 是一种简单而有效的前馈卷积神经网络注意模块。在给出一个中间特征映射后, CBAM 模块沿着通道和空间两个维度依次推导出注

意力映射,然后将注意力映射乘到输入特征映射上进行自适应的特征细化。由于 CBAM 是一个轻量级的模块且具有广泛适用性,它可以无缝地集成到任何 CNN 架构中,显著提升深度 CNN 分类任务的性能,而开销可以忽略不计,并且可以与基本的 CNN 一起进行端到端的训练。图 3 为 CBAM 模块结构示意图。

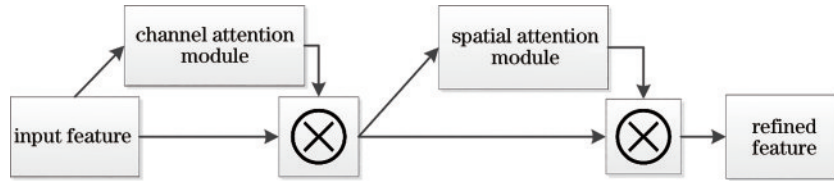


图3 CBAM 模块结构示意图

Fig. 3 Structure diagram of CBAM module

CBAM 模块包括通道注意力和空间注意力两个顺序子模块。由于卷积操作将跨通道注意力和空间注意力信息混合在一起提取信息特征,模块 CBAM 强调沿着通道和空间这两个主要维度的有意义的特征。CBAM 具体算法如下:给定中间特征图  $F \in \mathbf{R}^{C \times H \times W}$  作为输入, CBAM 依次推导出 一维通道注意力图  $M_c \in \mathbf{R}^{C \times 1 \times 1}$  及二维空间注意力图  $M_s \in \mathbf{R}^{1 \times H \times W}$ 。整个注意力过程可以描述为

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (3)$$

式中:  $\otimes$  表示元素的乘法。在乘法过程中,注意值被相应地传播:通道注意值沿着空间维度传播,空间注意

值沿着通道维度传播。 $F''$  是最终的精细化输出。

图 4 为通道注意模块。通道子模块在共享网络中同时利用最大池化输出和平均池化输出,通过平均池化和最大池化操作来聚合特征图的空间信息,生成两种不同的空间上下文描述符:  $F_{Avg}^c$  和  $F_{Max}^c$ , 分别为平均池化特征和最大池化特征。然后,这两个描述符被转发到一个共享网络,从而产生通道注意力映射  $M_c \in \mathbf{R}^{C \times 1 \times 1}$ 。共享网络由多层感知器(MLP)组成,其中包含一个隐含层。为了减少参数开销,隐藏的激活层大小设置为  $\mathbf{R}^{C/r \times 1 \times 1}$ , 其中  $r$  是缩减率。当共享网络应用于每个描述符后,使用元素的求和来合并输出的特征向量。通道注意力可描述为

$$M_c(F) = \sigma \left\{ \text{MLP} \left[ \text{AvgPool}(F) \right] + \text{MLP} \left[ \text{MaxPool}(F) \right] \right\} = \sigma \left\{ W_1 \left[ W_0(F_{Avg}^c) \right] + W_0(F_{Max}^c) \right\}, \quad (4)$$

式中:  $\sigma$  表示 sigmoid 函数;  $W_0 \in \mathbf{R}^{C/r \times C}$ ;  $W_1 \in \mathbf{R}^{C \times C/r}$ 。两个输入共享 MLP 权重  $W_0$  和  $W_1$ , ReLU 激活函数

后接  $W_0$ 。

图 5 为空间注意模块。空间子模块利用类似最大

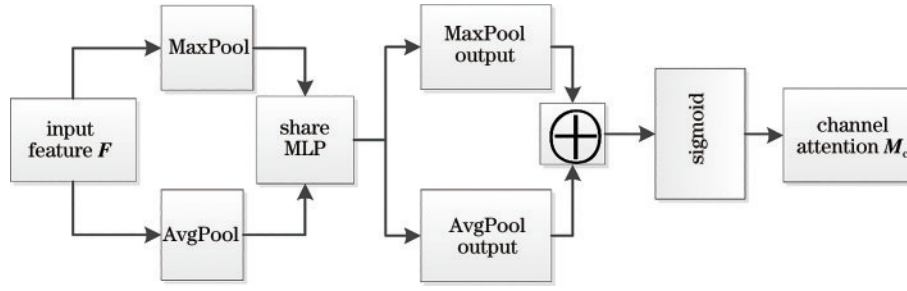


图 4 通道注意模块

Fig. 4 Channel attention module

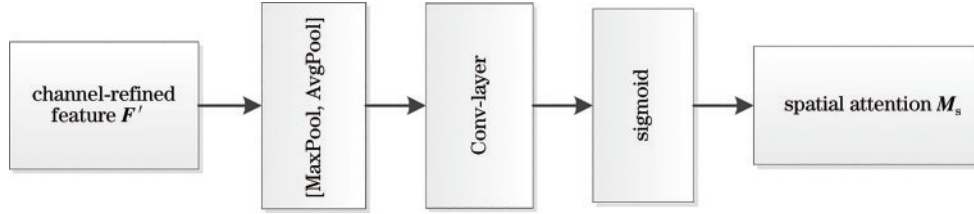


图 5 空间注意模块

Fig. 5 Spatial attention module

池化输出和平均池化输出的两个输出,沿着通道轴汇集,并将它们转发到卷积层;通过两个池化操作来聚合一个特征映射的通道信息,生成两个二维映射: $F_{\text{Avg}}^s \in \mathbf{R}^{1 \times H \times W}$ 和 $F_{\text{Max}}^s \in \mathbf{R}^{1 \times H \times W}$ ,表示通道上的平均池特征和最大池特征;然后通过一个标准的卷积层将这些信息连接起来并进行卷积,就产生了二维空间注意力图。空间注意力可描述为

$$M_s(\mathbf{F}) = \sigma \left\{ f^{7 \times 7} \left[ \left[ \text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F}) \right] \right] \right\} = \sigma \left[ f^{7 \times 7} (\mathbf{F}_{\text{Avg}}^s; \mathbf{F}_{\text{Max}}^s) \right], \quad (5)$$

式中: $f^{7 \times 7}$ 表示卷积核大小为 $7 \times 7$ 的卷积运算。

在网络模型中,通道注意力和空间注意力模块可以通过并行或者顺序的方式放置,但是顺序排列比平行排列可以得到更好的性能提升效果。CBAM作为能够提高深度卷积神经网络表示能力的方法,基于注意力的特征精细化方法,通过通道注意和空间注意两个不同的模块,在保持开销较小的情况下,可以获得相当大的性能提升。因此,本实验组将CBAM引入秦俑碎片深度分类模型中,使之成为网络的重要组成部分,可对分类任务的性能带来显著的提升。

### 2.3 CutMix 增强策略

区域 Dropout 方法通过覆盖黑色像素或随机噪声的补丁来去除训练图像上的信息像素,让模型不仅关注对象的最有区别的部分,而且关注整个对象区域,产生更好的分类性能并提高模型泛化性能。但这种移除会导致训练期间的图像信息出现丢失且效率低下。虽然区域 Dropout 策略在一定程度上提高了分类性能,通过置零或用随机噪声填充删除的区域,大大降低了信息像素在训练图像中的比例,但仍存在一些问题。CutMix 不是简单地删除像素,而是用另一张训练图片

上的区域来替换被删除的区域,地面真实标签也按组合图像的像素数比例混合;可以最大限度地利用删除的区域,保持区域 Dropout 方法的优势。

CutMix 策略的优点如下:1)在训练过程中不会出现非信息像素,能够提高训练效率,且保留了区域 Dropout 的优势,能够关注目标的非重视的部分;2)要求模型从局部视图识别对象,对裁剪区域中添加其他样本的信息,能够进一步增强模型的定位能力;3)不会有图像混合后不自然的情形,能够提升模型分类的表现;4)在训练过程中,CutMix 只会产生微不足道的额外训练成本,生成可能包含多个对象的组合样本提升训练样本的多样性。

在 CNN 内部特征上添加噪声或在结构上添加额外路径可以提高图像分类性能。CutMix 策略是对上述方法的补充,因为它在数据层上操作,而不改变模型内部表示或网络体系结构。基于上述理论,所提秦俑碎片深度分类模型引入 CutMix 策略,来进一步增强碎片分类模型的性能,提高分类的准确率。CutMix 具体算法如下:设 $\mathbf{x} \in \mathbf{R}^{C \times H \times W}$ 和 $y$ 分别表示训练图像及其标签。CutMix 的目标是通过结合两个训练样本 $(\mathbf{x}_A, y_A)$ 和 $(\mathbf{x}_B, y_B)$ 生成一个新的训练样本 $(\tilde{\mathbf{x}}, \tilde{y})$ 。生成的训练样本 $(\tilde{\mathbf{x}}, \tilde{y})$ 用其原始的损失函数来训练模型。即

$$\begin{cases} \tilde{\mathbf{x}} = \mathbf{M} \odot \mathbf{x}_A + (1 - \mathbf{M}) \odot \mathbf{x}_B \\ \tilde{y} = \lambda y_A + (1 - \lambda) y_B \end{cases}, \quad (6)$$

式中:二进制掩码 $\mathbf{M} \in \{0, 1\}^{W \times H}$ ,指从两个图像中删除和填充的位置,1是用一个填充的二进制掩码; $\odot$ 表示元素乘法;两个数据点之间的组合比 $\lambda$ 是从 $\beta$ 分布 $\beta(\alpha, \alpha)$ 中采样的。将 $\alpha$ 设置为1,即 $\lambda$ 是从均匀分布

(0, 1)中采样的。CutMix用另一个训练图像的补片替换一个被删除的图像区域。并且生成更多的局部自然图像。

为了对二进制掩码  $M$  进行采样,首先对边框坐标  $B=(r_x, r_y, r_w, r_h)$  进行采样,这表示  $x_A$  和  $x_B$  上的裁剪区域。 $x_A$  是中的区域  $B$  被移除并用  $x_B$  的  $B$  中切下的补丁填充。采样矩形掩模  $M$ ,其长宽比与原始图像成正比。盒坐标按下列方法进行均匀采样,

$$\begin{cases} r_x \sim \text{Unif}(0, W), r_w = W \sqrt{1-\lambda} \\ r_y \sim \text{Unif}(0, H), r_h = H \sqrt{1-\lambda} \end{cases}, \quad (7)$$

式中:  $\text{Unif}(\cdot)$  表示均匀采样。裁切面积比  $\frac{r_w r_h}{WH} = 1 - \lambda$ 。对于裁切的区域,二进制掩码  $M \in \{0, 1\}^{W \times H}$  由边框  $B$  内填充 0 决定,否则为 1。在每次训练迭代中,根据式(6)将随机选择的两个训练样本组合到一个批量中,生成一个 CutMix 样本  $(\tilde{x}, \tilde{y})$ 。CutMix 很容易实现,它的计算开销可以忽略不计,而且在各种分类任务上都非常有效。所提秦俑碎片深度分类网络模型引入

了 CutMix 增强策略,来提升模型的性能。

### 2.4 兵马俑碎片图像深度分类模型

所提深度分类模型将 CBAM 双通道注意力机制和 CutMix 增强策略两大优化模块集成到 ResNet18 中,组成了优化的 ResNet18 网络模型,来高效、高性能、高准确率地完成秦俑碎片分类任务。图 6 为集成 CBAM 的残差块结构示意图,从图 6 可以看出通道注意力模块和空间注意力模块集成在 ResBlock (残差块) 中的确切位置。对秦俑碎片深度分类模型 ResNet18 每个块的卷积输出都集成了双通道注意力机制 CBAM 模块。秦俑碎片的分类流程如图 7 所示。

从图 7 可以看出:所提模型首先对原有秦俑数据集通过 CGAN 数据增强的方法实现秦俑数据集的扩充;其次对扩充后的数据集进行统一的标签处理;然后将秦俑碎片数据集输入集成 CBAM 双通道注意力机制模块和 CutMix 增强策略模块的 ResNet18 分类架构中,得到完整基于数据增强的秦俑碎片深度分类模型,并通过训练测试得出最佳的分类结果。

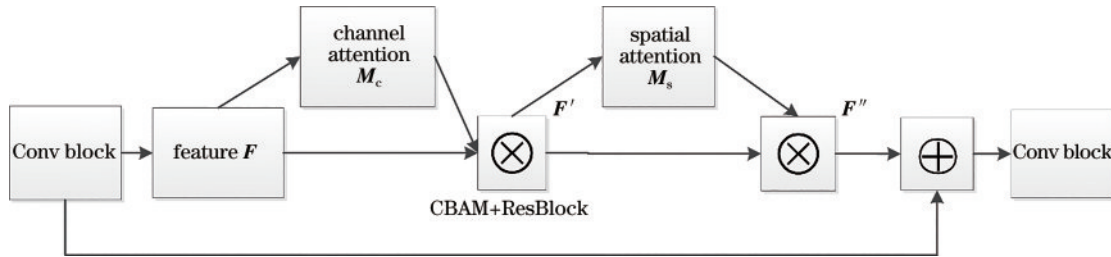


图 6 集成 CBAM 的残差块结构示意图

Fig. 6 Residual block structure diagram of integrated CBAM

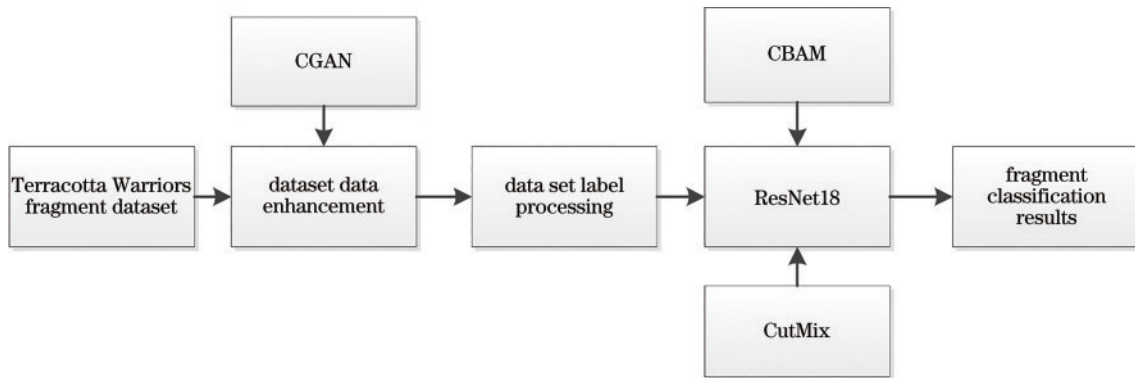


图 7 秦俑碎片分类流程图

Fig. 7 Classification flow chart of TerraCotta Warriors and horses fragments

## 3 实验结果及分析

### 3.1 实验设置

#### 3.1.1 实验环境

实验环境基于 Tensorflow-gpu1.14.0、pytorch1.2.0、torchvision0.4.0 实现秦俑碎片深度分类模型,硬件环境为 Intel Core i7 处理器、16 GB 内存、2 TB 硬盘、

Geforce 2080ti 显卡,软件环境为 Windows 10 操作系统,使用 Pycharm 编程。

#### 3.1.2 秦俑实验数据集

以作者所在文化遗产数字化国家地方联合工程研究中心所采集并保存的秦俑碎片图像数据作为实验数据集,所提 ResNet18 分类优化模型也是在该秦俑碎片图像实验数据集上进行训练及评估的。原始数据集包

含六大类秦俑碎片图像,碎片图像为  $512 \times 256$  的 RGB 彩色图像,输入 ResNet18 深度网络时将其裁切为  $256 \times 256$  的尺寸,分别包括 B1 胸脯(300)、B2 胳膊(300)、B3 手(300)、B4 裙摆(300)、B5 腿(100)、B6 脚及踏板(300)等共计 1600 张图像,通过 CGAN 进行数据增强处理后,实验数据集所包含的秦俑碎片图像数量变为 16000 张,并对数据增强后的数据集进行标签

处理。适当增大数据量可以增加深度卷积神经网络秦俑碎片分类模型的性能。将实验数据集按照五折交叉验证随机划分为训练集和测试集,即每次从数据集中随机选出 12800 张图像作为所提分类模型的训练集,剩余的 3200 张图像作为分类模型的测试集。图 8 给出了兵马俑碎片图像实验数据集的部分样本图像。



图 8 碎片样本图像

Fig. 8 Fragment sample images

### 3.2 训练过程

所提深度分类网络的训练方式为有监督训练,ResNet18 网络对 B1 胸脯、B2 胳膊、B3 手、B4 裙摆、B5 腿、B6 脚及踏板这六类进行分类,所以最终的分类器将得到一个结果为 6 维向量的分类标签。标签第  $K$  维中的“1”表示秦俑碎片的分类结果,否则为“0”。

ResNet18 主要包括两个阶段的训练过程,首先将预处理后的秦俑碎片图像输入 ResNet18 中,并通过计算得出碎片图像分类的标签。当得到的分类结果与预期不相符时,则需要对网络的权值进行初始化,紧接着计算网络输出的分类标签与实际分类标签两者之间的损失值。

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - \sum_{c=1}^M y_{ic} \log p_{ic}, \quad (8)$$

式中:  $M$  是类别的数量,数据集中碎片类别为 6 类,所以  $M=6$ ;  $y_{ic}$  是指示变量,如果该类别和样本  $i$  的类别相同就是 1,否则就是 0;  $p_{ic}$  是样本  $i$  属于类别  $c$  的预测概率。

如果得到的损失值不在容许范围内,则利用误差最小化原则对各神经元的权值进行更新,即

$$M(i+1) = \alpha \cdot \beta(i) \cdot N(i) + M(i), \quad (9)$$

式中:  $M(i+1)$  表示第  $i+1$  个训练样本的权值;  $\alpha$  表示网络学习速率;  $\beta$  为神经元的误差项;  $N$  表示神经元的输入;  $M(i)$  表示第  $i$  个训练样本的权值。当 ResNet18 的损失值小于期望值时,停止训练。

训练过程中参数批量大小设置为 8,优化器学习率设置为 0.25,模型训练周期为 200。网络结构与具体参数如表 1 所示。

表 1 网络结构及参数  
Table 1 Network structure and parameters

Layer	Output size	Size of conv kernel	Output channels	Stride
Input layer	$256 \times 256$		3	
conv1	$128 \times 128$	$7 \times 7$	64	2
max pool	$128 \times 128$	$3 \times 3$	64	2
conv2_x	$64 \times 64$	$\begin{bmatrix} 3 \times 3 \\ 3 \times 3 \end{bmatrix} \times 2$	$\begin{bmatrix} 64 \\ 64 \end{bmatrix} \times 2$	1
conv3_x	$32 \times 32$	$\begin{bmatrix} 3 \times 3 \\ 3 \times 3 \end{bmatrix} \times 2$	$\begin{bmatrix} 128 \\ 128 \end{bmatrix} \times 2$	1
conv4_x	$16 \times 16$	$\begin{bmatrix} 3 \times 3 \\ 3 \times 3 \end{bmatrix} \times 2$	$\begin{bmatrix} 256 \\ 256 \end{bmatrix} \times 2$	1
conv5_x	$8 \times 8$	$\begin{bmatrix} 3 \times 3 \\ 3 \times 3 \end{bmatrix} \times 2$	$\begin{bmatrix} 512 \\ 512 \end{bmatrix} \times 2$	1
Fc_6d	$1 \times 1$		6	

### 3.3 实验结果与分析

准确率是分类问题中最常见的指标,该指标指的是所有样本中分类正确的样本数量占总样本数量的比例。本实验组采用平均准确率(Accuracy-Avg)作为评价指标,Accuracy-Avg 对所有分类类别的准确率取算术平均值。在兵马俑碎片分类实验中,将设置两组实

验。第 1 组实验主要对比秦俑碎片传统分类方法 SIFT 特征、形状特征、SIFT 特征和形状特征及显著几何特征等四种方法与所提采用深度学习的分类方法的结果。实验结果如表 2 所示。

表 2 传统分类方法与所提分类方法效果对比

Table 2 Comparison of effects of traditional classification methods and proposed classification method

Method	Accuracy-Avg / %
SIFT Feature	78.42
Shape Feature	67.55
SIFT+Shape Feature	84.41
Salient geometric Feature	71.32
ResNet18+CBAM+CutMix	88.69

从表 2 可以看出,所提秦俑碎片深度分类模型所取得的分类结果明显优于传统方法。由于秦俑碎片形状不规则,表面纹理特征复杂,因此单独使用 SIFT 特征或者使用形状特征的方法,对于识别部分秦俑碎片胸脯及裙摆等比较相似的碎片的概率比较低,且显著几何特征不能有效识别表面平滑的碎片,因此分类效果不是很好。基于多种特征融合的办法可以有效提高碎片分类准确率,相较于以上四种传统的分类方法,本实验组将深度学习的方法引入秦俑碎片的分类工作中,实验结果表明,所提方法可以不仅自动提取秦俑碎片特征信息,且能更高效地提取更多有用的特征信息,显著提高了秦俑碎片分类准确率。

第 2 组实验比较了集成优化模块前的 ResNet18 模型应用于秦俑碎片分类的实验结果、集成了 CBAM 双通道注意力机制的 ResNet18 模型应用于兵马俑碎片分类的实验结果及所提秦俑碎片深度分类模型,即集成了 CBAM 双通道注意力机制及 CutMix 增强策略的 ResNet18 模型应用于秦俑碎片分类的实验结果。具体结果如表 3 所示。

表 3 ResNet18 优化前后分类方法效果对比

Table 3 Comparison of effects of classification methods before and after optimization of ResNet18

Method	Accuracy-Avg / %
ResNet18	84.76
ResNet18+CBAM	86.21
ResNet18+CBAM+CutMix	88.69

从表 3 可以看出,将深度学习的方法引入秦俑碎片分类工作中能够带来更高效、更准确的分类结果。相比加入 CBAM 双通道注意力机制前的 ResNet18 的分类结果,在加入 CBAM 后,ResNet18 模型的代表权力增加,模型在训练时更关注秦俑碎片的重要特征,抑制对分类结果无作用的不必要特征,增加的计算开销可以忽略不计,且使得秦俑碎片的分类结果得到了进一步的提升。而所提模型在 ResNet18 模型中同时集

成了 CBAM 双通道注意力机制及 CutMix 增强策略两大优化模块。CutMix 策略作用于数据层通过用另一张秦俑训练图像上的区域来替换秦俑训练图像被删除的区域,让模型不仅关注秦俑碎片最有区别的部分,而且关注整个秦俑碎片的图像区域,CutMix 增强策略模块还通过小批量剪切和粘贴区域生成新的秦俑碎片图像样本,而且引入的计算开销同样可以忽略不计,并使得模型能够产生更好的分类性能和泛化性能。所提方法与对比方法收敛性对比如图 9 所示。

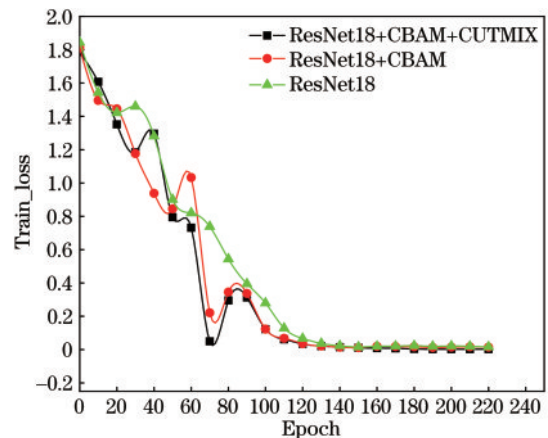


图 9 收敛性对比

Fig. 9 Contrast of convergence

从图 9 可以看出:在前 100 次的迭代训练中,所提模型的收敛性强于其他两个网络;在之后的 200 次迭代中,三个网络都逐渐趋于平缓,基本稳定,但所提模型的损失函数在迭代了 220 次后损失值降到了 0.003,而 ResNet18+CBAM 的损失值只降到了 0.015,ResNet18 损失值只降到了 0.017。即所提模型在秦俑碎片图像分类任务中取得了更高准确率、更高性能的秦俑碎片分类结果。此外,在 CIFAR-10 公共数据集上,对所提模型进行了测试,结果如表 4 所示。

表 4 ResNet18 优化前后在 CIFAR-10 分类结果对比

Table 4 Comparison of classification results of ResNet18 in CIFAR-10 before and after optimization

Method	Accuracy-Avg / %
ResNet18	89.25
ResNet18+CBAM	91.21
ResNet18+CBAM+CutMix	91.73

从表 4 可以看出,ResNet18 优化模型在公共数据集 CIFAR-10 上所取得的准确率高于在秦俑数据集的准确率。对比秦俑数据集,公共数据集 CIFAR-10 拥有更多的数据量及更丰富的样本类别,包括 10 种类别,各自 5000 张图像及 10000 张测试集图像,总计 60000 张图像。该数据集数据量更大且样本多样性丰富,因此在训练过程中可以得到更优的训练模型。实验结果表明,针对秦俑碎片的分类工作,进一步采集更



多类别的秦俑碎片图像数据来扩充秦俑碎片数据集的数据量及丰富该数据集的多样性,可以作为后续进一步提高分类准确率的有效方法之一。

## 4 结 论

针对秦始皇陵破碎秦俑碎片的复原工作中的碎片分类工作,手动提取特征效率低下、特征提取不充分、分类效率及准确率不佳等缺点,且由于秦俑文物碎片数据采集工作的特殊性,碎片数量较少等缺点,提出了一种基于数据增强的秦俑碎片深度分类模型。所提模型通过条件生成式对抗网络对现有的碎片数据进行有效的数据增强,使得秦俑碎片数据集规模满足一般深度网络训练所需要的数据集规模需求。将深度学习的方法引入秦俑碎片图像的分类工作中,深度卷积神经网络 ResNet18 通过卷积层和池化层可以更加充分且自动地提取秦俑碎片的特征,且在 ResNet18 分类网络基本保持计算开销不变的情况下引入 CBAM 双通道注意力机制和 CutMix 增强策略两大优化模块,这有效提升了分类模型的性能及泛化性能。实验结果表明,相比传统的秦俑碎片分类方法,所提方法更先进且具有更有效、更高准确率的秦俑碎片分类结果。将该分类模型应用于秦俑碎片的分类工作中,相较以往的秦俑碎片传统分类方法,为秦俑修复工作提供了一种高效率、高准确率的碎片分类方法,能够有效降低后续复原工作中的匹配、拼接等工作的复杂度,提高了秦俑复原工作的整体效率。希望所提方法有机会可以在其他类型文物保护领域的工作中得到应用,为洽谈文物碎片的修复复原工作,提供有力的帮助,贡献微薄之力。

未来将继续采集秦俑碎片数据来丰富实验的数据集,提升数据集的数据量及数据样本的多样性,使得现有的碎片样本数据集规模满足深度卷积神经网络模型训练所需要的数据集规模需求。并继续探索网络深度或宽度的加深是否能够带来更高准确率的分类结果。

## 参 考 文 献

- [1] 田静. 世界遗产视野下的秦陵秦俑研究[J]. 秦始皇帝陵博物院, 2018(00):1-17.  
Tian Jing. Study on the Terracotta Warriors of Qin Mausoleum from the Perspective of World Heritage [J]. Emperor Qin Shihuang Mausoleum Museum, 2018(00): 1-17.
- [2] 李岗, 田亚岐, 肖健一, 等. 2008—2017 年陕西秦汉考古综述[J]. 考古与文物, 2018, 229(5): 66-110.  
Li Gang, Tian Y Q, Xiao J Y, et al. Review of archaeological discoveries and research of the Qin-Han periods in Shaanxi(2008-2017) [J]. Archaeology and Cultural Relics, 2018, 229(5): 66-110.
- [3] 兰德省. 秦俑一号坑出土陶俑修复粘接剂的筛选与应用研究[J]. 文物保护与考古科学, 2019, 31(5): 49-59.  
Lan D S. Research on adhesive materials for restoration of painted pottery figurines excavated from Pit No.1 of the Qin Terracotta Warriors and Horses[J]. Sciences of Conservation and Archaeology, 2019, 31(5): 49-59.
- [4] 汤涛. 三维技术在物质文化遗产“修复”中的应用研究[D]. 南京: 南京艺术学院, 2018.  
Tang T. Research on the application of three-dimensional technology in material cultural heritage “restoration” [D]. Nanjing: Nanjing University of the Arts, 2018.
- [5] 赵思仲, 侯妙乐, 李爱群, 等. 融合多特征的兵马俑碎片分类技术研究[J]. 地理信息世界, 2019, 26(5): 14-21.  
Zhao S Z, Hou M L, Li A Q, et al. A classification technology of terra-cotta warriors' fragments with multiple features[J]. Geomatics World, 2019, 26(5): 14-21.
- [6] 李伟民. 面向破损文物虚拟修复系统的设计与实现[D]. 西安: 西北大学, 2020.  
Li W M. Design and implementation of virtual restoration system for damaged cultural relics[D]. Xi'an: Northwest University, 2020.
- [7] 魏阳. 基于多特征的兵马俑碎片图像分类及检索技术研究[D]. 西安: 西北大学, 2018.  
Wei Y. Research on image classification and retrieval technology of terra-cotta warriors fragments based on multi-feature[D]. Xi'an: Northwest University, 2018.
- [8] 康馨月. 计算机辅助秦俑碎片三维模型分类方法研究[D]. 西安: 西北大学, 2015.  
Kang X Y. Research on method of computer aided 3D models of terra-cotta warriors fragments classification[D]. Xi'an: Northwest University, 2015.
- [9] 王娜. 基于稀疏图的流形正则化多核半监督分类研究[D]. 西安: 西北大学, 2018.  
Wang N. Research on multi-kernel semi-supervised classification of manifold regularization based on sparse graphs[D]. Xi'an: Northwest University, 2018.
- [10] 贺格. 基于图的秦俑碎片半监督协同分类研究[D]. 西安: 西北大学, 2019.  
He G. Research on semi-supervised collaborative classification of terra-cotta warriors fragments based on graph[D]. Xi'an: Northwest University, 2019.
- [11] 王瑶瑶. 基于卷积神经网络优化模型的兵马俑碎片分类方法研究[D]. 西安: 西北大学, 2019.  
Wang Y Y. Research on the classification algorithm of terracotta warrior fragments based on the optimization model of convolutional neural network[D]. Xi'an: Northwest University, 2019.
- [12] 吴正文. 卷积神经网络在图像分类中的应用研究[D]. 成都: 电子科技大学, 2015.  
Wu Z W. Application research of convolution neural network in image classification[D]. Chengdu: University of Electronic Science and Technology of China, 2015.
- [13] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [15] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[M]//Fleet D, Pajdla T, Schiele

- B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8689: 818-833.
- [16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2021-05-06]. <https://arxiv.org/abs/1409.1556>.
- [17] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [19] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [20] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [21] 马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法[J]. 激光与光电子学进展, 2020, 57(24): 241001.  
Ma Y J, Liu P P. Convolutional neural network based on DenseNet evolution for image classification algorithm[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241001.
- [22] 张祥东, 王腾军, 杨耘. 基于多尺度残差网络的小样本高光谱图像分类[J]. 激光与光电子学进展, 2020, 57(16): 162801.  
Zhang X D, Wang T J, Yang Y. Classification of small-sized sample hyperspectral images based on multi-scale residual network[J]. Laser & Optoelectronics Progress, 2020, 57(16): 162801.
- [23] 冯凡, 王双亭, 张津, 等. 基于多特征融合和混合卷积网络的高光谱图像分类[J]. 激光与光电子学进展, 2021, 58(8): 0810010.  
Feng F, Wang S T, Zhang J, et al. Hyperspectral images classification based on multi-feature fusion and hybrid convolutional neural networks[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810010.
- [24] Mei S H, Jiang R Q, Ji J Y, et al. Invariant feature extraction for image classification via multi-channel convolutional neural network[C]//2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), November 6-9, 2017, Xiamen, China. New York: IEEE Press, 2017: 491-495.
- [25] Park J, Woo S, Lee J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018-07-17) [2021-05-06]. <https://arxiv.org/abs/1807.06514>.
- [26] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [27] 王雷全, 褚晓亮, 秦智超, 等. 基于多层次图像表征的层次化注意力图像描述[J]. 中国电子科学研究院学报, 2020, 15(1): 63-68.  
Wang L Q, Chu X L, Qin Z C, et al. Hierarchical attention for image captioning with multi-level image representations[J]. Journal of China Academy of Electronics and Information Technology, 2020, 15(1): 63-68.
- [28] 刘嘉敏, 谢文杰, 黄鸿, 等. 基于空间和通道注意力机制的目标跟踪方法[J]. 电子与信息学报, 2021, 43(9): 2569-2576.  
Liu J M, Xie W J, Huang H, et al. Spatial and channel attention mechanism method for object tracking[J]. Journal of Electronics & Information Technology, 2021, 43(9): 2569-2576.
- [29] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [30] Zhang H Y, Cisse M, Dauphin Y N, et al. mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25)[2021-05-08]. <https://arxiv.org/abs/1710.09412>.
- [31] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-08-15)[2021-05-08]. <https://arxiv.org/abs/1708.04552>.
- [32] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6022-6031.
- [33] Mirza M, Osindero S. Conditional generative adversarial nets[EB/OL]. (2014-11-06) [2021-06-08]. <https://arxiv.org/abs/1411.1784>.
- [34] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.