DOI: 10.13382/j.jemi.B2003449

分层双线性池化图像行为识别方法*

吴 伟 于嘉乐
(内蒙古大学 计算机学院 呼和浩特 010021)

摘 要:基于图像的行为识别由于受到类内图像背景信息的差异性和类间行为的相似性影响,至今仍然是一项极具挑战性的任务。某些行为类别在人物姿态、表情动作方面十分相似,因此对图像中各种富含语义信息的部位提取显著性特征对于提高行为识别的精度至关重要。借鉴双线性池化模型在细粒度分类中的优势,同时为避免该模型包含大量背景噪声而影响识别精度,提出一种改进的双线性池化模型用于图像行为识别。该模型利用通道和空间注意力机制关注图像中的重要目标,并通过集成多层注意力掩码图来生成 RoI,这可以有效抑制图像中的背景噪声信息,提高行为识别的准确性。最终提出的方法在 Stanford-40 dataset 获得了 85.24%准确率,同时在自定义的 60 类行为数据集上获得了 84.57%的准确率。

关键词: 图像行为识别;分层双线性池化;注意力机制;掩码聚合
中图分类号: TP37 **文献标识码**:A **国家标准学科分类代码**: 520.20

Hierarchical bilinear pooling method for image-based action recognition

Wu Wei Yu Jiale
(School of Computer, Inner Mongolia University, Hohhot 010021, China)

Abstract: Image-based action recognition is still a very challenging task because it is disturbed by the differences in the background information of the images in the class and the similarity of the behavior between the classes. Some action categories are very similar in terms of human poses and facial expressions, so extracting salient features from various parts of the image that are rich in semantic information is essential to improve the accuracy of action recognition. Drawing on the advantages of the bilinear pooling model in fine-grained image classification, and to avoid this model which containing a lot of background noise to affect the recognition accuracy, an improved bilinear pooling model is proposed for action recognition in the paper. The model uses channel and spatial-wise attention mechanism to focus on the important targets in the image, and generates RoI by integrating multi-layer attention mask, which can effectively suppress the background noise information in the image and improve the accuracy of action recognition. Our method achieves the accuracy of 85.24% on the Stanford-40 dataset, and the accuracy of 84.57% on the custom 60 kind of action dataset.

Keywords: image-based action recognition; hierarchical bilinear pooling; attention mechanism; mask aggregation

0 引 言

近年来,随着深度学习在计算机视觉领域中的成功应用,许多研究人员尝试利用卷积神经网络来处理图像中人类行为识别的问题。尽管目前已经取得了巨大的进步,但是仍然存在许多挑战。相比于视频行为识别[1-3],因为缺乏时间流信息,所以基于图像的行为识别方法通常只能提取图片的空间特征进行分类检测。尽管卷积神经网络可以有效地的提取图像中的有效信息,但是大量研究发现某些人体行为在身体姿势,外观上较为相似,神经网络往往不能很好的发现这些细小差别的图像特征。

为了处理这些细微的视觉差异,许多研究人员通过定位目标区域或者提取显著性特征以提高识别的精度。基于部件特征的模型[4-7]通常利用神经网络的卷积特征或者训练带有完全监督的边界框或区域注释的检测器进

收稿日期: 2020-08-31 Received Date: 2020-08-31
*基金项目:国家自然科学基金(61763035)、内蒙古自然科学基金(2020MS06006)资助项目

行设计。最近,文献[8-12]利用双线性池化模型对目标对象的局部显著性特征进行建模,并且在许多细粒度视觉任务中获得了出色的表现。其中,文献[12]提出了一个可以集成多个层间特征交互的双线性池化模型(hierarchical bilinear pooling, HBP),与文献[11]的方法相比,获得了卓越的性能。此外,注意力机制很容易学习通道和空间的权重,可以进一步发现显著性特征的位置,然后将较大的权重分配给显著性区域,从而增强模型的特征表达能力。为了更好地捕获图像中人类行为动作的有效信息,本文提出了一种分层注意力双线性池化模型,该模型采用通道和空间注意力机制挖掘不同卷积层的图像特征,增强模型对显著性特征的提取能力。

因为双线性池化方法需要在图像的所有区域(包括嘈杂的背景信息)上进行特征交互,所以减少有害背景信息的干扰并提取有用的(region of interest, RoI)特征信息对于行为识别十分重要。为了解决这一问题,一些研究人员利用手工标注的方式(如目标提取框或人体关键点)来提取前景目标以减少背景区域的干扰。但是这些方法代价昂贵,在实际生活中无法被广泛应用。因此。本文提出了一种具有掩码聚合的 Hierarchical Attention Bilinear Pooling 模型,用来提取 RoI 特征。

在两个具有挑战性的图像行为识别数据集(Stanford-40[13]和自定义的 60 类行为数据集)上评估了模型的性能。通过广泛的实验,模型在 Stanford-40 数据集上取得了最好的表现,平均精度(mean average precision, mAP)达到了 85.24%,而且在自定义的 60 类行为数据集上也达到了 84.57%的准确率。实验结果表明本文提出的方法要优于其他模型。

1　相关工作

近些年来,尽管图像行为识别的研究有所增加,但是仍然存在许多问题需要解决。大多数的工作都试图利用人物交互,身体部位或者人体姿态等高级线索来处理图像行为识别任务。在基于图像上下文信息的方法中,文献[13]利用了对象-动作交互的上下文信息进行行为检测。为了消除背景干扰,文献[14]介绍了利用分割潜在对象的交互区域的方法,从而准确地定位出前景对象。此外,现有的方法通常还需要在训练和测试阶段输入带有人工标注的目标提取框,这在表征人与物体之间的相互作用方面起着至关重要的作用。文献[15]采用候选目标对象框来寻找合适的交互对象。尽管已经获得令人满意的结果,但是人工标注仍然非常耗时而且成本昂贵。由于人体部位具有丰富的语义信息,文献[5]将姿势推断任务引入到原始的卷积神经网络(CNN),利用姿势估计作为辅助任务来提高动作识别的准确性。文献[6]结合 CNN 和 Poselets 进行人类行为和属性分类,通过利用滑动窗口对人体部位进行检测。但是姿势信息需要在身体的关键点上添加大量的注释,既费时又费力。

本文提出一种不需要使用昂贵的手工注释方法。相反,使用的双线性池化方法可以直接从图像中提取大量的显著性特征。仅仅使用单个卷积层的特征,通常不能完全表示对象的全部特征,因此采用整合多个中间卷积层的特征来有效地增强图像整体的特征表示。但是大量实验证明,直接提取卷积层的特征往往不能充分捕获图像里面重要的信息,因此一些研究人员试图通过引入注意力机制以获得更好的特征表示。注意力在人类视觉系统中起到非常重要的作用,人们可以快速的扫描整个场景并找到需要关注的区域,然后从目标区域中捕获更加详细的信息。受到人类视觉系统的启发,将注意力模块添加到分层双线性池化模型中,以提高模型的特征表示能力。

2　方法

注意力机制主要应用于空间和通道维度,以增强模型对显著性特征的提取能力;通过整合多个注意力特征图生成聚合掩码来减少有害背景信息的干扰。模型的总体架构如图 1 所示。

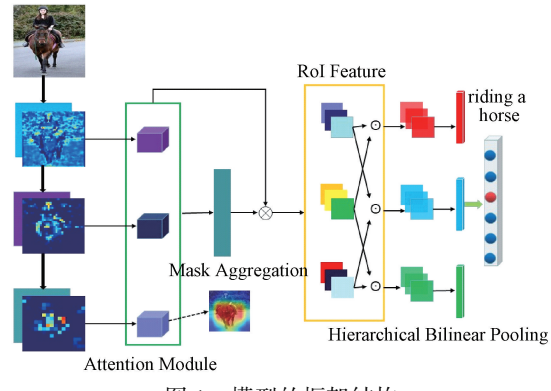


图 1　模型的框架结构

Fig. 1　Illustration of our model architecture

图 1 中,左边为 CNN 中不同层的特征图。将不同层的特征图分别输入到注意力模块中生成注意力激活图,以获取更加详细的特征,该模型通过集成多个注意力特征掩码图来提取感兴趣区域的特征,⊙代表元素点乘运算。

2.1　注意力模型

1)通道注意力

通道注意力的主要作用是发现重要的通道特征并为其添加权重,以增强有用的通道特征并压缩无用的通道信息。为了更加有效地计算通道注意力,需要压缩输入

的特征图。在之前的工作中,文献[16]使用全局平均池化的方法来计算注意力特征图的空间信息。尽管全局平均池化(GAP)可以充分利用通道的空间信息,而且可以减少网络参数防止过拟合现象的发生;但是全局最大池化(GMP)也可以利用全局最大响应来选择重要的通道信息。文献[17]证明了利用GMP和GAP的特征融合方法比单独使用全局平均池化更加有效。因此,本文也采用GAP和GMP信息融合的方法来构建通道注意力模块。

对于输入的图像特征图 $F^{W\times H\times C}$,首先使用全局平均池化和最大池化操作压缩特征维度,生成两个不同的特征图 F_{GAP} 和 F_{GMP},然后分别将两个特征图输入到两个卷积核大小为1的卷积层中,以实现通道维度的变化和信息融合。为了减少参数,中间特征通道的数量被设置为 $R^{C/r}$,其中 r 是缩小率。在实验中,将 r 设置为2的效果最佳。在经过两个1×1卷积层之后,我们使用逐像素相加的方式来合并输出的特征向量。最后,将合并后的特征通过Sigmoid激活函数,获得通道注意力特征图 M_C。

由于部分通道信息可能会在GAP和GMP操作过程中丢失,受到ResNet[18]残差学习的启发,设计了与注意力模块的连接,以保护原始输入信息的完整性。因此,最终的通道注意力模块的计算公式为:

$$F'_C = (1 + M_c) \otimes F = (1 + \sigma(F_{GAP} + F_{GMP})) \otimes F \tag{1}$$

式中:σ 表示Sigmoid函数;+表示逐元素加法;⊗为逐元素乘法。

2)空间注意力

通道注意力模型主要关注图片中"what"具有意义,而空间注意力主要发现"where"是一个信息丰富的区域。由于同一通道的不同像素点对分类结果的重要性不同,因此空间注意力模块的作用是为特征图的每个像素分配权重,以增强有用区域的特征并减弱无用的区域信息。为了有效的计算空间注意力,采用和通道注意力模块相同的方法,利用GAP和GMP来获得两个2D特征图,为 F_{sGAP} 和 F_{sGMP},之后将二者相加并通过一个大小为3的卷积层。最后,将注意力特征图与原始特征图相乘,得到最原始的空间注意力特征 M_S。类似地,将空间注意力特征作为侧分支添加到初始的图像特征。最终的空间注意力模块的输出特征表示为:

$$F'_S = (1 + M_S) \otimes F = (1 + \sigma(f^{3\times 3}(F_{sGMP} + F_{sGAP}))) \otimes F \tag{2}$$

式中:$f^{3\times 3}$ 表示卷积核大小为3的卷积运算;⊗表示元素乘法;σ 表示Sigmoid函数。

2.2 掩码聚合

$X \in R^{W_{5\_1}\times H_{5\_1}\times C_{5\_1}}$、$Y \in R^{W_{5\_2}\times H_{5\_2}\times C_{5\_2}}$ 和 $Z \in R^{W_{5\_3}\times H_{5\_3}\times C_{5\_3}}$ 分别表示从VGG-16的Relu5_1,Relu5_2和Relu5_3中获得的注意力特征图。掩码聚合模型的注意力HBP模型的定义如下:

$$\Theta = concat(B(X^*, Y^*), B(X^*, Y^*), B(Y^*, Z^*)) \tag{3}$$

式中:B 是双线性池化操作;X^*、Y^*、Z^* 分别代表了从 X、Y、Z 中获得的RoI特征掩码图。

对于注意力特征X,在X上生成的掩码模型定义如下:

$$S_X = \begin{cases} 1, & X_{ij} \rangle \lambda \\ \theta, & \text{其他} \end{cases};\ \lambda = \mu \frac{\sum_{ij} X_{ij}}{H_X \times W_X} \tag{4}$$

式中:λ 表示激活阈值;μ 是控制 λ 变化的参数;$\theta \in (0,1)$ 表示松弛值。因为每一个注意力特征都来自不同层,因此利用式(5)来处理不同尺度下的多个注意力特征图。更具体地说,采用Max pooling操作将注意力激活图的维度调整成相同大小,然后在掩码映射上使用元素点积生成最终的聚合掩码 S:

$$S = S_X \circ F_{maxpooling}(S_X, S_Y) \circ F_{maxpooling}(S_X, S_Z) \tag{5}$$

因为需要通过带有相应注意力特征的掩码图生成每个RoI特征图,所以为了适应每个对应注意力特征图的大小,采用双线性插值操作来调整聚合掩码图的比例。最终,多个注意力RoI特征定义如下:

$$X^* = X \circ S \tag{6}$$

$$Y^* = Y \circ S'_Y (S'_Y = \Gamma(Y \circ S)) \tag{7}$$

$$Z^* = Z \circ S'_Z (S'_Z = \Gamma(Z \circ S)) \tag{8}$$

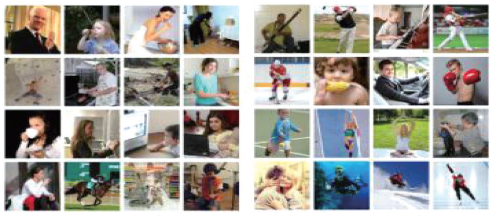
式中:X^*、Y^*、Z^* 分别表示从 X、Y、Z 上提取的RoI特征;Γ 表示双线性插值运算;∘表示元素点积。

3 实验

首先介绍实验的基本设置和数据集的详细信息。然后,在Stanford-40 dataset对两个关键组件的不同设置进行了详细的消融实验。最后将模型与最新的方法进行比较,结果证明,在stanford-40行为数据集上,模型获得了最好的性能;而且在自定义的60类行为数据集上也证实了所提出方法的有效性。

3.1 实验设置

Stanford-40 dataset一共包含9 532张图像,包含"鼓掌","骑马","刷牙"等40类行为。每个行为类别大约有180-300张图像。因为Stanford-40行为数据集类别相对较小,无法满足实际需求。因此,我们创建了一个新的行为数据集。该数据集包含60个类别,一共18 046张图片,数据集中人物姿态和图像背景更加复杂多样,与Stanford-40 dataset相比,大大增加了行为识别的难度。图2所示为Stanford-40和我们60类行为数据集。



(a) Stanford-40数据集的示例　(b) 自定义60 类数据集的样例

图 2　(a)Stanford-40 数据集的示例;
(b)自定义 60 类数据集的样例

Fig. 2　(a) Some examples on the Stanford 40 dataset;
(b) instances of our dataset

为了构建模型,在 VGG-16[4]网络中删除了最后的全连接层,并添加了几个新层来构建通道-空间注意力模块以及聚合掩码模块。同时使用在 ImageNet 数据集[19]上预先训练的权重来初始化模型。在实验中,首先冻结网络的其他层的权重来训练新添加的层,然后微调整个网络以更新所有层中的参数。输入图像的大小被设置为 224×224。在每个数据集上训练 80 个周期,并且 mini-batch 的大小设置为 16;同时采用随机梯度下降的优化方法进行网络优化,其中权重衰减和动量分别设置为 0.000 01 和 0.9。训练添加的层(整个网络)时,学习率开始设置为 0.01,之后每 10 个周期衰减 0.1。本文使用 PyTorch 构建模型,并在单个 NVIDIA Tesla P40 GPU 上进行实验。在整个实验过程中,主要采用 mAP 作为实验结果的评价指标。

3.2　消融实验

在本文实验中,首先探究了不同注意力模块对 HBP 模型的影响。之后,还设计了两种不同的通道和空间注意力模块的排列方式,顺序通道空间和顺序空间通道注意力模块,以研究不同排列方式对模型分类产生的效果。为了观察不同注意力模块作用以及不同排列方式带来的影响,在 Stanford-40 数据集上进行了大量的实验,结果如表 1 所示。

表 1　Stanford-40 上不同注意力模型的表现

Table 1　Performance (mean AP) of different attention methods on the Stanford-40 action dataset

Method	mAP/%
Baseline (VGG-16)	71.26
HBP	77.36
Spatial-wise attention	78.84
Channel-wise attention	79.31
Sequential Channel-Spatial attention	**83.58**
Sequential Spatial-channel attention	82.82

实验结果表明,使用注意力机制的 HBP 与基本模型 VGG-16 相比,具有更加突出的性能。首先,从表 1 可以看出,嵌入通道注意力模块的性能要优于空间注意力模块。接下来,将 Stanford-40 行为数据集上注意力模块的两种不同排列方式与最佳的通道注意力模块进行了比较。两种不同组合的注意力方式均比单个注意力模块获得更突出的表现。以上实验证明,提出的顺序通道空间注意力模型(sequential spatial-channel)可以更加有效地检测出显著性区域,从而提高行为识别的准确性。最后,表 2 为使用掩码聚合方法的实验结果。可以观察到,所有具有掩码聚合的模型都优于表 1 中的方法,尤其是具有掩码聚集的通道空间注意力模型,在 Stanford-40 行为数据集上获得了最佳表现。

表 2　Stanford-40 数据集上不同掩码聚合方法的表现

Table 2　Performance (mean AP) of different methods with mask aggregation on the Stanford-40 action dataset

Method	mAP/%
Baseline+Mask Aggregation	73.69
HBP+Mask Aggregation	78.35
Channel-wise attention+Mask Aggregation	80.73
Spatial-wise attention+Mask Aggregation	79.65
Channel-spatial attention+Mask Aggregation	**85.24**
Spatial-channel attention+Mask Aggregation	83.87

图 3 所示为 Stanford-40 行为数据集中生成聚合掩码的一些示例。从图 3 可以看出,聚合掩码可以有效地减少有害背景的影响并生成更加鲁棒的 RoI 区域。

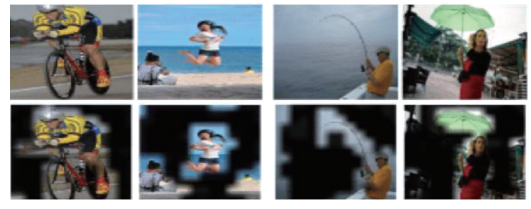


图 3　Stanford-40 中生成聚合掩码的示例

Fig. 3　Visualization of the aggregated mask on some samples in the Stanford 40 action dataset

3.3　与其他方法的对比试验

Stanford-40 行为数据集的分类结果如表 3 所示。可以看出,相比于其他方法,本文提出的方法获得了最佳的表现。文献[4]提出结合两个不同的 16 层和 19 层网络,并在“FC7”上训练线性支持向量机(SVM)进行行为分类。文献[20]介绍了一个深度融合模型,该模型将几种不同的网络特征连接起来,并将这些融合特征输入到 Random Forest 和 SVM 中进行行为识别。对于合并多个网络提取的特征方法,文献[4, 20]通过提取不同卷积网络的特征进行融合,但是模型仅使用单个网络的不同层间特征进行训练,并且本文提出的方法比这两种方法的准确率分别高出 4.1% 和 12.84%。文献[21]提出从图像中提取目标对象,并将对象区域与多示例学习相结合

的方式进行识别。文献[5]将姿势提示集成到 VGG-16 中,辅助网络进行分类。文献[14]主要利用人与物体的相互关系来分割目标区域进行分类。比较上述 3 种模型,可以清楚地了解到,模型不需要目标框等辅助信息即可获得良好的性能,其效果比文献[5, 14, 21]中所提出的算法分别高出 19.11%、4.55%和 4.44%。

表 3　不同模型在 Stanford-40 数据集上的表现

Table 3　Performance (mAP) of different models on the Stanford-40 action dataset

Method	Network	mAP/%
Baseline (VGG-16)	VGG-16	71.26
Cagdas *et al.*	AlexNet	66.13
Simonyan *et al.*	16-layer and 19-layer net	72.40
Qi *et al.*	VGG-16	80.69
Zhang *et al.*	VGG-16	80.80
Lavinia *et al.*	ResNet50, VGG-19, GoogleNet	81.146
Ours	VGG-16	**85.24**

接下来又了所提出的方法在不同行为类别中的表现,并将结果与文献[20-21]进行了比较。每种行为的平均精度如图 4 所示。从图 4 可以看出,在绝大多数行为类别中获得了较好的分类结果。但是在不同类别上的分类结果差异很大,范围从“射箭”的 98.613%到“挥手”的 62%。在“弹吉他”,“骑马”,“骑自行车”和“烹饪”行为类别中,模型获得了出色的表现,这是因为这些与人交互的目标物体(例如吉他,自行车或炊具)比较大,作为图像中的显著性目标很容易被检测。

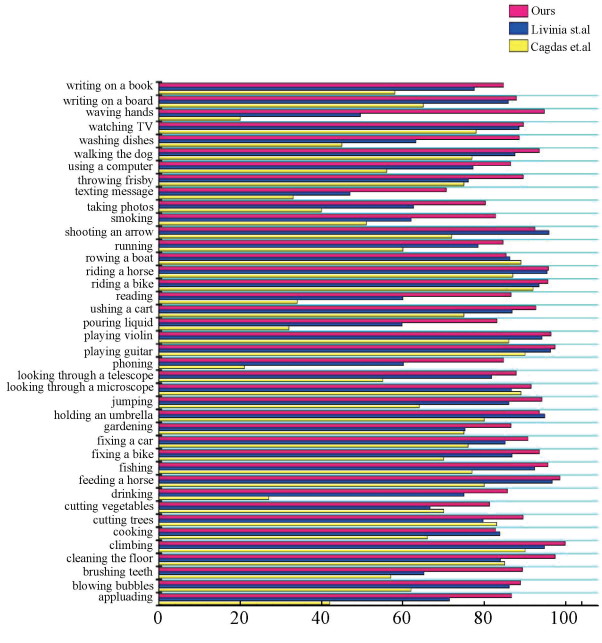


图 4　Stanford-40 数据集中每一类行为的比较

Fig. 4　The comparison of each class on the Stanford 40 action dataset

在自定义的 60 类行为数据集中模型性能的评估结果如表 4 所示。结果表明,所提出的方法分别高出基本网络(HBP)、注意力 HBP 和掩码聚合方法(HBP)5.94%、1.82%和 2.95%。此外,还分析了模型在识别特定行为类别时的准确性。在该数据集上,模型分类准确度最高的类别是“骑自行车”,其准确率为 97.68%,分类准确度最低的类别是“打电话”,为 57%。

表 4　不同方法在行为数据集上的表现

Table 4　Performance (mAP) of different methods on our action dataset

Method	mAP/%
Baseline (VGG-16)	70.53
HBP	78.63
Attention HBP (channel-spatial attention)	82.75
Mask Aggregation (baseline)	76.21
Mask Aggregation (HBP)	81.62
Ours	84.57

4　结　论

提出了一种新颖的具有掩码聚合的分层注意力双线性池化模型用于图像行为识别。该模型通过通道-空间注意力模块提取更多的显著性特征,并使用掩码聚合的方式生成 RoI 特征来抑制背景噪声的干扰。在两个行为数据集上进行了广泛的实验,实验结果证明了方法的有效性。消融实验又进一步验证了模型中所提出部件的重要作用。

参考文献

[1] SUN S, KUANG Z, OUYANG W, et al. Optical flow guided feature: a fast and robust motion representation for video action recognition [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1390-1399.

[2] MOLTISANTI D, FIDLER S, DAMEN D, et al. Action recognition from single timestamp supervision in untrimmed videos [C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 9907-9916.

[3] LI C, ZHONG Q, XIE D, et al. Collaborative spatio-temporal feature learning for video action recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7872-7881.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]. 3rd International Conference on Learning Representations (ICLR), 2015.

[5] QI T, XU Y, QUAN Y, et al. Image-based action

recognition using hint-enhanced deep neural networks [J]. Neurocomputing, 2017, 267(6): 475-488.
[6] GKIOXARI G, GIRSHICK R, MALIK J. Actions and attributes from wholes and parts [C]. 2015 IEEE International Conference on Computer Vision (ICCV), 2015: 2470-2478.
[7] ZHAO Z, MA H, YOU S. Single image action recognition using semantic body part actions [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2016: 3411-3419.
[8] KONG, SHU, and CHARLESS FOWLKES.: Low-rank bilinear pooling for fine-grained classification [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 7025-7034.
[9] GAO Y, BEIJBOM O, ZHANG N, et al. Compact bilinear pooling [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), 2016: 317-326.
[10] WEI X, ZHANG Y, GONG Y, et al. Grassmann pooling as compact homogeneous bilinear pooling for fine-grained visual classification [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 365-380.
[11] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear convolutional neural networks for fine-grained visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1309-1322.
[12] YU C J. Hierarchical bilinear pooling for fine-grained visual recognition [C]. Proceedings of the European Conference on Computer Vision (ECCV),2018.
[13] YAO B, JIANG X, KHOSLA A, et al. Human action recognition by learning bases of action attributes and parts [C]. 2011 International Conference on Computer Vision (ICCV), 2011: 1331-1338.
[14] ZHANG Y, CHENG L, WU J, et al. Action recognition in still images with minimum annotation efforts [J]. IEEE Transactions on Image Processing, 2016, 25(11): 5479-5490.
[15] GKIOXARI G, GIRSHICK R, DOLLAR P, et al. Detecting and recognizing human-object interactions [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8359-8367.
[16] HU J, SHEN L, ALBANIEl S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
[17] WOO S, PARK J, LEE J Y. Cbam: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
[19] 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 (CVPR), 2009: 248-255.
[20] LAVINIA Y, VO H H, VERMA A. Fusion based deep CNN for improved large-scale image action recognition [C]. 2016 IEEE International Symposium on Multimedia (ISM), 2016: 609-614.
[21] CAGDAS B, CEMIL Z, NAZLI I C. Using deep multiple instance learning for action recognition in still images [C]. 25th Signal Processing and Communications Applications Conference (SIU), 2017: 1-4.

作者简介

吴伟,2002 年于内蒙古大学获得学士学位,2005 年于内蒙古大学获得硕士学位,2014 年于内蒙古大学获得博士学位,现为内蒙古大学副教授,主要研究方向为人工智能和图形图像处理。
E-mail:cswuwei@ imu. edu. cn

Wu Wei received his B. Sc. degree from Inner Mongolia University in 2002, M. Sc. degree from Inner Mongolia University in 2005, Ph. D. degree from Inner Mongolia University in 2014. Now he is an associate professor at Inner Mongolia University. His main research interests include artificial intelligence and image processing.

于嘉乐,2018 年于内蒙古大学获得学士学位,现为内蒙古大学硕士研究生,主要研究方向为机器学习和计算机视觉。
E-mail:31809117@ mail. imu. edu. cn

Yu Jiale received his B. Sc. degree from Inner Mongolia University in 2018. Now he is a M. Sc. candidate at Inner Mongolia University. His main research interests include machine learning and computer vision.