

跨信道注意力权重最大区域掩盖的正则化方法

贾晓芬^{1,2}, 王景泰¹, 郭永存², 赵佰亭¹

(1. 安徽理工大学电气与信息工程学院, 232001, 安徽淮南;
2. 安徽理工大学省部共建深部煤矿采动响应与灾害防控国家重点实验室, 232001, 安徽淮南)

摘要: 针对 Dropout 的流动性和随机性可能引起欠拟合或过拟合的问题, 提出跨信道计算注意力权重并掩盖注意力权重最大区域的正则化方法(MARDrop)。设计跨信道注意力权重提取机制, 提取所有特征图的注意力权重, 并筛选出注意力权重较大的部分作为特征判别单元; 根据特征判别单元周围区域的注意力权重变化确定出注意力权重最大区域, 并将其掩盖即对应区域置为 0, 其余区域均置为 1; 将取值为 1 的注意力权重区域还原为特征图, 完成正则化处理。实验表明: MARDrop 具有较强的灵活性, 结合不同优化方法用于 Model2 模型后, 在 CIFAR-10 上的分类精度比基础模型提高了 19.55%; MARDrop 具有较强的泛化能力, 能灵活用于不同模型, 用于 VGG16 和 Res-Net18 模型后, 在 CIFAR-10 上的分类错误率比基础模型分别降低了 10.05% 和 9.96%, 用于 MoblieNetV3S 和 MoblieNetV3L 后, 在 ImageNet-C9 上的分类错误率比基础模型分别降低了 14.67% 和 96.90%。MARDrop 利用一维卷积跨信道交互计算注意力权重, 能有效降低提取的复杂度; MARDrop 通过准确掩盖造成过拟合的特征区域, 降低了模型对少数特征的依赖性, 进而增加了图像分类的准确率。

关键词: 正则化; 跨信道交互; 图像分类; Dropout

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202203005 **文章编号:** 0253-987X(2022)03-0045-09



Regularization Method for Masking the Maximum Attention Weight Region in Across Channels

JIA Xiaofen^{1,2}, WANG Jingtai¹, GUO Yongcun², ZHAO Baiting¹

(1. Institute of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China; 2. State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan, Anhui 232001, China)

Abstract: Aiming at the problem that Dropout mobility and randomness may cause under-fitting or over-fitting, a regularization method (MARDrop) is proposed to calculate the attention weight across channels and mask the region with the largest attention weight. A cross-channel attention weight extraction mechanism is designed to extract the attention weights of all feature graphs, and the parts with larger attention weights are selected as feature discrimination units. Then according to the change of attention weight of the area around the feature discrimination unit, the area with the largest attention weight is determined and masked, that is, the corresponding area is set to 0 and the other areas are set to 1. The attention weight region with the value of 1 is restored to the feature graph to complete the regularization process. Experiments show the strong

flexibility of MARDrop. Combined with different optimization methods, the classification accuracy on CIFAR-10 is 19.55% higher than that of the basic model. MARDrop has strong generalization ability and can be flexibly used in different models. After being used in VGG16 and ResNet18 models, the classification error rate on CIFAR-10 is reduced by 10.05% and 9.96% respectively compared with the basic model; the classification error rates of ImageNet-C9 for MoblieNetV3S and MoblieNetV3L are reduced by 14.67% and 96.90% respectively compared with the basic model. MARDrop adopts one-dimensional convolution to calculate attention weight interactively across channels, which can effectively reduce the complexity of extraction. MARDrop reduces the dependence of the model on a few features by accurately masking the feature areas bringing out over-fitting, and then increases the accuracy of image classification.

Keywords: regularization; interaction across channels; image classification; Dropout

卷积神经网络(CNN)凭借着其丰富的表达能力以及对复杂参数的处理能力在图像去噪^[1]、分类^[2]和超分辨率重建^[3]等领域得到爆发式发展。研究发现,CNN拟合情况跟训练集数目近似呈现对数比例关系。在训练样本不足时,由于CNN模型中有大量参数,在训练时会错把一些无法泛化的详细、繁杂的特征当作分类特征,从而出现过拟合现象^[4-6]。ResNet^[7]、VggNet^[8]、WRNNet^[9]、MobileNet^[10]、MoblieNetV3^[11]、EfficientnetV2^[12]、DenisNet-SE^[13]等在训练时都会有不同程度的过拟合现象出现。

在实际应用中,样本获取困难,多为小样本情况。常采用数据集增强(翻转、随机裁剪等)、模型集成、早停、正则化^[14]等方法来解决过拟合现象。其中,数据集增强对过拟合缓解有一定的限度,模型集成会使训练时间倍增,早停会使训练模型的准确率降低。正则化方法中 L_1/L_2 正则化^[15]、批归一化(BN)^[16]、Dropout^[17]等的应用较为普遍。 L_1/L_2 、BN正则化方法对缓解过拟合有一定限度,缓解过拟合效果小。Dropout虽然效果很好,但是存在很大的随机性和流动性,不普遍适用于现有的模型和数据集。

为了提高Dropout的普适性,学者们基于Dropout机制提出了DropPath^[18]、CorrDrop^[19]、DropFilterR^[20]、DropBlock^[21]、AttentionDrop^[22]和TargetDrop^[23]等改进方法。DropPath采用随机丢弃其中一些支路的策略,仅适用于Block支路较多的模型。DropBlock将随机丢弃特征单元的方案改进为自适应丢弃特征区域的策略,但随机丢弃的特点同样容易造成过拟合或欠拟合。CorrDrop和AttentionDrop的丢弃策略较为极端,均是屏蔽掉具有较明显特征的单元或具有特征不够明显的特征的区域。

TargetDrop的注意力提取方法含有降维、升维操作,效率较低且会影响预测准确性,该方法仅以最大特征判别单元来确定掩盖特征判别区域,存在特征区域注意力权重代表性较弱的问题。

为了解决这些问题和增强模型训练的拟合度,本文提出跨信道注意力权重最大区域掩盖的正则化方法(MARDrop),使用一维卷积对不同通道中跨信道交互信息处理,无需降维。同时,根据注意力权重最大的特征判别单元附近注意力权重变化情况来确定注意力权重最大区域并掩盖。MARDrop降低了注意力机制的计算复杂度,提取的注意力权重也更加符合理想的分类依据,使掩盖掉的区域更加符合网络模型依据的分类特征。

1 前期工作

Dropout^[17]通过忽略部分特征检测器的方式减少特征检测器间的相互作用,减小对局部特征的依赖性,从而减少过拟合。但是,Dropout的随机丢失以及仅对全连接层有效的特点限制了其效果。TargetDrop^[23]是根据Dropout原理提出的正则化方法,使用类似SENet^[24]的注意力提取机制。经注意力机制筛选后,选出部分注意力权重最大的单元作为特征判别单元,并以这些单元为中心掩盖周围区域。TargetDrop的注意力机制工作流程如图1所示。

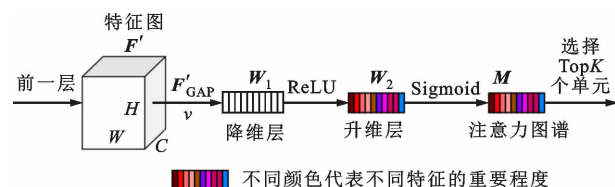


图1 TargetDrop中的注意力机制

Fig. 1 Attention mechanism in TargetDrop

注意力机制中主要包括全局平均池化(GAP)以及两个全连接层。第一个全连接层用于降维,然后采用 ReLU 函数激活。第二个全连接层用于恢复原始维度。 $\mathbf{M} \in \mathbb{R}^{H \times W \times C}$ 为通道注意力图谱,计算公式为

$$\mathbf{M} = \text{Sigmoid}(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{v})) \tag{1}$$

式中: $\mathbf{W}_1 \in \mathbb{R}^{(C/16) \times C}$; $\mathbf{W}_2 \in \mathbb{R}^{C \times (C/16)}$; $\mathbf{v}$ 为特征图 $\mathbf{F}'$ 经过 GAP 后的输出结果。

TargetDrop 通过选取特征单元中的最大特征点,并以其为中心掩盖周围的 $k \times k$ 区域,实现掩盖区域选择,选择过程为

$$\left. \begin{aligned} h_1 &= a - \left\lfloor \frac{k}{2} \right\rfloor; h_2 = a + \left\lfloor \frac{k}{2} \right\rfloor \\ w_1 &= b - \left\lfloor \frac{k}{2} \right\rfloor; w_2 = b + \left\lfloor \frac{k}{2} \right\rfloor \end{aligned} \right\} \tag{2}$$

式中: a, b 为特征单元中最大特征点的坐标值; k 为超参数,代表掩盖的正方形区域的边长, $k < H, k < W$; h_1, h_2 和 w_1, w_2 分别代表掩盖的正方形区域在 H 和 W 两个方向的顶点坐标值。

TargetDrop 按照这一方法,选取出全部最大特征点并逐个掩盖掉周围的 $k \times k$ 区域,但掩盖掉的区域并没有包含特征最明显的区域,从而引起某些重要特征的过度训练,使分类过于依赖这些特征,进而导致过拟合现象。同时,TargetDrop 使用全连接层进行降维和升维,此操作会影响通道注意力权重的预测,且在捕获通道之间依赖时也是低效的。

2 掩盖注意力权重最大区域的正则化

为了解决 TargetDrop 中出现的问题,本文提出了 MARDrop,结构见图 2。MARDrop 包括特征判别单元的提取、最大特征判别区域的选择、掩盖和还原共 3 部分。第 1 部分负责注意力权重的提取、注意力图谱的构建、注意力权重的排序以及 n 个权重最大特征单元的筛选;第 2 部分负责分析 n 个权重最大特征单元周围区域的注意力权重变化情况,选择最大特征判别区域;第 3 部分是对选出的最大特征判别区域的掩盖以及非最大特征判别区域对应原特征图的还原。

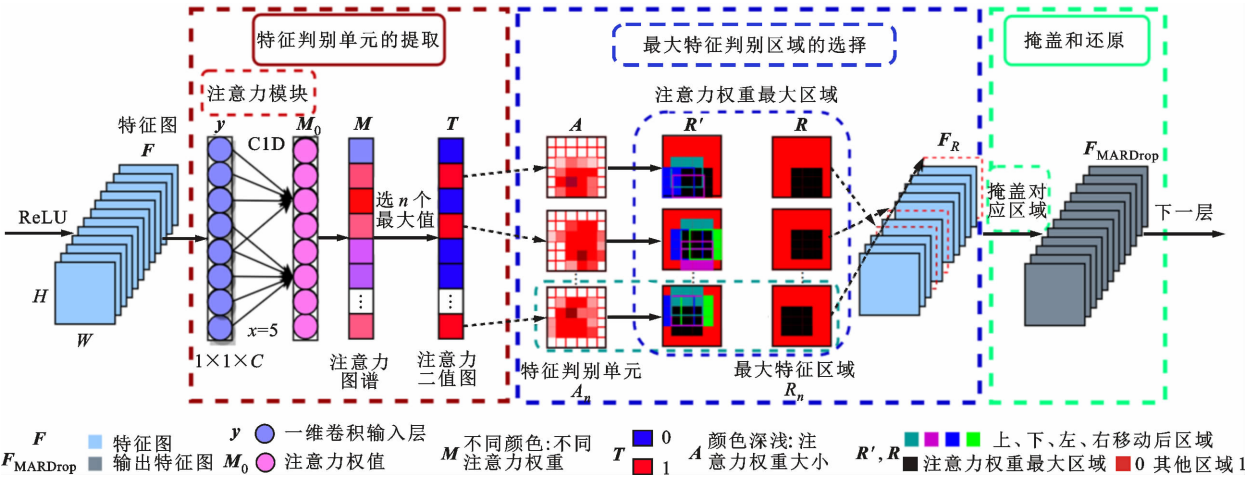


图 2 MARDrop 结构流程
Fig. 2 Structure flow chart of MARDrop

2.1 特征判别单元的提取

MARDrop 使用计算量更小、特征提取更准确的跨信道注意力机制对特征图 F 进行特征提取,得到注意力权重图谱 M ,然后从中筛选出部分注意力权重大的作为特征判别单元,筛选后得到注意力二值图 T 。

MARDrop 的提取特征判别单元结构见图 3。提取过程为:首先,对特征图 F 进行全局平均池化操作,得到一维卷积输入层 y ;然后,求出 y 的跨信道交互范围 x ,然后对输入层 y 进行一维卷积处理,求出注意力权值 M_0 ;使用 Sigmoid 函数激活 M_0 ,得

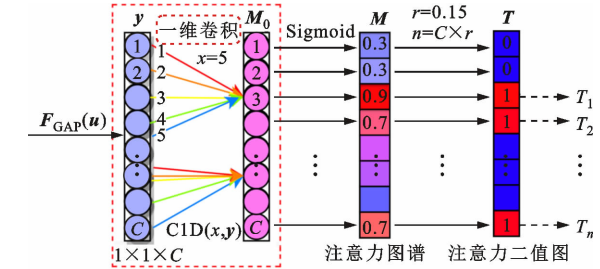


图 3 MARDrop 提取特征判别单元
Fig. 3 MARDrop extracting feature discrimination unit

到严格单调的注意力权重图谱 M (介于 $0 \sim 1$ 之间);最后,按照 M 中的注意力权重大小,选出 n 个

重要特征通道(n 为掩盖比例 r 与总通道数 C 的乘积),并将其注意力权重置为1,其余全部置为0。

2.1.1 注意力权重的提取 对卷积层、ReLU层处理后的特征图 F 进行全局平均池化处理,将特征图 F 之间的空间信息以及空间联系聚合到相应通道输入层 y 中。计算公式为

$$y = F_{\text{GAP}}(u) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W u_q(i, j) \quad (3)$$

式中: $y = \{y_1, y_2, \dots, y_C\} \in \mathbb{R}^{1 \times 1 \times C}$; $u = \{u_1, u_2, \dots, u_C\} \in \mathbb{R}^{H \times W \times C}$; u_q 为前一层输出 u 中第 q 个元素。

由式(1)可知,TargetDrop利用全连接层的先降维后升维方法提取注意力权重,降维时会降低特征图各信道之间的信息依赖,破坏原有空间结构,引起升维过程中注意力提取效率低且注意力权重提取不准确的问题。为此,本文提出一维卷积提取注意力机制。使用注意力机制适当捕获跨信道交互信息,无需降维,不会破坏原有的空间结构。注意力权重则通过结合前一层获取的其他特征信息来确定。一维卷积提取注意力机制的过程如下。

通过使用每个通道及 x 个邻近信道捕获跨信道交互信息(x 个邻近信道对应一个通道的注意力的预测),提取更准确的注意力权重。跨信道信息交互作用的覆盖范围 x 与通道数 C 成正比,通过自适应函数确定,计算式为

$$x = \phi(C) = \left\lfloor \frac{\ln C}{\gamma} + \frac{\beta}{\gamma} \right\rfloor_{\text{odd}} \quad (4)$$

式中: $\phi(C)$ 表示关于通道 C 的自适应函数; $\lfloor \cdot \rfloor_{\text{odd}}$ 表示向上取最近的奇数值操作;考虑到总通道数 C

不小于2、 x 不小于1,所以将 γ, β 分别设置为2、1。

确定一维卷积中局部跨信道交互范围 x 后,使用一维卷积提取每个单元对应的注意力数据,得到 M_0 。然后,使用Sigmoid函数得到连续、光滑的严格单调注意力权重图谱 M 。计算过程为

$$M_0 = \text{C1D}(x, y) \quad (5)$$

$$M = \text{Sigmoid}(M_0) \quad (6)$$

2.1.2 特征单元的筛选 根据注意力权重和丢弃比率 r ,在注意力图谱中选出前 n 个最重要特征通道。以第 n 个的注意力权重为筛选标准(1表示被选中,0表示未选中),将得到的相应值放入注意力二值图 T 中。计算过程为

$$T_p = \begin{cases} 1, & M_p \geq M_n \\ 0, & M_p < M_n \end{cases} \quad (7)$$

式中 T_p, M_p 分别为 T, M 中的第 p 个元素。

2.2 最大特征判别区域的选择

以所占权重最大的特征单元为中心,从中心向四周延伸时,注意力权重以不同变化率随之减小。TargetDrop使用式(2)方法,只依据特征判别单元来选择判别区域,未对周围单元权重进行考究,并且如果所选特征单元代表性不足,则选择的特征判别区域很可能不存在代表性。

为此,本文提出根据权重变化以及注意力权重特征图来确定最大特征判别区域的方法,具体过程见图4。根据特征判别单元图 A 中各特征单元的位置,求出这些点在特征图 F 上的位置坐标,以该坐标 (F_h, F_w) 为中心,计算给定边长为 k 的正方形区

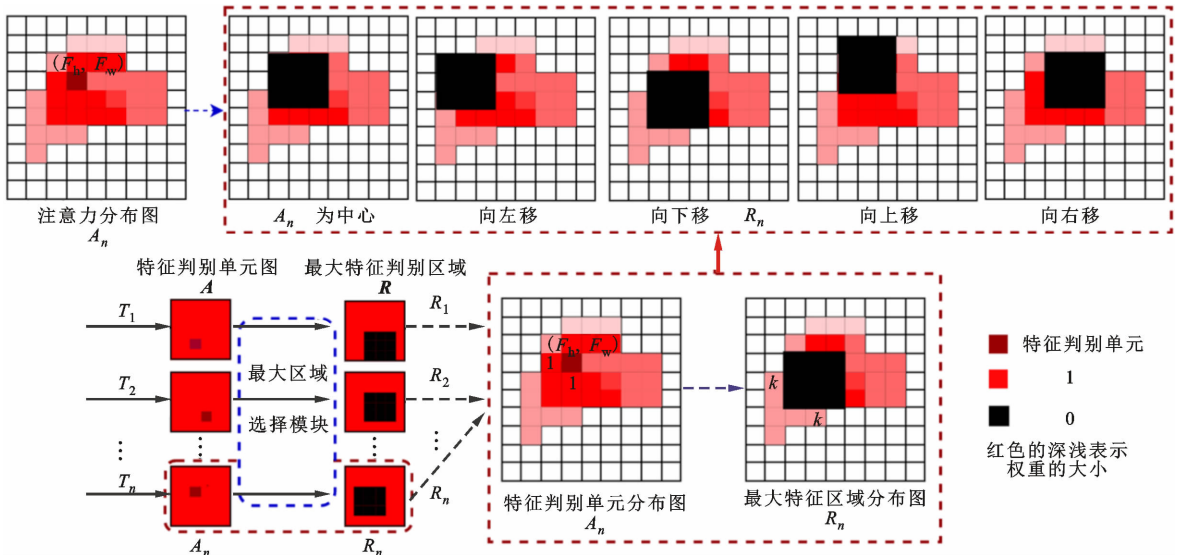


图4 最大特征判别区域的选择

Fig. 4 Selection of maximum feature discrimination region

域的注意力权重,以及该区域外向上、下、左、右平移后区域的注意力权重之和,比较中心区域与平移后的区域注意力权重,选出注意力权重最大区域即最大特征判别区域。 m_c 为初始的 $k \times k$ 区域对应的权重之和, m_u 、 m_d 、 m_l 、 m_r 分别为 $k \times k$ 区域外向上、下、左、右平移后区域的注意力权重之和,计算公式为

$$m_c = \sum_{i=M_w-\frac{k-1}{2}}^{M_w+\frac{k-1}{2}} \sum_{j=M_h-\frac{k-1}{2}}^{M_h+\frac{k-1}{2}} m_{(i,j)} \quad (8)$$

$$m_u = \sum_{i=M_w-\frac{k-1}{2}}^{M_w+\frac{k-1}{2}} \sum_{j=-\frac{k-1}{2}}^{\frac{k}{2}} m_{(i,M_h-\frac{k-1}{2}-j)} \quad (9)$$

$$m_d = \sum_{i=M_w-\frac{k-1}{2}}^{M_w+\frac{k-1}{2}} \sum_{j=-\frac{k-1}{2}}^{\frac{k}{2}} m_{(i,M_h+\frac{k-1}{2}+j)} \quad (10)$$

$$m_l = \sum_{i=0}^{k-1} \sum_{j=-\frac{k-1}{2}}^{\frac{k}{2}} m_{(M_w-\frac{k-1}{2}-j,M_h-\frac{k-1}{2}+ki)} \quad (11)$$

$$m_r = \sum_{i=0}^{k-1} \sum_{j=-\frac{k-1}{2}}^{\frac{k}{2}} m_{(M_w+\frac{k-1}{2}+j,M_h-\frac{k-1}{2}+ki)} \quad (12)$$

式中 $\mathbf{m} = [m_{(1,1)}, m_{(1,2)}, \dots, m_{(1,W)}; \dots, m_{(H,W)}]$ 。

求出各区域注意力权重后,选出最大特征判别区域的注意力权重 $R_{\max}$,筛选公式为

$$R_{\max} = \max\{m_c, m_u, m_d, m_l, m_r\} \quad (13)$$

$R_{\max}$ 对应的区域就是注意力权重最大区域,即最大特征判别区域。

2.3 掩盖和还原

确定出最大特征判别区域 $R_{\max}$ 后,需要将其掩盖,并还原全部未掩盖区域,得到特征图 $\mathbf{F}_{\text{MARDrop}}$ 。掩盖及还原过程见图5,图中最大特征判别区域 $\mathbf{R}$ 中黑色为掩盖区域,代表着最大特征判别区域,该区域元素置为0,其他红色区域为未掩盖区域,该区域元素置为1。将 $\mathbf{R}$ 中每个元素与原特征图 $\mathbf{F}$ 中对应元素相乘,得到特征图 $\mathbf{F}_{\text{MARDrop}}$ 。

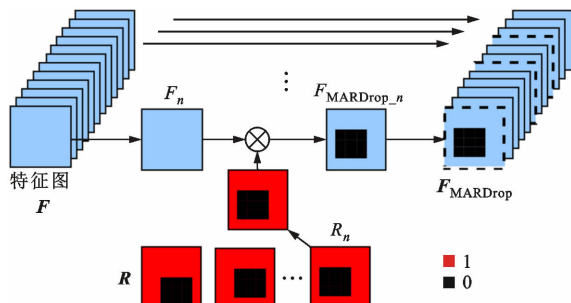


图5 特征图的掩盖与还原

Fig. 5 Masking and restoration of feature maps

3 实验结果与分析

3.1 实验设置

使用 Python 语言,在 Pytorch 环境开展相关实验。参数设置为:batchsize 为 256,优化器使用动量为 0.9 的 SGD,初始学习率为 0.1,在总时数比为 0.4、0.6、0.8 时衰减 $2e^{-1}$ 倍。

MARDrop 根据特征单元周围注意力权重的分布情况,选择出最大特征判别区域并进行掩盖,需要先确定掩盖比率和掩盖尺寸,得到 MARDrop 的模型结构后,再开展对比实验。

3.2 确定掩盖比率

掩盖比率是选取的掩盖特征所占总判别特征的比例,与模型对图像分类依据所占比例相关,与掩盖尺寸无关。实验固定掩盖尺寸为 5×5 像素,在基础模型 MoblieNetV3^[5]上分别加入掩盖比率为 0.05、0.1、0.15、0.2 和 0.25 的 MARDrop,对比其在数据集 CIFAR-10 上的分类准确度,结果见图6。可以看出:掩盖比率从 0.05 增大到 0.25 时,分类精度呈先增大后减小的趋势;掩盖比率为 0.15 时,分类精度为 87.25%;掩盖比率为 0.2 时,分类精度达到最大,为 87.27%。掩盖比率取 0.15 和 0.2 时效果相差很小,但掩盖比率小可以减少 MARDrop 的数据处理量,提高分类效率。因此,后续实验中取掩盖比率为 0.15。

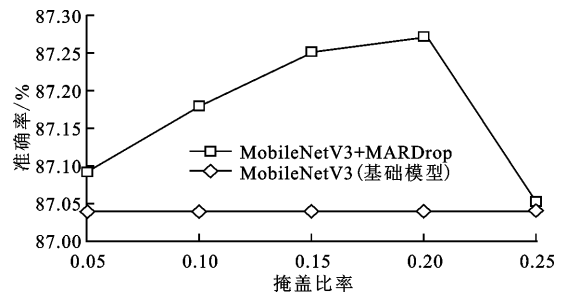


图6 MARDrop 掩盖比率对分类准确率的影响

Fig. 6 Influence of MARDrop masking ratio on classification accuracy

3.3 确定掩盖尺寸

掩盖尺寸是指特征判别区域尺寸,与图像被检测物体尺寸以及卷积提取的特征元素有关,与掩盖比例无关。

固定掩盖比例为 0.15,在 MoblieNetV3 上分别使用掩盖尺寸为 3×3 、 5×5 、 7×7 和 9×9 像素的 MARDrop,对比其在数据集 CIFAR-10 上的分类准确率,结果见图7。可以看出,随着掩盖尺寸的变

化,分类准确率先升高后降低,在掩盖尺寸为 5×5 像素时准确率达到最高,为 87.25%。后续实验中, MARDrop 模块的掩盖比率和掩盖尺寸分别设置为 0.15、 5×5 像素。

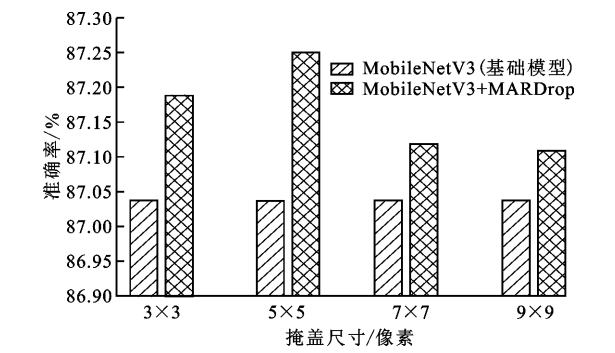


图 7 不同掩盖尺寸的 MARDrop 对基础模型的影响
Fig. 7 Influence of MARDrop with different mask sizes on the basic model

3.4 优越性验证

以 ResNet18 为基础模型,在 CIFAR-10 和 CIFAR-100 数据集上测试 MARDrop 的分类错误率,并与各种改进型的 Dropout 正则化方法,如 DropBlock、AttentionDrop、TargetDrop 进行对比分析,结果见表 1。

表 1 ResNet18 上不同改进型 Dropout 方法的分类错误率
Table 1 Classification error rate of different improved Dropout methods on ResNet18

正则化方法	错误率/%	
	CIFAR-10	CIFAR-100
基础模型	4.72	22.46
Dropout ^[17]	5.14	23.82
DropBlock ^[21]	4.59	21.95
AttentionDrop ^[22]	4.51	21.53
TargetDrop ^[23]	4.41	21.37
MARDrop	4.25	21.15

注:表中加粗数据为最优值。

由表 1 可知,加入 Dropout 正则化比原基础模型错误率更高,说明训练过程中过拟合更加严重,或者出现欠拟合现象。使用 DropBlock、AttentionDrop、TargetDrop 时的分类精度均比 Dropout 有所提升, MARDrop 在两个数据集上分类错误率均最小。相对于基础模型,加入 MARDrop 后在 CIFAR-10 上分类错误率降低了 9.96%,在 CIFAR-100 上错误率降低了 5.83%。相比于 2020 年提出

的 TargetDrop^[23], MARDrop 在 CIFAR-10 和 CIFAR-100 上分类错误率分别降低了 3.63% 和 1.03%。

3.5 泛化性验证

3.5.1 不同模型上的泛化性验证 在 ResNet20、VGG16、WRN28-10、MoblieNetV3S、MoblieNetV3L、EfficientnetV2 和 DenisNet-SE 共 7 个网络模型上,分别加 MARDrop、Dropout、TargetDrop 正则化方法,并在 CIFAR-10 数据集上测试分类准确率,结果见表 2。其中, EfficientnetV2 和 DenisNet-SE 两个模型的参数比较复杂,使用了原文中的参数,其他 5 个模型均采用 3.1 小节设置的参数。

表 2 不同模型使用不同正则化方法对 CIFAR-10 的分类错误率
Table 2 Classification error rate of different network models with different regularization methods on CIFAR-10

网络模型	错误率/%			
	基础模型	Drop-out	Target-Drop	MAR-Drop
ResNet20 ^[7]	8.21	7.80	7.61	6.75
VGG16 ^[8]	6.17	6.43	5.89	5.55
WRN28-10 ^[9]	4.02	4.04	3.68	3.60
MoblieNetV3S ^[11]	19.55	19.55	19.02	18.76
MoblieNetV3L ^[11]	12.96	12.96	12.85	12.73
EfficientnetV2 ^[12]	8.11	8.11		7.27
DenisNet-SE ^[13]	4.50			4.24

注:表中加粗数据为最优值。

由表 2 可知,加入 Dropout 后, VGG16、WRN2810 错误率不降反升, ResNet20 的错误率有所下降。加入 TargetDrop 后, 7 个网络模型分类错误率均有所下降。 MARDrop 在 7 个网络模型均获得了最低的分类错误率。对于 VGG16, 加入 MARDrop 后在 CIFAR-10 上分类错误比基础模型降低了 10.05%, 比使用 TargetDrop 错误率降低了 5.77%。在 2021 年提出的最新模型 DenisNet-SE^[13]上, 加入 MARDrop 后比基础模型分类错误率降低了 5.78%。

3.5.2 不同图像尺寸数据集上的泛化性验证 从 ImageNet 数据集中选前 9 类制作成 ImageNet_C9 数据集(训练集和测试集按照数据量 7:3 的比例划分), 分别在 MoblieNetV3S、MoblieNetV3L 上使用 MARDrop 正则化方法对 ImageNet_C9 数据集进行分类实验, MARDrop 参数设置以及实验结果见表 3。

表 3 MARDrop 在 ImageNet_C9 上的分类错误率
Table 3 MARDrop classification error rate on ImageNet_C9

模型	掩盖 比率	掩盖尺寸 /像素	错误率 /%
MoblieNetV3L			2.59
MoblieNetV3L+MARDrop	0.15	5×5	2.21
MoblieNetV3S			4.20
MoblieNetV3S+MARDrop	0.10	5×5	0.13

注:表中加粗数据为最优值。

由表 3 可知,使用 MARDrop 后,MoblieNetV3L 错误率降低了 14.67%。使用 MobileNetV3S 模型实验时掩盖比例调整为 0.10,加入 MARDrop 后分类错误率降低了 96.90%。

3.6 灵活性验证

以文献[25]中搭建的 Model2 为基础模型,在 CIFAR-10 数据集上测试 MARDrop 的优越性及其和不同优化器或者正则化方法相结合的灵活性,每组实验的参数设置相同,测试结果见表 4。表中,

表 4 Model2 使用不同训练方法时的测试结果
Table 4 Test results of Model2 using different training methods

训练方法	损失函数值		准确率/%	
	训练集	测试集	训练集	测试集
SGD	7.15×10^{-6}	2.699	100.00	76.49
Adam	2.00×10^{-10}	0.920	100.00	79.84
BN(Adam)	8.96×10^{-5}	0.405	100.00	82.89
L_2 正则化, $\lambda=10^{-5}$ (Adam)	2.09×10^{-2}	1.483	100.00	80.70
L_1 正则化, $\lambda=10^{-6}$ (Adam)	9.27×10^{-2}	1.339	100.00	76.40
Dropout+BN(Adam)	9.90×10^{-3}	0.904	99.66	86.33
DataAugm+BN+Dropout(Adam)	9.06×10^{-2}	0.519	96.78	88.67
MARDrop(SGD)	5.10×10^{-2}	0.522	98.27	87.81
MARDrop+BN(SGD)	2.35×10^{-2}	0.465	96.78	90.20
MARDrop+BN(ReLU6+SGD)	1.10×10^{-2}	0.374	99.75	91.44

注:表中加粗数据为前 2 个最优值。

4 结 论

为了解决现有正则化方法对过拟合现象处理效果不足的缺点,本文提出使用跨通道的注意力机制提取注意力权重,根据注意力权重选择特征单元并通过比较附近区域权重大小选择出注意力权重最大区域,最后将该区域掩盖并还原为特征图的正则化方法。为了证明该方法的有效性,本文将 MARDrop 方法用于不同 CNN 分类模型,在 CIFAR-10、CIFAR-100 数据集进行图像分类测试。测试结论

SGD 和 Adam 是 Model2 基础模型的测试结果。

由表 4 可知,相对于基础模型,使用正则化方法后,测试集与训练集的损失函数值明显减小,训练准确率与测试准确率差值减小。采用正则化方法组合的优化方案后,训练集准确率小于 100%,且测试集准确率远大于原模型准确率,说明优化后模型的拟合程度高于原模型拟合程度。从分类精度的角度分析,使用优化器 Adam 的基础模型效果较好。在此基础模型上加入不同正则化方法后,除了加入 L_1 正则化后分类精度降低外,其余均有所提升。使用优化器 SGD 的基础模型效果较差,比使用 Adam 时的分类精度低了 4.20%。为了更好地显示 MARDrop 的优势,在使用 SGD 的基础模型上分别引入 MARDrop、MARDrop+BN、MARDrop+BN+ReLU6,开展 3 个实验。可以看出,相对于仅使用 SGD 时的分类精度 76.49%,3 个实验的分类精度分别提高到了 87.81%、90.20% 和 91.44%,提升率分别为 14.79%、17.92% 和 19.55%。

如下。

(1)确定掩盖比率和掩盖尺寸的实验结果表明,可以通过调节 MARDrop 的掩盖比率和掩盖尺寸使拟合情况达到最优。

(2)MARDrop 与不同 Drop 机制正则化方法对比,错误率均最低。相比于对过拟合处理最好的 TargetDrop,使用 MARDrop 在 CIFAR-10、CIFAR-100 上分类错误率分别降低了 3.63%、1.03%。证明了该方法相对于同类方法的优越性。

(3)DenisNet-SE 等 7 种模型经 MARDrop 优

化后在 CAFIR-10 上的分类错误率均有所降低。使用 MoblieNetV3 在图像尺寸不同的 ImageNet-C9 数据集上分类错误率相对于基础模型分别降低了 14.67% 和 96.90%。这两组实验证明了 MARDrop 方法的泛化性。

(4) 使用包含 MARDrop 的多种正则化方法优化后, 模型在 CAFIR-10 上的拟合程度远高于使用其他正则化方法优化的模型, 其中使用 MARDrop + BN(ReLU6 + SGD) 方法优化后在 CIFAR-10 上的分类准确率比原模型提高了 19.55%, 证明了 MARDrop 方法的灵活性。

参考文献:

- [1] 吕永标, 赵建伟, 曹飞龙. 基于复合卷积神经网络的图像去噪算法 [J]. 模式识别与人工智能, 2017, 30(2): 97-105.
LÜ Yongbiao, ZHAO Jianwei, CAO Feilong. Image denoising algorithm based on composite convolutional neural network [J]. Pattern Recognition and Artificial Intelligence, 2017, 30(2): 97-105.
- [2] SONG Jia, GAO Shaohua, ZHU Yunqiang, et al. A survey of remote sensing image classification based on CNNs [J]. Big Earth Data, 2019, 3(3): 232-254.
- [3] 宋长明, 王赞. 融合低秩和稀疏表示的图像超分辨率重建算法 [J]. 西安交通大学学报, 2018, 52(7): 18-24.
SONG Changming, WANG Yun. Super resolution reconstruction algorithm combined low rank with sparse representation [J]. Journal of Xi'an Jiaotong University, 2018, 52(7): 18-24.
- [4] ZHANG Xiangyu, LI Jianqing, CAI Zhipeng, et al. Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection [J]. Medical & Biological Engineering & Computing, 2021, 59(1): 165-173.
- [5] 任义丽, 罗路. 卷积神经网络过拟合问题研究 [J]. 信息系统工程, 2019(5): 140, 142.
REN Yili, LUO Lu. Research on over-fitting of convolutional neural networks [J]. China CIO News, 2019(5): 140, 142.
- [6] 谢璐阳, 夏兆君, 朱少华, 等. 基于卷积神经网络的图像识别过拟合问题分析与研究 [J]. 软件工程, 2019, 22(10): 27-29, 26.
XIE Luyang, XIA Zhaojun, ZHU Shaohua, et al. Analysis and research of overfitting of image recognition based on convolutional neural networks [J]. Software Engineering, 2019, 22(10): 27-29, 26.
- [7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale visual recognition [C]// Proceedings of the International Conference on Learning Representations. London, UK: ICLR, 2015 [2021-05-01]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [9] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [C]// Proceedings of the 2016 British Machine Vision Conference (BMVC). Guildford, UK: BMVA Press, 2016: 1-15.
- [10] HOWARD A G, ZHU Menglong, CHEN Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2021-05-01]. <https://arxiv.org/abs/1704.04861>.
- [11] KAVYASHREE P S P, EL-SHARKAWY M. Compressed MobileNet V3: a light weight variant for resource-constrained platforms [C]// Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC). Piscataway, NJ, USA: IEEE, 2021: 0104-0107.
- [12] TAN Mingxing, LE Q V. EfficientNetV2: smaller models and faster training [C]// Proceedings of the 38th International Conference on Machine Learning. Princeton, NJ, USA: IMLS, 2021: 10096-10106.
- [13] YANG Mingming, MA Tinghuai, TIAN Qing, et al. Aggregated squeeze-and-excitation transformations for densely connected convolutional networks [J/OL]. The Visual Computer, 2021 [2021-05-01]. <https://doi.org/10.1007/s00371-021-02144-z>.
- [14] WANG Haojun, LONG Haixia, WANG Ailan, et al. Deep learning and regularization algorithms for malicious code classification [J]. IEEE Access, 2021, 9: 91512-91523.
- [15] LIU Ziqi, XU Yanbin, DONG Feng. L_1 - L_2 spatial adaptive regularization method for electrical tomography [C]// Proceedings of the 2019 Chinese Control Conference (CCC). Piscataway, NJ, USA: IEEE, 2019: 3346-3351.
- [16] ABBASI A N, HE Mingyi. Convolutional neural network with PCA and batch normalization for hyperspectral image classification [C]// Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium. Piscataway, NJ, USA: IEEE, 2019:

- 959-962.
- [17] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [18] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 8697-8710.
- [19] ZENG Yuyuan, DAI Tao, XIA Shutao. Corrdrop: correlation based dropout for convolutional neural networks [C]// Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2020: 3742-3746.
- [20] PAN Hengyue, NIU Xin, LI Rongchun, et al. Drop-FilterR: a novel regularization method for learning convolutional neural networks [J]. Neural Processing Letters, 2020, 51(2): 1285-1298.
- [21] GHIASI G, LIN T Y, LE Q V. DropBlock: a regularization method for convolutional networks [C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. Vancouver, Canada: NIPS, 2018: 10750-10760.
- [22] OUYANG Zhihao, FENG Yan, HE Zihao, et al. Attentiondrop for convolutional neural networks [C]// Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME). Piscataway, NJ, USA: IEEE, 2019: 1342-1347.
- [23] ZHU Hui, ZHAO Xiaofang. TargetDrop: a targeted regularization method for convolutional neural networks [EB/OL]. [2021-05-01]. <https://arxiv.org/abs/2010.10716>.
- [24] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [25] MARIN I, KUZMANIC SKELIN A, GRUJIC T, et al. Empirical evaluation of the effect of optimization and regularization techniques on the generalization performance of deep convolutional neural network [J]. Applied Sciences, 2020, 10(21): 7817.

(编辑 陶晴)