中国科学院软件研究所学术年会暨重点实验室科技活动周 2025 第十届
学术论文

Less Yet Robust Crucial Region Selection for Scene Recognition Models

张健琦*，王蒙轩*，王静瑶，司凌宇，郑昌文，徐帆江

IEEE International Conference on Acoustics, Speech, and Signal Processing

张健琦 jluzhangjianqi@163.com

概述

针对水下与遥感场景图像常见的模糊、过曝及局部噪声干扰，本文设计了一种嵌入式可学习掩码机制，自动在高层特征图上筛选判别性区域，并在损失中引入稀疏正则，降低类别间共享干扰影响。该机制无需额外推理开销，可与各类CNN主干无缝结合。实验在自建海底地质场景数据集UGS与公开UCM上均取得领先精度，且在人工加噪测试中保持稳健，验证了方法的通用性与实用价值。

动机与分析

水下与遥感图像常受散射、过曝等影响，图像中充斥着无关元素。使用基于卷积的场景识别方法时，模型会从图像的所有区域提取特征，可能将无关元素也纳入决策过程。在图 1 中，红框内的区域已足以判断图像类别，但 ResNet-18 却将注意力集中到了不相关的区域上。

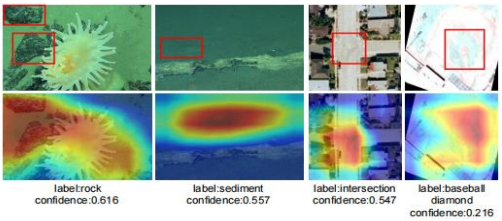

图 1：第一行展示输入图像，红框标出了足以判定图像类别的区域。第二行（通过 Grad-CAM生成）显示 ResNet-18 在预测这些图像时关注的区域。每幅图像底部给出了对应标签及其置信度分数。可以看出，ResNet-18 的注意区域明显超出了红框所示的判别性区域。

方法

我们提出只关注图像中少量但关键的区域，以帮助模型学习并消除噪声或遮挡对决策的影响。基于这一洞见，我们设计了一种新方法。具体而言，步骤 1： 使用基于 CNN 的特征提取器获得语义特征；步骤 2： 引入可学习的掩码矩阵来定位重要特征，并对该掩码施加稀疏约束，使模型在预测时仅利用尽可能少而关键的特征区域。

下图展示了所提方法的整体流程：

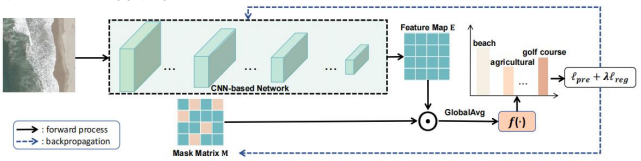

图 2：所提方法的整体框架。

实验

在对比实验中，我们采用了 ResNet 系列、MobileNet 系列、VIT 系列，以及我们提出的 ResNet+掩码矩阵 M 方法。为了验证掩码矩阵 M 的有效性，我们还将其引入到 MobileNet 中。为消除随机性的影响，每个模型在每个数据集上均测试五次，记录平均准确率、标准差以及最低准确率。表 I 展示了完整的实验结果，最佳结果用红色标出，次优结果用蓝色标出。

Models	FLOPs (G)	UGS acc	UGS min acc	UGS sparsity of M(%)	UCM acc	UCM min acc	UCM sparsity of M(%)
ResNet18	7.29	$0.900 \pm 7.5e^{-4}$	0.900	-	$0.964 \pm 1.5e^{-4}$	0.945	-
ResNet18*	7.29	$0.940 \pm 1.2e^{-3}$	0.900	48.98	$0.973 \pm 6.5e^{-5}$	0.959	95.92
ResNet50	16.53	$0.885 \pm 6.5e^{-4}$	0.850	-	$0.947 \pm 5.4e^{-4}$	0.909	-
ResNet50*	16.53	$0.920 \pm 1.4e^{-3}$	0.900	34.69	$0.948 \pm 2.9e^{-4}$	0.941	91.84
ResNet101	31.46	$0.860 \pm 3.2e^{-3}$	0.750	-	$0.946 \pm 9.0e^{-4}$	0.901	-
ResNet101*	31.46	$0.900 \pm 1.3e^{-3}$	0.850	32.65	$0.951 \pm 1.9e^{-4}$	0.932	91.84
MobileNet v2	1.30	$0.875 \pm 7.5e^{-4}$	0.850	-	$0.963 \pm 6.6e^{-5}$	0.951	-
MobileNet v2*	1.30	$0.880 \pm 6.0e^{-4}$	0.850	28.57	$0.972 \pm 4.7e^{-5}$	0.961	91.84
MobileNet v3 small	0.24	$0.680 \pm 1.7e^{-2}$	0.575	-	$0.957 \pm 4.1e^{-4}$	0.917	-
MobileNet v3 small*	0.24	$0.695 \pm 1.3e^{-2}$	0.575	59.18	$0.962 \pm 1.0e^{-4}$	0.946	95.92
MobileNet v3 large	0.93	$0.725 \pm 3.5e^{-2}$	0.575	-	$0.956 \pm 4.9e^{-4}$	0.920	-
MobileNet v3 large*	0.93	$0.765 \pm 9.2e^{-3}$	0.575	51.02	$0.966 \pm 2.3e^{-4}$	0.936	95.92
ViT Base 16	11.29	$0.835 \pm 4.7e^{-3}$	0.725	-	$0.930 \pm 9.2e^{-5}$	0.913	-
ViT Base 32	2.95	$0.880 \pm 1.1e^{-3}$	0.875	-	$0.939 \pm 1.9e^{-5}$	0.934	-
ViT Large 16	39.86	$0.780 \pm 4.6e^{-3}$	0.675	-	$0.886 \pm 6.8e^{-4}$	0.863	-
ViT Large 32	10.23	$0.815 \pm 4.2e^{-3}$	0.725	-	$0.945 \pm 4.4e^{-5}$	0.935	-

表 1：不同模型在 UGS 和 UCM 数据集上的性能对比。模型名称后带 * 表示该模型引入了我们提出的掩码矩阵 M。“acc” 表示模型在测试集上五次实验的平均准确率，括号内为对应的标准差； “min acc” 表示五次实验中测试集上的最低准确率。