

基于消融分析的卷积神经网络可解释性分析

李绍轩, 杨有龙

(西安电子科技大学 数学与统计学院, 西安 710126)

摘要: 针对基于类激活映射(CAM)的可解释性方法因受到与目标类别无关特征的干扰, 导致可视化结果中含有较多噪声及对目标物体定位精度较低的问题, 提出一种基于消融分析的卷积神经网络(CNN)可视化方法. 先通过消融实验考察深层网络特征与目标类别的相关性并计算特征融合权重; 再通过 ReLU 或 Softmax 函数对融合权重进行修正, 以减少无关特征的干扰, 得到定位精度更高的类激活图, 从而对网络决策做出有效说明. 在验证集 ILSVRC 2012 上使用多种评估指标进行验证, 实验结果表明, 该方法在各项指标上均取得了更好的模型解释能力.

关键词: 可解释性; 卷积神经网络; 消融分析; 深度学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)03-0606-09

Interpretability Analysis of Convolutional Neural Networks Based on Ablation Analysis

LI Shaoxuan, YANG Youlong

(School of Mathematics and Statistics, Xidian University, Xi'an 710126, China)

Abstract: Aiming at the problem that the interpretable method based on class activation mapping (CAM) was disturbed by features unrelated to the target class, which led to more noise in the visualization results and lower localization accuracy of target objects, we proposed a convolutional neural network (CNN) visualization method based on ablation analysis. Firstly, the correlation between deep network features and target classes was investigated and feature fusion weights were calculated through ablation experiments. Secondly, the feature fusion weights were corrected by ReLU or Softmax functions to reduce the interference of irrelevant features and obtain class activation map with higher localization accuracy, so as to make an effective description of network decisions. A variety of evaluation metrics were used for verification on the ILSVRC 2012 validation set, the experimental results show that the method achieves better model interpretation capability in all indicators.

Keywords: interpretability; convolutional neural network; ablation analysis; deep learning

0 引言

深度学习的可解释性任务旨在开发新的技术和方法解释神经网络的决策过程, 揭示网络内部的判

收稿日期: 2023-02-14.

第一作者简介: 李绍轩(1997—), 男, 汉族, 硕士研究生, 从事计算机视觉的研究, E-mail: 1195935783@qq.com. **通信作者简介:** 杨有龙(1967—), 男, 汉族, 博士, 教授, 博士生导师, 从事数据分析和概率图模型的研究, E-mail: ylyang@mail.xidian.edu.cn.

基金项目: 国家自然科学基金(批准号: 61573266)和陕西省自然科学基金基础研究计划项目(批准号: 2021JM-133).

别规则和学习策略. 卷积神经网络(convolutional neural networks, CNN)在图像识别^[1-2]、目标检测^[3-5]和语义分割^[6]等任务中性能优异, 而针对 CNN 的理解主要通过各种可视化方法得到. 文献[7-8]利用目标节点相对于输入图像的梯度信息计算每个像素对网络决策的贡献度; RISE 方法^[9]和 LIME 方法^[10]通过生成大量的随机掩码对原图像进行扰动, 然后根据扰动样本和网络决策结果判断图像中不同区域对目标类别的重要性, 但采样的特征空间非常大, 增加了算法的运行时间和计算机的内存损耗. Zhou 等^[11]提出了类激活映射(class activation mapping, CAM)方法对网络的决策做出说明, 将 CNN 的全连接层替换为全局平均池化层(global average pooling, GAP)^[12], 利用特定的网络结构生成类激活图. 如图 1 所示, 通过对最后一个卷积层的特征图进行加权融合计算类激活图, 但需要修改网络结构并重新训练模型, 限制了 CAM 技术的应用. 文献[13-15]通过特征图的梯度信息计算特征融合权重, 增强了 CAM 技术的泛化能力; Wang 等^[16]提出了 Score-CAM 可视化技术, 通过计算特征图对网络输出的贡献度作为类激活的权重, 该方法对多目标可视化、分类可视化的类激活图质量较已有方法均有提高; 文献[17]提出了 Ablation-CAM 算法, 通过特征图消融前后网络对目标类别的预测分数计算特征图的融合权重.

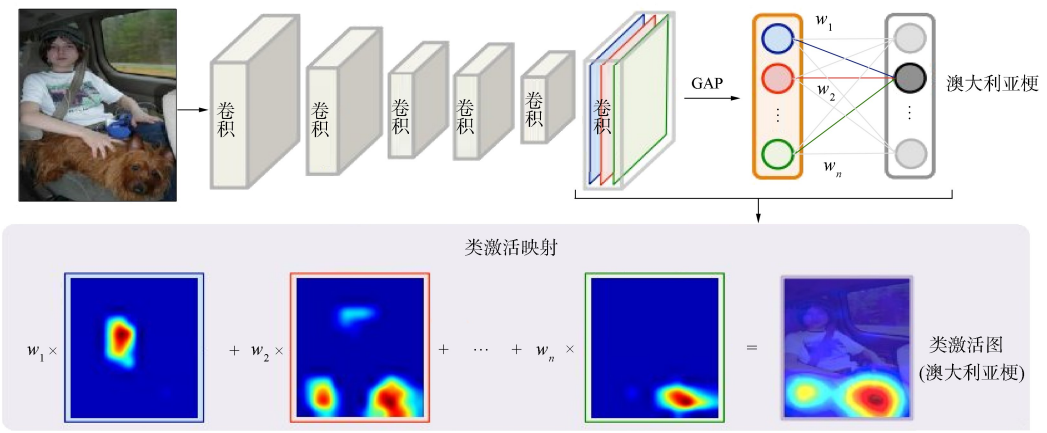


图 1 类激活映射流程
Fig. 1 Process of CAM

为解决类激活图中较多噪声和对目标物体定位不准确的问题, 本文基于消融分析^[18]和 Ablation-CAM 算法设计了 Ablation-CAM++ 和 Softmax Ablation-CAM(SA-CAM)两种改进算法. 首先通过消融实验考察深层网络特征对网络决策的影响, 进而衡量特征图与目标类别的相关性, 并使用网络的分类分数计算特征融合权重; 然后使用 ReLU 或 Softmax 函数对特征融合权重进行修正, 避免与目标类别无关的特征信息对类激活图造成干扰, 减少类激活图中的噪声并增强对目标物体的定位能力. 实验结果表明, 本文算法能对网络的决策做出有效说明, 有较强的模型解释能力.

1 算法设计

针对特定类别 c 生成的类激活图应只包含对 c 有积极影响的特征信息, 这样会使类激活图更专注于目标物体所在的图像区域. 基于 CAM 的算法通常选择 CNN 最后一个卷积层的所有特征图用于类激活图的计算, 忽略了不同特征信息与目标类别之间的相关性, 不仅导致算法的时间成本较高, 且可视化结果中含有较多噪声并对目标物体的定位不够精确, 无法对网络决策做出有效说明.

CNN 通过连续的卷积和池化操作, 得到拥有丰富语义信息的深层网络特征图, 然后全连接层根据众多的语义特征完成对输入图像的预测. 因此, 考察深层语义特征与网络预测结果的关系是一种加深模型理解的有效方法. 图 2 为将不同类别图像输入 CNN 后网络深层特征图的可视化结果, 其中第 1 行为原图像, 从左至右类别分别为柯基犬、恩特雷布赫山地犬、史宾格猎犬、热气球和鹈鹕; 第 2~4 行分别是不同卷积核提取特征的可视化结果. 由图 2 可见: 上述 3 种犬类之间存在很多相似的视觉特征, 如毛发颜色、形体等, 提取相关特征的卷积核对 3 种类别产生了相似的响应结果, 如图 2

中第 1~3 列所示;而这些卷积核从“热气球”和“鹈鹕”中提取到不规则特征或未产生响应,如图 2 中第 4,5 列可视化结果所示.表明网络深层提取的特征并不为某一类别所特有,而是具有归纳性的高级语义信息,是不同类别公共特征的抽象化.

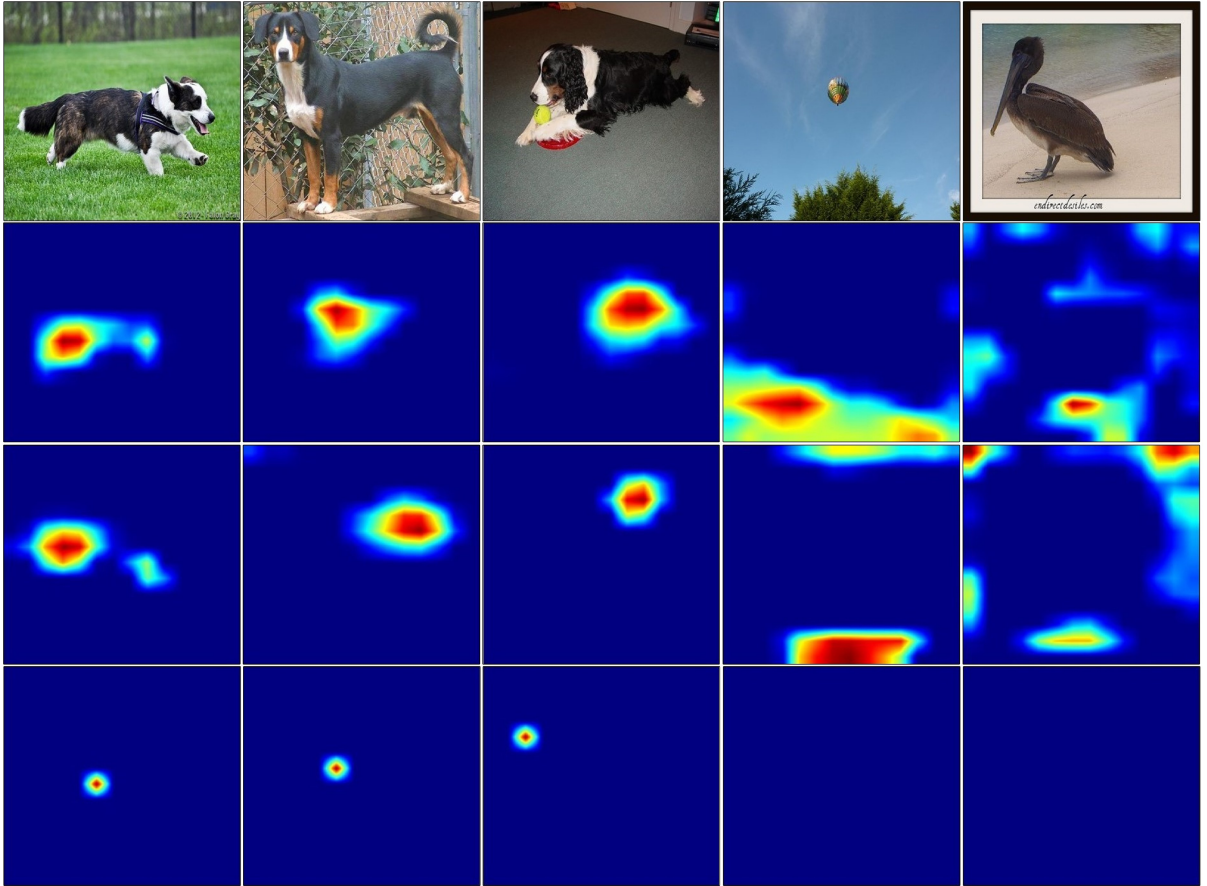


图 2 特征图可视化结果

Fig. 2 Visualization results of feature map

为改善类激活图质量,增强 CAM 算法的模型解释能力,本文基于消融分析和 Ablation-CAM 设计了 Ablation-CAM++ 和 SA-CAM 算法.首先通过消融实验考察深层网络特征与目标类别的相关性,并计算特征融合权重,然后采用 ReLU 或 Softmax 函数过滤无关的干扰信息,保证类激活图中只包含与目标类别相关的正向信息.图 3 为本文算法的流程,具体操作过程如下.

首先确定目标类别 c 和目标卷积层 l ,将原图像输入神经网络后获取网络对类别 c 的初始预测分数 y^c ,并提取卷积层 l 的输出特征图 $A = \{A_1, A_2, \dots, A_n\}$,如图 3 中流程一所示;然后沿通道维度,依次将每张特征图中所有神经元的值修改为 0,将消融处理后的特征图重新输入目标层 l 后的网络模块,获取网络对目标类别的消融分数 y_k^c ,并计算相应特征图的特征融合权重,用公式表示为

$$w_k^c = \frac{y^c - y_k^c}{y^c}, \quad (1)$$

其中 k 表示当前消融特征图的索引,该过程如图 3 中流程二所示.

通过上述消融实验得到每张特征图的权重,权重为正数表示特征的消融使得 CNN 对类别 c 的预测分数下降,表明该特征图中包含与类别 c 相关的正向信息,卷积核做出的响应会提高类别 c 的置信度;权重为负数表明类别 c 中不包含相应的特征,卷积核做出的响应会对网络决策做出干扰,导致类别 c 的置信度下降,因此应将该特征排除在类激活图的计算中.为消除无关特征的消极影响,本文采用 ReLU 或 Softmax 函数对特征融合权重做进一步处理.

ReLU 函数通常用作神经网络的激活函数,表达式为

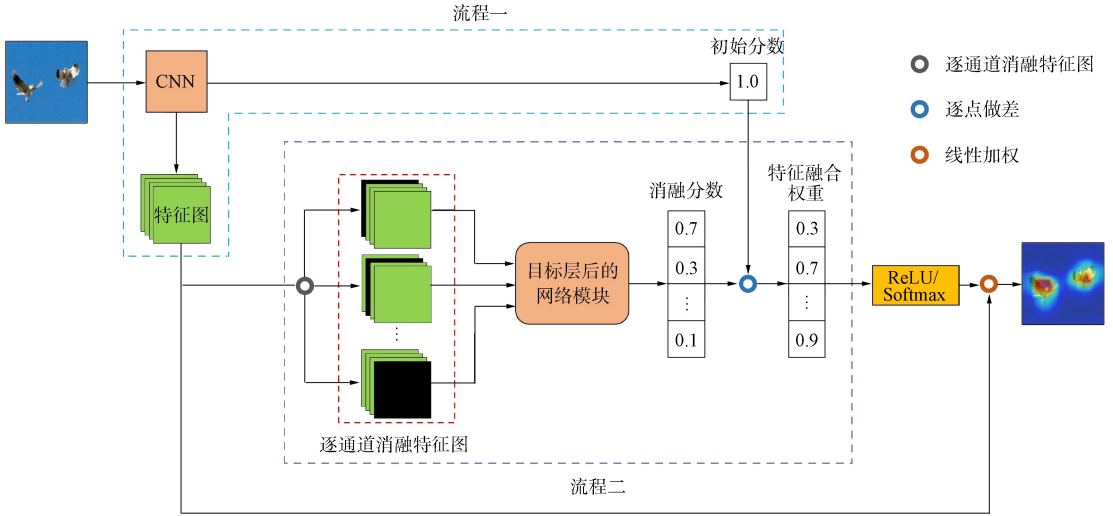


图 3 本文算法流程

Fig. 3 Flow chart of proposed algorithm

$$\text{ReLU}(x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases} \quad (2)$$

借助 ReLU 函数对负值权重进行修正, 且保持正值权重不变, 只采用与目标类别相关的特征信息进行加权融合, 可得到 Ablation-CAM++ 算法:

$$L_{\text{Ablation-CAM++}}^c = \sum_k \text{ReLU}(\omega_k^c) A_k. \quad (3)$$

CNN 输出层的分类分数经过 Softmax 函数可转化为不同类别的置信度, 计算过程为

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, 2, \dots, n. \quad (4)$$

本文使用 Softmax 函数对式(1)计算的权重进行处理, 作为最终的特征融合权重, 表示不同特征与目标类别相关的概率, 不仅可以将融合权重全部转化为非负值, 还可以放大主要特征的比重, 使其在最后的的结果中占主要作用, 增强对同类别多目标物体的识别和定位能力. 可得到 Softmax Ablation-CAM (SA-CAM) 算法:

$$L_{\text{SA-CAM}}^c = \sum_k \text{Softmax}(\omega_k^c) A_k. \quad (5)$$

2 实验

本文实验基于 PyTorch 深度学习框架, 操作系统为 Windows10 专业工作站版, 硬件平台为 Intel(R) Xeon(R) Silver 4114 CPU@2.20 GHz, 2.19 GHz 双处理器, 运行内存 64 GB, 配置 GeForce RTX 2080 Ti 显卡.

2.1 数据集和神经网络

本文使用当前图像分类、目标检测等任务中最常用的验证集 ILSVRC 2012^[19] 作为实验数据集, 该数据集共包含 5 万张高分辨率的自然图像, 1 000 个类别, 平均每个类别包含 50 个实例.

实验中使用预训练的 VGG-16^[20] 和 ResNet-34^[21] 神经网络对本文方法的有效性进行验证, 无需对网络进行重新训练, 统一提取网络最后一个卷积层的输出特征图用于生成类激活图. 原图像在输入 CNN 前, 将尺寸统一调整为 224 × 224, 并将像素中的值转换至 [0, 1] 内, 然后减去均值向量 (0.485, 0.456, 0.406) 和除以标准差向量 (0.229, 0.224, 0.225).

2.2 评估指标

解释图是一张与原图像有相同分辨率的图片, 但只包含原图像的一部分区域. 通过保留原图像中对 CNN 决策贡献度较高的像素点得到相应的解释图. 图 4 为基于 Ablation-CAM 算法生成的解释图,

保留 40% 像素.



图 4 解释图示例

Fig. 4 Instance of explanation map

2.2.1 Average Drop 和 Percent Increase 评估指标

Average Drop 是指将解释图和原图像分别输入 CNN 时,对目标类别预测得分的差距,计算公式为

$$\text{Average Drop} = \frac{1}{N} \sum_{i=1}^N \frac{\max\{0, Y_i^c - O_i^c\}}{Y_i^c}, \tag{6}$$

其中 Y_i^c 和 O_i^c 分别为 CNN 将原图像和解释图判定为类别 c 的分数, N 为数据量,使用 \max 函数排除解释图得分高于原图像的情况. Average Drop 数值越小说明算法效果越好.

Percent Increase 是指解释图得分高于原图像的样本占整个数据集的比重,计算公式为

$$\text{Percent Increase} = \frac{1}{N} \sum_{i=1}^N \text{Func}(Y_i^c < O_i^c), \tag{7}$$

其中 Func 为 Boolean 函数,当括号内表达式为真时返回 1,否则返回 0. Percent Increase 越大算法效果越好.

2.2.2 Deletion 和 Insertion 评估指标

Deletion(Insertion)^[9] 是指 Deletion(Insertion) 曲线下与坐标轴围成的面积(area under curve, AUC),图 5 为根据不同算法绘制的 Deletion 和 Insertion 曲线. Deletion(Insertion)数值越大(小)说明算法效果越好.

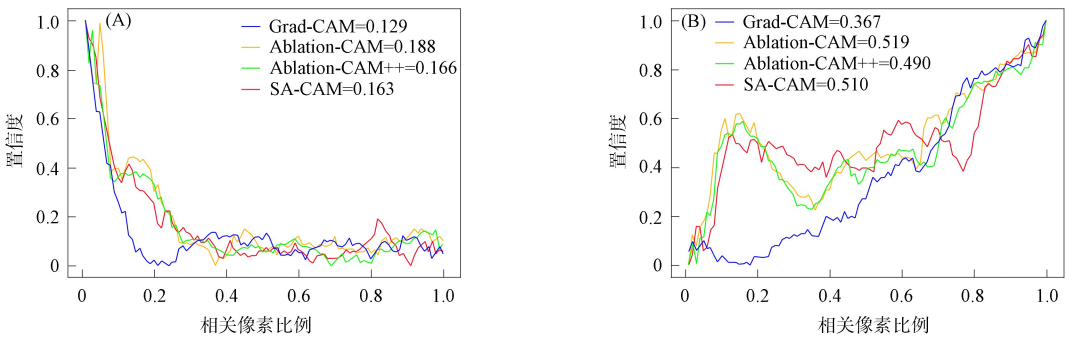


图 5 Deletion(A)和 Insertion(B)曲线示例

Fig. 5 Instance of Deletion curves (A) and Insertion curves (B)

2.3 结果与分析

为验证本文算法在验证集 ILSVRC 2012 上的性能,将 Ablation-CAM ++, SA-CAM 与 Grad-CAM^[13], Ablation-CAM^[17] 和 Score-CAM^[16] 算法进行对比分析.

首先从视觉上直观地进行对比分析,不同算法生成的类激活图可视化结果如图 6 所示,其中第 1 列为原图像,从上到下类别分别为豪猪、鹤、白鹤、澳洲青苹果和加拿大盘羊.由图 6 可见: Grad-CAM 受到梯度饱和问题影响以及 Ablation-CAM 受到无关特征的干扰,背景中含有较多噪声,不能准确地定位目标物体,特别是对同类别的多目标物体检测不精确,如图 6 中第 2,3 列的可视化结果所示; Score-CAM 及本文算法均能对目标物体实现准确识别和定位,避免噪声的干扰,且本文算法

略优于 Score-CAM 算法, 如图 6 中第 4~6 列可视化结果所示. 实验结果表明, 本文提出的 Ablation-CAM++ 和 SA-CAM 算法在识别和定位目标物体性能上优于其他算法.

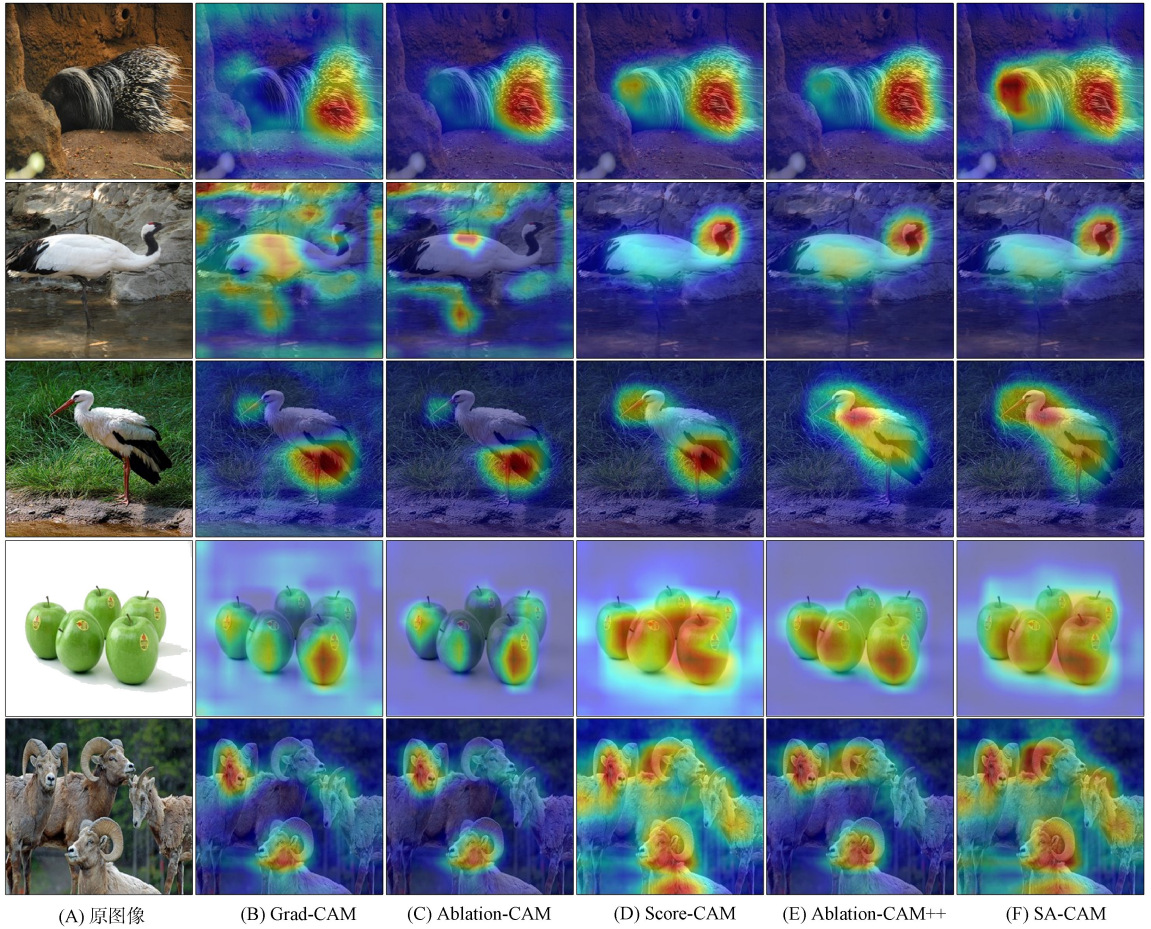


图 6 5 种类激活图的可视化结果

Fig. 6 Visualization results of five class activation maps

本文算法还有良好的类辨别能力, 即能在一张图像中对不同类别的物体进行识别和定位. 如图 7 所示, VGG-16 网络以 66.21% 的概率将原图像分类为挪威猎犬, 0.08% 的概率分类为蝴蝶犬. 虽然网络对两种类别的预测概率相差极大, 但 Ablation-CAM++ 和 SA-CAM 算法仍可以对相关区域进行突出显示, 准确定位目标物体在图片中的位置.

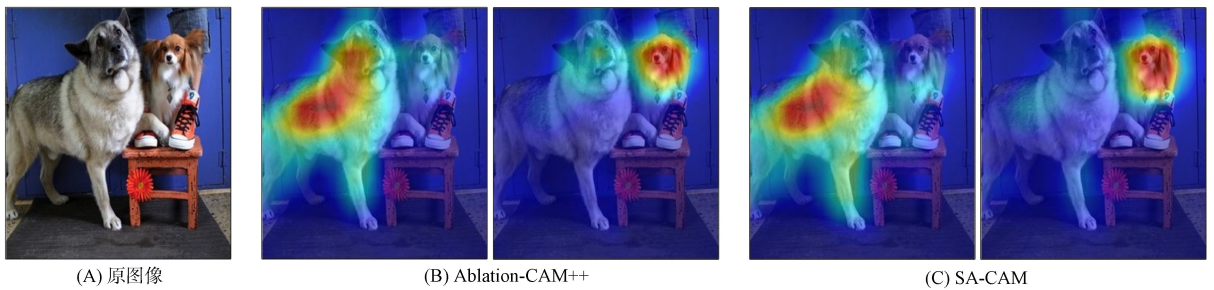


图 7 本文算法的类辨别结果

Fig. 7 Class discriminative results of proposed algorithm

下面进行定量对比实验分析. 表 1 和表 2 分别列出了不同算法在 VGG-16 和 ResNet-34 神经网络上的定量对比结果. 由表 1 可见: 本文算法均取得了各项评估指标的最优结果; 相比于其他算法中的最优结果, 本文算法在 Average Drop, Average Increase, Deletion 和 Insertion 指标上的最优结果分别提升了 1.31, 3.20, 0.24, 1.87 个百分点. 由表 2 可见: 在 Average Drop, Deletion 和 Insertion 指标上

本文算法均取得了最优结果,与其他算法中的最优结果相比,分别提高了 1.37,0.44,0.94 个百分点;在 Average Increase 指标上也取得了较好的结果.实验结果表明,本文算法具有优于其他算法的模型解释能力.

表 1 不同算法在 VGG-16 模型上的定量对比结果

Table 1 Quantitative comparison results of different algorithms on VGG-16 model

%

算法	Average Drop	Average Increase	Deletion	Insertion
Grad-CAM	35.65	10.15	33.13	58.17
Ablation-CAM	29.63	12.66	32.10	61.14
Score-CAM	29.54	10.40	30.13	61.46
Ablation-CAM++	28.23	15.86	31.18	63.33
SA-CAM	29.04	15.79	29.89	62.09

表 2 不同算法在 ResNet-34 模型上的定量对比结果

Table 2 Quantitative comparison results of different algorithms on ResNet-34 model

%

算法	Average Drop	Average Increase	Deletion	Insertion
Grad-CAM	30.15	17.19	34.95	66.22
Ablation-CAM	26.91	15.55	34.42	65.69
Score-CAM	29.87	15.79	35.02	65.78
Ablation-CAM++	25.54	16.35	34.49	67.16
SA-CAM	26.25	15.82	33.98	65.96

2.4 模型诊断

虽然 CNN 在大规模图像分类任务中表现优异,但在实际应用中仍有很多错误决策,影响用户对网络预测结果的可信度.本文算法除能对网络的正确决策进行解释外,还可以分析网络错误决策的原因,以便对网络进行诊断和改进,增加用户对模型的信任度.下面以 SA-CAM 算法为例(使用 Ablation-CAM++ 算法可得到相同结论)对 VGG-16 网络的错误决策进行分析和说明,考察网络失败的原因,以证明本文算法的实用性.

实验结果表明,神经网络误分类的原因主要有以下 3 点:

1) 不同类别物体会包含相同或者相似特征.如图 8 所示,VGG-16 将真实类别为“喜乐蒂牧羊犬”的图像误分类为“边境牧羊犬”;对比其他实例(图 8 中第 2,3 张图像)可见,两种犬类之间存在许多相似特征,如毛发颜色、体型等,网络根据这些相似特征进行决策时,易输出错误的预测结果.利用 SA-CAM 算法将两种类别进行可视化(图 8 中第 4,5 张图像),发现网络的关注区域基本重合.



图 8 归因分析 1) 示例

Fig. 8 Instance of attribution analysis 1)

2) 场景对网络决策的误导.某些类别总与特定的场景相关,网络的决策很大程度上会受场景影响,导致输出错误结果.如图 9 所示,VGG-16 将真实类别为“转轴”的图像误分类为“船桨”,而这两种类别物体大多出现在相似的场景中(图 9 中第 1~3 张图像).利用 SA-CAM 算法对原图像中的“转轴”可视化可见,除“转轴”所在区域外,网络对背景也有较高的关注度.

3) 同一张图像中包含不同类别物体.不同类别对象会出现在同一张图像中,而人工标注的图像真实类别只有一个.在这种情况下,网络从图像中提取到属于不同类别物体的特征,对多种类别产生响

应, 输出结果为响应最大的类别, 这通常会与图像的真实类别存在差异, 导致网络做出错误决策。



图 9 归因分析 2) 示例

Fig. 9 Instance of attribution analysis 2)

2.5 算法效率对比

下面对多种算法的运行速度进行对比分析, 进行如下实验: 首先从验证集 ILSVRC 2012 中随机挑选 3 个子集, 每个子集包含 500 张图片, 且任意两个子集之间没有公共样本; 然后统计 5 种算法在每个子集上的平均运行时间, 结果列于表 3. 由表 3 可见: Grad-CAM 算法只需进行一次前向传播和反向传播便可完成整个算法流程, 所需时间最短; Score-CAM 算法需要对所有特征图进行大量复杂操作, 算法复杂度较高, 运行时间最长; 本文算法和 Ablation-CAM 算法的平均运行时间均小于 0.1 s, 虽然运行效率略低于 Grad-CAM 算法, 但仍可以满足实时性要求, 且本文算法比 Grad-CAM 算法具有更好的可视化效果和模型解释能力。

表 3 不同算法的平均运行时间

Table 3 Average running time of different algorithms

编号	Grad-CAM 算法	Ablation-CAM 算法	Score-CAM 算法	Ablation-CAM++ 算法	SA-CAM 算法
1	0.080	0.095	5.64	0.100	0.094
2	0.079	0.098	5.45	0.095	0.101
3	0.083	0.094	5.62	0.099	0.098

综上所述, 针对基于类激活映射的可解释性方法生成的类激活图中含有较多噪声及对目标物体定位能力较差的问题, 本文根据卷积神经网络特征提取的特点, 通过消融实验考察深层视觉特征与目标类别的相关性, 并使用 ReLU 或 Softmax 函数对特征融合权重做进一步处理, 以减少无关特征的干扰, 增强类激活图的定位能力. 实验结果表明, 本文 Ablation-CAM++ 和 SA-CAM 算法可有效减少类激活图中的噪声, 有较强的目标定位能力和模型解释能力。

参 考 文 献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1-9.

[3] XUE B, TONG N N. Real-World ISAR Object Recognition Using Deep Multimodal Relation Learning [J]. IEEE Transactions on Cybernetics, 2020, 50(10): 4256-4267.

[4] XUE B, HE Y, JING F, et al. Robot Target Recognition Using Deep Federated Learning [J]. International Journal of Intelligent Systems, 2021, 36(12): 7754-7769.

[5] XUE B, TONG N N. DIOD: Fast and Efficient Weakly Semi-supervised Deep Complex ISAR Object Detection [J]. IEEE Transactions on Cybernetics, 2019, 49(11): 3991-4003.

[6] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3431-3440.

[7] SIMONYAN K, VEDALDI A, ZISSERMAN A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps [EB/OL]. (2014-04-19)[2023-04-01]. <https://arxiv.org/abs/1312.6034>.

- [8] SUNDARARAJAN M, TALY A, YAN Q. Axiomatic Attribution for Deep Networks [C]//International Conference on Machine Learning. New York: PMLR, 2017: 3319-3328.
- [9] PETSUK V, DAS A, SAENKO K. Rise: Randomized Input Sampling for Explanation of Black-Box Models [EB/OL]. (2018-09-25)[2023-04-04]. <https://arxiv.org/abs/1806.07421>.
- [10] RIBEIRO M T, SINGH S, GUESTRIN C. “Why Should I Trust You?” Explaining the Predictions of Any Classifier [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1135-1144.
- [11] ZHOU B L, KHOSLA A, LAPEDRIZA A, et al. Learning Deep Features for Discriminative Localization [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2921-2929.
- [12] LIN M, CHEN Q, YAN S C. Network in Network [EB/OL]. (2014-03-04)[2023-04-10]. <http://arxiv.org/abs/1312.4400>.
- [13] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 618-626.
- [14] CHATTOPADHAY A, SARKAR A, HOWLADER P, et al. Grad-CAM ++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ: IEEE, 2018: 839-847.
- [15] JIANG P T, ZHANG C B, HOU Q, et al. Layercam: Exploring Hierarchical Class Activation Maps for Localization [J]. IEEE Transactions on Image Processing, 2021, 30: 5875-5888.
- [16] WANG H F, WANG Z F, DU M N, et al. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2020: 24-25.
- [17] RAMASWAMY H G. Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-Free Localization [C]//Proceeding of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2020: 983-991.
- [18] ZHOU B L, SUN Y Y, BAU D, et al. Revisiting the Importance of Individual Units in CNNs via Ablation [EB/OL]. (2018-06-07)[2023-04-13]. <http://arxiv.org/abs/1806.02891>.
- [19] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [20] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [EB/OL]. (2015-04-10)[2023-04-17]. <https://arxiv.org/abs/1409.1556>.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778.

(责任编辑:韩 啸)