

# 基于可重参的自适应激活函数算法

吴限, 陶青川

(四川大学电子信息学院, 四川成都 610065)

**摘要:** 经典轻量化网络的特征提取能力存在一定不足, 在推理阶段几乎不引入额外计算的基础上, 对可重参自适应激活函数进行研究, 提出RepAct激活函数。RepAct的自适应可重参多分支结构通过不同特征信息的分支权重进行自适应学习, 显著提升了轻量化网络的学习能力。在各经典轻量化CNN和ViT网络中, RepAct可以有效提升分类任务的精度。在ImageNet100上, MobileNetV3-Small的Top-1准确率提升了6.9%; 在Cifar100上, MobileNetV3-Small的Top-1准确率提升了5.71%, 超越了其他主流的激活函数。此外, 通过GradCAM可视化, 揭示了RepAct提升网络特征提取能力的原因。

**关键词:** 自适应激活函数; 可重参结构; 轻量化网络

**DOI:** 10.11907/rjdk.241937

**中图分类号:** TP183

**文献标识码:** A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)002-0083-06



## Adaptive Activation Function Algorithm Based on Reparameterility

WU Xian, TAO Qingchuan

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

**Abstract:** The feature extraction capability of classic lightweight networks exhibits certain limitations. To address this, we investigate reparameterizable adaptive activation functions and propose the RepAct activation function, which introduces negligible additional computational cost during the inference stage. RepAct's adaptive reparameterizable multi-branch structure leverages branch-specific weights to adaptively learn diverse feature information, significantly enhancing the learning capacity of lightweight networks. RepAct demonstrates significant accuracy improvements for classification tasks across various classic lightweight CNN and ViT networks. For instance, on ImageNet100, the Top-1 accuracy of MobileNetV3-Small increases by 6.9%; on CIFAR100, the Top-1 accuracy improves by 5.71%, surpassing other mainstream activation functions. Furthermore, Grad-CAM visualizations reveal the mechanism behind RepAct's enhancement of the network's feature extraction capabilities.

**Key Words:** adaptive activation function; reparameterizable structure; lightweight networks

## 0 引言

在经典的神经网络设计中, 不同网络不同层中激活函数的设计及选择通常依赖手工设计经验<sup>[1-2]</sup>或NAS搜索<sup>[3-4]</sup>进行适配。自适应激活函数的出现有效提升了网络性能<sup>[4-6]</sup>。自适应激活函数的自适应参数常作用在特征图的不同维度下, 引入的参数与计算量虽小, 但逐元素操作内存开销往往形成轻量化的网络推理瓶颈<sup>[7-10]</sup>。本文注意到可重参的卷积结构凭借其训练中的多分支结构提升了网络的特征捕捉能力, 在推理阶段利用线性计算特性恢

复为单分支模型进行部署, 在提升网络特征提取能力的同时, 兼顾了网络的推理速度, 使各经典网络结构重焕生机<sup>[11-15]</sup>。受此启发, 本文提出可重参自适应激活函数RepAct, 训练时多分支激活, 推理时单分支激活, RepAct在网络不同层中会随梯度下降自适应调整各分支权重。

## 1 相关工作

### 1.1 激活函数

不同类型的激活函数对神经网络的特征提取能力及梯度传播有着较大影响, 如ReLU<sup>[16]</sup>、SoftPlus<sup>[17]</sup>、ELU<sup>[18]</sup>、

收稿日期: 2024-11-05

扫描二维码阅读全文:



**作者简介:** 吴限(2000-), 男, 四川大学电子信息学院硕士研究生, 研究方向为计算机应用与模式识别; 陶青川(1972-), 男, 博士, 四川大学电子信息学院副教授、硕士生导师, 研究方向为模式识别与智能系统。本文通讯作者: 陶青川。

Mish<sup>[19]</sup>等激活函数及其变体<sup>[6-7,20-21]</sup>。已有研究主要关注激活函数的连续可微性以及是否存在上下界饱和的问题。在不同数据集中,对于不同网络,激活函数的设计仍然依靠手工经验<sup>[1-2]</sup>或者NAS搜索<sup>[3-4]</sup>。每次尝试的代价往往是高昂的,因此如何设计适合当前数据分布及网络结构的激活函数是一个长期困扰网络设计的问题。MobileNetV2讨论了倒残差结构中特征图通道维狭窄时激活函数过强的危害,印证了激活层与网络层级结构的设计空间相辅相成<sup>[1]</sup>。

在自适应激活函数中,PRReLU<sup>[5]</sup>作为LReLU的自适应改进,通过引入一个可学习斜率参数参与训练,在几乎没有额外推理成本情况下,使网络在ImageNet数据集的分类任务中达到了接近人类视觉水平的分类性能<sup>[22]</sup>。FReLU引入2D漏斗空间门控机制使激活函数拥有动态的空间建模激活能力<sup>[8]</sup>。DYReLU通过类似SE<sup>[23]</sup>的辅助模块对当前特征图进行动态计算后再进行激活,显著提升了轻量化网络的特征提取能力<sup>[7]</sup>。ARReLU作为可学习的激活函数,具有梯度选择性放大的能力,可以促进神经网络快速收敛<sup>[6]</sup>。ACON通过自适应地选择激活神经元,在轻量化网络上具有良好精度,并提出系列不同粒度与动静态组合的设计<sup>[24]</sup>。

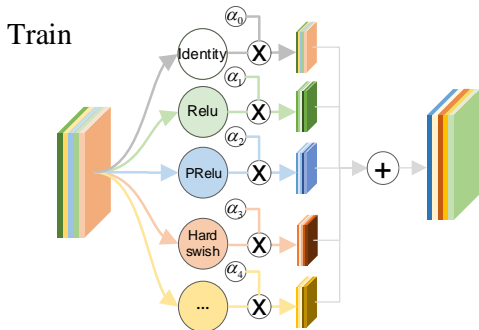


Fig. 1 RepAct multi-branch training structure and single-branch inferring structure

图1 RepAct多分支训练结构与单分支推理结构

RepAct结构是通用型结构,将常见的激活函数进行多分支组合,并将各分支计算特征图相加,使不同非线性分支下激活后映射的特征信息融合,以增强网络的特征提取能力。

RepAct结构在训练与部署中呈现两套不同的范式,训练时为多分支结构,部署时重参为单分支结构。当选取分支的激活函数均为幂函数形式时,各分支分段的可重参程度更高。因此,本文RepAct均选用适合轻量化网络部署的幂函数型激活函数。RepAct的函数分段范围为各幂函数型激活函数分段范围的叠加。在重参恢复单分支后,推理阶段的单分支访存消耗仍为非动态计算类激活函数所需的访存消耗,即只需遍历访存特征图一次。重参后RepAct的计算复杂度近似各分支激活函数的最大计算复杂度,如式(1)所示。

$$O(\text{RepAct}) \approx \max(O(\text{RepAct}_n)) \quad (1)$$

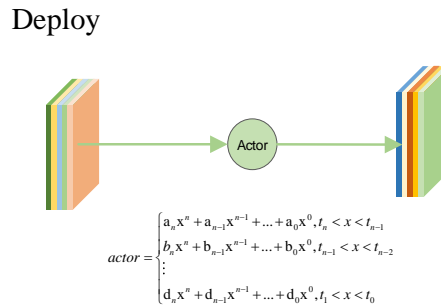
其中, $O(\text{RepAct}_n)$ 为RepAct第n个分支的计算复杂度。

## 1.2 可重参结构

可重参卷积结构将网络训练多分支与部署单分支的网络结构分离,在训练阶段尽可能获得更多信息,在部署阶段则利用线性特点等对其网络结构进行分支融合。苹果MobileOne对MobileNet网络结构进行重参数化,得到移动端推理时长为1ms级别的主干网络<sup>[13]</sup>。RepViT对ViT角度进行重参设计,超越了现有的最先进的轻量级ViTs<sup>[15]</sup>。目前可重构结构的设计及应用研究仍聚焦于密集线性计算模块中,还未充分关注激活函数的可重参形态,激活函数占据一个网络模型中非线性运算的重要组成部分<sup>[25-26]</sup>。因此,探索激活函数的可重参形态对于网络特征提取能力的提升十分重要。

## 2 RepAct模型

本文提出的RepAct是一种可重参自适应激活函数,在训练中利用常见的激活函数进行多分支组合训练,通过梯度下降自适应调整各分支权重因子。在推理阶段,将各分支权重因子在不同分段下进行重参,恢复单分支结构。具体结构如图1所示。



在部分网络结构中,非线性运算会破坏层间的特征信息。在MobileNetV2的设计中,验证了在某些层中激活函数的非线性行为会破坏特征信息的传递,导致网络性能下降<sup>[1]</sup>。因此,在RepAct结构中,将恒等映射也作为多分支中的一个分支,通过各分支的可学习系数,充分解耦线性和非线性特征,自适应平衡网络不同层间的非线性特征表达能力及线性特征保留能力。

选取Identity、ReLU、PReLU<sup>[27]</sup>、HardSwish 4种幂函数型激活函数组成RepAct的4个不同分支,RepAct在训练阶段的激活函数表示为式(2)。

$$X' = \alpha_0 \text{Identity}(X) + \alpha_1 \text{ReLU}(X) + \alpha_2 \text{PReLU}(X) + \alpha_3 \text{HardSwish}(X) \quad (2)$$

在网络训练阶段,各分支权重因子为可学习自适应参数 $\alpha$ 。在网络推理阶段, $\alpha$ 作为静态常数进行重参融合。因此,RepAct激活函数推理阶段可转化为式(3),恢复推理阶段的单分支幂函数分段形式。在该组合中,RepAct重参

后单分支分段区间为各分支分段区间的并集,  $\delta_n$  为重参后的参数。

$$X' = \text{RepAct}(X) = \begin{cases} x^*(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3), & x \geq 3 \\ x^*x^*(\alpha_3/6) + x^*(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3/2), & 0 \leq x < 3 \\ x^*x^*(\alpha_3/6) + x^*(\alpha_0 + 0 + (\text{PReLU.T}*\alpha_2) + \alpha_3/2), & -3 \leq x < 0 \\ x^*(\alpha_0 + 0 + (\text{PReLU.T}*\alpha_2) + 0), & x < -3 \end{cases} \quad (3)$$

$$= \begin{cases} x^*\delta_1, & x \geq 3 \\ x^*x^*\delta_2 + x^*\delta_3, & 0 \leq x < 3 \\ x^*x^*\delta_2 + x^*\delta_4, & -3 \leq x < 0 \\ x^*\delta_5, & x < -3 \end{cases}, \text{其中} \begin{cases} \delta_1 = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 \\ \delta_2 = \alpha_3/6 \\ \delta_3 = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3/2 \\ \delta_4 = \alpha_0 + 0 + (\text{PReLU.T}*\alpha_2) + \alpha_3/2 \\ \delta_5 = \alpha_0 + 0 + (\text{PReLU.T}*\alpha_2) + 0 \end{cases}$$

在 RepAct 重参后,模型在每层 RepAct 中仅增加了 5 个  $\delta$  参数。RepAct 各分支中  $\max(O(\text{RepAct}_n))$  为 HardSwish, 如式(4)、式(5)所示,RepAct 相较于  $\max(O(\text{RepAct}_n))$  的计算量仅增加了一个 if 运算和两个乘法运算。

$$\text{HardSwish}(X) = \begin{cases} x, & x \geq 3 \\ x^*x^*\frac{1}{6} + x^*\frac{1}{2}, & -3 \leq x < 3 \\ 0, & x < -3 \end{cases} \quad (4)$$

$$\text{RepAct}(X) = \begin{cases} x^*\delta_1, & x \geq 3 \\ x^*x^*\delta_2 + x^*\delta_3, & 0 \leq x < 3 \\ x^*x^*\delta_2 + x^*\delta_4, & -3 \leq x < 0 \\ x^*\delta_5, & x < -3 \end{cases} \quad (5)$$

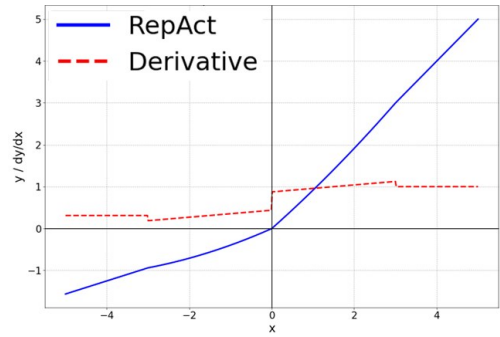
在部署阶段, GPU 或 AI 加速边缘设备中激活函数通常与卷积计算融合, 以减少频繁的内存访问。卷积过程通常是并行或部分并行计算, 因此对融合后的分段幂级数类激活函数, 推理速度的瓶颈为各分段最大计算量分段。RepAct 重参后各个分段中最大计算量的分段计算复杂度仍为  $O(ax^2 + bx)$ , 与  $\max(O(\text{RepAct}_n))$  一致, 可以认为  $O(\text{RepAct}) \approx \max(O(\text{RepAct}_n))$ 。

在 RepAct 各分支权重因子初始化, 简单地将分支权重  $\alpha_n$  平均初始化使之和为 1, 保证在网络训练的最初阶段特征值和梯度不会因过小而消失, 从而使各分支能够有效地传递不同的激活特征。初始化后的 RepAct 激活函数及其一阶导数如图 2 所示。图 2(a) — 图 2(c) 分别展示了各分支在  $[0, 1]$  范围内的激活映射及其对应的一阶导数。

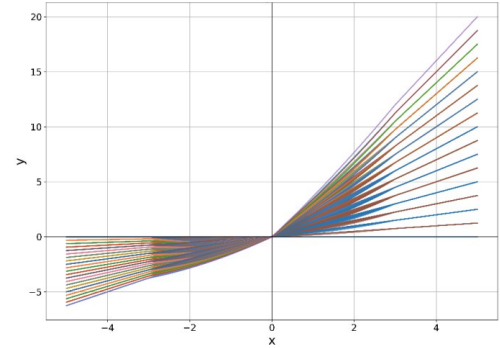
### 3 实验与分析

#### 3.1 实验环境

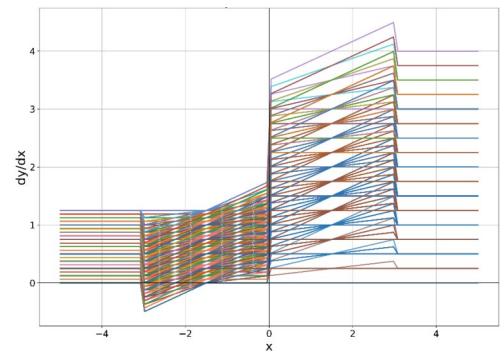
实验在 GPU NVIDIA GeForce 3090Ti (24 GB 显存)、32GB RAM 的硬件环境下进行, 操作系统为 Ubuntu 22.04, CUDA 版本为 11.7。本文采用 ImageNet100 和 Cifar100 数据集进行实验, 其中 ImageNet100 是从 ImageNet2012-1k 数据集中抽取 100 个类别构成的子集<sup>[22]</sup>。针对各数据集均采用各自经典的训练参数进行训练, 对于 ImageNet100 数据集, 图像大小设为  $224 \times 224$ , batch size 为 32, 初始学习率为 0.01; 对于 Cifar100 数据集, 图像大小设为  $32 \times 32$ , batch size 为 128, 初始学习率为 0.1, 固定随机种子为 0, 以



(a) Initialized RepAct activation function and its first derivative  
(a) 初始化 RepAct 激活函数及其一阶导数



(b) RepAct activation function in the  $\alpha_n [0-1]$  range  
(b)  $\alpha_n [0-1]$  范围下 RepAct 激活函数



(c) RepAct first derivative in the  $\alpha_n [0-1]$  range  
(c)  $\alpha_n [0-1]$  范围下 RepAct 一阶导数

Fig. 2 Initialization and mapping characteristics of RepAct activation function and its first-order derivative

图 2 RepAct 激活函数及其一阶导数初始化与映射特性

确保实验可复现。

### 3.2 对比实验

轻量化模型在边缘应用中具有更多机会,为轻量化模型设计合适的激活函数对边缘侧的应用部署具有重要意义<sup>[28]</sup>。在经典轻量化CNN模型<sup>[10,29]</sup>及小型ViT模型<sup>[30,31]</sup>上简单地替换各原始网络激活函数为RepAct进行对比实验,实验结果如表1—表3所示。在Baseline中,均使用原始网络的激活函数,也是各自适应激活函数替换的部分。其中,R代表ReLU,H代表HardSwish,G代表GELU。

**Table 1 Comparison of Top-1 accuracy between original activation networks and RepAct networks on ImageNet100**

表1 在ImageNet100上原始激活网络与RepAct网络Top-1acc.对比

ImageNet100/ Top-1 acc.	Baseline	PReLU <sup>[5]</sup>	ACONC <sup>[24]</sup>	RepAct
ShuffleNetV2x0.5 <sup>[10]</sup>	73.0(R)	+3.64	+3.96	<b>+5.54</b>
ShuffleNetV2x1.0 <sup>[10]</sup>	80.8(R)	+0.82	+0.78	<b>+1.92</b>
MobileNetV3-Small <sup>[28]</sup>	72.88(H)	+6.88	+6.34	<b>+6.9</b>
MobileNetV3-Large <sup>[28]</sup>	79.78(H)	+3.8	+3.22	<b>+4.38</b>
SwinTinyPatch4Window7 <sup>[30]</sup>	70.14(G)	/	/	+9.54
VitVasePatch16 <sup>[29]</sup>	58.92(G)	/	/	+3.74

**Table 2 Comparison of Top-1 accuracy between original activation networks and RepAct networks on Cifar100**

表2 在Cifar100上原始激活网络与RepAct网络Top-1acc.对比

Cifar100/ Top-1 acc.	Baseline	PReLU <sup>[5]</sup>	ACONC <sup>[24]</sup>	RepAct
ShuffleNetV2x0.5 <sup>[10]</sup>	66.71(R)	+0.72	-0.58	+0.04
ShuffleNetV2x1.0 <sup>[10]</sup>	70.82(R)	+0.36	+0.59	+0.43
MobileNetV3-Small <sup>[28]</sup>	67.35(H)	+1.77	+3.96	<b>+5.71</b>
MobileNetV3-Large <sup>[28]</sup>	70.85(H)	+2.23	+2.25	<b>+3.74</b>

在表1中,各经典轻量化CNN网络得益于RepAct在训练中各分支特征自适应的组合方式,在ImageNet100上显

**Table 3 Comparison of Top-1 accuracy of different activation functions in MobileNetV3-small on Cifar100**

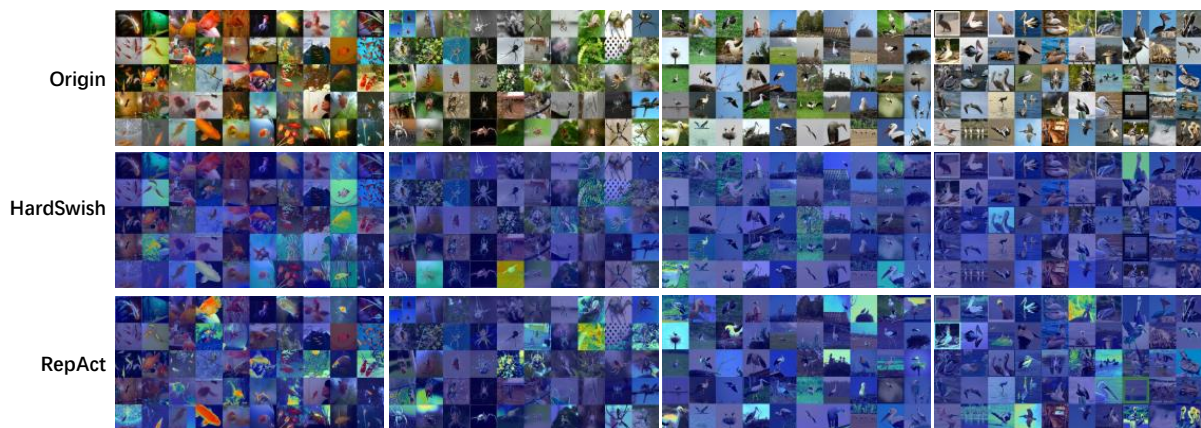
表3 MobileNetV3-Small在Cifar100中各激活函数Top-1acc.对比

Act	Top-1 acc.	Act	Top-1 acc.	Act	Top-1 acc.
HardSwish <sup>[28]</sup>	67.35	ReLU <sup>[16]</sup>	+1.46	Swish <sup>[4]</sup>	+2.72
RepAct	<b>+5.71</b>	LReLU <sup>[32]</sup>	+2.98	Mish <sup>[19]</sup>	+2.05
ELU <sup>[18]</sup>	-1.83	PReLU <sup>[5]</sup>	+1.77	FReLU <sup>[8]</sup>	+3.56
SELU <sup>[33]</sup>	-2.37	ARReLU <sup>[6]</sup>	+2.00	DYReLU <sup>[7]</sup>	+0.94
Identity	-35.90	Softplus <sup>[34]</sup>	-14.42	ACONC <sup>[24]</sup>	+3.96
GELU <sup>[35]</sup>	+3.62	CELU <sup>[36]</sup>	-1.75	SiLU <sup>[35]</sup>	+1.90

著提升了分类精度,且在推理阶段恢复单分支,不引入额外的特征图访存操作。这一点对于更加轻量的ShuffleNetV2x0.5和MobileNetV3-Small显得尤为明显,Top1准确率分别提升了5.54%和6.9%。在表2中,由Cifar100数据集可以看出,RepAct相较于其他自适应激活函数在MobileNetV3类型网络上的提升更加显著。在表3中,RepAct相较于其他的主流激活函数展现出最佳性能。另外,在表1中,RepAct还在VitVasePatch16及SwinTinyPatch4Window7上进行了验证,在同样的训练条件下,RepAct也显著提升了网络的分类精度,验证了RepAct的广泛适用性。

### 3.3 GradCAM可视化

在图3的GradCAM可视化对比图中,RepAct网络浅层提取的纹理信息更加靠近该分类类别所需的局部纹理信息,这与RepAct前向过程中对有效信息特征放大,以及反向传播中带有类别信息的梯度有效缩放回传吻合<sup>[37]</sup>。反向链式求导中网络深层梯度将更多类别的信息回传到网络浅层,网络浅层学习捕捉到了具有更多类别信息的纹理特征,从而更高效地驱使模型来适应任务需求和数据分布。



**Fig. 3 Comparison of MobileNetV3-Large-First-Feature GradCAM visualization results**

图3 MobileNetV3-Large-First-Feature GradCAM可视化效果对比

在图4中,原始网络在深层卷积中只关注了类别物体的局部特征,且网络注意热力图与类别物体的位置往往有一些偏移。RepAct则对整个分类类别的语义级关注能力更加优异,显著聚焦了整个类别物体,同时对物体周围的环境也有一定关注。

## 4 结语

本文提出的可重参自适应RepAct激活函数利用多分支训练单分支推理的特性,在几乎不引入额外参数和额外

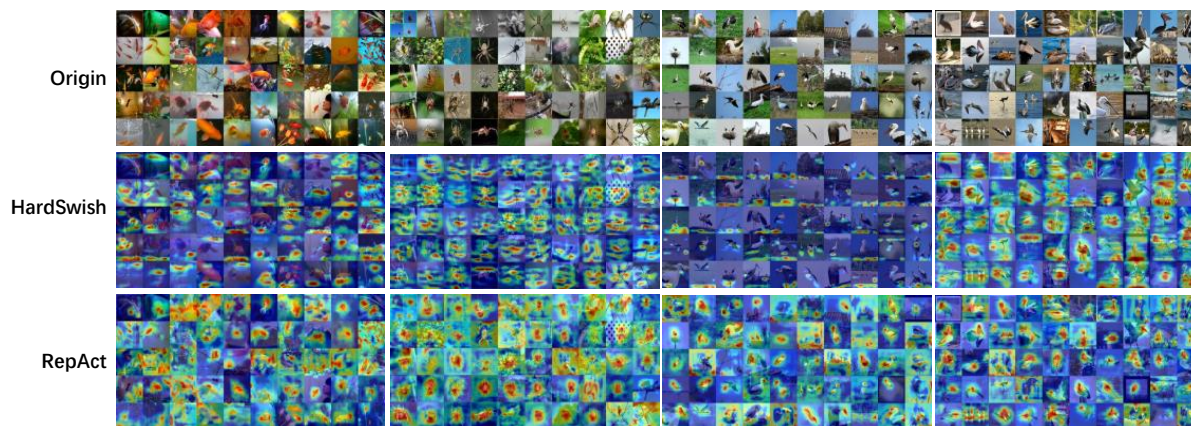


Fig. 4 Comparison of MobileNetV3-Large-Last-Feature GradCAM visualization results

图4 MobileNetV3-Large-Last-Feature GradCAM 可视化效果对比

访存计算的情况下,增强了轻量化网络学习能力,并且在不同的经典轻量化网络与经典公开数据集中均得到了验证。然而,对于 ResNet 等具有良好梯度设计的网络结构,RepAct 可能会导致网络过拟合<sup>[38]</sup>。因此,如何平衡 RepAct 的学习能力与非轻量网络中的过拟合问题仍有待未来继续研究。

#### 参考文献:

- [1] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510–4520.
- [2] XU B, WANG N, CHEN T, et al. Empirical evaluation of rectified activations in convolutional network [DB/OL]. <https://arxiv.org/pdf/1505.00853>.
- [3] PHAM H, GUAN M, ZOPH B, et al. Efficient neural architecture search via parameters sharing [C]// International Conference on Machine Learning, 2018: 4095–4104.
- [4] RAMACHANDRAN P, ZOPH B, LE Q V. Searching for activation functions [DB/OL]. <http://arxiv.org/pdf/1710.05941>.
- [5] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification [C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026–1034.
- [6] CHEN D, LI J, XU K. Arelu: attention-based rectified linear unit [DB/OL]. <https://arxiv.org/pdf/2006.13858>.
- [7] CHEN Y, DAI X, LIU M, et al. Dynamic relu [C]// European Conference on Computer Vision, 2020: 351–367.
- [8] MA N, ZHANG X, SUN J. Funnel activation for visual recognition [C]// Proceedings of 16th European Conference on Computer Vision, 2020: 351–368.
- [9] CHEN H, WANG Y, GUO J, et al. VanillaNet: the power of minimalism in deep learning [DB/OL]. <https://arxiv.org/pdf/2305.12972>.
- [10] MA N, ZHANG X, ZHENG H, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C]// Proceedings of the European Conference on Computer Vision, 2018: 116–131.
- [11] DING X, ZHANG X, MA N, et al. RepVGG: making VGG-style ConvNets great again [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733–13742.
- [12] DING X, GUO Y, DING G, et al. ACNet: strengthening the kernel skeletons for powerful CNN via asymmetric convolution blocks [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1911–1920.
- [13] VASU P K A, GABRIEL J, ZHU J, et al. MobileOne: an improved one millisecond mobile backbone [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7907–7917.
- [14] DING X, ZHANG X, HAN J, et al. Scaling up your kernels to 31x31: revisiting large kernel design in CNNs [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11963–11975.
- [15] WANG A, CHEN H, LIN Z, et al. Replit: revisiting mobile CNN from vit perspective [DB/OL]. <https://arxiv.org/pdf/2307.09283>.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2017, 60: 84–90.
- [17] DUGAS C, BENGIO Y, BÉLISLE F, et al. Incorporating second-order functional knowledge for better option pricing [C]// Proceedings of the 14th International Conference on Neural Information Processing Systems, 2000: 451–457.
- [18] CLEVERT D, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (elus) [DB/OL]. <https://arxiv.org/pdf/1511.07289>.
- [19] MISRA D. Mish: a self regularized non-monotonic activation function [DB/OL]. <https://arxiv.org/pdf/1908.08681>.
- [20] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C]// International Conference on Machine Learning, 2013: 3.
- [21] MASTROMICHALAKIS S. ALReLU: a different approach on leaky ReLU activation function to improve neural networks performance [DB/OL]. <https://arxiv.org/pdf/2012.07564>.
- [22] 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, 2009: 248–255.
- [23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Pro-

- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.
- [24] MA N, ZHANG X, LIU M, et al. Activate or not: learning customized activation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8032–8042.
- [25] LI S C, WANG J, SONG Y F, et al. TriCh-LKRepNet: large-kernel convolutional malware classification network combining three-channel mapping and structural re-parameterization [J]. Journal of Electronics, 2024, 52(7): 2331–2340.  
李思聪, 王坚, 宋亚飞, 等. TriCh-LKRepNet: 融合三通道映射与结构重参数化的大核卷积恶意代码分类网络[J]. 电子学报, 2024, 52(7): 2331–2340.
- [26] YE H M, LU S Q, CHENG X H, et al. Attention mechanism algorithm based on re-parameterization [J]. Computer Engineering and Design, 2024, 45(10): 2960–2969.  
叶汉民, 陆泗奇, 程小辉, 等. 基于重参数化的注意力机制算法[J]. 计算机工程与设计, 2024, 45(10): 2960–2969.
- [27] DUBEY S R, SINGH S K, CHAUDHURI B B. Activation functions in deep learning: a comprehensive survey and benchmark [J]. Neurocomputing, 2022, 503: 92–108.
- [28] SHEN K T, MIN T Y, HU D M. Infrared image super-resolution reconstruction combining lightweight ViT and CNN [J]. Software Guide, 2023, 22(2): 21–27.  
沈恺涛, 闵天悦, 胡德敏. 融合轻量级 ViT 与 CNN 的广范围红外图像超分辨率重建[J]. 软件导刊, 2023, 22(2): 21–27.
- [29] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenet3 [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314–1324.
- [30] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [DB/OL]. <https://arxiv.org/pdf/2010.11929>.
- [31] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012–10022.
- [32] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines [C]// Proceedings of the 27th International Conference on Machine Learning, 2010: 807–814.
- [33] KLAMBAUER G, UNTERTHINER T, MAYR A, et al. Self-normalizing neural networks [DB/OL]. <https://arxiv.org/abs/1706.02515>.
- [34] ZHENG H, YANG Z, LIU W, et al. Improving deep neural networks using softplus units [C]// 2015 International Joint Conference on Neural Networks, 2015: 1–4.
- [35] HENDRYCKS D, GIMPEL K. Gaussian error linear units (gelu) [DB/OL]. <https://arxiv.org/pdf/1606.08415>.
- [36] BARRON J T. Continuously differentiable exponential linear units [DB/OL]. <https://arxiv.org/pdf/1704.07483>.
- [37] 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, 2017: 618–626.
- [38] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.

(责任编辑:黄健)