

自适应局部注意力遮挡图像的情绪识别

陶煦¹, 黄文豪², 张萍¹, 丁凯², 李增辉²

1. 皖江新兴产业技术发展中心, 安徽 铜陵 244000

2. 安徽大学 电子信息工程学院, 安徽 合肥 230601

DOI: 10.61369/TACS.2025140044

摘 要 : 针对真实场景中面部遮挡及复杂环境导致情绪识别率下降的问题, 本文提出了一种自适应局部注意力算法。该方法摒弃了传统的图像级裁剪策略, 通过在特征层引入局部注意力单元进行逻辑分块与编码, 精准捕捉细粒度情感表征; 同时设计相对置信度模块, 利用动态权重分配机制自适应聚焦非遮挡区域并抑制噪声干扰。多个主流数据集的实验结果证实, 该算法在识别精度与抗遮挡鲁棒性上均优于现有主流方法。

关 键 词 : 自适应局部注意力; 动态权重分配; 相对置信度

Adaptive Local Attention Occlusion for Image Emotion Recognition

Tao Xu¹, Huang Wenhao², Zhang Ping¹, Ding Kai², Li Zenghui²

1. Wanjiang Emerging Industry Technology Development Center, Tongling Anhui 244000

2. School of Electronic and Information Engineering, Anhui University, Hefei, Anhui 230601

Abstract : To address the degradation of emotion recognition rates caused by facial occlusion and complex environments in real-world scenarios, this paper proposes an Adaptive Local Attention algorithm. Abandoning traditional image-level cropping strategies, this method introduces local attention units at the feature level to perform logical partitioning and encoding, thereby accurately capturing fine-grained emotional representations. Furthermore, a Relative Confidence Module is designed to utilize a dynamic weight allocation mechanism, enabling the network to adaptively focus on non-occluded regions while suppressing noise interference. Experimental results on multiple mainstream datasets confirm that the proposed algorithm outperforms existing mainstream methods in both recognition accuracy and occlusion robustness.

Keywords : adaptive local attention; dynamic weight allocation; relative confidence

一、面部情绪识别研究现状

情绪是人类主观认知的一系列表达, 表现为对客观事物的态度和相应的行为反应^[1]。作为情感计算领域的核心分支, 面部表情识别 (FER) 不仅是人机交互智能化的关键一环, 在医疗辅助与驾驶监控等场景中也发挥着重要作用。面部表情识别 (FER) 正从传统手工特征向深度语义学习演进。为捕捉细微表情, Ruan 等^[2]提出 FDR 框架, 通过特征分解与重构提升模型感知力; Xue 等^[3]利用 MAD 机制迫使网络关注多样化局部区域, 抑制背景噪声。针对更复杂的面部遮挡与姿态问题, 主流方案分为 GAN 与注意力机制。Wang 等^[4]尝试利用 GAN 重构全脸以校正偏差, 但存在网络复杂及伪影风险。相比之下, Wang 等^[5]与 Li 等^[6]提出的注意力加权及区域注意力网络 (RAN), 通过自适应聚

焦高判别力的非遮挡区域, 展现出更优的鲁棒性。然而, 传统区域注意力方法多依赖图像级裁剪与独立特征提取, 导致计算开销巨大。

二、自适应局部注意力遮挡面部情绪识别

在计算机视觉领域, 面部遮挡会导致关键视觉特征的缺失与畸变, 如何有效缓解其对情绪识别精度的负面影响, 始终是一项极具挑战性的课题。本文提出自适应局部注意力算法 (Adaptive Part Ware, APW), 算法框架如图1所示。

不同于传统方法对原始图像进行裁剪, APW 策略性地在深层特征图上执行区域划分, 避免了重复卷积带来的算力冗余。该算法协同全局与局部双重注意力机制: 全局通路维持整体语义连贯

性,局部通路则对逻辑子区域进行深度编码与独立约束。此外,本文设计了相对置信度评估模块,通过动态量化区域有效性,赋予非遮挡区域高置信度权重并抑制遮挡噪声,引导网络自适应聚焦于高价值的视觉特征区域。

(一) 全局注意力模块

为了充分挖掘深层全局语义,本文引入模态缓解单元(MAM)^[7]对特征图进行二次精炼。首先利用实例归一化IN对输入特征进行标准化处理,随后通过通道注意力图进行调制。该过程的数学表达如公式所示:

$$F_g = w_c \odot M + (1 - w_c) \odot \text{IN}(M) \quad (1.1)$$

其中, $\odot$ 表示逐元素相乘, F_g 为精炼后的输出特征。

$w_c \in \mathbb{R}^{C \times d}$ 代表通道注意力掩膜,其计算过程旨在建模通道间的依赖关系,公式如下:

$$w_c = \sigma(W_2 \delta(W_1 \text{GAP}(F))) \quad (1.2)$$

$\text{GAP}(\cdot)$ 表示全局平均池化操作, W_1 和 W_2 分别对应两个可学习全连接层的权重参数, δ 和 σ 分别代表 ReLU 修正线性单元与 Sigmoid 激活函数。

为了消除单张图像内部因遮挡或光照不均造成的对比度差异,本模块采用实例归一化(IN)对特征进行标准化处理。其定义如公式所示:

$$\text{IN}(x) = \frac{x - \mu_m}{\sqrt{\sigma_m^2 + \epsilon}} \quad (1.3)$$

其中, ϵ 为防止分母为零的数值稳定常数, μ_m 和 σ_m^2 分别代表当前单样本特征图在空间维度上的均值与方差。

经过 MAM 模块的特征重校准后,为了获得最终用于分类的全局语义向量,特征流将依次通过卷积层进行特征融合、批归一化层(Batch Normalization, BN)进行分布调整,最后经由全局平均池化层压缩为一维向量。该后处理阶段可表示为:

$$f_g = \text{GAP}(\text{BN}(\text{Conv}(F_g))) \quad (1.4)$$

(二) 局部注意力模块

为捕捉局部纹理细节,模型在深层特征图上实施规则化的滑动窗口裁剪。针对裁剪特征语义关联不足的问题,本文构建了基于自注意力机制^[8]的局部特征编码器。局部注意力矩阵的计算公式定义如下:

$$A_{pi} = \text{softmax}(q(M_{pi})k(M_{pi})^T) \quad (1.5)$$

其中 $q(\cdot)$, $k(\cdot)$ 分别表示两个不同参数的全连接层, $(\cdot)^T$ 表示矩阵转置操作; $\text{softmax}(\cdot)$ 函数用于将相关性分数转化为概率分布,从而突出关键特征点。随后,利用该注意力图对原始局部特征进行加权重构,并通过第三个全连接层进行特征映射,最终结合残差结构输出增强后的局部特征 f_{pi} :

$$f_{pi} = g(F_{pi}) = g(\text{BN}(A_{pi}v(M_{pi})) + M_{pi}) \quad (1.6)$$

此处, F_{pi} 即为第 i 个区域经过编码器深度精炼后的输出,它融合了区域内的全局上下文信息,对局部遮挡具有更强的鲁棒性。

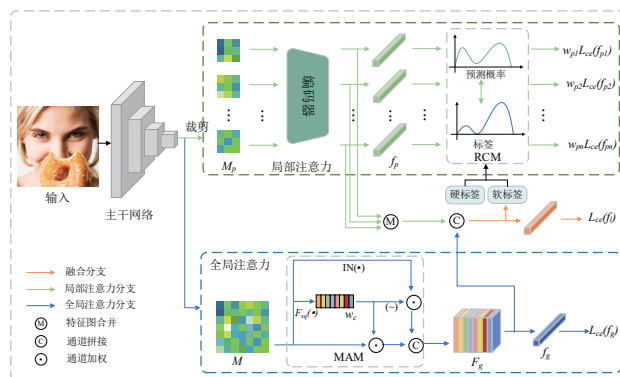


图1 APW 算法框架图

表1 在遮挡面部情绪识别数据集的对比实验(%)

方法	Occlusion- RAF-DB	Occlusion- FERPlus	Occlusion- AffectNet	FED-RO	OF-Emo(Our)	MFLOPs
Baseline ^[9]	84.06	84.96	59.00	67.50	48.35	1823.53
RAN ^[5]	82.72	83.63	58.50	67.98	46.15	14588.20
RUL ^[10]	86.38	84.30	61.39	66.83	39.56	1826.80/1825.15
APW(Our)	86.51	85.62	61.93	69.75	50.55	1930.27

(三) 实验及分析

为了验证 APW 算法的实际性能,本文首先选取了多个主流的遮挡场景面部表情数据集,将其与当前先进的识别方法进行了对比实验。

由于目前专门针对面部遮挡识别的算法相对较少,为了丰富对比维度,本文将处理标签模糊问题的 RUL^[10]方法也纳入了比较范围。为了确保对比的公平性,所有的实验模型均统一使用了在

MSCeleb-1M 数据集上预训练的 ResNet-18 作为主干网络。

从表1的实验结果可以看出,APW 算法在五个遮挡情感数据集上均取得了目前的最佳成绩。具体而言,相比于基准(Baseline)模型,本方法在 Occlusion-RAF-DB、Occlusion-AffectNet、FED-RO 以及 OF-Emo 数据集上的准确率分别提升了 2.45%、2.93%、2.25% 和 2.20%。我们注意到,APW 在 Occlusion-FERPlus 数据集上的提升幅度相对较小(0.66%)。

经过分析，这主要是因为该数据集的原始图像分辨率过低（仅为 48×48 ），导致特征分块后的区域包含的有效信息过少，从而限制了区域注意力模块的性能发挥。尽管如此，整体结果依然证明了 APW 在解决遮挡问题上的有效性。

三、结论

针对真实场景面部遮挡难题，本文提出高效的自适应局部注

意力算法（APW）。该算法构建特征层级感知框架，替代高负载的图像级裁剪，在显著降低运算复杂度的同时实现语义精细挖掘。通过引入多尺度局部注意力模块（MLA）捕捉多层次上下文，并设计基于混合监督的相对置信度模块（RCM）动态量化区域可靠性，模型实现了对未遮挡区域的自适应聚焦与噪声抑制。实验证实，APW 在识别精度与计算效率上均超越现有先进方法，有效解决了遮挡环境下的特征畸变问题。

参考文献

[1](America) 施塔, (America) 卡拉特. 情绪心理学 [M]. 中国轻工业出版社. 2015.

[2]Ruan D., Yan Y., Lai S., et al. Feature decomposition and reconstruction learning for effective facial expression recognition[A]. Proceedings of 45th the IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. Greek: IEEE, 2021: 7660–7669.

[3]Xue F., Wang Q., Guo G.. Transfer: Learning relation-aware facial expression representations with transformers[A]. Proceedings of 45th the IEEE/CVF International Conference on Computer Vision[C]. IEEE, 2021: 3601–3610.

[4]Wang C., Wang S., Liang G.. Identity-and pose-robust facial expression recognition through adversarial feature learning[A]. Proceedings of the 27th ACM international conference on multimedia[C]. Nice, France: ACM, 2019: 238–246.

[5]Wang K., Peng X., Yang J., et al. Region attention networks for pose and occlusion robust facial expression recognition[J]. IEEE Transactions on Image Processing, 2020, V29: 4057–4069.

[6]Li Y., Zeng J., Shan S., et al. Occlusion aware facial expression recognition using CNN with attention mechanism[J]. IEEE Transactions on Image Processing, 2018, V28(5): 2439–2450.

[7]Wu Q., Dai P., Chen J., et al. Discover cross-modality nuances for visible–infrared person re-identification[A]. Proceedings of 45th the IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. IEEE, 2021: 4330–4339.

[8]Dosovitskiy A., Beyer L., Kolesnikov A., et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. arXiv:2010.11929, 2020.

[9]He K., Zhang X., Ren S., et al. Deep residual learning for image recognition[A]. Proceedings of 40th the IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. Las Vegas: IEEE, 2016: 770–778.

[10]Zhang Y., Wang C., Deng W.. Relative uncertainty learning for facial expression recognition[J]. Advances in Neural Information Processing Systems, 2021, V34: 17616–17627.