

中图法分类号:(此号在中国图书馆分类法中查) 文献标识码:A 文章编号:1006-8961(年)-

论文引用格式:

# 一种用于面部表情识别的双向注意混合特征网络

许笑颜<sup>1</sup>

1.中国科学技术大学 人工智能与数据科学学院,安徽省合肥市 230026;

**摘要:** 近年来,面部表情识别(Facial Expression Recognition, FER)在计算机视觉研究领域引起了广泛关注。本文提出了一种创新网络——双向注意混合特征网络(Dual-Direction Attention Mixed Feature Network, DDAMFN),该网络兼具鲁棒性和轻量化特性。其网络架构主要由两部分组成:作为骨干的混合特征网络(Mixed Feature Network, MFN)和作为头部的双向注意网络(Dual-Direction Attention Network, DDAN)。为增强MFN的特征提取能力,网络利用混合尺寸的卷积核提取鲁棒特征。此外,提出了一种全新的双向注意(Dual-Direction Attention, DDA)头部模块,用于生成双向注意力图,帮助模型有效捕获长距离依赖关系。为了进一步提高准确性,本文在DDAN中引入了一种新颖的注意力损失机制,不同头部聚焦于输入的不同区域。通过在RAF-DB上的实验评估,结果表明DDAMFN相比其他现有模型具有显著优势,确立了其在FER领域的最新性能基准。

**关键词:** MobileFaceNets; 坐标注意力; 面部表情识别; 混合卷积(MixConv)

## A dual-direction attention mixed feature network for facial expression recognition

Xiaoyan Xu<sup>1</sup>

1.School of Artificial Intelligence and Data Science, University of Science and Technology of China;

**Abstract: Objective** In recent years, facial expression recognition (FER) has garnered significant attention within the realm of computer vision research. This paper presents an innovative network called the DualDirection Attention Mixed Feature Network (DDAMFN) specifically designed for FER, boasting both robustness and lightweight characteristics. The network architecture comprises two primary components: the Mixed Feature Network (MFN) serving as the backbone, and the Dual-Direction Attention Network (DDAN) functioning as the head. To enhance the network's capability in the MFN, resilient features are extracted by utilizing mixed-size kernels. Additionally, a new Dual-Direction Attention (DDA) head that generates attention maps in two orientations is proposed, enabling the model to capture long-range dependencies effectively. To further improve the accuracy, a novel attention loss mechanism for the DDAN is introduced with different heads focusing on distinct areas of the input. Experimental evaluations on several widely used public datasets, including AffectNet, RAF-DB, and FERPlus, demonstrate the superiority of the DDAMFN compared to other existing models, which establishes that the DDAMFN as the state-of-the-art model in the field of FER.

**Key words:** MobileFaceNets; coordinate attention; facial expression recognition; MixConv

## 0 引言

面部表情在人类交流中起着重要作用，是理解情绪和态度的关键信号。因此，计算机需要获得辨识和解释面部表情的能力。[1-3]阐明了视觉感知、环境映射算法与基于生物识别认证计算机视觉的面部表情识别（FER）之间的关系，因此使用深度学习方法解决 FER 问题具有重大的科研意义。FER 网络的主流架构通常由主干和头部组成。然而，最近的方法主要集中在头部或颈部区域，并仅使用 VGG [4] 或 ResNet [5] 作为其主干。这些主干最初是为更大的数据集而设计的，可能会从图像中提取冗余信息，在相对较小的数据集中导致过度拟合。本文提出了一种创新的主干网络，称为混合特征网络（MFN）。MFN 建立在 MobileFaceNets [6] 的基础上，这是一个专门为面部验证任务量身定制的轻量级网络。通过引入混合深度卷积核 [7] 来增强 MFN，这些卷积核利用了不同尺寸卷积核的优势。此外，坐标注意 [8] 被引入到 MFN 架构中，以促进对长距离依赖的捕捉。因此，提取了用于 FER 的有意义的特征。

此外，FER 面临着两个重要挑战：类间差异小和类内差异大。要解决这些挑战，建立各种面部区域之间的连接，如嘴巴、眼睛、鼻子等，至关重要。在这方面，注意机制提供了潜在的解决方案。具体来说，提出了双向注意（DDA）头，该头应用于所提出的方法，旨在根据提取的特征信息构建注意图。根据以前的工作 [8]，设计了从垂直和水平方向生成注意图的注意头。随后，将从双向注意网络（DDAN）获得的注意图与输入特征图相乘，得到一个新的特征图。该特征图经过线性全局深度卷积（GDCConv）层 [6]，然后进行重塑操作。采用全连接层生成最终结果。集成所提出的 DDA 头和后续处理步骤可以提升模型性能

最后，本文将 MFN 和 DDAN 集成在一起，提出了一种名为双向注意混合特征网络（DDAMFN）的新模型。为了直观地展示 DDAMFN 的有效性，进行了涉及 ResNet\_50、MFN 和 DDAMFN 模型的比较分析。所有模型都在 AffectNet-7 数据集上进行了训练，并在相同的测试数据上进行了测试。Grad-CAM [9] 应用于捕获各自主干架构提取的特征的见解。该技术通过基于梯度的定位，促进了通过热图突出显示图像中的重要区域进行预测。该分析结果

见图 1。很明显，MFN 关注的区域比 ResNet\_50 更特定。对于 DDAMFN，DDAN 允许 MFN 定位更合适的区域。

此外，当在 RAF-DB 数据集上进行了大量实验时，DDAMFN 模型表现出了显著的性能。研究的贡献可以总结如下：

（1）为了增强 FER 的提取特征的质量，本研究提出了一种名为 MFN 的新型主干网络。MFN 利用不同内核尺寸的利用，从而有助于获得强大的特征。此外，在 MFN 架构中包含坐标注意层，可以捕捉长距离依赖性，进一步增强了其在 FER 任务中的有效性。

（2）为了有效检测不同面部表情之间的细微变化，引入了 DDAN。通过从两个不同方向生成关注，DDAN 可以全面捕捉相关的面部区域，并提高 FER 的区分能力。

（3）应用了一种新颖的注意力损失机制，以确保 DDAN 的注意力聚焦在不同区域，从而显著提高了模型的整体性能和区分能力。

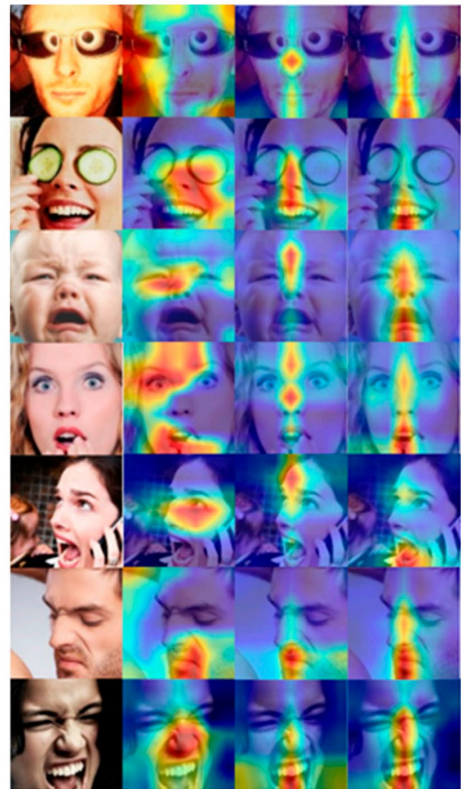


图 1. 七张图像的热力图：中性（第 1 行）、高兴（第 2 行）、悲伤（第 3 行）、惊讶（第 4 行）、恐惧（第 5 行）、厌恶（第 6 行）和愤怒（第 7 行）。第 1 列：原始图像。第 2 列：ResNet\_50。第 3 列：MFN。第 4 列：

DDAMFN (MFN + DDAN)。显然, MFN 比 ResNet\_50 更加集中于特定区域。DDAN 使得 MFN 能够定位更多的关键信息区域。

## 1 相关工作

### 1.1 面部表情学习

面部表情识别 (Facial Expression Recognition, FER) 作为一个重要的研究领域, 已受到学术界数十年的关注。传统的 FER 方法依赖于手工特征或浅层学习技术, 如非负矩阵分解 (Non-Negative Matrix Factorization, NMF) [10]、局部二值模式 (Local Binary Patterns, LBP) [11] 以及稀疏学习 [12]。然而, 这些方法在处理实际场景中的模糊和遮挡等复杂问题时往往表现不足。

近年来, 深度学习技术彻底变革了计算机视觉领域, 为 FER 带来了显著进展。卷积神经网络 (Convolutional Neural Networks, CNNs)、递归神经网络 (Recurrent Neural Networks, RNNs) 以及生成对抗网络 (Generative Adversarial Networks, GANs) 被广泛用于应对 FER 的复杂挑战。值得注意的是, 诸如 VGG [4] 和 ResNet [5] 等最新的分类模型, 作为 FER 系统的骨干网络, 展现了卓越的性能。

在此基础上, 本文提出了一种新颖的骨干网络——混合特征网络 (Mixed Feature Network, MFN), 专门设计用于从多种卷积核尺寸中提取信息。MFN 在具有轻量化特性的同时, 在 FER 任务中表现出卓越的性能, 这标志着该领域的显著进步。MFN 骨干网络在提升 FER 精度和效果方面具有重要的潜力。

### 1.2 注意力机制

近年来, 注意力机制在各种视觉任务中的应用引起了广泛关注。研究人员尝试将注意力机制与深度卷积神经网络 (Convolutional Neural Networks, CNNs) 相结合, 以从图像中提取更多有用的特征。其中, 代表性的工作包括 [13] 提出的“Squeeze-and-Excitation”模块, 该模块专注于从不同通道中提取有用信息; SGE 方法 [14] 通过将空间特征划分为多个组来建立空间连接并增强特征表示; 此外, CBAM [15] 将通道注意力和空间注意力机制结合在一个统一框架中, 从而捕获更丰富的信息。

近年来, Transformer [16] 作为一种强大的新范式在许多任务中表现卓越, 其多头注意力机制使其在多个领域超越了传统的递归神经网络 (Recurrent

Neural Networks, RNNs) 和卷积神经网络 (CNNs)。这一成功促使研究人员探索并改进基于 Transformer 的方法, 用于解决各种视觉任务。

在面部表情识别 (FER) 领域, 也有研究引入了注意力机制。例如, [17] 提出利用多个非重叠区域的注意力机制, 从不同的面部区域捕获信息; [18] 则探讨了 Transformer 在 FER 任务中的应用。然而, 这些现有方法在从面部图像中学习全面信息方面仍然存在局限性。基于 [6] 的研究, 本文在 DDAMFN 网络中应用了一种新颖的双向注意力头部模块 (Dual-Direction Attention Head)。该模块通过建模长距离依赖关系, 能够捕获全面且具有上下文关联的面部信息。通过将双向注意力头部与 MFN 骨干网络相结合, 本文有效克服了现有方法的局限性, 从而更全面地学习面部图像中的信息, 实现更优的特征提取效果。

## 2 主要方法

DDAMFN 的整体架构如图 2 所示, 主要由两部分组成: 混合特征网络 (MFN) 和双向注意网络 (DDAN)。首先, 将面部图像输入 MFN, 生成基础特征图作为输出。接着, 通过 DDAN 在垂直和水平方向生成注意力图。最终, 注意力图被重塑为特定维度, 并通过全连接层预测图像的表情类别。

### 2.1 混合特征网络 (MFN)

考虑到在小规模 FER 数据集上使用复杂网络架构可能导致过拟合问题, 本文采用了轻量化网络 MobileFaceNet [6] 作为基础网络。如图 3 所示, MFN 由两种主要构建模块组成: 残差瓶颈块和非残差块。

残差瓶颈块旨在捕获复杂特征, 并促进网络中的信息传播。通过引入残差连接, 该模块能够缓解退化问题, 并在训练过程中改善梯度流动。

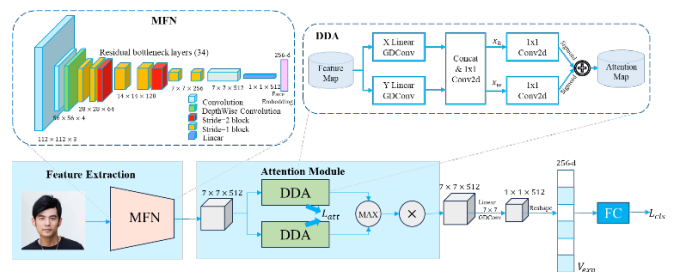


图 2. DDAMFN 的整体结构。“GDCConv”指的是全局深度卷积 (Global Depthwise Convolution)。该方法包含两个主要的创新步骤。首先, 使用 MFN 从人脸图像中提取基础特征。接下来, 构建 DDAN 模块, 包含多个 DDA 头, 可以

从水平和垂直方向生成注意力图。经过 DDAN 模块后，特征经过线性 GDConv 层，并重新调整特征图的形状。最后，采用全连接层生成分类结果。

非残差块则通过非残差连接增强模型的表示能力，使 MFN 能够捕获多样化且具有区分性的面部特征，从而更有效地完成 FER 任务。这种架构设计在模型复杂性与泛化能力之间实现了平衡，使其非常适合 FER 任务。图 2 的左上部分展示了 MFN 的主要结构，而表 1 全面列出了各层的具体规格。

基于[7]的研究，本文在网络的瓶颈模块中引入了 MixConv 操作。MixConv 包含多种尺寸的卷积核，其排列如图 4 所示。这种配置使 MFN 能够从输入图像中有效提取多样化且信息丰富的特征，性能优于原始的 MobileFaceNet 架构。

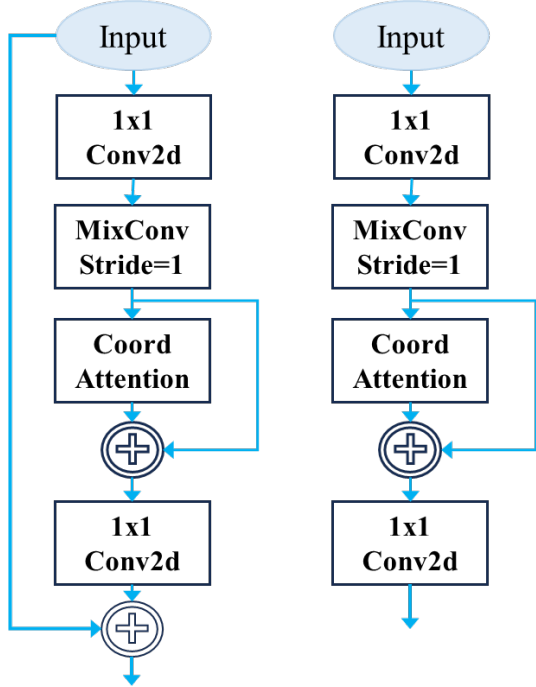


图 3. MFN 的瓶颈部分，左侧为步幅为 1 的模块，右侧为步幅为 2 的模块，见表 1。

此外，本文采用 PreLU 作为激活函数，其在人脸特征提取方面优于 ReLU [6]。同时，本文对网络深度[19]进行了精细调整，并在 MFN 骨干网络的每个瓶颈模块中引入了坐标注意力机制[8]。与[19]中使用的通道与空间注意力模块 (CBAM) 相比，该注意力机制能够更好地建模长距离依赖关系，并生成更准确的位置信息。

MFN 通过不同尺寸的卷积核捕获面部图像中的多样化空间信息，这种设计灵感来源于[7]。另一方面，DDAN 模块整合了一系列高效的双向注意力

头部，每个注意力头部生成一个注意力图，并通过对比分析选出最具信息量的注意力图。

DDAMFN 框架结合了 MFN 的特征提取能力与 DDAN 注意力机制的判别能力，通过整合这两个组件，DDAMFN 在 FER 任务中取得了性能提升。

## 2.2 双向注意网络 (DDAN)

DDAN 由多个独立的双向注意 (DDA) 头组成，每个头都用于捕获网络中的长距离依赖关系。DDAN 模块的基础结构采用了坐标注意力机制[8]。

DDAN 的详细结构如图 2 右上部分所示。与[8]中的方法类似，注意力头最初从水平方向和垂直方向生成方向感知特征图。然而，本文用线性深度卷积 (GDConv) 替代了传统的平均池化操作，这种修改可以在不同的空间位置学习到截然不同的重要性[6]。这一设计增强了对不同面部区域的关注，从而提高了注意力机制的判别能力。

随后，DDA 头通过与[8]相同的结构，分别为两个方向生成两个注意力图 ( $x_h, x_w$ )。最终注意力图通过对两个方向的注意力图进行逐元素相乘得到，其过程如下：

$$a_i = x_h * x_w$$

生成多个双向注意力图后，比较这些注意力图并选择显著性最高的一个是关键步骤。最终生成最优注意力图  $a_m$ ，如图 2 所示。接着，DDAN 的输出通过输入特征图与  $a_m$  进行逐元素相乘获得。

为了确保每个双向注意力头专注于不同的面部区域，本文为 DDAN 模块引入了一种新颖的损失函数，称为注意力损失 (Attention Loss)。

### 注意力损失

注意力损失通过计算不同双向注意力头生成的注意力图之间的均方误差 (MSE) 定义。具体来说，注意力损失是这些 MSE 损失之和的倒数，其数学表达式为：

$$L_{att} = (\sum_{i=0}^n \sum_{k=0, k \neq i}^n MSE(a_i, a_k))^{-1}$$

其中， $n$  表示注意力头的数量， $a_i$  和  $a_k$  分别是两个不同头生成的注意力图。该损失函数的设计旨在鼓励每个注意力头专注于不同区域，提高网络捕获全局和局部信息的能力，从而增强 FER 任务的表现。

## 2.3 损失函数的构造

如图 2 所示，从 DDAN 获得的特征图大小为  $7 \times 7 \times 512$ ，该特征图经过线性深度卷积 (GDConv)



层和线性层的处理后，转换为 512 维向量。通过全连接层获得类别置信度。

在损失函数方面，训练过程中采用标准的交叉熵损失。该损失函数有效地衡量了预测类别概率与真实标签之间的差异，有助于优化模型参数。

整体损失函数可以表示为：

$$L = L_{cls} + \lambda_a L_{att},$$

其中， $L_{cls}$  表示标准的交叉熵损失， $L_{att}$  是注意力损失， $\lambda_a$  是超参数，默认值为 0.1。该损失函数综合了分类损失和注意力损失，确保模型不仅能够准确预测类别，还能够有效地关注图像中的重要区域，从而提高 FER 任务的表现。

### 3 实验结果

首先开展了消融实验，分析了 DDAMFN 中各个组件的独立贡献，评估了每个组件在提升 DDAMFN 整体性能方面的作用。随后，DDAMFN 与其他最先进的网络进行了对比分析，验证了其优越性。

#### 3.1 RAF-DB 数据集

RAF-DB 是一个真实世界的面部表情数据库，来源于互联网。该数据集包含 29,672 张面部图像，标注了七种基本情感标签和十一种复合情感标签。在 RAF-DB 中，训练集包括 12,271 个样本（惊讶 1290、恐惧 281、厌恶 717、开心 4772、悲伤 1982、愤怒 705 和中性 2524），测试集包含 3,068 个样本用于 FER 任务。

#### 3.2 数据预处理

在预处理阶段，使用 RetinaFace [24] 来检测 RAF-DB 数据集中的面部和关键点（包括两只眼睛、鼻子和两个嘴角）。所有的面部图像都被对齐并调整为标准化的 112×112 像素大小。为了减少过拟合，对于 RAF-DB 使用了水平翻转、随机旋转和擦除来进行数据增强，提高了 DDAMFN 在训练过程中的鲁棒性和泛化能力。

#### 3.3 预训练和训练配置

为了确保与其他骨干架构的公平对比，MFN 骨干网络在 Ms-Celeb-1M 数据集 [25] 上进行了预训练。这一预训练步骤确保了 DDAMFN 的统一基准测试。

所有实验均使用 PyTorch 2.1.0 进行，模型训练在配备 TESLA P40 24G GPU 的服务器上进行。所有

任务均训练了 40 个 epoch，而 DDAN 模块中的注意力头数量设置为默认值 2。在训练过程中使用了 ADAM 作为优化器，学习率设置为 0.01。这些参数设置旨在为不同数据集提供高效且有效的模型优化。

### 3.4 消融实验

#### 3.4.1 MFN 的有效性

为了评估 MFN 骨干网络的有效性，开展了系列对比实验。表 2 展示了不同骨干网络在 RAF-DB 数据集上的性能。

表 2. 不同网络在 RAF-DB 数据集上的评估结果

方法	准确率 (%)	参数量 (Params)	FLOPs
MobileFaceNet	87.52	1.148 M	230.34 M
ResNet-18	87.47	16.78 M	2.6 G
ResNet-50	89.63	41.56 M	6.31 G
MFN (our backbone)	90.32	3.973 M	550.74 M
DDAMFN (our model)	91.35	4.106 M	551.22 M

结果表明，MFN 在 RAF-DB 数据集上的准确率为 90.32%，超越了其他三个骨干网络的表现。此外，MFN 骨干网络只有 3.973 百万个参数，在所有比较的骨干网络中排名第二，显著小于 ResNet-18 和 ResNet-50。同时，MFN 的计算复杂度（以浮点运算次数 FLOPs 为度量）仅为 550.74 百万，虽然高于 MobileFaceNet，但远低于其他两个网络。

#### 3.4.2 DDAN 的有效性

为了验证 DDAN 的有效性，进行了消融实验，评估了 MFN 和 DDAN 在 RAF-DB 和 AffectNet-7 数据集上的影响。结果见表 3。

表 3. MFN 与 DDAN 在 RAF-DB 和 AffectNet-7 数据集上的评估结果

MFN	DDAN	RAF-DB (%)	AffectNet-7 (%)
√	√	91.35	67.03
√	-	90.32	66.19

表 3 的结果表明，DDAN 的引入使得模型在 RAF-DB 和 AffectNet-7 数据集上的性能分别提升了 1.06% 和 1.04%。这些结果表明，DDAN 在增强 MFN 的性能方面起到了至关重要的作用，通过从提取的特征中生成更全面的注意力图，使模型能够更好地关注相关区域并捕获关键信息，从而提升了识别准确率。

#### 3.4.3 注意力头数的影响

为了研究 DDA 头数量对模型性能的影响,进行了关于 DDA 头数量变化的实验,实验结果见表 4。

表 4. 不同数量的 DDA 头在 RAF-DB 数据集上的评估结果

DDA 头数量	RAF-DB (%)
0	90.32
1	90.67
2	91.35
3	91.11
4	91.21

从表 4 可以看到,使用两个 DDA 头的模型在 RAF-DB 数据集上取得了显著优于其他 DDA 头数量的表现。使用两个 DDA 头的 DDAN 架构在识别准确率方面超越了使用不同数量 DDA 头的模型。

#### 3.4.4 注意力损失函数的有效性

为了评估 DDAN 中使用的损失函数的有效性,进行了消融实验,结果见表 5。

表 5. DDAN 中损失函数的消融实验结果

方法	RAF-DB (%)
无	90.86
注意力损失	91.35

从表 5 的结果可以看出,创新的注意力损失函数对 DDAN 性能具有显著影响。引入该损失函数后,模型的表现得到了明显提升,强调了引导注意力头聚焦于不同区域的重要性,并帮助模型在面部表情识别中实现更好的区分。这些结果凸显了注意力损失函数在优化 DDAN 中注意力机制方面的有效性。通过鼓励注意力头关注不同的面部区域,损失函数提升了模型的辨识能力和整体性能。

这些消融研究结果表明,DDAMFN 的设计和各个组成部分(包括 MFN 骨干、DDAN 模块和注意力损失函数)都为面部表情识别任务的高性能做出了重要贡献。

#### 3.5 与现有方法的对比

为了证明提出的模型在 RAF-DB 数据集上的有效性,与 7 个 baseline 进行了对比,对比结果如下图所示,发现 DDAMFN 在 RAF-DB 数据集上具有最佳的表现。此外,我们对模型进行了交叉验证,结果见表格 9。

表 8. 不同模型在 RAF-DB 数据集上的表现

Methods	Accuracy (%)
RAN[30]	86.90

SCN[31]	87.03
DACL[28]	87.78
MViT[27]	88.62
PSR[32]	88.98
DAN[17]	89.70
TransFER[18]	90.91
DDAMFN	91.35

可以看出,在不同的交叉验证过程中,达到了平均 90.635 的验证水平。

Table 9. K-Fold 交叉验证

数据集	RAF-DB
Fold 1	90.69
Fold 2	90.23
Fold 3	89.89
Fold 4	90.82
Fold 5	89.83
Fold 6	91.02
Fold 7	91.82
Fold 8	90.55
Fold 9	90.55
Fold 10	90.95
平均准确率 (%)	90.635

## 4 结 论

本文提出了一种新颖且有效的面部表情识别(FER)方法——DDAMFN。DDAMFN 包含两个关键组件:MFN(多尺度特征网络)和 DDAN(双重依赖注意力网络)。MFN 通过利用不同尺寸的卷积核,生成全面且具有辨别性的特征,以用于表情分类;同时,DDAN 通过新引入的 DDA 头部,捕获长程依赖关系。

通过在 RAF-DB 数据集上进行广泛的实验,DDAMFN 展现了领先的性能,在 RAF-DB 上取得了 91.35%,领先第二名 0.44%,验证了 DDAMFN 在 FER 领域的有效性和优越性。

通过提出一种新颖的更轻量级的骨干网络,并在模型中应用 Mixconv 和坐标注意力机制,DDAMFN 为面部表情识别的网络结构发展做出了贡献,且为计算机视觉任务的未来发展提供了推动力。此外也期待 DDAMFN 模型能够在人工智能的各个领域得到应用,促进各类应用的进步,并推动

AI 驱动技术的进一步发展。

## 参考文献(References)

1. Andronie, M.; Lăzăroiu, G.; Karabolevski, O.L.; S, tefănescu, R.; Hurloiu, I.; Dijmărescu, A.; Dijmărescu, I. Remote Big Data Management Tools, Sensing and Computing Technologies, and Visual Perception and Environment Mapping Algorithms in the Internet of Robotic Things. *Electronics* 2023, 12, 22. [CrossRef]
2. Pelău, C.; Dabija, D.C.; Ene, I. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics on the acceptance of artificial intelligence in the service industry. *Comput. Hum. Behav.* 2021, 122, 106855. [CrossRef]
3. Dijmărescu, I.; Iatagan, M.; Hurloiu, I.; Geamănu, M.; Ruscescu, C.; Dijmărescu, A. Neuromanagement decision making in facial recognition biometric authentication as a mobile payment technology in retail, restaurant, and hotel business models. *Oeconomia Copernic.* 2022, 13, 225–250. [CrossRef]
4. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
5. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
6. Chen, S.; Liu, Y.; Gao, X.; Han, Z. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Proceedings of the Chinese Conference on Biometric Recognition*, Urumqi, China, 11–12 August 2018; pp. 428–438.
7. Tan, M.; Le, Q.V. Mixconv: Mixed depthwise convolutional kernels. In *Proceedings of the 30th British Machine Vision Conference 2019*, Cardiff, UK, 9–12 September 2019.
8. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 19–25 June 2021; pp. 13713–13722.
9. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 618–626.
10. Zhi, R.; Flierl, M.; Ruan, Q.; Kleijn, W.B. Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition. In *Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Anchorage, AK, USA, 9–12 October 2011; pp. 38–52.
11. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* 2009, 27, 803–816. [CrossRef]
12. Zhong, L.; Liu, Q.; Yang, P.; Liu, B.; Huang, J.; Metaxas, D.N. Learning active facial patches for expression analysis. In *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, 16–21 June 2012; pp. 2562–2569.
13. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
14. Li, X.; Hu, X.; Yang, J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks. *arXiv* 2019, arXiv:1905.09646.
15. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 8–14 September 2018; pp. 3–19.
16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30, 5998–6008.

## 作者简介

许笑颜，女，2002 年出生于湖北省十堰市，研究方向：PDE Foundation Model，邮箱：[xyy1101@mail.ustc.edu.cn](mailto:xyy1101@mail.ustc.edu.cn)