

基于标签分布学习和数据增强的面部情绪识别网络设计

摘要

面部情绪识别作为计算机视觉和情感计算的重要研究方向，应用于人机交互、医疗健康、市场分析等多个领域。然而，目前面部情绪识别技术存在模型计算复杂、特征追溯能力不足以及情绪复杂性完善模型的挑战。本文设计了一个面部情绪识别网络，以 EfficientNet 为基础，并将标签分布学习作为情绪复杂性完善手段，通过图像风格化技术解决 RAF-DB 数据集中实际情绪的不均衡问题。实验结果显示，该网络在实际场景情绪识别中呈现出了较高的出临能力和识别准确性。

关键词：面部情绪识别；标签分类；风格化；轻量级

1 引言

近年来，面部情绪识别（Facial Expression Recognition, FER）作为计算机视觉和情感计算的一个重要领域，旨在利用计算机对面部表情的分析和解读进而判断个体的情绪状态。面部表情作为人类情感表达的重要方式之一，能够反映一个人的情绪变化，例如喜悦、悲伤、愤怒、厌恶、生气等情绪。随着深度学习的发展，计算机视觉也得到相应的发展。面部情绪识别通过深度学习、图像处理和模式识别等算法，计算机能够自动提取图像或者视频中的面部表情特征，并对采集到的特征进行分类，从而实现对个体的情绪识别。

面部情绪识别技术在多个实际应用领域取得了显著进展。与传统的主要依赖键盘和鼠标进行文字输入的人机交互方式相比，面部情绪识别的引入使得机器能够通过感知个体的情绪变化做出更为智能的反应，从而丰富了人机交互的形式，提高了交互系统的智能化水平。此外，面部情绪识别在智能安防领域也发挥了重要作用。在超市、电影院等人流密集的公共场所，过去由于个体情绪失控可能导致的人员伤亡和公共财物损失难以得到有效控制。而通过面部情绪识别技术对人群进行实时监控，可以及时发现异常情绪反应，进而有效预防类似事件的发生，增强社会安全性。除了在安防领域的应用，面部情绪识别技术还在心理健康、市场营销等多个领域展现了其巨大的潜力，它的应用不仅提高了技术对人类情感的理解能力，还为提升服务质量和用户体验提供了强有力的支持。

但是，面部情绪识别仍然面临着诸多挑战。首先，深度学习模型参数庞大，计算复杂，使其难以在资源受限的设备中部署，限制了其实际应用范围。其次，轻量化模型如 MobileNet [8] 和 ShuffleNet [11] 等虽然减小了参数量，但在特征提取能力上较弱，难以应对表情识别对细节特征的高要求，直接应用于真实场景中的面部情绪识别，可能导致较差的准确性和鲁棒性。

此外，情绪通常是多种基本情绪的混合，但传统面部情绪识别方法多采用硬标签 (one-hot) 训练方式 [7]，忽略了真实情绪中模糊性和复合性的特点。除此之外，常用的面部情绪识别数据集 RAF-DB [10] 存在着类别不均衡问题，例如：高兴类别样本较多。生气、嫌弃等类别样本较少，这样就模型倾向于预测样本较多的类别，导致小类别的分类准确率较低，最终影响整体性能。

针对上述问题，本文旨在设计一个轻量化、鲁棒性强的面部情绪识别网络，能够在实际场景下高效、准确地识别复杂情绪。具体目标包括：

第一，设计一个轻量化的深度学习网络，减少计算复杂度并提升情绪特征的提取能力；

其次，引入标签分布学习机制，建模情绪复杂性，优化模型预测的情绪分布，使其更加贴近真实情绪的表达；

最后，采用图像风格化技术解决数据集类别不均衡问题，增加小类别样本数量，同时引入视觉多样性，减少模型识别中出现的偏置现象。

2 相关工作

本文所复现的基于标签分布的轻量级面部情绪识别网络以 EfficientNet 为基础，并将标签分布学习作为情绪复杂性完善手段。为解决经典面部情绪识别数据集 RAF-DB 的类别不均衡问题，本文将采取风格化对数据集进行优化，增强网络对小类别数据的学习能力，从而提高网络的整体性能。因此，本节将分别从轻量级面部情绪识别网络、标签分布学习、风格化技术三个方面出发，简要介绍它们的相关工作。

2.1 轻量级面部情绪识别网络

随着深度学习技术的发展，面部情绪识别 (FER) 在各种实际场景中得到了广泛应用。然而，传统的 FER 模型通常依赖于复杂的深度卷积神经网络，它们虽然在准确率方面表现优异，但由于参数量大、计算开销高，难以满足资源受限设备的实际需求。为了解决这一问题，研究者们提出了一系列轻量级 FER 网络。

2018 年，Hewitt 和 Gunes 设计了三种轻量级 FER 模型 [6]，分别基于 AlexNet、VGGNet 和 MobileNet，专门用于移动设备上的情绪识别任务。2020 年，Barros 等人提出了名为 FaceChannel 的轻量级 FER 模型 [1]，该模型包含 10 个卷积层和 4 个池化层，并在最后一层应用了抑制场机制，从而在一定程度上提升了模型的效率和性能。此外，同年 Ferro-Pérez 和 Mitre-Hernandez 针对资源受限系统设计了名为 ResMoNet 的 FER 模型 [2]，该网络由五种模块组成，在减少参数量和计算量的同时保持了较高的识别精度。然而，这些轻量级 FER 模型主要在实验室环境下测试，缺乏对真实场景的鲁棒性评估，导致其在复杂场景中的性能仍然有限。

2.2 标签分布学习

标签分布学习 (Label Distribution Learning, LDL) 是一种新的机器学习范式，尤其适用于多标签分类任务。在标签分布学习中，标签不再是单一的离散值，而是通过一个概率分布来表示每个标签的可能性，这种方法可以更好地处理标签之间的相关性和依赖关系。

2016 年, Geng 等人首次提出了标签分布学习的基本框架, 并将其应用到多标签分类任务中 [5]。随着深度学习技术的兴起, 标签分布学习与深度神经网络结合已成为一种新的研究趋势, 2017 年, Gao 等人利用深度神经网络对标签分布进行建模, 能够进一步提升模型的表示能力和学习效率 [3]。

2.3 风格化技术

数据增强是应对深度学习任务中数据不足和类别不平衡的重要手段, 而图像风格化技术近年来在数据增强领域得到了广泛应用。通过将图像的内容和特定的风格特征相结合, 风格化技术能够生成多样化的训练数据, 从而提升模型对小类别的识别能力。

2016 年, Gatys 等人首次提出了基于卷积神经网络的图像风格化方法 [4], 该方法利用预训练网络提取图像的内容特征和风格特征, 通过最小化内容损失和风格损失实现图像重构。基于此, 同年 Johnson 等人进一步提出了快速图像风格化方法 [9], 通过将风格迁移过程转换为一次性前向传播, 大幅提升了风格化效率。

3 本文方法

3.1 本文方法概述

本文基于 EfficientNet 设计了一种轻量化的面部情绪识别网络, 用于解决传统 FER 模型在复杂度和鲁棒性方面的不足问题。EfficientNet 的设计核心是复合缩放策略 (Compound Scaling), 通过同时调整网络的深度、宽度和分辨率, 实现在参数量较少的情况下获得较强的特征提取能力。EfficientNet 在标准图像分类任务中的性能优异, 使其成为轻量化 FER 模型的理想选择。

在本研究中, EfficientNet 被选为基础骨干网络, 并针对 FER 任务进行了一定的改进。传统的 FER 训练方法使用 one-hot 硬标签作为监督信号, 每张图片被强制归类到一个固定的情绪类别。这种方式忽略了情绪的复杂性, 例如一个人的表情可能同时包含“惊讶”和“高兴”等多种情绪成分。为解决这一问题, 本文引入了标签分布学习, 通过生成每张图片的情绪概率分布来替代硬标签。

标签分布的生成由一个独立的标签分布生成器 (Label Distribution Generator) 完成。在输入图片后, 生成器根据图片的情绪特征输出一个 7 维的情绪分布向量, 表示每种情绪的概率。同时, 本文使用 KL 散度 (Kullback-Leibler Divergence) 作为损失函数以监督模型训练, 将模型预测的概率分布与生成的标签分布进行比较, 优化网络参数。

具体而言, EfficientNet 的网络结构在保持轻量化的基础上, 通过上述改进提升了 FER 任务的鲁棒性和准确性。本文的模型架构如图 1 所示:

3.2 损失函数定义

在基于标签分布的轻量级面部情绪识别网络中, 使用 KL 散度 (Kullback-Leibler Divergence) 作为损失函数, 主要是为了衡量网络输出的标签分布与真实标签分布之间的差异。KL 散度是信息论中的一种度量方式, 用于衡量两个概率分布之间的距离。在情绪识别中, KL 散度可以有效地比较网络输出的情绪标签概率分布与实际标签分布之间的差异。

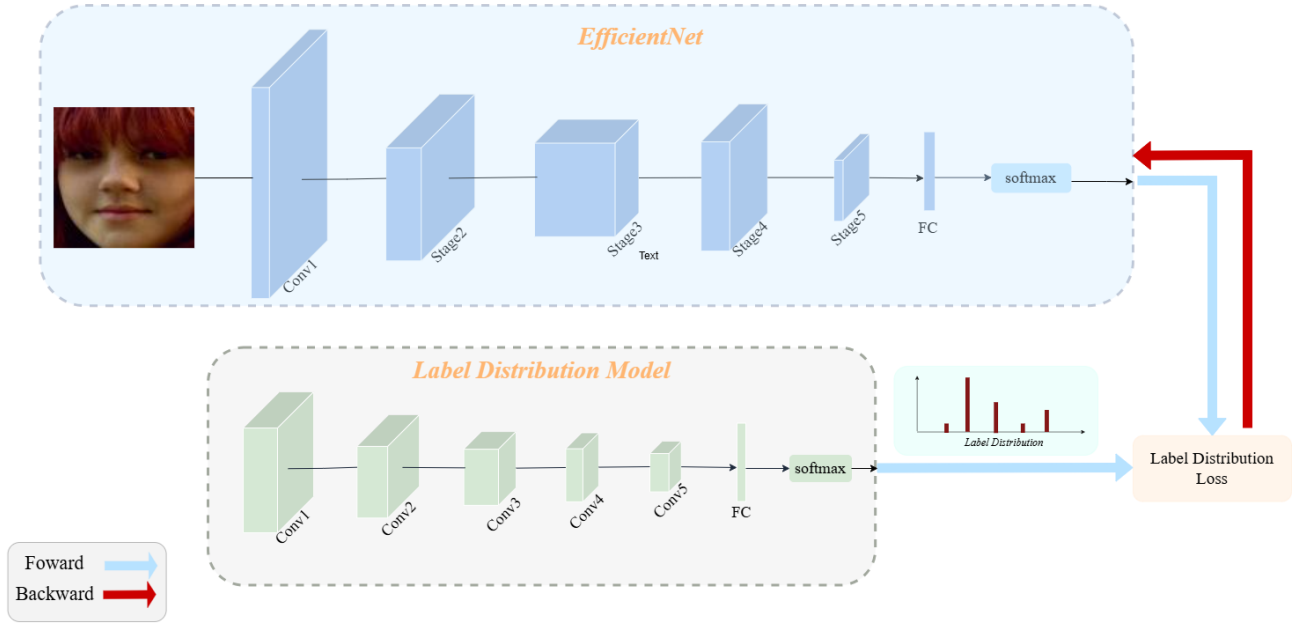


图 1. 基于标签分布的轻量级面部情绪识别网络架构

在给定两个概率分布 P 和 Q 的情况下, KL 散度可以定义为

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (1)$$

其中, $P(i)$ 表示真实标签分布中的第 i 类标签的概率, $Q(i)$ 表示模型预测的标签分布中的第 i 类标签的概率, i 表示标签类别。

在情绪识别任务中, 真实标签通常以概率分布的形式表示, 而不是单一的标签。例如, 如果情绪识别任务有多个情绪类别 (如 “快乐”、“悲伤”、“愤怒” 等), 那么每个情绪类别的概率会根据实际标签分布给定。这种标签分布可以通过人工标注或者通过某些先验知识生成 (例如, 假设某个情绪的标签更为确定时, 给出较高的概率)。假设我们有 C 个情绪类别, 并且真实标签为 $P = (p_1, p_2, \dots, p_C)$, 预测标签分布 $Q = (q_1, q_2, \dots, q_C)$ 。那么对于每个样本, 我们的目标是最小化 KL 散度损失, 即:

$$\ell_{KL} = D_{KL}(P \parallel Q) = \sum_i^C p_i \log \frac{p_i}{q_i} \quad (2)$$

4 复现细节

4.1 与已有开源代码对比

在复现基于标签分布的轻量级面部情绪识别网络过程中, 本文严格依照论文所提供的架构设计 [12], 成功复现了其中的 Label Distribution Generator 模块和标签分布损失函数 (Label Distribution Loss), 并在 RAF-DB 数据集上进行了训练与测试。由于原论文中 EfficientFace 模块的源代码未公开, 且在复现过程中我未能成功复现此段代码, 因此我在复现过程中选择使用 EfficientNet 作为替代模型来完成特征提取部分的功能。通过对模型进行适当的调优与实验优化, 我的复现模型在 RAF-DB 数据集上取得了 84.63% 的准确率, 相较原论文中报道的 88.94% 结果有所下降, 但仍表现出了较强的鲁棒性和性能可行性。

在复现过程中,本文未参考任何与 EfficientFace 相关的开源代码。标签分布生成模块的实现完全基于论文中的设计和描述,KL 散度作为损失函数也由本文独立完成。由于 EfficientFace 的缺失,本文引用了开源的 EfficientNet 模型库,调整了其结构以适应面部情绪识别任务的特定需求,同时进行了多次实验调优,包括学习率的选择、优化器的设置以及损失函数超参数的调整。这些调整使得 EfficientNet 能够有效替代 EfficientFace 的功能,同时保持了网络的轻量化特点。

与原论文相比,复现模型在准确率上略有下降,这可能是由于 EfficientFace 专为该任务设计,而 EfficientNet 虽然是一个高效的通用特征提取网络,但在处理特定情绪识别任务时可能未能完全发挥针对性特性。此外,原论文中部分训练细节未明确说明,可能对结果产生一定影响。然而,即便如此,复现模型依然与原始性能差距不大,这表明本文替代设计的合理性与实验方法的有效性。

在模型替代的基础上,为了弥补由于使用 EfficientNet 带来的性能下降,同时考虑到 RAF-DB 数据集存在类别不均衡的问题,本文采用了数据增强策略,特别是对小类别情绪样本应用了风格化技术进行数据扩充。风格化技术通过对原始图像施加不同风格的转换,生成多样化的样本,以提高模型对小类别的辨识能力。在应用数据增强后,复现模型的性能得以进一步优化,准确率从最初的性能下降恢复到 86.25%,这也表明数据增强对模型效果的提升具有显著作用。

本文的主要贡献在于,在无法获得原论文核心模块代码的情况下,成功使用 EfficientNet 作为替代设计,独立实现了标签分布生成模块,完成了对标签分布学习任务的复现,并通过实验验证了模型的性能。同时,通过针对类别不均衡问题的风格化数据增强策略,本文进一步提升了模型性能,使得复现结果达到了与原论文相差不大的水平。

4.2 实验环境搭建

4.2.1 数据集

本研究采用 RAF-DB (Real-world Affective Faces Database) 作为面部情绪识别的主要实验数据集。RAF-DB 包含 7 种基本情绪类别:高兴、愤怒、悲伤、惊讶、恐惧、厌恶、中性,覆盖多种真实场景下的面部表情数据。其中数据集总计包含 15,339 张人脸图像,分为训练集 (12,271 张) 和测试集 (3,068 张)。数据集中类别分布不均匀,如“高兴”类别占比超过 50%,而“厌恶”和“恐惧”等小类别数据量明显不足,这导致了模型在小类别上的识别能力较弱。

4.2.2 评估指标

实验中,以 EfficientNet-B0 作为基础模型,并将其与 ResNet-50 模型进行对比,分析本文方法的轻量化性能和识别能力。同时,本文以总体准确率,即模型在所有测试样本上的分类准确率作为评估指标,用于衡量模型的整体性能。

4.3 创新点

在模型替代的基础上,为了弥补由于使用 EfficientNet 带来的性能下降,同时应对 RAF-DB 数据集存在类别不均衡的问题,本文在复现过程中设计并采用了一种基于数据增强的策略,特别针对小类别情绪样本数量较少的问题,使用风格化技术进行数据扩充。RAF-DB 数

据集中，不同情绪类别的样本分布不均匀，小类别样本的数量显著少于大类别样本，这导致模型在训练时对小类别的辨识能力较弱，从而影响整体性能。针对这一问题，本文通过风格化数据增强技术，生成了更加多样化的小类别样本，以有效提升模型对这些类别的识别能力。

风格化技术是一种基于图像风格转换的方法，通过将原始图像与不同的艺术风格或图像纹理结合，生成具有风格化特征的样本。在本文中，风格化增强具体实现为利用现有的风格迁移算法对小类别情绪样本进行处理，将其转换为多种风格版本的图像。这些风格化的图像保留了原始样本的关键特征，同时通过增加样本的视觉多样性，增强了模型在训练过程中对小类别的泛化能力和鲁棒性。具体地，本文选取了一些常用的风格迁移模板（如经典艺术风格、抽象纹理风格等），针对每个小类别生成了多种风格化版本，并将这些扩充后的样本与原始样本一起用于训练。

本文中将小类别样本风格化为水彩画风格和毕加索风格。风格化后的数据样例如图2和图3所示：

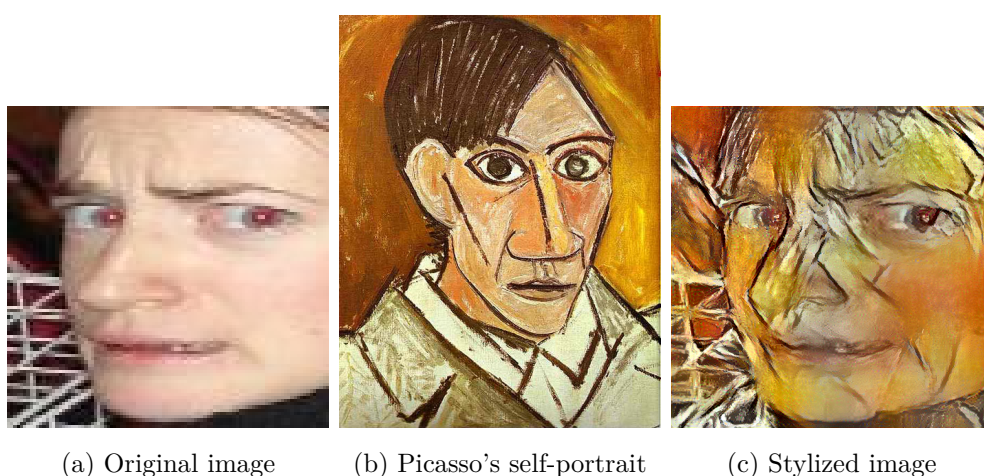


图 2. 对于恐惧类别中的某一张图片应用毕加索自画像风格后的图片

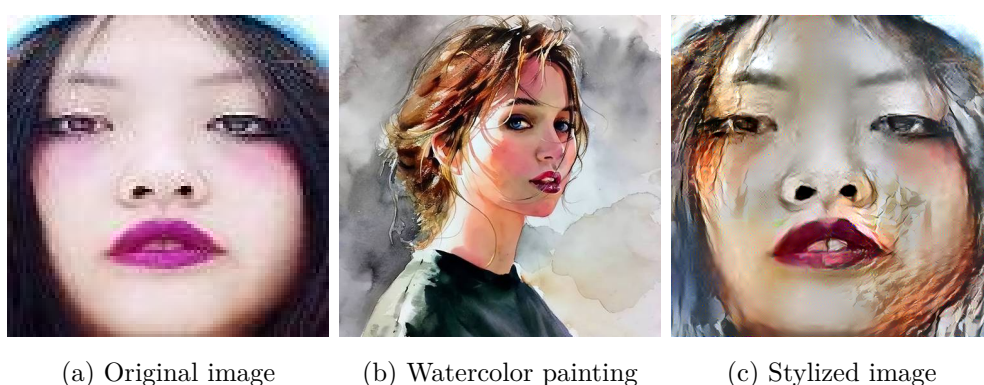


图 3. 对于中立类别中的某一张图片应用水彩画风格后的图片

通过风格化技术对小类别样本进行数据增强后，RAF-DB 数据集相对于原始数据集来说类别更加均衡，具体结果如表1所示：

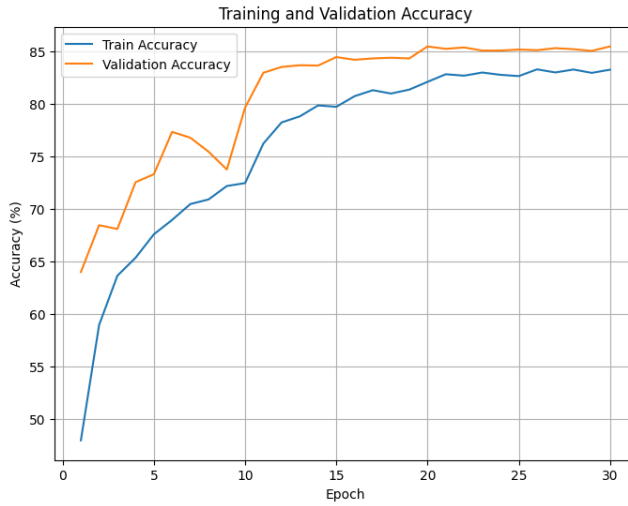


图 4. 基于标签分布和数据增强的面部情绪识别结果

表 1. 原始数据集数目和风格化后数目对比

情绪类别	原始数目	风格化后数目
Neutral	1234	1534
Disgust	281	843
Fear	705	1305
Surprise	1004	1604

数据增强策略的引入显著缓解了由于类别不均衡而导致的性能下降问题。在应用风格化增强之前，复现模型的准确率受到小类别情绪识别效果不足的影响，明显低于原论文的结果。而通过风格化数据扩充，模型得以在小类别样本上获得更充足的训练，提高了对这些类别的判别能力。最终实验结果显示，复现模型的准确率从应用数据增强前的较低水平恢复到 86.25%，与原论文的 88.94% 结果接近。这一性能提升证明了风格化数据增强技术在类别不均衡问题上的有效性，也表明该方法能够显著提升基于标签分布学习的面部情绪识别模型的整体性能。

本文的创新点主要体现在以下几个方面：首先，通过结合风格化技术进行数据增强，提出了一种针对小类别情绪样本不足的有效解决方案；其次，该方法通过增加小类别样本的多样性，提高了模型对类别不均衡数据集的适应能力；最后，在模型替代的前提下，数据增强策略不仅弥补了由于特征提取模块差异导致的性能下降，还进一步提升了模型的准确性。这种风格化数据增强方法的成功应用，不仅验证了其在 RAF-DB 数据集上的有效性，还为解决其他类别不均衡问题提供了一种通用的思路，体现了本文的技术贡献与创新价值。

5 实验结果分析

5.1 整体实验结果

按照本文所设计的基于 EfficientNet 的面部情绪识别网络和实验设置，实验结果如图4所示：

5.2 基线实验

在不使用任何优化的情况下，基于 EfficientNet-B0 的模型相比 ResNet-50 表现出了更轻量化的特点，但在准确率上仍存在一定不足。具体结果如表2所示：

表 2. 模型性能对比表

模型	参数量 (M)	FLOPs (G)	总体准确率 (OA)
ResNet-50	25.6	4.1	80.67%
EfficientNet-B0	5.3	0.4	81.52%

5.3 标签分布学习的效果

将传统的 one-hot 标签替换为标签分布后，模型在总体准确率上有显著提升。这表明标签分布学习机制有效建模了情绪的复杂性。具体结果如表3所示：

表 3. 标签类型与模型性能对比

模型	标签类型	总体准确率 (OA)
EfficientNet-B0	one-hot	81.52%
EfficientNet-B0	标签分布学习	84.63%

5.4 数据增强的效果

通过图像风格化技术对小类别样本进行增强后，模型在小类别上的表现进一步提升，总体准确率达到 86.25%。这一结果表明，风格化技术不仅能够缓解数据不平衡问题，还能提高模型对低频类别的鲁棒性。具体结果如表4所示：

表 4. 数据增强的效果

模型	数据增强	总体准确率 (OA)
EfficientNet-B0	无数据增强	84.63%
EfficientNet-B0	风格化增强	86.25%

5.5 消融实验

为验证各个优化模块的独立贡献，本文进行了消融实验，逐步移除标签分布学习和数据增强模块以观察性能变化。结果如表5所示：

表 5. 消融实验的结果

模型设置	总体准确率 (OA)
基础模型 (无优化)	81.52%
基础模型 + 标签分布学习	84.63%
基础模型 + 标签分布学习 + 风格化增强	86.25%

消融实验表明，标签分布学习在情绪复杂性建模方面贡献显著，同时数据增强通过缓解类别不均衡问题进一步提升了小类别识别能力，进而提高了网络的整体识别能力。

6 总结与展望

本文提出了一种基于 EfficientNet 的轻量化面部情绪识别网络，通过引入标签分布学习机制有效建模情绪的复杂性，并结合图像风格化技术缓解数据不均衡问题，提升了小类别的识别能力。在 RAF-DB 数据集上的实验结果表明，与基线模型相比，本文方法在总体准确率性能上取得了显著的提升，验证了本文设计的有效性和鲁棒性。

未来的研究将进一步关注面部情绪识别的跨领域适应性，探讨该方法在不同场景中的泛化能力，以应对实际应用中情绪表达形式因环境变化而引发的复杂性。此外，面部情绪识别的全面性仍有提升空间，未来将结合语音、文本等多模态信息进行融合分析，从多个角度综合判断情绪状态，从而获得更为准确的情绪识别结果。与此同时，小类别数据生成与增强也将成为重要的研究方向，未来将探索生成对抗网络（GAN）或扩散模型等先进技术生成高质量的小类别样本，进一步缓解数据不均衡问题，并提升模型在真实场景下的表现力和鲁棒性。这些研究方向的拓展不仅能够进一步优化面部情绪识别技术，还将为情感计算领域的发展提供新的视角和技术支撑。

参考文献

- [1] Pablo Barros, Nikhil Churamani, and Alessandra Sciutti. Facechannel: A lightweight neural network for facial expression recognition. In *Proceedings of the 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–7, 2020.
- [2] Armando Ferro-Pérez and Heriberto Mitre-Hernandez. Resmonet: A lightweight convolutional neural network for facial expression recognition. *IEEE Access*, 8:158460–158470, 2020.
- [3] Bin-Bin Gao, Chao Xing, Chen-Wei Xie, Jianxin Wu, and Xin Geng. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, 26(6):2825–2838, 2017.
- [4] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016.
- [5] Xin Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1734–1748, 2016.
- [6] Christopher Hewitt and Hatice Gunes. C3ae: Exploring the limits of compact model for facial expression recognition. In *Proceedings of the 2018 ACM Multimedia Conference*, pages 564–572, 2018.
- [7] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [8] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [9] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 694–711. Springer, 2016.
- [10] Shan Li and Weihong Deng. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2852–2861, 2019.
- [11] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6848–6856, 2018.

- [12] Zengqun Zhao, Qingshan Liu, and Feng Zhou. Robust lightweight facial expression recognition network with label distribution training. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 3510–3519, 2021.