

文献综述：基于 CNN 与 Transformer 混合模型的口罩人脸识别方法

1. 文献信息

- **作者：** 姜绍忠，姚克明，陈 磊，王中洲，郭复澳
- **发表年份：** 2022
- **文献标题：** 基于CNN与Transformer混合模型的口罩人脸识别方法
- **期刊/会议名称：** 传感器与微系统 (transducer and Microsystem Technologies) ?
- **DOI/网址：** 10.13873/J.1000-9787(2023)01-0144-05

2. 研究背景与目的 (Why)

本研究聚焦疫情背景下，因佩戴口罩导致普通人脸识别算法失效的问题（鼻子、嘴巴这 2 个重要人脸特征被破坏，人脸关键点位置信息带来大量噪声），提出一种基于卷积神经网络与Transformer混合模型的口罩人脸识别方法，在处理戴口罩和不戴口罩的人脸识别任务中，均使模型拥有较高的识别精度，更好的性能及泛化能力。

以往文献提及：

文献 [3] 通过混合戴口罩和不戴口罩的人脸数据集来对模型进行训练以提高现有人脸识别模型性能

文献 [4] 利用 Triplet loss 对在混合数据集上预训练的模型进行微调，进一步提高人脸识别模型对口罩人脸识别的准确率

文献 [5] 采用裁剪，并结合注意力模块使网络更多地关注眼部周围地特征，该方法能够很好地提高对口罩人脸识别的准确率，但却降低了无遮挡场景下人脸识别的性能

文献 [6] 采用双路的训练策略，提出上半人脸分块注意力，通过这种训练策略引导网络更多地关注到上半人脸。

3. 研究方法与技术路线 (How)

作者在**主干特征提取网络**使用的是 CNN 与Transformer 的混合模型CoAtNet，使用自注意力的输入同时具有自适应加权和全局感受野特性以及 CNN 的平移不变性的最优架构。

在卷积模块引入**空间注意力机制**模块，由于通道注意力部分能使更多有利于识别任务的特征通道的权重得以增加，使模型能够更多地关注到未被遮挡的人脸部分，加强特征提取网络。

同时在构建分类器时使用了 **Sub-center ArcFace 损失函数**，为每个身份使用子类的思想，放宽了类内约束，可以有效的提高噪声下的鲁棒性，提高模型的分类精度。

最后在LFW-Masked和RMF数据集上做了消融实验，证明添加了空间注意力机制后网络的准确率，使用了 Sub-center ArcFace 和CoAtNet后模型的准确率均有所上升。

数据集：实验使用CASIA-WebFace作为训练集，LFW-Masked 作为验证集，RMF真实口罩脸作为测试集。

4. 核心内容与创新点 (What)

混合模型设计：采用 CNN + Transformer 混合结构 CoAtNet，能够很好地处理局部和全局信息，使模型拥有更好的性能及泛化能力。

空间注意力机制引入：引入了空间注意力（Spatial Attention）机制来加强特征提取网络。

改进损失函数：使用 Sub-center ArcFace的添加口罩版本，相比 ArcFace 更适应大规模噪声数据，更好地对口罩人脸进行分类。

5. 研究结论与实际意义

本文提出的方法在RMF、LFW、LFM-Maked数据集上识别精度达到98.54%，97.68% 和 93.36%，相比于 MobileFaceNet、ResNet50、ViT 等模型，模型的参数量不大，计算复杂度较低，且总体性能更优。该方法对实际口罩佩戴环境下的人脸识别技术有较强的应用价值。

6. 本文的不足与未来展望

依赖合成数据训练

（未考虑不同口罩种类、光照变化、动态表情等因素对识别的影响）

7. 个人思考与选题关联

本文启发我对“**遮挡场景下的人脸识别**”方向的兴趣，后续可能考虑探究基于 **Transformer 的图像局部增强机制**，尤其是在**资源受限场景下（如车载终端、移动设备）**如何改变模型参数并保持识别性能。结合我对**智能终端和边缘识别系统**的兴趣，未来研究可考虑“**轻量化Transformer人脸识别模型**”作为潜在选题方向
