基于肌肉运动引导网络的微表情识别论文复现

汪子晗

摘要

面部微表情（ME）是一种非自愿的面部动作，它揭示了人们的真实感受，在早期干预精神疾病、国家安全和许多人机交互系统中发挥着重要作用。因此，对于微表情的研究是非常有必要的。我们所复现的文章是发表在 IJCAI 2022 上的一篇关于微表情识别的文章，文章认为现有的微表情数据集是有限的，并且通常对训练好的分类器提出一些挑战。为了对细微的面部肌肉运动进行建模，文章提出了一种鲁棒的微表情识别（MER）框架，即肌肉运动引导网络（MMNet）。具体来说，通过引入连续注意力（CA）块，专注于用很少的身份信息来建模局部细微肌肉运动模式，这与大多数先前的方法不同，先前的方法直接从具有大量身份信息的完整视频帧中提取特征。此外，还设计了一个基于 Vision Transformer 的位置校准（PC）模块。通过在两个分支的末端添加由 PC 模块生成的面部位置嵌入，可以对面部肌肉运动模式特征添加位置信息从而进行微表情识别。在三个公共微表达数据集上的大量实验表明，文章的方法在很大程度上优于最先进的方法。同时，我们在新数据集上进行了实验，并对原文方法进行了改进，取得了优于原文的效果。

关键词：微表情；注意力

1 引言

面部表情是一种流行的非言语交流方式，在反映一个人的情绪状态方面起着重要作用。面部肌肉运动的不同组合最终代表了特定类型的情绪。根据心理学家的说法，无论种族或文化如何，人们都会以同样的方式在脸上描绘某些特定的情绪[1]。此外,[2]验证了盲人和正常人在面部肌肉运动对情绪刺激的反应方面没有差异。换句话说，面部表情是普遍的。他们通常可以分为六类情绪：快乐、悲伤、恐惧、愤怒、厌恶和惊讶。

通常，面部表情分为两种类型，即宏表情和微表情。宏表情通常持续 0.5s 到 4s 之间，肌肉运动可能同时发生在面部的多个部位。因此，人类在实时对话中很容易感知到宏表情。在过去的几十年中，自动宏表情识别分析的研究一直是一个活跃的话题。迄今为止，开发的许多识别系统实现了 95 % 以上的表情分类准确率[3][4]，其中一些甚至达到了几乎 100 % 的完美识别性能。然而，需要注意的是，宏表情并不能准确地暗示一个人的情绪状态，因为它很容易被伪造。因此，有必要从肌肉运动中探究更深层次的情绪状态。

在几种类型的非言语交流中，微表情被发现更有可能揭示一个人的真实情绪。微表情通常发生在 0.5s 以内并且它们可能只出现在面部的几个小区域中。此外，他们并不是自愿发生的，这意味着人们无法控制自己的微表情。由于其潜在暴露一个人真实情绪的特性，它可以应用于国家安全、警察审讯、商业谈判、社交互动和临床实践。

我们所复现的文章是发表在 IJCAI 2022 上的一篇关于微表情识别的文章[5]，该文章所提出的方法将微表情识别的精度提升到了一个新的高度。尽管近年来的方法逐渐提高了自动微表情识别算法的性能。然而，他们中的大多数直接将原始视频帧或视频帧的手工特征输入到深度网络中用于提取 ME 的

特征，这使得 DNN 易于引入样本的身份信息。显然，与表情无关的身份信息对面部表情识别（FER）任务有害。这个问题可能对具有丰富训练数据的宏观表情识别任务影响不大。然而，由于收集和标记微表情数据的成本极高，我们仍然没有可与宏表情数据集相比的大规模微表达数据集。MEs 主要取决于面部肌肉运动的位置和肌肉运动模式（例如，两侧唇角的轻微上翘可能表示幸福）。因此，MER 的关键是学习面部肌肉运动的位置和模式，而不是直接从整个视频帧中学习。那么这篇文章的主要贡献如下：

- 提出了一种新的双分支 MER 范式，该范式分别通过主分支和子分支提取肌肉运动模式特征和面部位置嵌入。然后，在网络末端融合这两种特征进行分类。
- 设计了一个肌肉运动引导网络（MMNet）来实现上述两分支 MER 范式。用于提取运动模式特征的主分支由所提出的连续关注（CA）块组成，用于生成位置嵌入的子分支通过设计的位置校准（PC）模块实现。
- MMNet 在三个流行的微表情数据集（即 CASME II、SAMM 和 MMEW）上大大优于最先进的方法。大量实验证明了所提出的 MMNet 的有效性。

2 相关工作

根据特征提取方法，微表情识别技术可大致分为两类：传统手工特征方法和基于深度学习的方法。

2.1 传统手工特征方法

对于传统方法，通常使用定向梯度直方图（HOG）、光流直方图（HOOF）和局部二值模式-三正交平面（LBP-TOP）来提取 ME 特征。Le Ngo[6]等人通过处理 LBP-TOP 使其学习到具有稀疏性约束的时间和光谱结构。Li[7]等人采用了 MER 的图像梯度方向 TOP（HIGO-TOP）直方图。Happy[8]等人提出了一种基于模糊的 HOOF（FHOOF）特征提取技术，该技术仅考虑 MER 的肌肉运动方向。然而，由于 ME 的持续时间短且运动不明显，手工特征通常无法可靠地表示不同微表情之间的差异，这对微表情识别不利。

2.2 基于深度学习方法

近年来，随着深度学习技术的发展，越来越多的研究人员通过设计深度神经网络（DNN）来处理 MER 任务，取得了令人满意的结果。例如，Gan[9]等人引入了一种特征提取器，该提取器结合了手工（即光流）特征和数据驱动（即卷积神经网络，CNN）特征。Song[10]等人提出了一种三流卷积神经网络（TSCNN），通过学习 ME 视频的三个关键帧中的 ME 辨别特征来识别 ME。Xie[11]等人通过结合动作单元（AU）和情绪类别标签开发了 MER 方法，同时 Lei[12]等人设计了一个图时间卷积网络（graph TCN）来提取 ME 的局部肌肉运动特征。Xia[13]等人设计了一个框架，利用宏表情样本作为 MER 的指导。在最新的结果中，Xia[14]等人分别从空间域和时间域通过两个辅助任务设计了宏表情到微表情的转换框架。

3 本文方法

3.1 本文方法概述

现有的 MER 方法通常侧重于设计深度网络的结构以提高识别精度，但通常忽略了找到更好的方法来利用 ME 的视频帧。大多数方法将带有身份信息的原始视频帧输入到单个分支网络[15][12]，或从帧中提取手工特征（例如，光流或 LBP），然后通过多分支网络将其融合[16]。由于主分支的输入在不同程度上包含样本的身份信息，网络很可能学习与 MER 任务无关的身份相关特征，特别是当 ME 数据不足时。为了解决这个问题，我们提出了一个新的双分支 MER 范式。主分支被设计为处理运动模式，而轻量级子分支从主要包含面部位置信息而没有任何表情相关信息的低分辨率起始帧中提取面部位置嵌入。以这种方式，即使低分辨率起始帧包含某些身份信息，因为它只被馈送到轻量级子分支中，所以整个网络仍然可以专注于学习运动模式而不是身份。

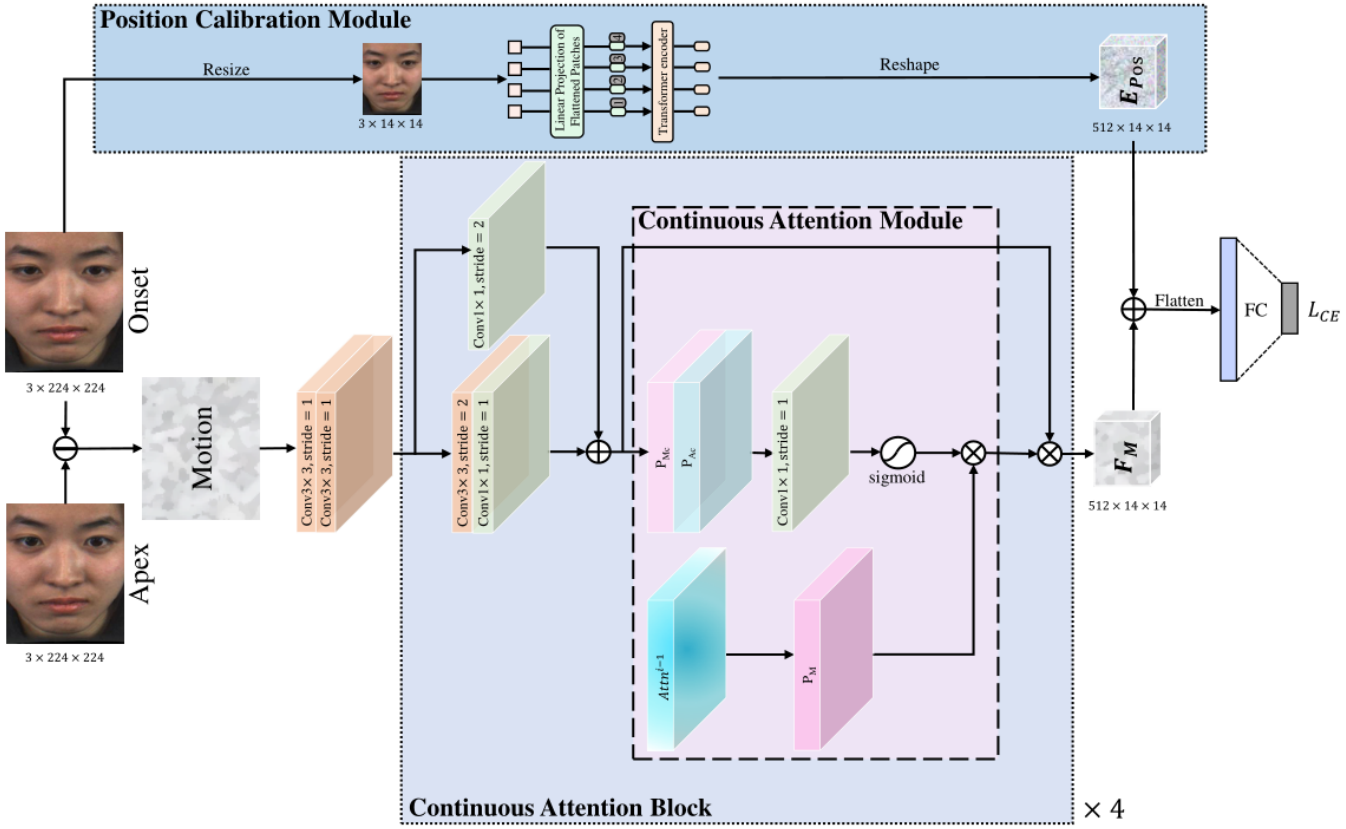

图 1: MMNet 的主要模型结构

如图 1所示，我们提出的 MMNet 主要由两个分支组成。主分支以顶点帧和起始帧之间的差异作为输入提取运动模式特征，子分支用于以低分辨率起始帧作为输入生成面部位置嵌入。最后，将面部位置嵌入添加到运动模式特征，以将运动模式映射到特定面部区域。

3.2 持续注意力模块

与宏表情相比，微表情往往具有更小的肌肉运动和更多的局部活动区域，这使得传统的注意力模块很难准确地聚焦于细微的面部肌肉运动。为了缓解这一问题，一些现有的工作[11]试图借助动作单元（AU）标签[17]来模拟肌肉运动和 MEs 之间的关系。然而在不引入额外监督的情况下获取精确的注意力图仍然是一个巨大的挑战。为了解决上述问题，我们设计了一个连续注意力块，通过引入前一层的注意力图作为先验知识来生成当前层的注意力地图。受图 2（a）所示卷积块注意力模块（CBAM）的

空间注意力模块的启发[18]，我们利用最大池输出和平均池输出来计算空间注意力图。如图 2（b）所示，我们将前一层的注意力图作为先验知识来获得当前层的注意力图，并使用更小的卷积核（即 1×1）来获得更多的局部注意力图。CA 模块形式上可以定义为

$$\begin{aligned} Attn^{i} &= \mathrm{M}^{\mathrm{i}}\left(F_{\mathrm{conv}}^{i}, Attn^{i-1}\right) \\ &= \sigma\left(f_{1\times1}^{i}\left(\left[\mathrm{P}_{\mathrm{Mc}}\left(F_{\mathrm{conv}}^{i}\right); \mathrm{P}_{\mathrm{Ac}}\left(F_{\mathrm{conv}}^{i}\right)\right]\right)\right) \bigotimes \mathrm{P}_{\mathrm{M}}\left(Attn^{i-1}\right) \end{aligned} \tag{1}$$

其中

$$F_{\mathrm{conv}}^{i} = f_{1\times1}^{i}\left(f_{3\times3}^{i}\left(F^{i}\right)\right) + f_{1\times1}^{i}\left(F^{i}\right) \tag{2}$$

其中 M^{i} 是关注肌肉运动区域的第 i 个 CA 块的 CA 模块，$Attn^{i-1}$ 是第（i−1）层的注意力图。$F_{conv}^{i} \in \mathrm{R}^{2\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$ 表示由第 i 层的前两个卷积层提取的特征，作为 CA 模块的输入。σ 表示 S 形函数。$f_{1\times1}^{i}$ 和 $f_{3\times3}^{i}$ 分别表示来自第 i 层的卷积核大小为 1 和 3 的卷积运算。$\mathrm{P}_{\mathrm{Mc}}\left(F_{\mathrm{conv}}^{i}\right) \in \mathrm{R}^{1\times\mathrm{H}\times\mathrm{W}}$ 和 $\mathrm{P}_{\mathrm{Ac}}\left(F_{\mathrm{conv}}^{i}\right) \in \mathrm{R}^{1\times\mathrm{H}\times\hat{\mathrm{W}}}$ 分别表示最大池化特征和平均池化特征。P_M 表示在第（i−1）层的注意力图上的最大池操作，以匹配当前层注意力图的大小，而 ⊗ 表示用于将最后一层的注意力图作为先验知识引入的元素乘积。$F^{i} \in \mathrm{R}^{\mathrm{C}\times2\mathrm{H}\times2\ \mathrm{W}}$ 代表第 i 个 CA 块的输入。

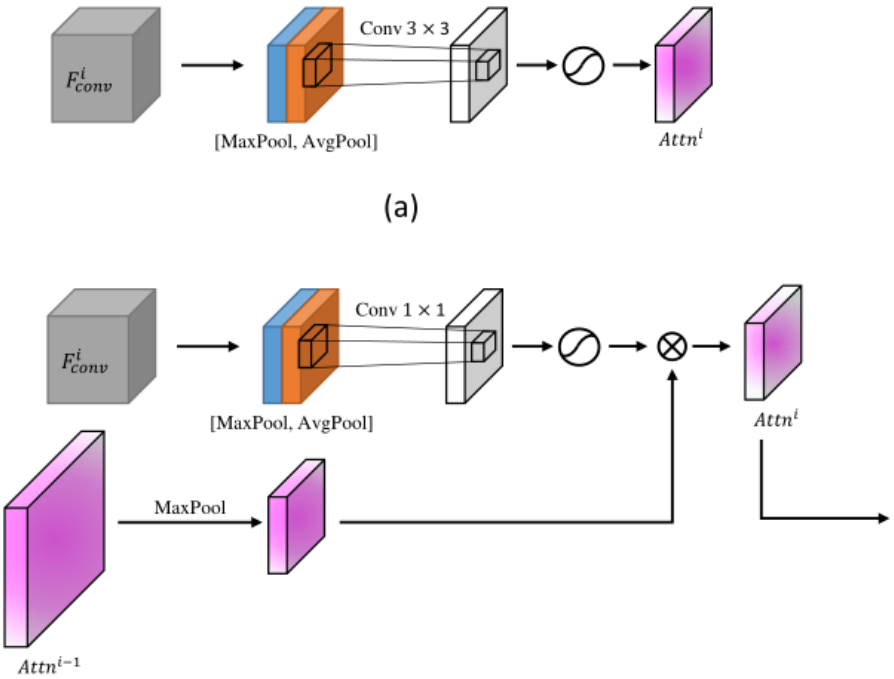

图 2: CBAM 和 CA 模块图。（a）CBAM 模块的空间注意力模块和（b）文章的 CA 模块。

通过将相邻层之间的注意力机制关联起来，CA 块可以逐渐且鲁棒地聚焦于具有细微运动的区域，而不是聚焦于不同层中的面部的不同区域，这可以使网络学习 ME 相关区域和不相关区域。我们使用四个 CA 块构成 MMNet 的主要分支，以学习大小为 512×14×14 的细微肌肉运动模式特征 F_M。由 CA 模块和两个卷积层组成的 CA 块可以表示为

$$\mathrm{CA}\left(F^{i},\ \mathrm{Attn}^{\,i-1}\right) = F_{\mathrm{conv}}^{i} \circledast \mathrm{M}^{\mathrm{i}}\left(F_{\mathrm{conv}}^{i},\ \mathrm{Attn}^{\,i-1}\right) \tag{3}$$

其中 CA 表示所提出的连续关注块，⊛ 表示广播元素对位乘法，这意味着 $F_{\mathrm{conv}}^{i} \in \mathrm{R}^{2\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$ 的每个信道将乘以空间注意图 $\mathrm{M}^{\mathrm{i}}\left(F_{\mathrm{conv}}^{i},\ \mathrm{Attn}^{\,i-1}\right) \in \mathrm{R}^{1\times\mathrm{H}\times\mathrm{W}}$，以关注感兴趣的区域。我们的 CA 模块有助于集中于准确的 ME 相关区域。如 1所示，我们将峰值帧和起始帧之间的差异作为主分支的输入，以学习

运动模式特征。

3.3 位置校准模块

由于微表情数据集中不同人的不同外观，瞳孔间距离不同、鼻子大小不同等原因，很难严格对齐所有人脸。因此，相同的人脸区域可能对应于图像的不同像素位置，这使得网络很难准确了解细微运动发生的位置。为了将位置信息准确地添加到由主分支提取的运动模式特征，我们提出了一个位置校准模块作为 MMNet 的子分支，以生成用于将运动模式特征映射到面部特定区域的面部位置嵌入。

由于面部特征的相对位置是物理确定的（例如，鼻子通常位于两只眼睛的中间以下），建模长距离依赖关系可以有效地帮助定位面部各个部分的位置并生成鲁棒的位置嵌入。最近，ViT 将自注意机制应用于长距离依赖性建模，并在图像分类任务上取得了很好的结果，而卷积神经网络（CNN）通常需要许多卷积层来获得全局感受野，这不利于建模面部位置信息。因此，我们介绍了基于浅层 ViT 的 PC 模块。如 1所示，我们利用峰值帧和起始帧之间的差异来学习运动模式特征 F_M，并利用低分辨率起始帧来学习面部位置嵌入 E_{pos}。由于我们只需要学习显著区域（例如，眼睛、嘴巴和鼻子）的位置，而不是与样本身份相关的详细纹理（例如，皱纹和肤色），因此我们将起始帧缩放为与 F_M 大小匹配的子分支的输入相同的 14×14。然后，我们将缩放的起始帧重塑为 196 个大小为 1×1×3 的平坦 2D 面片 I_p 序列，并通过可训练线性投影来得到它们的映射，以获得 512 维的面片嵌入 E_p，这与 F_M 的信道维度相匹配。添加 ViT 的位置嵌入以保留位置信息后，这些 patch 被发送到 encoder 中以学习 patch 之间的关系。最后，将尺寸为 196×512 的 ViT 的输出整形为 512×14×14，以获得用于位置校准的 E_{pos}。然后将位置嵌入 E_{pos} 添加到运动模式特征 F_M 中，用于将运动模式映射到 MER 的特定面部区域。

4 复现细节

4.1 与已有开源代码对比

文章已有部分开源代码，代码地址为 https://github.com/muse1998/MMNet[5]. 在微表情识别任务中，人脸预处理包括人脸裁剪，人脸对齐等步骤，这些重要步骤都会直接影响识别的性能，原始代码中并没有给出对数据集进行预处理的程序，所以对于人脸的预处理全部由我自己完成。其次，微表情识别任务的评估采用留一法，即每次采取一个对象的所有样本作为测试集，剩余所有对象的样本作为训练集，最后在所有对象都被作为测试集后，计算总体的准确度和 F1 分数，那么原文代码的评估代码并不完善，也并不完全正确，因此评估模块的代码也全部由我自己完成。与此同时，我还尝试搭建了一些 baseline，例如 resnet18，resnet50 等，可以对文章方法的性能做一个参照。并且我还在今年新发布的数据集上进行了实验，进一步验证了该方法的有效性。最后，对模型进行了一些改进尝试，例如将差值输入改为光流输入，通过融合深度信息提高性能，通过改变 loss 解决数据样本不平衡的问题等等。

4.2 人脸预处理

在对数据进行正式实验之前,我们对数据集中原始样本片段进行了三步预处理。假设 $S = [s_i \mid s \in S, i =$ 是一组微表情片段。第 i 个样本 $s_i = [f_{i,j} \mid f \in s_i, j = 1, \ldots, k_i]$，$k_i$ 是序列 s_i 的帧数。

首先，选择具有中性表情的正面面部图像 M 作为模板脸。使用主动形状模型[19]检测到模板脸的 68 个面部标志点 $\psi(M)$。

其次，使用局部加权平均（LWM）[20]变换将每个微表情片段 $f_{i,1}$ 的第一帧归一化为标准脸，并且

变换矩阵 T 被写为：

$$T_i = LWM\left(\psi(M), \psi\left(f_{i,1}\right)\right), i = 1, \ldots, n \tag{4}$$

其中 $\psi(f_{i,1})$ 是微表情样本 s_i 的第一帧的 68 个界标点的坐标。然后使用相同的矩阵 T_i 对 s_i 的所有帧进行归一化。我们仅在第一帧而不是在所有帧上检测 ASM 界标点有两个原因。第一个原因是，由于微表情的持续时间很短，因此可以忽略持续时间内的刚性头部运动。第二个原因是 ASM 检测到的界标点可能不够准确；如果应用在一系列帧上，即使面部完全不移动，同一点的位置也可能有很大的偏差。归一化图像 f' 是原始图像的 2D 变换：

$$f'_{i,j} = T_i \times f_{i,j}, j = 1, \ldots, k_i \tag{5}$$

$f'_{i,j}$ 是归一化后微表情序列 s'_i 的第 j 帧。

第三，定位每个标准化微表情序列 $f'_{i,j}$ 的第一帧的眼睛坐标 E_i，然后使用由眼睛位置 E_i 确定的矩形裁剪出 s'_i 的每一帧的面部。

4.3 创新点

我们对论文所提出的方法进行了改进尝试。

本文中作者希望将起始帧和峰值帧的差值输入到网络中，这样即消除了人的身份信息，同时也专注于人脸的运动，然而我们认为这样直接将两帧做差的方法不足以展现两帧之间的差异以及两帧之间的运动。于是我们希望提取两帧之间的光流特征来代替原始的差值输入。其中光流表达了图像的变化, 由于它包含了目标运动的信息, 因此可被观察者用来确定目标的运动情况。

之前包括本文的方法都仅仅基于 2D 的 RGB 图像来进行微表情识别，而深度信息可以帮助人们建立更好的面部特征，从而使立体视觉中的面部更接近现实中的人脸。因此，深度信息使面部特征更容易被识别。所以，深度信息是有助于增强人类对微表情的感知的，受此启发，我们将深度信息引入了微表情识别，这种多模态的感知更加贴近人的行为，也更容易使模型获得更好的性能。

由于微表情数据集中的类别样本十分不均衡，一般来说，负面情绪的样本居多，而正面情绪的样本较少，那么在这种情况下去训练模型，会导致样本量少的那些类别性能较差，那么文章并没有考虑到这一点，因此我们希望在 loss 上给每个类别赋予不同的权重，以此来平衡这个问题。

5 实验结果分析

为了验证 MMNet 的有效性，原文对三个流行的微表情数据集进行了广泛的实验，包括 CASME II、SAMM 和 MMEW。我们首先介绍这三个数据集和实现细节。由于权限问题，我们没有要到 SAMM 数据集，因此我们只对 CASME II 和 MMEW 数据集进行了实验，我们将实验结果与原文进行对比。最后，我们在新数据集 $CAS(ME)^3$[21]上进行了基本实验和改进实验。

5.1 数据集介绍

CASME II[22]包含 26 名受试者的 256 个微表情视频，帧率为 200 fps，裁剪后大小为 280×340。与之前的大多数方法一致，只使用了五种典型表情的样本，即快乐、厌恶、压抑、惊讶和其他。SAMM[23]以 200 帧/秒的速度从来自 13 个不同种族的 32 名受试者那里获得了 159 个微表情片段。实验中使用了五种微表情（快乐、愤怒、轻蔑、惊讶和其他）。MMEW[24]包括从同一受试者中取样的宏表情和微表情，供

研究人员探索它们之间的关系。它包含 300 个微表情和 900 个宏表情样本，分辨率更高（1920×1080），每秒 90 帧。我们在消融研究中使用了四种微表情（快乐、惊讶、厌恶和其他）。与之前的大多数工作一致，在所有实验中使用了留一（LOSO）交叉验证，这意味着每个受试者依次作为测试集，其余受试者作为训练数据。对于所有实验，准确度和 F1 分数用于性能评估。

5.2 实现细节

在我们的实验中，我们首先检测人脸上的关键点，并根据这些关键点裁剪图像。所有数据集上的裁剪图像大小调整为 224×224。为了避免过度拟合，我们从标记的起始帧和顶点帧周围的四个帧中随机选择一个帧作为起始帧和峰值帧进行训练。还采用了水平翻转、随机裁剪和颜色抖动。我们使用四个 CA 块来构成主分支，并使用一个浅层 ViT 来构建子分支。在训练阶段，我们采用 AdamW 来优化批量大小为 32 的 MMNet。对于交叉熵损失函数，学习率被初始化为 0.0008，在 70 个 epoch 内以指数速率降低。所有实验都是基于 PyTorch 并在单张 NVIDIA Tesla P100 卡上进行的。

5.3 实验结果展示

我们在 CASME II 和 MMEW 数据集上进行实验，实验结果如表 1所示，并将我们所复现的结果与原文中报告的结果进行对比，可以看到我们基本上完成了论文的复现，在 CASME II（3 类）的测试上我们的 Acc 和 F1 都超过了原文，而 CASME II（5 类）和 MMEW 上与原文只有极其微小的差距，那么可能的原因是数据集本身样本量太小，而训练模型过程中随机性较大，所以最后结果的随机性也较大，不过可以通过大量实验来消除这种影响，最后取得的最优结果是和原文基本一致的。

	CASME II (3 类)		CASME II(5 类)		MMEW(4 类)	
	Accuracy(%)	F1	Accuracy(%)	F1	Accuracy(%)	F1
原文	95.51	0.9494	**88.35**	**0.8676**	**87.45**	**0.8635**
复现	**96.15**	**0.9570**	87.15	0.8652	87.17	0.8620

表 1: CASME II 和 MMEW 上实验结果

同时，我们对原文的方法在新数据集 $CAS(ME)^3$ 上进行了实验，也对我们尝试的的改进方法进行了实验，实验结果如表 2。首先，我们先用传统方法提取出两帧之间的光流特征，再将光流特征作为网络的输入，实验证明该方法在 $CAS(ME)^3$ 上得到的 Acc 是优于原文的，而 F1 分数是略低于原文。其次我们将深度信息引入，将两帧的深度图做差作为网络的输入，以及将两帧的 RGB-D 图做差作为网络的输入，结果显示这两种方法并没有原文的效果好。

method	$CAS(ME)^3$ (4 类)	
	Accuracy(%)	F1
RGB difference	79.14	**0.6799**
Optical flow	**79.38**	0.6674
Depth difference	68.98	0.5193
RGBD difference	77.68	0.6385

表 2: $CAS(ME)^3$ 上实验结果

6 总结与展望

我们复现了发表在 IJCAI 2022 上的一篇关于微表情识别的文章，基于肌肉运动引导网络的微表情识别。这篇文章将微表情识别的精度提升到了一个新的高度。那么本次报告我通过介绍微表情的概念，意义，应用场景，同时介绍了近年来的一些相关工作，并详细阐述了文章中的微表情识别方法，将论文的核心内容完全体现。后续，我又介绍了我的整个复现工作以及我的改进思路，并且展示了我的实验结果，可以说较好地完成了本次复现工作。

那么对于本篇文章方法的改进工作还是可以继续的，深度信息作为一个重要的模态，对微表情的识别一定是有积极作用的，那么如何将深度信息融合到这样一个网络中是一个关键的问题。同时，由于面部微表情的局部性和微弱性，我们可以考虑采用多尺度的卷积核来提取特征，这样会帮助我们提取到更加细微的特征。

参考文献

[1] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion.[J]. Journal of personality and social psychology, 1971, 17(2): 124.

[2] MATSUMOTO D, WILLINGHAM B. Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals.[J]. Journal of personality and social psychology, 2009, 96(1): 1.

[3] LOPES A T, DE AGUIAR E, DE SOUZA A F, et al. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order[J]. Pattern recognition, 2017, 61: 610-628.

[4] WANG Z, RUAN Q, AN G. Facial expression recognition using sparse local Fisher discriminant analysis [J]. Neurocomputing, 2016, 174: 756-766.

[5] LI H, SUI M, ZHU Z, et al. MMNet: Muscle motion-guided network for micro-expression recognition [J]. arXiv preprint arXiv:2201.05297, 2022.

[6] LE NGO A C, SEE J, PHAN R C W. Sparsity in dynamics of spontaneous subtle emotions: analysis and application[J]. IEEE Transactions on Affective Computing, 2016, 8(3): 396-411.

[7] LI X, HONG X, MOILANEN A, et al. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods[J]. IEEE transactions on affective computing, 2017, 9(4): 563-577.

[8] HAPPY S, ROUTRAY A. Fuzzy histogram of optical flow orientations for micro-expression recognition [J]. IEEE Transactions on Affective Computing, 2017, 10(3): 394-406.

[9] GAN Y S, LIONG S T, YAU W C, et al. OFF-ApexNet on micro-expression recognition system[J]. Signal Processing: Image Communication, 2019, 74: 129-139.

[10] SONG B, LI K, ZONG Y, et al. Recognizing spontaneous micro-expression using a three-stream convolutional neural network[J]. IEEE Access, 2019, 7: 184537-184551.

[11] XIE H X, LO L, SHUAI H H, et al. Au-assisted graph attention convolutional network for micro-expression recognition[C]//Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2871-2880.

[12] LEI L, LI J, CHEN T, et al. A novel graph-tcn with a graph structured representation for micro-expression recognition[C]//Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2237-2245.

[13] XIA B, WANG W, WANG S, et al. Learning from macro-expression: a micro-expression recognition framework[C]//Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2936-2944.

[14] XIA B, WANG S. Micro-Expression Recognition Enhanced by Macro-Expression from Spatial-Temporal Domain.[C]//IJCAI. 2021: 1186-1193.

[15] LI Y, HUANG X, ZHAO G. Joint local and global information learning with single apex frame detection for micro-expression recognition[J]. IEEE Transactions on Image Processing, 2020, 30: 249-263.

[16] LIONG S T, GAN Y S, SEE J, et al. Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition[C]//2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019). 2019: 1-5.

[17] EKMAN P, FRIESEN W V. Facial action coding system[J]. Environmental Psychology & Nonverbal Behavior, 1978.

[18] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[19] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models-their training and application [J]. Computer vision and image understanding, 1995, 61(1): 38-59.

[20] GOSHTASBY A. Image registration by local approximation methods[J]. Image and Vision Computing, 1988, 6(4): 255-261.

[21] LI J, DONG Z, LU S, et al. CAS (ME) 3: A Third Generation Facial Spontaneous Micro-Expression Database with Depth Information and High Ecological Validity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[22] YAN W J, LI X, WANG S J, et al. CASME II: An improved spontaneous micro-expression database and the baseline evaluation[J]. PloS one, 2014, 9(1): e86041.

[23] DAVISON A K, LANSLEY C, COSTEN N, et al. Samm: A spontaneous micro-facial movement dataset [J]. IEEE transactions on affective computing, 2016, 9(1): 116-129.

[24] BEN X, REN Y, ZHANG J, et al. Video-based facial micro-expression analysis: A survey of datasets, features and algorithms[J]. IEEE transactions on pattern analysis and machine intelligence, 2021.