面部表情识别场景中的无三元组损失中心损失采样策略

Hossein Rajoli*, Fatemeh Lotfi*, Adham Atyabi†, Fatemeh Afghah*
*克莱姆森大学 Holcombe 电气与计算机工程系，美国南卡罗来纳州克莱姆森。电子邮箱：{hrajoli, flotfi, fafghah}@clemson.edu,
†科罗拉多大学科罗拉多斯普林斯分校计算机科学系，美国科罗拉多州科罗拉多斯普林斯。电子邮箱：aatyabi@uccs.edu

arXiv:2302.04108v1 [cs.CV] 8 Feb 2023

摘要—面部表情传达大量信息，在情绪表达中发挥着至关重要的作用。深度神经网络（DNN）结合深度度量学习（DML）技术，可提升模型在面部表情识别（FER）应用中的判别能力。仅采用交叉熵等分类损失函数的深度神经网络，无法像加入深度度量学习支持的损失项那样压缩类内特征变化或拉开类间特征距离。三元组中心损失（TCL）函数作用于样本在嵌入空间中嵌入的所有维度。在本研究中，我们提出了三种策略：完全合成、半合成和基于预测的负样本选择策略。为取得更好的结果，我们引入了一种选择性注意力模块，利用输入样本的高语义深层特征，结合像素级和逐元素注意力系数。我们在高度不平衡的数据集 RAF-DB 上评估了所提出的方法。实验结果表明，与基线相比，三种负样本选择策略均取得了显著提升。

索引词—**深度度量学习、三元组损失、人脸表情识别、负样本选择。**

I. 引言

面部表情识别（FER）一直在众多日常应用的发展与普及过程中发挥着重要作用。自动驾驶汽车、人机交互（HCI）、实验质量测量、法证领域、医疗保健与治疗、欺骗检测等领域，都广泛使用自动面部表情识别作为监测或评估工具 [1]。计算能力的提升、新型复杂深度神经网络（DNN）的出现以及大规模数据集的提供，推动计算机视觉这一热门领域取得了长足进步。面部所表达的情绪取决于面部呈现的三类主要信号，即静态、缓慢和快速信号。所有判别特征都属于这三类信号：例如，肤色和个人的面部骨骼形状属于静态信号，而肌肉张力、皮肤纹理和眉毛快速上扬则分别属于缓慢信号和快速信号。因此，在同一类面部情绪表达中，样本具有很高的多样性，会因年龄、智力水平、种族、性别等因素而有所不同 [1, 2]。

本材料基于空军科学研究办公室（项目编号 FA9550-20-1-0090）和国家科学基金会（项目编号 CNS-2232048 和 CNS-2204445）资助的研究工作。

通常，使用深度神经网络执行标准面部表情识别任务主要分为三个步骤：提取输入样本的空间深度特征，将高度语义化的特征空间映射到嵌入空间，然后使用 Softmax激活估计所有类别上的概率分布。与此同时，深度度量学习（DML）也能发挥关键作用。通过促使嵌入空间中的类内样本紧密聚集，并使类间样本尽可能远离，深度度量学习能够改善监督分类任务 [3]。一种有效的基于深度度量学习的损失函数是无损三元组损失 [4]，其设计旨在通过反向传播利用所有样本的信息。

在普通三元组损失中，如果嵌入空间中的样本表示与其对应中心（代表正样本）之间的距离，比其与负样本之间的距离小至少一个预设间隔，那么该样本便不会参与反向传播。为提升性能，本文提出了一种特殊形式的 TC3L，即自适应间隔*T3L*（*AMTC3L*）。与对所有输入采用固定间隔的普通 TC3L 相比，AMTC3L 会为每个样本采用自适应间隔。所提出的损失函数着重突出那些特征有助于提升网络判别能力的样本，并使其在反向传播过程中发挥更大作用。

引入了两种不同的注意力模块：逐像素模块和逐元素包含/排除模块。其核心思想是，仅筛选出面部判别性区域中的判别特征，用于识别所表达的情绪。通过这种方式，模型能够基于已提取的深层特征，学习区分噪声特征和噪声区域。

本文的主要贡献可概括如下；

- 我们提出了一种新形式的三元组损失（TL），即 AMTC3L。在该方法中，网络能够学习适合自身的间隔，并通过以注意力方式为每个样本分配适当的间隔，使模型从某些样本中获得比其他样本更多的学习效果。
- 为更加突出判别特征，我们提出了一种注意力网络，从深层特征中生成排除/包含权重。
- 为了将 AMTC3L 应用于面部表情识别场景，我们提出了三种不同策略，以提供数量最多的负样本——

本选择策略。它们分别是完全合成、半合成和基于预测的负样本选择策略。

- 源代码将公开发布于以下 GitHub 仓库：AMTC3L 采样策略。

II. 相关工作

本文主要介绍三种策略，以在深度度量学习方法的支持下，在面部表情识别场景中定制并利用三元组损失。因此，我们从两个角度考察已有研究：深度度量学习在面部表情识别中的应用，以及分类场景中的三元组损失采样策略。

A. 分类中的三元组损失采样策略

三元组损失能够同时增强类间紧凑性和类内可分性。[5] 中的作者开发了一种按类别配对的间隔识别机制，通过结合离群点检测器提升了面部表情识别应用的性能。他们提出了一种离群点检测器，以防止被遮挡和容易混淆的样本对选定负样本的贡献。在对多标签样本进行分类时，[6] 提出了一种两步样本选择机制。第一步，从小批量中选出一组信息量最大且具有多样性的样本作为锚点；第二步，在选择过程中根据图像的相关性、难度和多样性来选取正样本和负样本。为应对三维检索挑战，[7] 采用了三元组中心损失（TCL），并引入类中心作为锚点样本，同时采用最近负样本采样作为负样本选择策略。由于在不平衡数据集中，作为负样本的少数类通过三元组损失增加了其对反向传播作出贡献的机会，Lei 等人 [8] 利用这一优势，通过三元组损失和三元组中心损失函数对不平衡数据集进行分类。在他们的分类场景中，作者将所有类中心而非仅最近的类中心都用作负样本。Huang 等人 [9] 针对高光谱图像分类的挑战开展了研究；由于这类图像在特征空间中占据较大范围，因此容易出现类间不可分的问题。为利用三元组损失函数，他们在每个小批量中随机选取一部分类别，并从每个选定类别对应的小批量样本中随机抽取一部分。为每个输入样本选取一组难度最高的正样本和负样本，可避免离群值影响训练过程。三元组损失也被用于应对知识蒸馏挑战。Georgescu 和 Ionescu [10]将一个在全脸数据集上训练的网络的知识蒸馏并迁移至一个在遮挡人脸数据集上训练的网络。在所提出的场景中，嵌入空间中的遮挡样本和负样本来自学生网络，而全脸网络提供正样本。在正样本和负样本采样过程中，都会考虑锚点与随机选取的可用样本中十分之一之间的距离。随后，选择距离最远和最近的样本分别作为正样本和负样本。Zhou 等人 [11] 在面部表情识别场景中采用了带固定间隔的 TCL 在面部表情识别场景中。在这项工作中，将输入样本的中心视为正样本，将最近的类中心（输入样本所属类的中心除外）视为负样本。 [12]将卡方距离应用于面部表情识别场景中的三元组损失函数。作者在负样本选择时考虑身份标注（如果有），以降低计算成本及身份特征在训练过程中的影响。

B. 面部表情识别场景中的深度度量学习

利用神经网络进行深度度量学习，是一种将输入样本投影到流形空间中的方法，在该空间中，可以使用欧氏距离、余弦距离或卡方距离等距离度量来衡量两个点之间的距离 [13]。对于面部情绪识别，Fard 和 Mahoor [14] 提出了自适应相关性损失，在训练过程中引导网络变得更具判别性。他们根据网络在嵌入空间中为每个类别生成的 k嵌入向量的均值，定义了一个自适应损失函数。该损失函数会惩罚相关性较低的类内均值向量以及相关性较高的类间均值向量。Farzaneh 和 Qi [15] 利用中心损失与注意力网络来提高类内紧凑性。Huang 等人 [16] 提出了马氏度量距离，并通过全连接网络将其应用于更高层次的语义特征。为达到姿态感知和身份不变的效果，[17]的作者提出了五元组深度度量学习，其中包含两组三元组，即具有共同锚点的正样本三元组和身份三元组，随后使用该深度度量学习方法训练网络。Liu 等人 [18] 提出了（$N+M$）元组簇损失：从同一身份中选择 N 个不同的表情作为负样本，并动态搜索 M正样本；随后，在该损失与 Softmax 的共同作用下对网络进行联合优化，以缓解不同身份带来的属性变化。

III. 提出的模型

本节将详细阐述所提出的方法，并介绍所有模块，即作为骨干网络的 ResNet-50、分类器、多任务损失函数，以及净化模块。该净化模块能够消除嵌入向量中的离群特征，同时为输入样本提供自适应边界。在我们的方法中，利用净化网络生成的权重，边界将计算为 $0 \leq \{\alpha_i = \sum_{j=1}^{C_d} w_j\} \leq C_d$。给定一个包含 m 个样本的小批次—— $D_m = \{(x^{(i)}, y^{(i)}) \mid i = 1, 2, \ldots, m\}$，其中 $x^{(i)} \in \mathbb{X}$ 是训练集中的训练样本，$y^{(i)} \in \{1, 2, \ldots, K\}$ 是其对应的类别标签。骨干网络 $\mathcal{N}_{\mathcal{BB}}$ 将小批次输入样本映射为其特征空间表示 $x^* \in \mathbb{R}^{C_f \times H_f \times W_f}$，其中 C_f、H_f 和 W_f 分别表示特征空间中张量的通道数、行数和宽度。通过利用两个深度卷积层，分类器网络 $\mathcal{N}_{\mathcal{C}}$ 获取样本的特征空间表示，并提取深层高语义上下文 $x^d \in \mathbb{R}^{C_d \times H_f \times W_f}$。随后，利用全局平均池化层 $\mathcal{P}$，将张量转换为嵌入



图 1：RAF-DB 面部表情数据集样例。

$e \in \mathbb{R}^{C_d \times 1 \times 1}$。最后，另一个深度卷积层将嵌入映射为 $x^z \in \mathbb{R}^{K \times 1 \times 1}$ 的行分数。然后，通过在通道维度上应用 Softmax，得到关于 K 个通道的概率分布 $Pr(y = j \mid x^{(i)})$。最后，使用交叉熵（CE）损失函数 $\mathcal{L}_{CE}$ 计算预测值 $\hat{y}$ 与真实标签 $y^{(i)}$ 之间的差异，具体如下：

$$\mathcal{L}_{CE} = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{K}\log y^{(i)} Pr(y^{(i)} = j \mid x^{(i)}). \quad (1)$$

A. 多任务损失

为计算差异，对于任意训练样本，都会计算以下损失函数:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{AMTC3L}, \quad (2)$$

其中，(1) 计算 $\mathcal{L}_{CE}$，而 $\mathcal{L}_{AMTC3L}$ 衡量自适应边际三元组中心损失。

1) 自适应边际三元组无中心损失（*AMTC3L*）：AMTC3L 的功能依赖于两个子模块：边际估计器和负类中心选择器。如 [19] 中所述并在 (3) 中所示，通常，边际三元组损失包含三个需要谨慎选择的主要项：正样本、负样本和间隔。为提高类内紧凑性和类间可分性，将嵌入空间中的样本类中心视为正样本，而负样本则从其余 $k-1$ 个类别中心中选取。

$$\mathcal{L}_{TL} = \max\Big(0, \frac{1}{2m}\sum_{i=0}^{m}\sum_{j=1}^{c_d}\|(e_j^{(i)}) - (c^p_{y^{(i)}j})\|_2^2 - \|(e_j^{(i)}) - (c^n_{y^{(i)}j})\|_2^2 + \alpha_i\Big), \quad (3)$$

其中，$\sigma[e_1^{(i)}, e_2^{(i)}, \ldots, e_d^{(i)}]^T$ 表示 Sigmoid 激活函数，$e_j^{(i)} \in e^{(i)}$ 表示输入小批次中样本嵌入的元素。对应的类中心 $c^p_{y^{(i)}j} \in c^p_{y^{(i)}} = [c^p_{y^{(i)}1}, c^p_{y^{(i)}2}, \ldots, c^p_{y^{(i)}d}]$ 定义了最正嵌入，而 $[c^n_{y^{(i)}1}, c^n_{y^{(i)}2}, \ldots, c^n_{y^{(i)}d}]$ 表示最负嵌入的类中心。请注意，α_i 是来自自适应边际生成器的小批次中第 i 个样本的间隔。

当嵌入落入足够接近其类别原型、且远离最负原型的区域时，$\mathcal{L}_{TL} = 0$。因此，样本在反向传播中的贡献将为零。在训练过程中，AMG 学会以适当的方式为样本分配 α_i，从而忽略那些简单的样本

MS-NSS 算法：数学合成负样本选择

输入：小批次的嵌入张量及其对应的标签，$D_m = \{(e^{(i)}, y^{(i)}) \mid i = 0, 1, \ldots, m;\ e^{(i)} \in \mathbb{R}^{C_d \times 1 \times 1}\}$；嵌入张量的中心，$c = \{c_k \mid k = 1, 2, \ldots, K;\ c_{(k)} \in \mathbb{R}^{1 \times C_d}\}$。**输出：** $c^n = \{c_i^n \mid i = 1, 2, \ldots, m\}$。

1: **当**未收敛**时，执行**
2: 计算簇间欧氏距离：
$\eta_{kj} = \{\|\sigma((e_j^{(i)}) - \sigma((c_{ckj})\|_2^2 \mid i = 1, \ldots, m;\ j = 1, \ldots, C_d;\ \forall c_k \mid k \neq y^{(i)}) \quad k = 1、2、\ldots\ldots、K\}$。
3: $c_j^n = \{$ 最小化$_{k \in \{1,2,\ldots,C_d\}}$ $(\eta_{kj}) \mid j = 1, 2, \ldots, C_d\}$
4: 根据（4）计算三元组中心损失。
5: **结束循环**

算法 NS-NSS：网络语义负样本选择

输入：小批次的嵌入张量及其对应的标签，$D_m = \{(e^{(i)}, y^{(i)}) \mid i = 0, 1, \ldots\ldots, m;\ e^{(i)} \in \mathbb{R}^{C_d \times 1 \times 1}\}$；小批次的预测，$\hat{y}^{(i)}$；嵌入张量的中心，$c = \{c_k \mid k = 1, 2, \ldots, K;\ c_{(k)} \in \mathbb{R}^{1 \times C_d}\}$。**输出：** $c^n = \{c_i^n \mid i = 1, 2, \ldots, m\}$。1: **当**尚未收敛时**执行**2: **如果**$\hat{y}^{(i)} \neq y^{(i)}$ **则**:
3 $c^n = \{c_{\hat{y}^{(i)}} \mid i \in \{i \mid y^{(i)} \neq y^{(i)}\}\}$：^: **否则**^: 5:
4 $y^{(i)} = y^{(i)}$
$\gamma^{(i)}$ =argmax$\{v(\hat{y}^{(i)} = k) \mid k \neq y^{(i)}\}$ 这里 $v()$ 表示频率函数。6 $c^n = \{c_{\gamma^{(i)}} \mid i \in \{i \mid y^{(i)} = y^{(i)}\}\}$：^: **结束**
7
如果8：根据（4）计算三元组中心损失。9: **结束循环**

算法 MM-NSS：混合方法负样本选择

输入：小批次的嵌入张量及对应标签，$D_m = \{(e^{(i)}, y^{(i)}) \mid i = 0, 1, \ldots, m;\ e^{(i)} \in \mathbb{R}^{C_d \times 1 \times 1}\}$，嵌入张量的中心，$c = \{c_k \mid k = 1, 2, \ldots, K;\ c_{(k)} \in \mathbb{R}^{1 \times C_d}\}$。
输出： $c^n = \{c_i^n \mid i = 1, 2, \ldots, m\}$。

1: **当**未收敛**时**
2: **如果**$\hat{y}^{(i)} \neq y^{(i)}$ **则**:
3: $c^n = \{c_{\hat{y}^{(i)}} \mid i = 1, \ldots, m\}$
4: **否则**$\hat{y}^{(i)} = y^{(i)}$:
5: 计算簇间欧几里得距离：
$\eta_{kj} = \{\|\sigma((e_j^{(i)}) - \sigma((c_{ckj})\|_2^2 \mid i = 1, \ldots\ldots, m;\ j = 1, \ldots, C_d;\ \forall c_k \mid k \neq y^{(i)}) \quad k = 1, 2, \ldots, K\}$.
6: $c_j^n = \{$ 最小化$_{k \in \{1,2,\ldots,C_d\}}$ $(\eta_{kj}) \mid j = 1, 2, \ldots, C_d\}$
7: **结束 if**
8: 根据公式（4）计算三元组中心损失。
9: **结束 while**

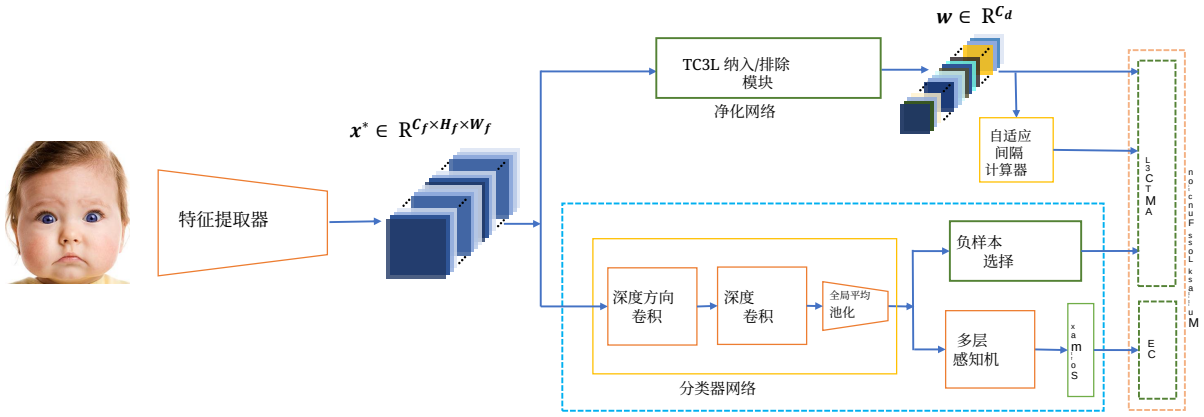


图 2：模型的详细结构。

特征，并对具有最难判别特征的样本进行难例挖掘。

为了提高簇内紧凑性，在公式（4）中，我们采用与公式（3）类似的方法，将簇中心作为最正原型，并依据 MS-NSS 算法、NS-NSS 算法和 MM-NSS 算法合成最负嵌入。这些算法分别对应数学合成、网络语义和混合方法负样本选择算法。

第一种算法完全基于欧氏距离合成最负原型；第二种算法则根据模型输出，为每个嵌入选择最负的原型。对于误分类样本，最负的类别原型恰好是错误预测对应的类中心。然而，对于正确预测，算法会查看记录，并选择模型曾经更难以应对的类别所对应的类中心。第三种算法结合了前两种算法：对于错误预测，其行为类似于 NS-NSS 算法；对于正确预测，其行为类似于 MS-NSS 算法。

具体来说，算法 MS-NSS：以所有小批次的嵌入、对应的标签和类别中心向量为输入，并返回一个合成的类别中心向量，作为最负样本的原型 $c^n_{y^{(i)}j}$。对于第 i 个嵌入的第 j 个维度，该算法计算其与所有类别中心的第 j' 个维度之间的欧氏距离，然后选取距离最小的一个。这样，每个维度上的合成最负原型都是所有类别中心向量在该维度上的投影的线性组合。注意，线性组合的系数均为零，且只有一个系数为一。

算法 NS-NSS：根据网络预测执行。它以嵌入、对应的标签和类别中心矩阵为输入。随后，它比较预测和标签，找出误分类样本，并据此创建一个新的误分类统计方阵 **S**，用于展示每个类别标签对应的误分类预测类别分布。对于误分类样本，预测类别的类中心将被视为负样本。然而，对于正确分类的样本，训练期间该小批次中每个类别的最常见分类将决定负类别。

训练期间该小批次中每个类别的分类结果将决定负类别。

对于分类正确的样本，算法 MM-NSS：的工作方式与算法 MS-NSS：类似；对于分类错误的样本，其工作方式则与算法 NS-NSS：相同。

Farzaneh 和 Qi [15] 认为，嵌入中的各个元素对判别能力的影响并不相同。受此启发，我们认为，并非所有样本的嵌入对判别能力的影响都相同。因此，AM 模块提供适当的 α_i，以调整每个样本误差值，如下所示：

$$\mathcal{L}_{AMTC3L} = \frac{1}{2m}\sum_{i=0}^{m}\Big[\sum_{j=1}^{c_d} w_j\Big(\|\sigma(e_j^{(i)}) - \sigma(c^{p}_{y_j^{(i)}})\|_2^2\Big) - w_j\Big(\|\sigma(e_j^{(i)}) - \sigma(c^{n}_{y_j^{(i)}})\|_2^2\Big)\Big] + \alpha_i, \quad (4)$$

$$\text{s.t.},\ 0 < \alpha_i \le c_d, \quad (5)$$

其中，σ 表示 Sigmoid 函数。这样，每个维度上的距离投影值都介于 0 和 1 之间

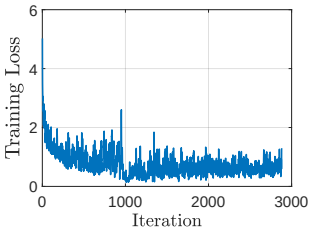
[4] $\alpha_i = \{\sum_{j=1}^{C_d} w_j \mid w_j \in$ 。因 $\mathcal{L}_{AMTC3L} > 0$ $\{0,1\}\}$

此，如果……，则可确保在所有情况下，该样本都会参与反向传播。然而，被乘以 0 的离群特征数量会导致间隔因样本而异。

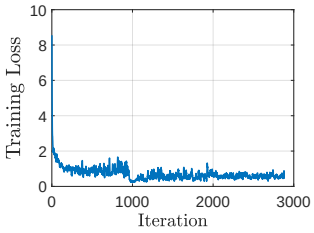
B. 分类器

为了发挥全卷积网络的优势，在分类器网络中使用平均池化层是不可避免的，$\mathcal{N}_C$。如前所述，$\mathcal{N}_C$ 的基本架构由两个深度方向 1×1 卷积层组成，它们依次缩减通道数，并映射特征空间表示。表I：AMTC3L辅助交叉熵与普通交叉熵及当前最先进方法在RAF-DB数据集上的性能比较。请注意，我们方法的结果基于五折交叉验证。

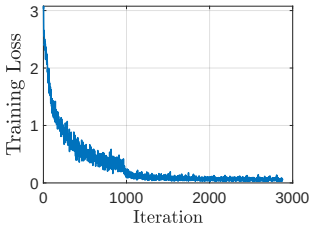
数据集	模型	Sur.	Fea.	Dis.	快乐 悲伤		Ang.	中性 平均值
RAF-DB	基线	0.85	0.60	0.57	0.94	0.82	0.74	0.86 0.7685
	MS-NSS	0.85	0.64	0.59	0.95	0.87	0.81	0.84 0.7929
	NS-NSS	0.86	0.66	0.62	**0.96**	0.82	0.78	0.86 0.7943
	MM-NSS	**0.87**	**0.66**	**0.64**	0.95	**0.89**	**0.80**	**0.87 0.8114**



(a) MS-NSS：收敛



(b) NS-NSS：收敛



(c) MM-NSS：算法收敛性

图 3：所提出的三种算法中训练损失随迭代次数的变化。

$\mathbb{R}^{C_f \times H_f \times W_f}$，映射到上下文空间，$\mathbb{R}^{C_d \times H_f \times W_f}$。随后，全局平均池化层从输入样本的上下文空间表示中生成嵌入张量 $\mathbb{R}^{C_d \times 1 \times 1}$。最后，全连接层将嵌入映射为行得分向量。

1）选择性注意力模块：通常，并非所有空间特征对分类过程的贡献都相同。如 [20] 所示，按通道堆叠后的每个特征图并不一定都能体现独特的语义特征。因此，利用注意力层的优势是合理的。受 [20–23]，启发，我们对嵌入向量、锚点和负样本向量的所有投影应用选择性纳入/排除注意力机制，以防止离群值对深度度量学习损失函数产生影响。如前所述，该模块还定义了自适应边界。

IV. 实验结果

本节首先介绍本研究使用的数据集 RAF-DB。随后，我们通过可视化比较本研究开发的三种负样本选择策略的性能。图1展示了 RAF-DB 中的一些样本。

A. RAF-DB

真实世界情感人脸数据库（RAF-DB）包含 29672 张真实世界图像，涵盖两组简单情绪和复合情绪，被认为是最丰富的面部表情数据库之一 [24]。第一组标签包含恐惧、快乐、惊讶、悲伤、愤怒、厌恶和自然等基本情绪。该数据集于2017年收集。数据集中的复杂情绪标签包括恐惧惊讶、愤怒惊讶等12种标签。基本情绪在所有图像中共有15,339个样本，其中12,271个为训练样本，3,068个为测试样本。

B. 实现细节

我们使用预训练的 ResNet-50 作为骨干网络，它在提取深层特征方面表现出色。所有提出的算法和基线均采用标准随机梯度下降（SGD）优化器进行训练，动量为 0。9，权重衰减为 5×10^{-4}。为了增强输入样本，我们采用多尺度随机裁剪、随机水平翻转、随机擦除，并根据均值和标准差进行归一化。为应用多尺度随机裁剪，在将一个小批量样本输入网络之前，我们先对其进行随机缩放，再随机裁剪出尺寸为 224×224 的图像块并输入网络。通过这种方式，模型能够学习局部高判别性特征以及全局特征。需要注意的是，恰当的数据增强对提升基线和所提技术的性能都有显著作用。我们在 RAF-DB 上训练模型 60 个训练轮次：基线的学习率为 0.01（仅用于交叉熵），三种算法的学习率均为 0.05（用于交叉熵和 AMTC3L）。学习率每经过 20 个训练轮次便衰减为原来的十分之一。在我们的研究中，超参数 λ 需要针对不同算法取不同值，以帮助模型达到最佳状态。在我们的架构中，深层特征为 $x^* \in \mathbb{R}^{2048 \times 4 \times 4}$，嵌入为 $e \in \mathbb{R}^{64 \times 1 \times 1}$。

图3展示了所提算法的收敛能力。此外，表I和表 II给出了所提算法及基线的平均值和总体准确率。根据表 II，所有所提算法在采用 AMTC3L辅助交叉熵后均取得了显著提升。然而，若采用能够排除离群值对反向传播影响的自适应模块方案，效果还可进一步改善。如图3-a所示，由数学合成的负样本导致损失下降过程中出现剧烈的噪声波动。MS-NSS 会探索类别中心，并根据其他类别中距离最近的元素，逐个构建负样本的维度。在所提出的三种方法中，这是最难的负采样场景。对于算法 NS-NSS，收敛平稳的原因在于，我们的选择策略完全基于模型在当前状态之前所呈现的性能。该方法倾向于压缩类内特征并分离类间特征，但效果不如 MS-NSS 算法。结合两种方法优势的 MS-NSS 算法呈现出平稳的波动，并且收敛速度最快。

五、结论

本文提出了一种深度度量学习注意力辅助机制，该机制利用了基于铰链的损失函数 TCL 的优势。为解决如何选择最难负样本这一挑战，我们提出了三种新方法，分别考虑数学距离最近的合成样本、统计上最近的样本，或两类样本的混合。相应地，我们发现，使用两类样本混合的方法（MM-NSS）能够实现更稳健的方案。仿真结果表明，在 RAF-DB数据集上，与基线相比，准确率最高提升了 3.12%。

表 II：消融研究

设置	MS-NSS			NS-NSS			MM-NSS		
基线	✓	✗	✗	✓	✗	✗	✓	✗	✗
流程线 a（消融）	✓	✓	✗	✓	✓	✗	✓	✓	✗
流程 b（消融）	✓	✗	✓	✓	✗	✓	✓	✗	✓
基线（准确率）	% 85.78			% 85.78			% 85.78		
流程线 a（准确率）	% 87.12			% 87.29			% 88.53		
流程 b（准确率）	**% 87.77**			**% 87.81**			**% 88.90**		

此外，结果凸显了在基于深度度量学习的无损三元组损失场景中，负采样策略的重要性。

参考文献

[1] I Michael Revina 和 WR Sam Emmanuel。人脸表情识别技术综述。《沙特国王大学学报：计算机与信息科学》，33 628，2021 6 619( ): – .[2] Niloufar Mehrabi 和 Sayed Pedram Haeri Boroujeni。利用混合特征和粒子群优化进行基于面部图像的年龄估计。发表于*2021*年第*11*届计算机工程与知识国际会议（*ICCKE*），第412–418页。IEEE，2021年。[3] Sungyeon Kim、Dongwon Kim、Minsu Cho 和 Suha Kwak。通过标签松弛进行嵌入迁移，以改进度量学习。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第3967–3976页，2021年。[4] Gábor Kertész。用于度量相似性学习中无损三元组损失的不同三元组采样技术。发表于*2021*年*IEEE*第*19*届应用机器智能与信息学世界研讨会（*SAMI*），第000449–000454页。IEEE，2021年。[5] Weicheng Xie、Haoqian Wu、Yi Tian、Mengchao Bai 和 Linlin Shen。结合多阶段离群值抑制和类对间隔的三元组损失，用于面部表情识别。*IEEE*视频技术电路与系统汇刊，32（2）：690–703，2021年。[6] Gencer Sumbul、Mahdyar Ravanbakhsh 和 Begöm Demir。用于多标签遥感图像检索的信息性和代表性三元组选择。*IEEE Transactions on Geo- science and RemoteSensing*，2022年。[7] Xinwei He、Yang Zhou、Zhichao Zhou、Song Bai 和 Xiang Bai。用于多视图三维物体检索的三元组中心损失。发表于*IEEE*计算机视觉与模式识别会议论文集，第1945–1954页，2018年。[8] Weixian Lei、Rong Zhang、Yang Yang、Ruixuan Wang 和 Wei-Shi Zheng。涉及类中心的三元组损失，用于不平衡数据上的皮肤病分类。发表于*2020*年*IEEE*第*17*届生物医学成像国际研讨会（*ISBI*），第1–5页。IEEE，2020年。[9] Ke-Kun Huang、Chuan-Xian Ren、Hui Liu、Zhao-Rong Lai、Yu-Feng Yu 和 Dao-Qing Dai。通过具有改进三元组损失的判别性卷积神经网络进行高光谱图像分类。《模式识别》，112：107744，2021年。[10] Mariana-Iuliana Georgescu 和 Radu Tudor Ionescu。通过教师-学生训练和三元组损失实现遮挡条件下的面部表情识别。发表于*2020*年第*25*届模式识别国际会议（*ICPR*），第2288–2295页。IEEE，2021年。[11] Linyi Zhou、Xijian Fan、Tardi Tjahjadi 和 Sruti Das Choudhury。用于自然场景下面部表情识别的判别性注意力增强特征学习。*Neural Computing and Applications*，34（2）：925–936，2022年。[12] Hamid Sadeghi 和 Abolghasem-A Raie。Histnet：基于直方图的卷积神经网络，结合卡方深度度量学习，用于面部表情识别。*Information Sciences*，608：472–488，2022年。[13] Weifeng Ge。采用层次三元组损失的深度度量学习。发表于欧洲计算机视觉会议论文集（*ECCV*），第269–285页，2018年。[14] Ali Pourramezan Fard 和 Mohammad H Mahoor。Ad-corre：用于自然场景下面部表情识别的自适应相关性损失。*IEEE Access*，10：26756–26768，2022年。[15] Amir Hossein Farzaneh 和 Xiaojun Qi。通过深度注意力中心损失实现自然场景下面部表情识别。发表于*IEEE/CVF*冬季计算机视觉应用会议论文集，第2402–2411页，2021年。[16] Wei Huang、Siyuan Zhang、Peng Zhang、Yufei Zha、Yuming Fang 和 Yanning Zhang。基于合成图像的深度度量学习实现身份感知面部表情识别。*IEEETransactionsonMultimedia*，2021年。[17]Yuanyuan Liu、Wei Dai、Fang Fang、Yongquan Chen、Rui Huang、Run Wang 和 Bo Wan。用于联合姿态感知与身份不变面部表情识别的动态多通道度量网络。*Information Sciences*，578 213，2021 195: – .[18] Xiaofeng Liu、BVK Vijaya Kumar、Jane You 和 Ping Jia。用于身份感知面部表情识别的自适应深度度量学习。发表于*IEEE*计算机视觉与模式识别会议研讨会论文集，第20–29页，2017年。[19] Florian Schroff、Dmitry Kalenichenko 和 James Philbin。FaceNet：用于人脸识别和聚类的统一嵌入。发表于*IEEE*计算机视觉与模式识别会议论文集，第815–823页，2015年。[20] Meng-Hao Guo、Tian-Xing Xu、Jiang-Jiang Liu、Zheng-Ning Liu、Peng-Tao Jiang、Tai-Jiang Mu、Song-Hai Zhang、Ralph R Martin、Ming-Ming Cheng 和 Shi-Min Hu。计算机视觉中的注意力机制：综述。*Computational Visual Media*，第1–38页，2022年。[21] Fatemeh Lotfi、Omid Semiari 和 Walid Saad。无线蜂窝网络中的语义感知协同深度强化学习。发表于*ICC 2022-IEEE*国际通信会议，第5256–5261页。IEEE，2022年。[22] Fatemeh Lotfi、Omid Semiari 和 Fatemeh Afghah。用于O-RAN动态切片管理的进化式深度强化学习。发表于*2022*年*IEEE Globecom*研讨会（*GC Wkshps*），第227–232页。IEEE，2022年。[23] Fanglei Xue、Qiangchang Wang 和 Guodong Guo。Transfer：利用变换器学习关系感知面部表情表征。发表于*IEEE/CVF*国际计算机视觉会议论文集，第3601–3610页，2021年。[24] Shan Li 和 Weihong Deng。通过可靠众包和深度局部性保持学习实现非受限面部表情识别。*IEEE Transactions on Image Processing*，28（1）：356–370，2019年。