arXiv:2602.01833v1 [cs.MM] 2 Feb 2026

面向鲁棒多模态情感分析的缺失模态解耦专家混合

Xiang Li[a]、Xiaoming Zhang[a,*]、Dezhuang Miao[a]、Xianfu Cheng[b]、Dawei Li[a]、Honggui Han[c]、Zhoujun Li[b]

[a]北京航空航天大学网络空间安全学院，北京，*100191*，中国[b]北京航空航天大学计算机科学与工程学院，北京，*100191*，中国[c]北京工业大学信息科学与技术学院，北京，*100124*，中国

摘要

多模态情感分析（MSA）整合多种模态以推断人类情感，但现实世界中的噪声常常会导致数据缺失或损坏。然而，现有的特征解耦方法难以应对不确定缺失情况下异质信息的内部变化，因此很难从退化的模态中学习有效的多模态表征。为了解决这一问题，我们提出了**DERL**，一种面向鲁棒多模态情感分析的**解**耦**专**家**表**征**学**习框架。具体而言，DERL采用混合专家，将多模态输入自适应地解耦到正交的私有和共享表示空间中。我们进一步设计了一种多层级重构策略，以提供协同监督，增强所学习表征的表达能力和鲁棒性。最后，解耦特征作为各司其职的模态专家，生成重要性感知融合结果。在两个MSA基准测试上进行的大量实验表明，在各种模态缺失条件下，DERL的表现优于最先进的方法。例如，在模态内缺失情况下，我们的方法在MOSI上的二分类准确率提高了2.47%，平均绝对误差降低了2.25%。

关键词： 表示学习，解耦专家，多模态情感分析，缺失模态

*通讯作者：yolixs@buaa.edu.cn

1. 引言

多模态情感分析（MSA）旨在通过整合文本、视觉和音频模态来解读人类情感 [1, 2]。尽管多模态学习迅速发展，大多数现有的多模态情感分析方法仍假设训练和推理期间所有模态都完整可用[3, 4, 5, 6, 7]。在现实场景中，由于传感器故障或背景噪声，多模态数据往往不完整，导致在完整数据条件下训练的模型性能大幅下降 [8, 9]。近期的多模态情感分析研究 [10, 11, 12, 13, 14]尝试处理缺失模态。这些方法主要侧重于学习联合多模态表征 [12, 13] ，或通过特征重建来恢复缺失语义 [10, 11]。例如，LNLN [12]将文本作为主导模态，并利用其他模态作为互补来源，从而提升模型在数据不完整情况下的鲁棒性。然而，这些方法未能充分利用每种模态中的独特信息。由于异构模态以不同方式表达情感，因此包含私有与共享情感线索 [15, 16]。近期研究探索了解耦表征学习 [3, 8]，以捕捉异构情感线索并增强多模态融合。例如，HRLF [8]通过跨模态变换和语义重构来分解情感表征。

尽管取得了进展，当前的特征解耦方法仍无法适应内部特征退化所导致的异质信息变化；这种退化发生于缺失情况不确定时，并带来三大主要挑战。❶ 模态解耦缺乏灵活性：随机缺失会降低模态质量和强度，扰乱私有与共享信息之间的内部平衡。以往的静态方法 [17, 18] 缺乏自适应解耦的灵活性，因此难以在缺失情况下分解异构表征。❷特征重建不足：仅基于退化后的解耦特征进行自回归重构 [17, 19]，约束不足。这种单层策略会放大重构误差，导致表征欠佳，且无法充分补偿缺失模态。❸ 多模态融合失衡：不同模态对融合结果的贡献并不相同 [4, 20]，，随机缺失会进一步加剧这种失衡。某些模态中关键情感信息的缺失会削弱它们在多模态融合中的作用。鲁棒的融合策略需要考虑缺失场景下不同模态的相对重要性。

为解决这些问题，我们提出一种面向缺失模态鲁棒多模态情感分析的**解**耦**专**家**表**征**学**习（**DERL**）框架。DERL 首先使用带有可学习瓶颈令牌的统一 Transformer 块对齐各模态。随后引入三个专门组件来处理不完整模态：混合专家解耦（HED）、多级协同重建（MLCR）以及模态路由与融合（MRF）。首先，针对❶，HED 通过解耦专家混合，自适应地分离模态私有和共享情感特征，从而能够在模态缺失场景下灵活捕捉异构情感线索。其次，为补偿退化，MLCR 从多个视角重建各模态，强化语义丰富度并提供全面监督，以应对❷。通过协同重建，MLCR 提升了所学习表征的质量、完整性和鲁棒性。第三，针对❸，解耦特征充当模态专家，MRF 则学习各专家的贡献，实现重要性感知融合。它强调主导模态与其他模态之间的交互，以改善融合结果。我们的工作主要有以下三方面贡献：

- DERL 提出了一种将解耦表征学习与混合专家架构相结合的新方法，用于鲁棒多模态情感分析，以应对不确定缺失下的模态信息变化。
- 我们提出 HED 模块以解耦异构情感线索，MLCR 模块以恢复多层次退化表征，以及模态路由与融合模块以实现重要性感知多模态融合。
- 在两个多模态情感分析基准测试上的大量实验表明，DERL 优于特征解耦基线方法以及其他强基线方法，展现了其在各种模态缺失场景下的鲁棒性能。

本文其余部分安排如下。第2节回顾鲁棒多模态情感分析、解耦多模态学习和专家混合架构的相关工作。第3节介绍所提出的 DERL 框架，详细阐述其解耦专家混合表示学习机制。第4节介绍实验设置，并报告模态缺失条件下的实验结果。最后，第5节总结全文并探讨未来研究的潜在方向。

2. 相关工作

2.1. 鲁棒多模态情感分析

多模态情感分析（MSA）整合来自多个模态的情感信息进行预测。现有方法大致可分为基于上下文的方法和鲁棒噪声感知方法。基于上下文的模型 [4, 21, 19, 22, 23] 假设多模态输入完整，并着重通过跨模态交互学习统一表示。例如，DMD [18] 对多模态输入进行解耦，并执行模态级蒸馏，以提升表示质量。然而，当一个或多个模态缺失或受损时，此类模型的性能往往会显著下降。

相比之下，鲁棒噪声感知多模态情感分析关注的是数据不完整或退化的现实场景。现有解决方案大致可分为基于重建的方法和联合表示学习方法。基于重建的方法 [10, 11, 9]旨在利用可用模态提供的互补线索恢复缺失的语义。例如，TFR-Net [10]采用基于Transformer的重建模块来恢复缺失的多模态特征，而 CorrKD [9] 则通过知识蒸馏增强跨模态表征。联合表示学习方法 [12, 24, 25] 捕捉模态间相关性，以构建紧密统一的共享嵌入。例如，LNLN [12] 优先利用文本线索，通过学习语言主导的联合表示来增强模型的鲁棒性。近年来，特征解耦方法 [26, 8] 被用于解耦不完整的多模态输入。然而，这些方法往往忽视了随机噪声对内部模态特征产生的不确定影响。本文提出一种自适应解耦框架，对退化模态中的异质信息进行解耦，以实现鲁棒多模态情感分析。

2.2. 解耦多模态学习

固有的跨模态异质性是多模态学习面临的核心挑战 [3, 16]。以往研究致力于解耦模态特有信息与共享信息，以获得全面表示，相关方法大致可分为基于空间的方法和基于互信息的方法。基于空间的方法 [17, 15, 19] 通过衡量特征表示之间的相似性和差异性来实现解耦。相比之下，基于互信息的方法 [3, 27] 通过优化互信息界或其时间变体来促进解耦

从而提取跨模态的独立且互补的特征。近期研究 [26, 8] 将解耦学习扩展至模态缺失场景，以实现稳健的多模态分析。然而，随机缺失模态会引入不确定性，扰乱私有与共享信息之间的比例，导致难以在不完整条件下实现自适应特征解耦。

2.3. 混合专家

混合专家（MoE）模型通过将网络组件划分为多个专家实例 [28]，实现专业化。MoE 的核心思想是利用专家之间的协作与竞争来提升模型容量。然而，同时激活所有专家会带来高昂的计算成本。稀疏混合专家 [29] 通过仅激活部分专家来提高效率，并已应用于多任务学习，以实现任务特定专家选择 [30]和模型压缩 [31]。在多模态情感分析中，混合专家架构已被用于替代稠密层以提高效率[32] ，或处理多种模态 [20, 33]。例如，EMOE [20] 将每种模态建模为一个专家，并采用基于路由的跨模态融合；而 MoMKE [33] 则通过整合单模态特征与联合特征来改进模态表示。然而，现有大多数基于 MoE 的多模态情感分析方法都假设输入完整，且未能保留用于多模态融合的模态特定线索。与以往工作不同，我们将混合专家架构融入解耦表征学习框架，从而能够根据不完整或含噪的多模态特征，自适应地分配模态私有和共享情感信息。

3. 方法

3.1. 问题定义

本研究考虑三种模态：文本（t）、视觉（v）和音频（a）。每种模态均表示为特征向量序列： $\boldsymbol{X}_t \in \mathbb{R}^{T_t \times D_t}$， $\boldsymbol{X}_v \in \mathbb{R}^{T_v \times D_v}$， $\boldsymbol{X}_a \in \mathbb{R}^{T_a \times D_a}$，其中 T_m 表示时间长度， D_m 表示模态 $m \in \{t, v, a\}$的特征维度。鲁棒多模态情感分析的目标是在不确定模态缺失或受损时，预测话语的情感极性或强度 $\boldsymbol{Y}$。根据任务设定，情感标签 $\boldsymbol{Y}$ 定义为表示情感强度的连续标量（$\boldsymbol{Y} \in \mathbb{R}$），也可以离散为 C个预定义的情感类别（$\boldsymbol{Y} \in \mathbb{R}^C$），用于分类任务。

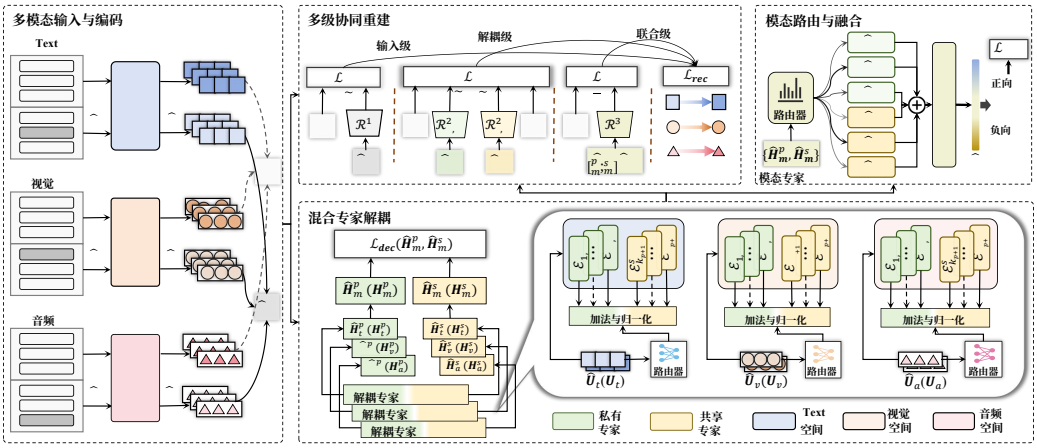


图 1：所提出的 DERL 框架概览。白色方块表示完整输入，深灰色方块表示缺失输入。经过多模态编码后，DERL 引入三个核心组件来处理不完整模态：1 混合专家解耦（HED）、2 多级协同重建（MLCR）和 3 模态路由与融合（MRF）。

3.2. 框架概览

所提出的 DERL 框架整体架构如图 1 所示。给定多模态输入，DERL 首先执行数据缺失模拟，以生成不完整样本。随后采用统一的、基于 Transformer 的编码器，确保各模态之间的预对齐。对于每种模态，我们引入混合式私有与共享解耦专家，自适应地分解模态特定和共享的情感表征。随后，重建模块协同优化原始特征与解耦特征，以提升完整性和鲁棒性。最后，重要性感知融合模块通过动态专家路由聚合解耦表征，进行情感预测。

3.3. 多模态输入与编码

随机输入缺失。为模拟真实世界中的多模态场景，我们有意破坏完整模态输入 $\boldsymbol{X}_m \in \mathbb{R}^{T_m \times D_m}$。遵循先前的研究 [12], ，文本模态中的缺失词元会被替换为 BERT [34], 中的特殊词元 [UNK] ，以模拟转录或翻译错误。对于视觉和音频模态，则用零向量替代缺失帧，以模拟传感器故障或信号退化。具体而言，对于

每个样本，我们会针对每种模态独立随机丢弃 0% 至 100% 的特征：

$$\{\hat{\boldsymbol{X}}_t, \hat{\boldsymbol{X}}_v, \hat{\boldsymbol{X}}_a\} = \text{Random-Missing}(\{\boldsymbol{X}_t, \boldsymbol{X}_v, \boldsymbol{X}_a\}), \tag{1}$$

ˆ其中， $\boldsymbol{X}_m$ 表示模态 m 的受损序列。

统一 Transformer 编码。 为统一多模态特征，我们首先通过线性投影 $\mathcal{P}$ 统一特征维度，再使用 Transformer 编码器捕获序列依赖关系。我们引入一组可学习共享瓶颈令牌 $\boldsymbol{U}_b \in \mathbb{R}^{N\times D}$ 作为预对齐的桥梁。通过自注意力，这些令牌吸收一致的线索，并将其嵌入公共空间。编码后的不完整特征计算如下：

$$\hat{\boldsymbol{U}}_m = \text{Unified-Transformer}_m([\boldsymbol{U}_b; \mathcal{P}_m(\hat{\boldsymbol{X}}_m)]), \tag{2}$$

其中， $[\cdot;\ \cdot]$ 表示拼接，ˆ $\boldsymbol{U}_m \in \mathbb{R}^{N\times D}$， D 表示模态 m 的统一表示所对应的公共特征维度。同样，对完整模态输入 $\boldsymbol{X}_m$ 进行编码，以获得用于后续处理的统一表示$\boldsymbol{U}_m = \{\boldsymbol{U}_t, \boldsymbol{U}_v, \boldsymbol{U}_a\}$ 。

3.4. 混合专家解耦

鉴于模态退化具有不确定性，每种模态内部的异质信息会以不可量化的方式发生变化。受混合专家启发，本文提出的混合专家解耦（HED）模块将输入自适应地分配给可学习的私有专家和共享专家，从而能够依据观测到的模态缺失程度，以数据驱动的方式进行解耦。

HED 保留三组混合解耦专家：

$$\{\mathcal{E}^p_{1,m}, \mathcal{E}^p_{2,m}, \ldots, \mathcal{E}^p_{k_p,m}, \mathcal{E}^s_{k_p+1}, \mathcal{E}^s_{k_p+2}, \ldots, \mathcal{E}^s_{k_p+k_s}\}.$$

私有专家 $\{\mathcal{E}^p_{k,m}\}^{k_p}_{k=1}$ 捕捉模态特定线索和分布特征，而共享专家 $\{\mathcal{E}^s_k\}^{k_p+k_s}_{k=k_p+1}$学习在各模态间保持一致的模态不变部分。每个专家均由一个两层多层感知机构成，共享专家的参数则在所有模态间共享。随后，软路由网络$\mathcal{G}^e_m$ 生成逐令牌门控分数。

对于每个词元，路由器会输出 ($k_p + k_s$) 维分数向量，并使用温度缩放 Softmax进行归一化：

$$\boldsymbol{W}_m^e = \sigma(\frac{\mathcal{G}_m^e(\hat{\boldsymbol{U}}_m)}{\tau_e}), \tag{3}$$

其中， $\sigma(\cdot)$表示 Softmax 函数， τ_e 是可学习温度，初始化为 $\tau = 1/(k_p+k_s)$，用于控制专家分配的尖锐程度。私有与共享表示通过混合专家输出获得：

$$\begin{aligned} \hat{\boldsymbol{H}}_m^p &= \sum_{k=1}^{k_p} \boldsymbol{w}_{m,k}^e \, \mathcal{E}_{k,m}^p(\hat{\boldsymbol{U}}_m), \\ \hat{\boldsymbol{H}}_m^s &= \sum_{k=k_p+1}^{k_p+k_s} \boldsymbol{w}_{m,k}^e \, \mathcal{E}_k^s(\hat{\boldsymbol{U}}_m), \end{aligned} \tag{4}$$

其中$\boldsymbol{w}_{m,\ k}^e$ 表示第 k个专家的路由权重。

对于完整输入 $\boldsymbol{U}_m$，应用相同的 HED编码器以确保特征分解一致，从而生成解耦特征 $\boldsymbol{H}_m^p$ 和 $\boldsymbol{H}_m^s$。为增强解耦表示的可判别性，我们施加软正交约束，鼓励私有与共享特征减少子空间间的冗余：

$$\mathcal{L}_{dec} = \sum_{m \in \{t,v,a\}} \cos(\hat{\boldsymbol{H}}_m^p, \hat{\boldsymbol{H}}_m^s), \tag{5}$$

其中，cos(·，·)表示余弦相似度。

3.5. 多级协同重建

为提升不完整输入下模态表示的质量，我们提出多级协同重建（MLCR）机制。MLCR 在三个层级进行特征重建：输入层、解耦层和联合层。它通过多视图监督恢复退化的语义，从而增强表示对缺失或噪声模态的鲁棒性。

输入级重建。 第一级旨在从受损输入中重建完整特征：

$$\tilde{\boldsymbol{U}}_m = \mathcal{R}_m^1(\hat{\boldsymbol{U}}_m), \tag{6}$$

其中， $\mathcal{R}_m^1(\cdot)$ 表示第一级重建网络，例如一个两层多层感知机。相应的重建损失使用 L1 范数定义：

$$\mathcal{L}_{rec}^1 = \sum_{m\in\{t,v,a\}} \|\tilde{\boldsymbol{U}}_m - \boldsymbol{U}_m\|_1. \tag{7}$$

解耦级重建。 在第二级，分别重建每个不完整模态的私有和共享解耦表示，使其与各自的完整对应项对齐：

$$\tilde{\boldsymbol{H}}_m^p = \mathcal{R}_{p,m}^2(\hat{\boldsymbol{H}}_m^p),\ \tilde{\boldsymbol{H}}_m^s = \mathcal{R}_{s,m}^2(\hat{\boldsymbol{H}}_m^s), \tag{8}$$

其中 $\mathcal{R}_{p,\ m}^2$ 和 $\mathcal{R}_{s,\ m}^2$ 是用于私有部分和共享部分的独立重建网络。相应的重建损失计算如下：

$$\mathcal{L}_{rec}^2 = \sum_{m\in\{t,v,a\}} (\|\tilde{\boldsymbol{H}}_m^p - \boldsymbol{H}_m^p\|_1 + \|\tilde{\boldsymbol{H}}_m^s - \boldsymbol{H}_m^s\|_1). \tag{9}$$

联合级重建。 最后，将缺失模态的解耦特征结合起来，共同重建编码后的完整特征：

$$\bar{\boldsymbol{U}}_m = \mathcal{R}_m^3([\hat{\boldsymbol{H}}_m^p; \hat{\boldsymbol{H}}_m^s]). \tag{10}$$

此处 $\mathcal{R}_m^3(\cdot)$表示联合级重建网络。该级别的重建损失为：

$$\mathcal{L}_{rec}^3 = \sum_{m\in\{t,v,a\}} \|\bar{\boldsymbol{U}}_m - \boldsymbol{U}_m\|_1. \tag{11}$$

总体重建损失结合了全部三个级别：

$$\mathcal{L}_{rec} = \frac{1}{3}\left(\mathcal{L}_{rec}^1 + \mathcal{L}_{rec}^2 + \mathcal{L}_{rec}^3\right). \tag{12}$$

3.6. 模态路由与融合

在模态不完整的情况下，各模态的相对重要性会变得更加多变。为此，稳健的融合策略应强调主导模态，同时利用其他模态。因此，我们提出模态路由与融合（MRF）模块，该模块能够自适应地识别主导模态，并在融合过程中为各模态分配适当的权重。

ˆ ˆ给定 $\boldsymbol{H}_m^p$ 和 $\boldsymbol{H}_m^s$，每种模态都作为一个独立的专家，使模型能够突出较强的模态，同时利用其他模态提供的互补线索。首先将特征组合起来，形成专家序列：

$$\hat{\boldsymbol{E}}_c = \{\hat{\boldsymbol{H}}_t^p, \hat{\boldsymbol{H}}_v^p, \hat{\boldsymbol{H}}_a^p, \hat{\boldsymbol{H}}_t^s, \hat{\boldsymbol{H}}_v^s, \hat{\boldsymbol{H}}_a^s\}. \tag{13}$$

接下来，双层 MLP 路由网络 $\mathcal{G}^r$ 会为所有专家计算词元级路由权重：

$$\boldsymbol{W}^r = \sigma(\frac{\mathcal{G}^r(\hat{\boldsymbol{E}}_c)}{\tau_r}) = [\boldsymbol{w}_t^p, \boldsymbol{w}_v^p, \boldsymbol{w}_a^p, \boldsymbol{w}_t^s, \boldsymbol{w}_v^s, \boldsymbol{w}_a^s], \tag{14}$$

其中，τ_r 初始化为 $1/k_e$，且 $k_e = 6$，分别对应三种模态的私有成分和共享成分。随后通过加权聚合得到增强后的多模态序列：

$$\hat{\boldsymbol{H}}_e = \sum_{m\in\{t,v,a\}} \boldsymbol{w}_m^p \hat{\boldsymbol{H}}_m^p + \sum_{m\in\{t,v,a\}} \boldsymbol{w}_m^s \hat{\boldsymbol{H}}_m^s. \tag{15}$$

这种自适应加权使模型能够突出专家序列中的主导模态。

最后，采用基于 Transformer 的融合层来捕获更深层次的跨模态交互：

$$\hat{\boldsymbol{H}}_f = \text{Fusion-Transformer}(\hat{\boldsymbol{H}}_e), \tag{16}$$

ˆ其中，$\boldsymbol{H}_f$ 表示最终的融合结果，在缺失情况不确定时对不平衡模态具有鲁棒性。

3.7. 目标优化

ˆ 对于任务学习，对融合表示 $\boldsymbol{H}_f$进行均值池化，ˆ随后将其输入多层感知机，以生成回归分数 $\boldsymbol{Y}$。训练目标是最小化预测ˆ分数 $\boldsymbol{Y}$ 与真实值 $\boldsymbol{Y}$ 之间的均方误差（MSE），批次大小为 B：

$$\mathcal{L}_{task} = \frac{1}{B}\sum_{i=1}^{B} \|\hat{\boldsymbol{Y}}_i - \boldsymbol{Y}_i\|_2^2, \tag{17}$$

整体目标由三个相辅相成的部分构成：回归任务损失 $\mathcal{L}_{task}$、解耦正则化损失 $\mathcal{L}_{dec}$和重建损失 $\mathcal{L}_{rec}$：

$$\mathcal{L} = \mathcal{L}_{task} + \mathcal{L}_{dec} + \mathcal{L}_{rec}. \tag{18}$$

4. 实验

4.1. 数据集与评估指标

数据集。 我们在两个广泛使用的多模态情感分析基准测试上进行实验：MOSI [35] 和 MOSEI [36]。**MOSI** 包含 2,199 条来自 YouTube 的话语，其情感分数按 7 分制标注，范围从 -3 （非常负向）到 $+3$ （非常正向），并划分为 1,284 个训练样本、229 个验证样本和 686 个测试样本。**MOSEI** 在 MOSI 的基础上扩展，包含来自 1,000 多位说话者、约 250 个主题的 22,856 个带标注视频片段。该数据集包含 16,326 个训练样本、1,871 个验证样本和 4,659 个测试样本，情感分数采用相同的 7 分制进行标注。有关特征提取和数据处理的更多细节，请参见附录A。

评估指标。 遵循以往研究 [11, 12], ，我们评估模型在模态内缺失条件下执行回归和分类任务时的鲁棒性。对于回归任务，我们采用平均绝对误差（MAE）和皮尔逊相关系数（Corr），分别衡量预测偏差以及与真实值的相关性。对于分类任务，我们将连续的情感分数离散化为 7、5 和 2 个类别。对于二分类，我们在两种配置下报告二分类准确率和 F1 分数：负向与非负向（“/”左侧），以及负向与正向（“/”右侧）。除 MAE 外，数值越高表示性能越好。对于模态间缺失实验，我们遵循 [14] 并使用 F1分数作为评估指标。

4.2. 实现细节

所有实验均使用 PyTorch 框架实现。模型训练采用带余弦退火学习率调度的 AdamW优化器。所有模态的特征均对齐至 128 维空间。初始学习率设为 $1e^{-4}$，批大小为 64，总训练轮数为 200。所有超参数均通过网格搜索确定。对于 MOSI 和 MOSEI，共享瓶颈长度分别设为 4 和 8。模型采用 {1, 3} 个私有专家和 {3, 3} 个共享专家，统一/融合Transformer模块的深度分别配置为 {1, 1} 和 {2, 2}。

为确保公平比较，我们遵循 LNLN [12]中的缺失模态协议。训练期间，每个 epoch 随机选取 50% 的样本，并以 0% 至 100% 的随机比例丢弃模态特征，以模拟模态缺失条件。选择在缺失率为 $r = 0.5$ 时测试的模型进行评估。

表1：在模态内缺失设置下，MOSI 和 MOSEI 上的性能比较，结果为所有缺失率下的平均值。标记为 $*$ 的结果是在相同设置下根据公开代码库复现的，其余结果来自 [12, 25]。

方法	MAE	Corr	二分类准确率	F1	五分类准确率	七分类准确率
			MOSI			
MISA	1.085	0.524	70.33 / 71.49	70.00 / 71.28	33.08	29.85
Self-MM	1.070	0.512	69.26 / 70.51	67.54 / 66.60	34.67	29.55
MMIM	1.077	0.507	67.06 / 69.14	64.04 / 66.65	33.77	31.30
TFR-Net	1.200	0.459	66.35 / 68.15	60.06 / 61.73	34.67	29.54
CENET	1.080	0.504	67.73 / 71.46	64.85 / 68.41	33.62	30.38
ALMT	1.083	0.498	68.39 / 70.40	71.80 / 72.57	33.42	30.30
DMD*	1.097	0.511	71.11 / 71.89	71.15 / 72.03	37.14	33.25
LNLN	1.062	0.503	70.11 / 71.91	70.02 / 71.71	36.25	32.53
EMOE*	1.114	0.513	70.92 / 71.65	70.69 / 71.71	36.55	32.90
DERL	**1.038**	**0.530**	**72.19**/ **73.69**	**71.95** / **73.57**	**39.48**	**34.50**
			MOSEI			
MISA	0.780	0.503	75.82 / 71.27	68.73 / 63.85	39.39	40.84
Self-MM	0.695	0.498	77.42 / 73.89	72.31 / 68.92	45.38	44.70
MMIM	0.739	0.489	75.89 / 73.32	70.32 / 68.72	41.74	40.75
TFR-Net	0.697	0.489	77.23 / 73.62	71.99 / 68.80	34.67	46.83
CENET	0.685	0.535	77.34 / 74.67	74.08 / 70.68	**47.83**	**47.18**
ALMT	0.674	0.481	77.54 / 76.64	78.03 / 77.14	41.64	40.92
DMD*	0.679	0.566	75.33 / 77.48	76.04 / 77.08	47.19	46.70
LNLN	0.692	0.530	78.19 / 76.30	**79.95**/ **77.77**	46.17	45.42
EMOE*	0.679	0.565	73.88 / 76.65	74.42 / 76.36	45.57	46.16
DERL	**0.665**	**0.585**	**78.60**/ **78.16**	77.43 / **77.82**	47.64	46.84

4.3. 基线模型

我们将 DERL 与多种具有代表性的最先进方法进行比较，包括模态完整方法和缺失模态方法。前者包括 MISA [17], Self-MM [15], MMIM [37], CENET [38],CubeMLP [39], ALMT [4], DMD [18], 和 EMOE [20]。后者的基线方法包括 MCTN [40], TransM [41], TFR-Net [10], SMIL [42], GCNet [43],CorrKD [9], 和 LNLN [12]。我们开展了全面实验，以评估 DERL 在模态内和模态间缺失条件下的鲁棒性。详细说明见附录 B。

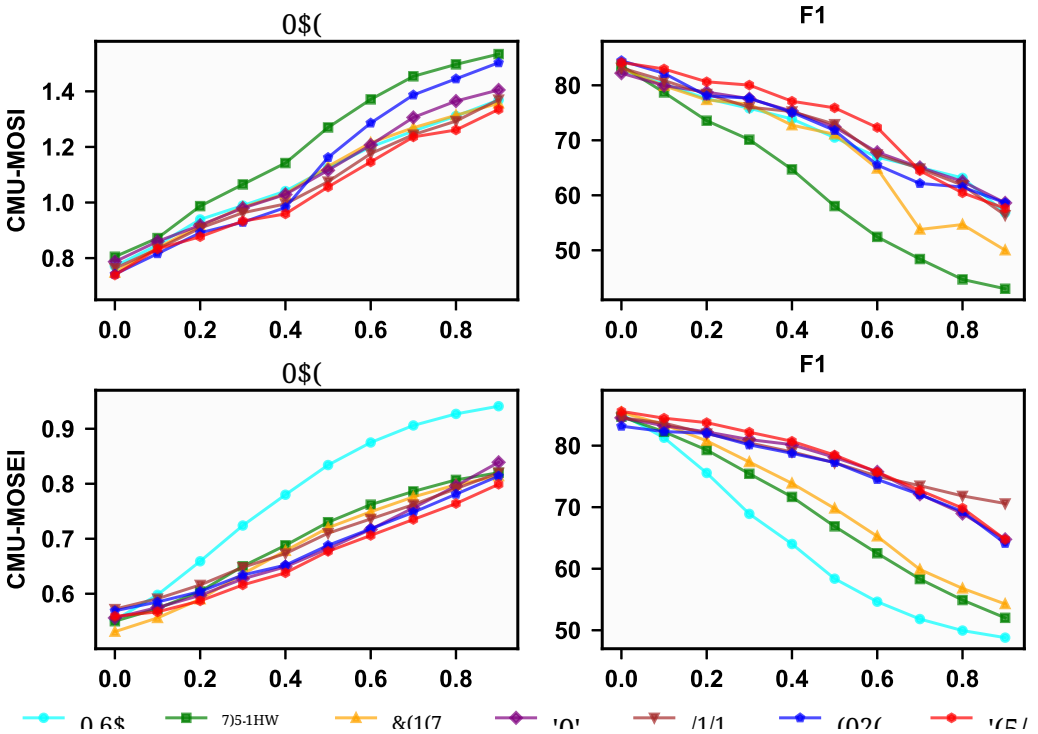


Figure 2：不同缺失率下代表性模型的性能趋势 o n MOSI 和 MOSEI 数据集。

4.4. 鲁棒性比较

模态内缺失的结果。根据 [12], ，我们通过以 $r \in [0, 0.9]$ 的速率随机屏蔽各个模态（增量为 0.1），评估 DERL 在模态内缺失条件下的表现。极端情况 $r = 1.0$（所有模态均缺失）因问题定义不适定而被排除。我们对所有缺失率下的性能取平均，以评估模型在不同噪声水平下的鲁棒性。MOSI 和 MOSEI 上的结果汇总于表1，图2则展示了不同缺失率下的性能趋势。总体而言，DERL 优于现有方法，表现出更强的噪声鲁棒性，并在大多数指标上取得了最佳的竞争性结果。这些结果揭示了三个关键观察结果：(i) 随着缺失率增加，模型性能下降，表明模态内信息的丢失削弱了多模态特征的表达能力以及联合嵌入的质量。(ii) 与 MISA、DMD 和 EMOE 等使用完整模态的方法相比，DERL 在模态内缺失情况下表现出更强的鲁棒性。这一优势

表2：MOSI 和 MOSEI 上跨模态缺失与完整模态测试条件下的性能。基线方法的结果来自 [9]，$\{m\}$ 表示可用模态。

方法	{t}	{v}	{a}	{t,v}	{t,a}	{v,a}	{t,v,a}	Avg.
				MOSI				
Self-MM	67.80	38.52	40.95	74.97	69.81	47.12	**84.64**	56.53
CubeMLP	64.15	43.24	38.91	65.12	63.76	47.92	84.57	53.85
DMD	68.97	42.26	43.33	68.45	70.51	50.47	84.50	57.33
MCTN	75.21	58.57	59.25	74.82	77.81	64.21	80.12	68.31
TransM	77.64	56.48	63.57	80.90	82.07	67.24	82.57	71.32
SMIL	78.26	59.67	67.69	79.15	79.82	71.24	82.85	72.64
GCNet	80.91	58.70	65.07	83.58	**84.73**	70.02	83.20	73.84
CorrKD	81.20	**60.72**	**66.52**	82.41	83.56	**73.74**	83.94	**74.69**
DERL	**84.18**	50.52	50.52	**84.32**	84.18	50.52	84.16	69.77
				MOSEI				
Self-MM	71.53	37.61	43.57	74.62	75.91	49.52	83.69	58.79
CubeMLP	67.52	32.58	39.54	70.06	71.69	48.54	83.17	54.99
DMD	70.26	39.84	46.18	72.45	74.78	52.70	84.78	59.37
MCTN	75.50	59.46	62.72	77.13	76.64	64.84	81.75	69.38
TransM	77.98	58.67	63.68	78.61	80.46	62.24	81.48	70.27
SMIL	76.57	60.57	65.96	76.24	77.68	66.87	80.74	70.65
GCNet	80.52	61.83	**66.54**	81.15	81.96	69.21	82.35	73.54
CorrKD	80.76	**62.30**	66.09	81.28	81.74	**71.92**	82.16	74.02
DERL	**85.19**	60.46	60.41	**85.57**	**85.16**	60.08	**85.57**	**74.63**

源于其通过混合专家实现的自适应解耦，能够有效捕捉不完整输入中的异构情感线索。(iii) 与 LNLN 等最先进的模态缺失方法相比，DERL 取得了显著提升：在 MOSI 上将二分类准确率提高了 2.47%，在 MOSEI 上提高了 2.43%；同时分别将回归误差降低了 2.25% 和 3.90%。这些结果源于 DERL 能够在不同模态退化情况下自适应地解耦并整合异质信息，同时多层次重构增强了从不完整输入中恢复关键语义的能力。

跨模态缺失实验结果。 接下来，我们使用F1分数评估DERL在跨模态缺失条件下移除一个或多个

模态时的表现。MOSI和MOSEI上的结果如表2所示，其中“平均值”表示六种缺失配置下的平均性能。我们得出以下主要观察结果：(i) 跨模态缺失通常会降低模型性能，凸显了利用跨模态线索增强融合表示的重要性。包含文本模态的模型始终优于其他组合，证实文本信息能够传递更丰富、更直接的情感线索。(ii) 与大多数基线方法相比，DERL在跨模态缺失条件下表现出更强的鲁棒性。在MOSI上，其平均F1分数比可比较的解耦方法DMD高21.6个百分点。在MOSEI上，DERL的平均F1分数比次优模型CorrKD高0.82%。(iii) 在模态缺失条件下训练的方法在鲁棒性测试中表现良好，而在完整模态上训练的模型在完整输入上表现出色；DERL则同时实现了较强的鲁棒性和具有竞争力的完整模态性能。例如，在完整模态MOSEI评估中，DERL达到了最佳报告F1分数的85.57%，凸显了基于MoE的解耦框架的有效性。(iv) 我们还观察到，在高噪声条件下，DERL在不同缺失配置下会收敛到相近的结果。这一现象可能源于惰性预测，因为噪声增加会阻碍有意义情感信息的学习与解耦。此外，仅在模态内缺失条件下训练的模型可能无法充分泛化到跨模态缺失场景。在多种缺失条件下提升模型鲁棒性，仍然是鲁棒多模态情感分析中的一项关键挑战。

表 3：MOSI 上模态内缺失情况下不同组件和策略的消融分析。

模型	MAE	Corr	二分类准确率	F1	五分类准确率	七分类准确率
不含 HED	1.065	0.518	71.14 / 72.13	70.97 / 72.07	37.39	33.62
不使用 MLCR	1.073	0.506	70.13 / 71.65	69.94 / 71.58	37.89	33.88
使用 $\mathcal{L}_{rec}^{1}$	1.061	0.510	70.99 / 71.84	71.02 / 71.78	35.95	31.80
使用 $\mathcal{L}_{rec}^{2}$	1.070	0.501	71.07 / 71.75	70.92 / 71.60	38.75	**34.78**
使用 $\mathcal{L}_{rec}^{3}$	1.055	0.523	70.41 / 72.91	70.29 / 72.11	36.72	32.30
不含 MRF	1.066	0.497	70.66 / 71.45	70.69 / 71.58	37.22	33.64
DERL	**1.038**	**0.530**	**72.19 /73.69**	**71.95 /73.57**	**39.48**	34.50

表 4：MOSI 上模态间缺失情况下不同组件和策略的消融分析。

模型	{t}	{v}	{a}	{t,v}	{t,a}	{a,v}	{t,v,a}	Avg.
不含 HED	83.82	33.87	33.00	83.82	83.82	33.29	83.82	62.20
不含 MLCR	83.34	48.16	48.16	83.19	83.34	48.16	83.34	68.24
含 $\mathcal{L}_{rec}^{1}$	84.07	37.45	37.45	84.07	84.07	37.45	84.07	64.09
含 $\mathcal{L}_{rec}^{2}$	82.15	42.26	38.14	82.47	82.28	41.61	83.04	64.56
含 $\mathcal{L}_{rec}^{3}$	83.43	42.31	42.31	83.43	83.43	42.31	83.43	65.81
不含 MRF	83.30	48.16	48.16	83.30	83.30	48.16	83.30	68.24
DERL	**84.18**	**50.52**	**50.52**	**84.32**	**84.18**	**50.52**	**84.16**	**69.77**

4.5. 消融研究

我们在 MOSI 上进行了全面的消融研究，以探究 DERL 中各个关键组件在模态内和跨模态缺失场景下的作用。消融结果汇总于表 3 和表 4，并揭示了以下几点：（i）移除 HED 模块会导致模型在各种缺失场景下的性能持续下降，这表明自适应专家解耦对于有效处理不完整模态至关重要。（ii）同样，移除 MLCR 模块会导致性能明显下降，说明特征重建对于在缺失条件下保持鲁棒性至关重要。此外，仅使用单层重建会产生次优结果，凸显了多级协同重建在改善特征恢复和语义丰富度方面的价值。（iii）移除 MRF 机制会导致性能明显下降，凸显了在模态缺失条件下采用重要性感知融合的必要性。完整的 DERL 模型取得了最佳性能，证明了所提出组件的有效性和鲁棒性。

4.6. 进一步分析

缺失率的影响。为探究预定义缺失率对模型性能的影响，我们在不同缺失率下评估模型 r。表5总结的结果展示了模型在不同程度的不完整性下的稳健性。随着缺失率增加，模型性能先提升后下降。在缺失程度较低时，不完整输入与完整输入之间的差距

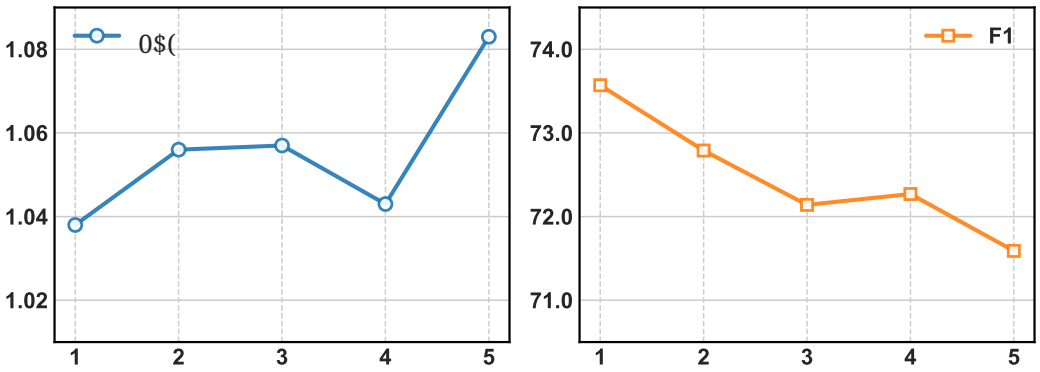
表5：MOSI 上不同预设缺失率下的结果。

方法	MAE	Corr	二分类准确率	F1	五分类准确率	七分类准确率
$r = 0.1$	1.099	0.481	70.89/72.35	70.73/72.14	38.80	34.23
$r = 0.3$	**1.038**	**0.530**	71.63/73.40	71.17/73.08	38.80	34.23
$r = 0.5$	**1.038**	**0.530**	**72.19**/**73.69**	**71.95**/**73.57**	**39.48**	34.50
$r = 0.7$	1.054	0.511	71.52/72.79	71.44/72.38	**39.48**	**34.78**
$r = 0.9$	1.054	0.512	70.47/73.14	70.51/72.98	36.49	33.78

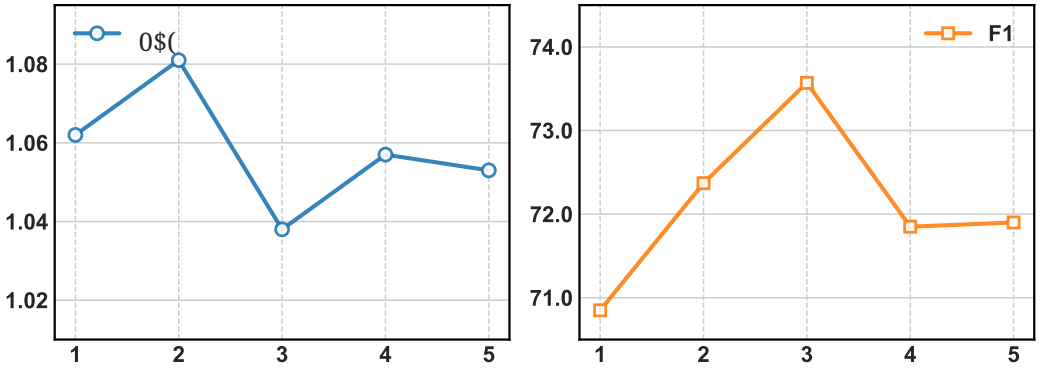
较小，这限制了模型从缺失模态中学习信息丰富的表示的能力。相反，过高的缺失率会严重降低输入质量，使模型难以解耦并重建有意义的情感线索。适中的缺失率能够实现最佳权衡，促使模型学习更稳健、均衡的跨模态表示，同时避免模型性能过度下降。

专家数量敏感性。我们进行敏感性分析，以考察混合专家的数量如何影响模型性能。私有专家和共享专家的数量分别在 1 至 5 之间变化，并在模态内缺失条件下使用 MOSI 数据集评估模型。图3展示了私有专家和共享专家数量变化对关键指标（包括平均绝对误差和 F1 分数）的影响。我们发现，当私有专家数量超过 4 个时，性能会明显下降，这可能是过拟合所致。增加共享专家的数量可以增强跨模态相关性学习，并提升模型在模态缺失场景下的鲁棒性。然而，过多的共享专家会造成信息冗余，降低鲁棒性。总而言之，专家数量过少会限制模型解耦异构表征的能力，而数量过多则会增加过拟合风险和计算负担。在 MOSI 数据集上，当私有专家和共享专家的数量适中且搭配均衡时（例如，一个私有专家和三个共享专家），模型性能最佳，能够有效捕获模态特定信息和共享语义信息。

混淆矩阵可视化。 为进一步评估所提出的DERL 框架的有效性，我们在 MOSI 数据集上将其与特征解耦基线 DMD 进行比较。在不同的预设缺失率（$r = 0.1$，0。5，0。9）下，我们对两个模型的混淆矩阵进行了可视化



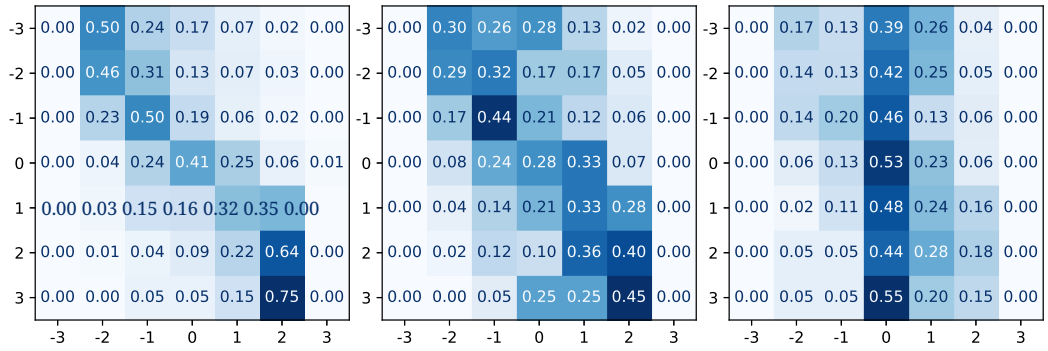
(a) 私有专家的影响



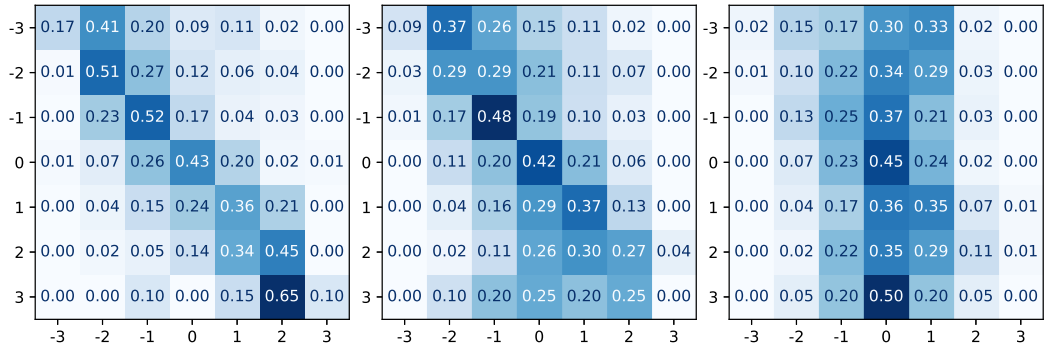
(b) 共享专家的影响

图3：私有专家和共享专家数量对 MOSI 模型性能的敏感性分析。

在具有挑战性的七分类情感分类任务中。如图4所示，随着缺失率增加，两个模型的性能均有所下降，而二者之间的差异也变得更加明显。在较低的缺失率（0。1）下，两个模型的表现相近，但 DERL 呈现出更清晰的对角线模式和更强的类别可分性。当缺失率增加到 0。5 时，DERL 仍保持稳定优势。在极端设置（0。9）下，对比最为明显：DMD 表现出一种怠惰现象，大多数预测集中在中性类别附近。相比之下，DERL 尽管受到缺失信息的轻微影响，仍保持了更加均衡的预测分布，并保留了对极端情感类别（−3和 3）的敏感性。除具备更强的鲁棒性外，DERL 相较于 D MD 还实现了具有竞争力的效率（见表6），能够在不确定的缺失模态下有效捕捉与情感相关的信息。



(a) DMD 的七分类混淆矩阵



(b) DERL 的七分类混淆矩阵

图4：MOSI 上 DERL 和 DMD 在不同缺失率下的七分类混淆矩阵。从左到右的结果分别对应 $r = 0.1$、 $r = 0.5$ 和 $r = 0.9$。

表6：MOSI 上模态内缺失条件下的效率比较。括号中的数值表示不包括文本 BERT 编码器在内的总参数量。

模型	参数量（M）	时间/轮次（秒）	MAE	F1	七分类准确率
TFR-Net	124.33 (14.85)	11.53	1.200	60.06 / 61.73	29.54
DMD	122.06 (12.58)	35.97	1.097	71.15 / 72.03	33.25
LNLN	115.97 (6.48)	11.11	1.062	70.02 / 71.71	32.53
DERL	**112.38 (2.89)**	**7.29**	**1.038**	**71.95 / 73.57**	**34.50**

4.7. 参数效率分析

表6从模型规模、训练效率和性能方面比较了 DERL 与具有代表性的鲁棒多模态情感分析方法在 MOSI 上的表现。括号中的数值表示移除文本 BERT 编码器后的参数量，以便公平比较多模态建模复杂度。

DERL 在完整设置和无 BERT 设置下始终保持较小的参数规模。与 TFR-Net 和 LNLN 相比，DERL 通过结构解耦专家建模减少了冗余的跨模态交互。相较于 DMD，DERL 进一步压缩了模态交互参数，在不使用文本编码器时仅保留 2.89M 个核心参数。在计算效率方面，DERL 在所有对比方法中训练速度最快，每轮训练时间为 7.29 秒，同时在七分类准确率、F1 分数和平均绝对误差等关键指标上仍取得最佳或具有竞争力的表现。这表明，DERL 通过紧凑且原理清晰的设计，有效兼顾了效率与性能。

5. 结论

本文提出一种新颖的解耦专家表示学习（DERL）框架，用于鲁棒多模态情感分析，能够在不确定缺失条件下处理异构信息变化。具体而言，我们提出一种混合专家解耦策略，自适应地从不完整输入中解耦私有与共享表示。此外，引入多级特征重建机制，通过协同监督恢复缺失语义线索。最后，模态路由与融合方案强调主导模态，同时利用其他模态，以实现重要性感知多模态融合。在多种缺失条件下的大量实验证明了所提方法的有效性。DERL的一个局限在于其关注预定义的缺失情形，将其扩展到现实场景中更复杂的缺失模式仍具有挑战性。

致谢

本工作得到国家自然科学基金委员会（编号62272025和U22B2021）以及软件开发环境国家重点实验室基金的资助。

附录A. 特征提取

MOSI 和 MOSEI 的详细统计数据见表 A.7。特征提取工具包的详情如下。

文本模态。 对于 MOSI 和 MOSEI，使用 bert-base-uncased 模型对原始话语进行编码 [34]。生成的文本特征为 768 维，两个数据集的输入序列长度均设为 50。

音频模态。 通过 COVAREP [44],提取低级声学特征，包括基频、梅尔频率倒谱系数等。MOSI 的音频特征维度为 5，MOSEI 为 74，序列长度分别为 375 和 500。

视觉模态。 使用 Facet系统 [45]获取面部特征，包括面部动作单元和头部姿态。MOSI 的特征维度为 20，MOSEI 为 35，两个数据集的序列长度均为 500。

表 A.7：对于每个数据划分，三个数字分别对应负向（<0）、中性（=0）和正向（>0）样本的数量。

数据集	#训练集	#验证集	#测试集	#总计
MOSI	552/53/679	92/13/124	379/30/277	2199
MOSEI	4738/3540/8048	506/433/932	1350/1025/2284	22856

附录 B. 基线模型

我们将模型与几种具有代表性的先进方法进行比较，包括完整模态方法和缺失模态方法。

MISA [17] 将每种模态分解为共享和私有潜在子空间，从而有效地对跨模态共性和模态特定属性进行建模。

Self-MM [15]采用自监督目标进行多模态和单模态学习，在保留模态独特性的同时捕捉模态间的共享模式。

MMIM [37] 最大化模态间的互信息，以学习具有相关性和判别性的表示，从而增强融合效果。

CENet [38] 在以语言为中心的框架中整合视觉和声学线索，从而强化文本嵌入。

CubeMLP [39] 通过基于多层感知机的交互模块，实现高效的多模态通信与融合。

ALMT [4] 以文本线索引导融合过程，抑制来自视觉和声学模态的无关或冲突信号。

DMD [18] 引入解耦多模态蒸馏，以实现灵活且自适应的跨模态知识迁移，同时增强单模态特征。

EMOE [20] 通过混合专家机制动态调整各模态的贡献，并通过知识蒸馏保持单模态可区分性。

MCTN [40] 通过模态间变换学习稳健的联合嵌入，并利用循环一致性损失保留跨模态信息。

TransM [41] 利用基于 Transformer 的模态间变换，对多种模态中的情感信息进行联合编码。

TFR-Net [10]采用由 Transformer 驱动的融合与重建范式，以提升不完整模态设置下的情感预测能力。

SMIL [42] 运用贝叶斯元学习扰动潜在空间，使单模态嵌入在模态缺失的情况下能够近似完整模态的表示。

GCNet [43]通过双图神经模块捕捉对话中的说话人层面和时间依赖关系，并联合优化分类与重建。

CorrKD [9] 基于跨模态相关性和知识蒸馏学习可靠的多模态表征，从而减轻不确定的缺失模态带来的影响。

LNLN [12] 将语言视为主导模态，并通过主导性校正和多模态对齐，增强其在缺失场景下的鲁棒性。

参考文献

[1] L. Zhu, Z. Zhu, C. Zhang, Y. Xu, X. Kong，基于融合方法的多模态情感分析：综述，Information Fusion 95 (2023) 306–325。[2] L. Wang, J. Peng, C. Zheng, T. Zhao, L. Zhu，一种基于多任务学习的跨模态层次融合多模态情感分析方法，Information Processing & Management 61 (3) (2024) 103675。[3] D. Yang, S. Huang, H. Kuang, Y. Du, L. Zhang，多模态情绪识别的解耦表示学习，载于：

第30届ACM国际多媒体会议，2022，第1642–1651页。[4] H. Zhang, Y. Wang, G. Yin, K. Liu, Y. Liu, T. Yu，学习语言引导的自适应超模态表示以进行多模态情感分析，载于：2023年自然语言处理实证方法会议论文集，2023，第756–767页。[5] X. Li, H. Zhang, Z. Dong, X. Cheng, Y. Liu, X. Zhang，通过词元级对齐学习用于多模态情感分析的细粒度表示，Expert Systems with Applications 269 (2025) 126274。[6] X. Zhao, X. Li, R. Jiang, B. Tang，用于多模态情感分析的解耦跨属性相关网络，Information Fusion 117 (2025) 102897。[7] C. Wang, F. Ji, B. Liu，Tpin：基于文本的并行交互网络，结合模态共有与模态特定特征进行多模态情感分析，Information Fusion (2025) 104087。[8] M. Li, D. Yang, Y. Liu, S. Wang, J. Chen, S. Wang, J. Wei, Y. Jiang, Q. Xu, X. Hou 等，基于层次表示学习实现鲁棒的不完整多模态情感分析，Advances in Neural Information Processing Systems 37 (2024) 28515–28536。[9] M. Li, D. Yang, X. Zhao, S. Wang, Y. Wang, K. Yang, M. Sun, D. Kou, Z. Qian, L. Zhang，面向不完整模态多模态情感分析的相关性解耦知识蒸馏，载于：IEEE/CVF计算机视觉与模式识别会议论文集，2024，第12458–12468页。[10] Z. Yuan, W. Li, H. Xu, W. Yu，面向鲁棒多模态情感分析的基于Transformer的特征重建网络，载于：第29届ACM国际多媒体会议，2021，第4400–4407页。[11] L. Sun, Z. Lian, B. Liu, J. Tao，高效多模态Transformer与双级特征恢复，用于鲁棒多模态情感分析，IEEE Transactions on Affective Computing 15 (1) (2023) 309–325。

[12] H. Zhang, W. Wang, T. Yu，面向不完整数据的鲁棒多模态情感分析，《神经信息处理系统进展》第37卷（2024），55943–55974。[13]Z. Guo, T. Jin, Z. Zhao，面向情感分析和情绪识别的缺失模态多模态提示学习，载于：计算语言学协会第62届年会论文集（第1卷：长篇论文），2024，第1726–1736页。[14] A. Zhu, M. Hu, X. Wang, J. Yang, Y. Tang, N. An，代理驱动的基于不完整数据的鲁棒多模态情感分析，载于：计算语言学协会第63届年会论文集（第1卷：长篇论文），2025，第22123–22138页。[15] W. Yu, H. Xu, Z. Yuan, J. Wu，利用自监督多任务学习，为多模态情感分析学习模态特定表示，载于：AAAI人工智能会议论文集，第35卷，2021，第10790–10797页。[16]Y. Zeng, W. Yan, S. Mai, H. Hu，用于多模态情感分析的解耦翻译网络，《Information Fusion》102（2024），102031。[17] D. Hazarika, R. Zimmermann, S. Poria，MISA：用于多模态情感分析的模态不变与模态特定表示，载于：第28届ACM国际多媒体会议论文集，2020，第1122–1131页。[18] Y. Li, Y. Wang, Z. Cui，用于情绪识别的解耦多模态蒸馏，载于：IEEE/CVF计算机视觉与模式识别会议论文集，2023，第6631–6640页。[19] P. Wang, Q. Zhou, Y. Wu, T. Chen, J. Hu，Dlf：以语言解耦为核心的多模态情感分析，载于：AAAI人工智能会议论文集，第39卷，2025，第21180–21188页。[20] Y. Fang, W. Huang, G. Wan, K. Su, M. Ye, Emoe：模态特定增强动态情绪专家，载于：计算机视觉与模式识别会议论文集，2025，第14314–14324页。

[21] C. Zhu, M. Chen, S. Zhang, C. Sun, H. Liang, Y. Liu, J. Chen, Skeafn：用于多模态情感分析的情感知识增强注意力融合网络，《Information Fusion》100（2023），101958。[22] X. Li, Z. Dong, X. Cheng, D. Miao, H. Zhang, T. Wang, X. Zhang, Z. Li，用于多模态情感分析的多尺度表示与多层次决策学习网络，《Expert Systems with Applications》（2025），129341。[23] L. Zhu, H. Zhao, Z. Zhu, C. Zhang, X. Kong，结合单模态标签生成与模态分解的多模态情感分析，《Information Fusion》116（2025），102787。[24] Z. Yang, Q. He, M. Yu, N. Du, Y. Lu，Tctr：针对多模态情感分析中缺失模态的文本引导对比学习与词元级重建网络，《Information Fusion》126（2026），103571。[25] X. Li, X. Cheng, D. Miao, X. Zhang, Z. Li，Tf-mamba：面向鲁棒多模态情感分析的缺失模态文本增强融合 Mamba，arXiv预印本 arXiv:2505.14329（2025）。[26] H. Wang, Y. Chen, C. Ma, J. Avery, L. Hull, G. Carneiro，利用共享-特定特征建模的缺失模态多模态学习，载于：IEEE/CVF计算机视觉与模式识别会议论文集，2023，第15878–15887页。[27] H. Sun, Z. Niu, H. Wang, X. Yu, J. Liu, Y.-W. Chen, L. Lin，基于互信息的解耦表示学习多模态情感分析，《IEEE Transactions on Affective Computing》（2025）。[28] M. I. Jordan, R. A. Jacobs，分层专家混合与EM算法，《Neural computation》6（2）（1994），181–214。[29]N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, J. Dean，规模惊人的大型神经网络：稀疏门控专家混合层，arXiv预印本 arXiv:1701.06538（2017）。[30] T. Chen, X. Chen, X. Du, A. Rashwan, F. Yang, H. Chen, Z. Wang, Y. Li, Adamv-moe：自适应多任务视觉专家混合，载于：IEEE/CVF国际计算机视觉会议论文集，2023，第17346–17357页。

[31] P. Li, Z. Zhang, P. Yadav, Y.-L. Sung, Y. Cheng, M. Bansal, T. Chen，先合并，再压缩：借鉴其路由策略，揭开高效稀疏混合专家的奥秘，arXiv预印本 arXiv:2310.01334（2023）。[32] S. Zhang, J. Zhang, Z. Zhang, L. Li，用于情感分析和情绪识别的多模态低秩专家混合，arXiv预印本 arXiv:2505.14143（2025）。[33] W. Xu, H. Jiang, X. Liang，利用模态专家知识进行不完整多模态学习，载于：第32届ACM国际多媒体会议论文集，2024年，第438–446页。[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova，BERT：用于语言理解的深度双向Transformer预训练，载于：2019年计算语言学协会北美分会会议：人类语言技术论文集，第1卷（长篇与短篇论文），2019年，第4171–4186页。[35] A. Zadeh, R. Zellers, E. Pincus, L.-P. Morency，视频中的多模态情感强度分析：面部动作与言语信息，《IEEE Intelligent Systems》31（6）（2016）82–88。[36] A. B. Zadeh, P. P. Liang, S. Poria, E. Cambria, L.-P. Morency，真实场景中的多模态语言分析：CMU-MOSEI数据集与可解释的动态融合图，载于：第56届计算语言学协会年会论文集（第1卷：长篇论文），2018年，第2236–2246页。[37] W. Han, H. Chen, S. Poria，利用层次化互信息最大化改进多模态情感分析中的多模态融合，载于：2021年自然语言处理实证方法会议论文集，2021年，第9180–9192页。[38] D. Wang, S. Liu, Q. Wang, Y. Tian, L. He, X. Gao，用于多模态情感分析的跨模态增强网络，《IEEE Transactions on Multimedia》25（2022）4909–4921。[39] H. Sun, H. Wang, J. Liu, Y.-W. Chen, L. Lin, CubeMLP：用于多模态情感分析和抑郁程度估计的基于多层感知机的模型，载于：第30届ACM国际多媒体会议论文集，2022年，第3722–3729页。

[40] H. Pham, P. P. Liang, T. Manzini, L.-P. Morency, B. Póczos，译中发现：通过模态间的循环转换学习鲁棒的联合表征，载于：AAAI人工智能会议论文集，第33卷，2019年，第6892–6899页。[41] Z. Wang, Z. Wan, X. Wan，跨模态：一种结合Transformer的端到端融合方法，用于多模态情感分析，载于：2020年网络会议论文集，2020年，第2514–2520页。[42] M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, X. Peng, SMIL：在模态严重缺失情况下的多模态学习，载于：AAAI人工智能会议论文集，第35卷，2021年，第2302–2310页。[43] Z. Lian, L. Chen, L. Sun, B. Liu, J. Tao，GCNet：用于对话中不完整多模态学习的图补全网络，《IEEE Transactions on Pattern Analysis and Machine Intelligence》45（7）（2023）8419–8432。[44] G. Degottex, J. Kane, T. Drugman, T. Raitio, S. Scherer，COVAREP——面向语音技术的协作式语音分析资源库，载于：2014年IEEE声学、语音与信号处理国际会议（ICASSP），IEEE，2014年，第960–964页。[45] iMotions，面部表情分析（2017）。