摘要

1. 引言

2. 相关工作

现有对话情感识别（ERC）研究的核心在于如何有效建模对话中的上下文依赖与说话人交互结构，大体可归纳为序列建模与图结构建模两条技术路线。序列建模方法早期以 RNN 为主，近期则更多采用 Transformer 架构，凭借其全局注意力机制在长距离上下文捕捉上展现出显著优势。相比之下，图模型则侧重于显式刻画对话的结构化特征，例如 DialogueGCN 首次将对话建模为图结构以捕捉说话人依赖。为了进一步细化交互结构，DSIG 引入了篇章结构与说话人引导机制，通过显式建模话语间的篇章关系来增强语义表示。然而，无论是序列模型还是复杂的图结构模型，虽然在“语义理解”与“结构传递”方面取得了显著进展，却往往忽略了语用层面的功能性约束：对话中的情感表达往往不完全取决于字面语义，而是与说话人的交际意图（如反讽、命令、安抚等）紧密相关。仅依赖上下文结构的隐式传播，难以在缺乏意图指引的情况下精准捕捉这些深层的情感线索。

鉴于情感表达与交际意图之间的强相关性，越来越多的研究开始探索情感识别与对话行为（或意图）识别的联合建模，期望通过任务间的互补性提升模型鲁棒性。除了早期的硬参数共享外，近期涌现了多种复杂的联合框架：MpGLIF 提出了一种多视角的全局-局部交互框架，试图同时捕捉粗粒度的全局上下文与细粒度的局部依赖；DARER 则侧重于利用时序关系递归推理网络来强化联合预测中的依赖建模。此外，TSCL 探索了另一种范式，通过两阶段学习（Two-stage Learning）与集成策略来分步优化两个任务。尽管这些工作验证了联合建模的有效性，但仍存在局限：基于交互的方法（如 MpGLIF、Co-GAT）多采用“对称式”交互，未能明确体现意图作为辅助线索对情感识别的“定向引导”作用；而基于两阶段或级联式的方法（如 TSCL）由于存在先后依赖，容易面临误差传播问题，即当意图识别不准时，噪声会直接影响情感分类的性能。

针对上述联合建模中存在的交互方向不明与噪声干扰问题，设计高效的特征

融合与去噪机制显得尤为关键。近期已有研究开始关注这一问题，例如 Liu 等人（2024，Generative Dialogue Sentiment and Act Recognition with Feature Denoising and Set Prediction）提出了一种基于生成式框架的联合识别方法，并通过特征去噪（Feature Denoising）与集合预测来缓解任务间的干扰。然而，在判别式框架下，如何实现任务间的精准对齐仍是难点。交叉注意力（Cross-Attention）机制提供了一种可行的思路，它能够以主任务特征为查询，从辅助任务中检索相关的键值，从而将任务交互从简单的“共享”提升为具有明确方向性的“语义对齐”。进一步地，为了解决辅助信息可能带来的噪声（即负迁移），门控机制（Gating Mechanism）被证明能有效根据上下文动态调节信息流的注入强度。因此，结合交叉注意力的定向检索能力与门控机制的自适应调节能力，构建一种显式的、意图引导的情感识别框架，能够在充分利用语用线索的同时有效抑制噪声干扰，这为在判别式框架下实现更稳健的任务联合建模提供了一个有效的视角。

3. 提出的方法

我们提出了一种基于意图引导的对话情感识别模型。该模型的整体架构如图 1 所示。主要包括三个核心模块：上下文语义编码模块，意图引导的情感识别模块和输出模块。给定一段包含 N 句话语的对话 $D=\{u_1,u_2,...,u_n\}$ 其中每句话语 u_i 由说话人 s_i 发出，模型首先利用预训练语言模型提取每句话语的语义特征；然后通过图神经网络捕捉对话中的上下文依赖关系和说话人交互模式；接着引入意图信息作为辅助信号，通过交叉注意力机制引导情感特征的学习；最终输出每句话语的情感类别预测。

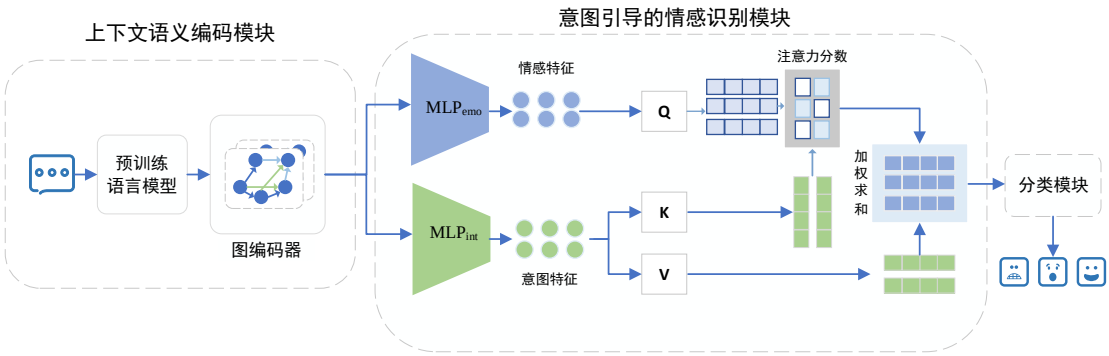

图 1 模型总体架构

3.1 特征提取

对话情感识别的首要任务是将文本话语转换为稠密的语义表示。我们采用预训练语言模型 RoBERTa-Large（下文简称 RoBERTa）作为特征提取器。在预训练特征提取阶段，首先使用分词器将对话中的每句话语u_i转换为子词序列：

$$\text{Token}(u_i) = [CLS, w_1, w_2, ..., w_m, SEP] \tag{1}$$

其中，$[CLS]$和$[SEP]$是特殊标记，$w_1, ..., w_m$为话语的子词表示。将该序列输入 RoBERTa 模型，提取$[CLS]$位置的隐藏状态作为整句话语的语义表示：

$$x_i = RoBERTa(Tokens(u_i))[CLS] \tag{2}$$

其中，x_i为话语u_i的初始特征向量。在特征投影阶段，为了降低计算复杂度并适配后续图神经网络模块，我们通过全连接层将高维特征投影到较低维度的隐藏空间:

$$h_i^o = PRELU(W_o x_i + b_o) \tag{3}$$

其中，W_o, b_o为可学习参数，d_h为隐藏维度，$PRELU$为参数化修正线性单元激活函数。投影后的特征h_i^o既保留了原始语义信息，又具有更紧凑的表示形式。对于包含N句话语的对话，可得到初始特征矩阵$H^o = [h_1^o, h_2^o, ..., h_n^o]^T \in \mathbb{R}^{N \times d_h}$，作为图神经网络模块的输入。我们使用 RoBERTa 对每个话语进行离线特征提取，取[CLS]向量作为输入特征x_i，情感模型训练阶段不更新参数。

3.2 基于说话人信息的图神经网络

对话是一个动态的交互过程，话语之间存在复杂的依赖关系。为了有效建模这种结构化信息，我们将对话建模为图结构，并采用图注意力网络（GAT）进行特征传播和聚合。在对话图构建阶段，为刻画对话中话语间的结构依赖关系，我们将长度为N的对话建模为有向图$G = (V, E)$，其中$N = \{1, ..., N\}$表示话语节点。为保证第i个话语的表征仅利用其历史上下文，我们采用因果连接方式构建边集：

$$N(i)=\{j \mid 1 \le j \le i\} \tag{4}$$

即仅允许从历史节点 j 指向当前节点 i 的连接，由此得到下三角邻接矩阵 $A \in \mathbb{R}^{N\times N}$，其中，当且仅当 $j \in N(i)$ 时 $A_{ij}=1$，否则为 0。该设置用于统一信息传播方向，并在后续实验中保持一致。另外，为显式编码说话人身份差异，定义说话人掩码矩阵 $M \in \mathbb{R}^{N\times N}$：

$$M_{ij}=\begin{cases}1, s_i = s_j \\ 0, s_i \ne s_j\end{cases} \tag{5}$$

其中，s_i, s_j 代表说话人，该掩码用于区分同/异说话人交互，在后续图注意力聚合中对两类邻居采用不同的特征变换，以建模“自我延续”和“对话交互”的差异。对于说话人感知的图注意力网络的设计上，标准的图注意力网络对所有邻居节点采用相同的特征变换，可能会导致忽略说话人身份这一重要信息。这里我们采用 DialogueGCN 的说话人感知图注意力聚合对邻居信息进行加权汇聚。对每个节点 i，先在其邻居集合 $N(i)$ 上计算注意力权重 α_{ij}，再根据说话人掩码 M_{ij} 对同/异说话人邻居采用不同的线性变换，最终得到更新表示：

$$h_i' = \sum_{j\in N(i)} \alpha_{ij} W_{M_{ij}} h_j + W_{self} h_i \tag{6}$$

由于图神经网络能够有效捕捉对话的结构化信息，但其聚合过程可能引入噪声或削弱原始语义。为融合语义特征 x_i 和图聚合特征 m_i，我们借鉴 Gan 等人提出的 denoiseGNN，引入双通道门控融合单元，对两路信息进行自适应加权并输出融合表示。该单元分别从 $[h, x_i]$ 与 $[h, m_i]$ 计算输入门与候选更新，通过两路输入门对候选更新进行加权融合，得到融合细胞 C_i，最终拼接各位置 C_i 得到融合表示 H_{lstm}：

原始特征通道：

$$i_i^{\,x} = \sigma(W^{x}{}_{i} x_i + b_i^{\,x}) \tag{7}$$

$$u_i^{\,x} = \tanh(W_u^{\,x} x_i + b_u^{\,x}) \tag{8}$$

图特征通道：

$$i_i^m = \sigma(W_i^m m_i + b_i^m) \tag{9}$$

$$u_i^m = \tanh(W_u^m m_i + b_u^m) \tag{10}$$

联合更新阶段：

$$C_i = i_i^x \odot u_i^x + i_i^m \odot u_i^m \tag{11}$$

$$h_{lstm} = [C_1, C_2, ..., C_N] \in \mathbb{R}^{N \times d_h} \tag{12}$$

其中，i_i^x, i_i^m 为两路输入门；u_i^x, u_i^m 代表候选更新；C_i 代表融合细胞更新；$\odot$ 为逐元素乘；H_{lstm} 代表融合表示输出。

3.3 意图引导的情感识别

情感表达往往与说话人的交际意图密切相关。例如，“询问”意图常伴随疑惑或好奇的情感，而“命令”意图则可能表现为愤怒或坚定。然而，现有方法大多忽略了意图信息对情感识别的指导作用。

我们将对话意图识别作为辅助任务引入模型，并构建了一种意图引导的情感识别框架：首先通过任务特定投影获得独立的情感/意图表征，在标准交叉注意力机制上，将情感作为 Query、意图作为 Key/Value 的信息流，以实现意图线索对情感表征的定向增强；随后采用门控融合对意图注入强度进行自适应调节。该框架主要包括三个子模块：任务特定投影、交叉注意力机制和自适应融合策略。首先是任务特定投影模块。为了保持情感和意图特征的独立性，避免早期融合导致的特征纠缠，我们采用了两个并行的多层感知机（MLP）分支，分别从融合特征 H_{lstm} 中提取任务特定的表示：

$$H_{emo} = MLP_{emo}(H_{lstm}) \tag{13}$$

$$MLP_{emo} = P\mathrm{Re}LU(W^K{}_{emo}(P\mathrm{Re}LU(W^1{}_{emo}))) \tag{14}$$

$$H_{\mathrm{int}} = MLP_{\mathrm{int}}(H_{lstm}) \tag{15}$$

$$MLP_{\mathrm{int}} = P\mathrm{Re}LU(W^K{}_{\mathrm{int}}(P\mathrm{Re}LU(W^1{}_{\mathrm{int}}))) \tag{16}$$

其中，MLP_{int} 具有与 MLP_{emo} 相同的层数和维度，但参数完全独立。这种双分

支设计可以使每个分支专注于提取各自任务的判别性特征，而参数独立的设置避免了两个任务在特征层面的相互干扰，还为后续的交叉注意力机制提供清晰的语义边界。虽然意图识别是辅助任务，但其特征提取过程与情感分支完全对等，这确保了意图信息的充分表达能力。在提取了独立的情感和意图特征后，关键问题是如何让意图信息有效地引导情感识别。

在获得了独立的情感表征 H_{emo} 和意图表征 H_{int} 后，我们进一步基于标准交叉注意力构建了一种意图引导的特征交互方式。我们将情感表征作为查询（Query），将意图表征作为键（Key）和值（Value），使模型能够在意图空间中检索与当前情感判别最相关的意图线索，并将其聚合为意图上下文表示，从而增强情感表征。查询-键-值的线性映射定义如下：

$$Q = W_Q H_{emo} \tag{17}$$

$$K = W_K H_{\text{int}} \tag{18}$$

$$V = W_V H_{\text{int}} \tag{19}$$

其中，W_Q, W_K, W_V 为可学习的投影矩阵。随后计算缩放点积注意力分数与注意力权重：

$$S = \frac{QK^T}{\sqrt{\mathrm{d}_h}} \tag{20}$$

$$A = soft\max(S) \tag{21}$$

其中，$S \in \mathbb{R}^{N \times N}$ 为原始注意力分数矩阵，A_{ij} 表示第 i 句话语的情感特征对第 j 句话语的意图特征的关注程度。基于该权重对意图值向量进行加权聚合，得到意图上下文表示：

$$C_{\text{int}} = AV \tag{22}$$

$$C_{i_{\text{int}}} = \sum_{j=1}^{N} A_{ij} v_j \tag{23}$$

其中，$C_{\text{int}} \in \mathbb{R}^{N \times d_h}$ 为意图感知的上下文向量，$C_{i_{\text{int}}}$ 为第 i 句话语的意图上下文向量。该机制能够在对话范围内自适应选择与当前情感判别相关的意图信息，使

意图线索以“上下文增强”的方式注入情感表征，为后续门控融合与情感分类提供更具判别性的特征支持。在获得了意图上下文C_{int}后，需要将其与原始情感特征H_{emo}进行融合。直接相加或拼接的固定融合方式无法适应不同对话场景的需求。例如，当意图信号明确且与情感高度相关时，应增大其权重；而当意图模糊或与情感矛盾时，应降低其影响。为此，我们采用一种基于门控机制的自适应融合策略：

$$\alpha = \sigma(W_g[H_{emo} \oplus C_{\text{int}}] + b_g) \tag{24}$$

$$\tilde{H}_{emo} = H_{emo} + \alpha \odot C_{\text{int}} \tag{25}$$

其中，$\oplus$表示特征拼接，W_g, b_g为可学习参数，σ为 sigmoid 函数，$\tilde{H}_{emo}$为融合后的特征。在获得意图上下文C_{int}后，我们采用该门控机制对意图信息的注入强度进行自适应调节，以避免固定融合在不同场景下的失配。门控系数计算如式(24)，并按式(25)得到增强后的情感表示。其中$\alpha \in [0,1]$表示逐话语的融合比例，α越大表示越依赖意图引导信号。

3.4 输出层与训练

经过意图引导模块处理后，模型分别输出情感和意图的预测结果。情感分类器的公式如下：

$$P_{emo} = soft\max(W_{emo}\tilde{H}_{emo} + b_{emo}) \tag{26}$$

其中，W_{emo}, b_{emo}为可学习参数，C_{emo}为情感类别数。$P_{emo} \in \mathbb{R}^{N \times C_{emo}}$为情感类别的概率分布，$\tilde{H}_{emo}$为经过意图引导增强后的情感特征。意图分类器公式如下：

$$P_{\text{int}} = soft\max(W_{\text{int}}H_{\text{int}} + b_{\text{int}}) \tag{27}$$

其中，$W_{\text{int}}, b_{\text{int}}$为可学习参数，$C_{\text{int}}$为意图类别数。$P_{\text{int}} \in \mathbb{R}^{N \times C_{\text{int}}}$为意图类别的概率分布。意图分类器直接作用于原始意图特征H_{int}，而非增强后的情感特征，这确保了意图分支的独立性，避免情感信息反向污染意图预测，并且独立训练的意图分类器能够为交叉注意力机制提供更纯粹、更判别性的意图表示。我们采用

多任务学习框架，通过联合优化情感识别和意图识别两个任务，实现意图信息对情感识别的有效引导。总损失函数定义为：

$$\ell_{total} = \lambda_{emo} \cdot \ell_{emo} + \lambda_{\text{int}} \cdot \ell_{\text{int}} \tag{28}$$

其中，λ_{emo} 和 λ_{int} 为任务权重系数，用于平衡两个任务的重要性。在情感识别任务中，对话数据集往往存在不同程度的类别不平衡问题。为了缓解这一问题，避免模型过度偏向多数类而忽略少数类情感，我们在所有数据集上统一采用加权交叉熵损失函数。具体定义如下：

$$\ell_{emo} = -\sum_{i=1}^{N}\sum_{c=1}^{C_{emo}} \omega_c \cdot y_{i_c} \log(p_{i_c}^{emo}) \tag{29}$$

$$\omega_c = \sqrt{\frac{N_{total}}{C \cdot N_c}} \tag{30}$$

其中，ω_c 为类别 c 的权重，N_c 为类别 c 的样本数，N_{total} 为总样本数。相比于逆频率加权（$W_c = N_{total} / N_c$），平方根逆频率加权能够在提升少数类性能和保持多数类性能之间取得更好的平衡。意图识别损失（辅助任务）如下：

$$\ell_{\text{int}} = -\sum_{i=1}^{N}\sum_{c=1}^{C_{\text{int}}} y_{i_c} \log(p_{i_c}^{\text{int}}) \tag{31}$$

其中，y_{i_c} 为意图标签。在最后的测试阶段，模型仅输出情感预测结果：

$$y_i^{emo} = \arg\max(P_{emo}[i]) \tag{32}$$

意图识别损失函数在训练中扮演着双重关键角色。一方面，通过优化意图分类损失，模型强制 H_{int} 学习具有判别性的意图语义；高质量的 H_{int} 确保了交叉注意力机制能够检索到准确的意图线索，从而对情感特征进行有效增强。另一方面，意图任务作为辅助监督信号，通过共享底层表示，促使模型学习更鲁棒的通用对话语义，从而减少对特定情感样本的过拟合，提升了模型整体的泛化能力。

4.实验

4.1 数据集和评估指标

我们选用 Mastodon 和 DailyDialog 两个数据集对模型进行实验验证。Mastodon 数据集来源于社交媒体平台上的真实用户对话，文本表达自然、情感多样，具有较强的上下文依赖性，能够反映复杂的真实对话情感交互场景。DailyDialog 数据集由多轮日常对话构成，对话结构清晰、情感标注规范，广泛应用于对话情感识别研究。通过在这两个具有不同特点的数据集上进行实验，能够较为全面地评估模型的性能与泛化能力。

在评估指标方面，我们在两个数据集上统一采用 Macro Precision、Macro Recall 和 Macro F1 作为性能评价指标。宏平均指标通过对各情感类别分别计算指标后取平均，有效减弱了类别不平衡对实验结果的影响，能够更加客观地反映模型在不同情感类别上的整体识别能力。

4.2 实验细节与对比模型

我们所有实验均在 Ubuntu Linux 平台上完成，使用单张 NVIDIA GeForce RTX 3090（24GB 显存）进行训练与测试，CUDA 版本为 13.0。实验采用 Python 版本为 3.9.23。深度学习框架为 PyTorch 2.7.1。我们在两个数据集上均采用 RoBERTa-large 作为文本编码器，并在此基础上构建后续图结构与多任务模块。为缓解类别不均衡问题，我们使用基于频次的加权策略 sqrt_inverse_freq。基本的训练参数如表 1 所示。

表 1 每个数据集使用的参数

Hyperparameter	Dailydialog	Matodon
hidden_dim	300	512
gnn_layers	2	2
mlp_layers	2	3
num_heads	4	4
attention_dropout	0.3	0.1
dropout	0.2	0.3
batch_size	32	32
epochs	70	200
learning rate	1e-5	1e-5

为验证所提出方法的有效性，我们在 Mastodon 与 DailyDialog 两个数据集上与多种代表性方法进行对比。所选方法均为近年来在“情感/情绪识别 + 对话行为/意图识别（Dialogue Act Recognition, DAR）”联合建模方向中具有代表性的工作，覆盖从早期多任务框架到图结构交互、时序推理与两阶段训练等不同建模范式。具体对比方法如下：

JointDAS：经典的联合学习框架，同时建模对话行为与情感信息，是该任务设定下的早期基线之一。

DCR-Net：通过显式关系建模刻画对话轮次间的共交互（co-interaction）信息，实现对 DAR 与情感任务的联合增强。

Co-GAT：基于图注意力网络构建 utterance 级交互图，并通过协同交互机制融合两任务信息，属于图结构联合建模的代表方法。

DARER：侧重时序关系与递归推理，通过时间与关系的联合推理提升联合识别性能。

MpGLIF：采用多视角的全局–局部交互机制，同时捕获对话级全局上下文与局部依赖关系，以提升联合识别效果。

TSCL：两阶段学习/集成式训练范式，通过阶段化优化策略提升联合任务表现。

为保证对比的可读性与一致性，我们对两数据集均报告 macro-Precision、macro-Recall 与 macro-F1（见表 2），并在相同数据集与评测协议下对比各方法的性能差异。

表 2 对比表

Model	Mastodon			DailyDialog		
	F1	R	P	F1	R	P
JointDAS	37.6	41.6	36.1	31.4	28.8	31.2
DCR-Net	45.1	47.3	43.2	59.96	49.51	53.42
Co-GAT	48.1	53.2	44	**61.8**	51.1	55.9
DARER	59.6	63.3	56	48.6	45.7	55.2
MpGLIF	64.9	67.3	62.7	51.9	49.2	56.6
TSCL	51.6	58.7	46.1	45.4	40.1	56
ours	**68.76**	**71.48**	**66.94**	54.40	**53.35**	**57.63**

4.3 实验结果和分析

为验证所提出模型的有效性，我们在 Mastodon 与 DailyDialog 两个公开数据集上进行实验，并与现有代表性方法进行对比。需要说明的是，并非所有对比模型都在两个数据集上给出了完整的实验结果或采用完全一致的评测协议，因此不同数据集上的对比方法集合可能存在轻微差异。下文将分别给出各数据集上的实验结果，并对性能差异及其可能原因进行进一步分析与讨论。

4.3.1 在 Mastodon 数据集上的对比

表 2 给出了我们的方法与现有代表性模型在 Mastodon 数据集上的对比结果。可以观察到，我们的方法在 macro-F1（68.76）、macro-Recall（71.48） 与 macro-Precision（66.94） 三项指标上均取得最优表现，整体优于所有对比方法。其中，相较于当前的基线 MpGLIF（macro-F1=64.9），我们在 macro-F1 上提升 3.86 个百分点，同时在 macro-Recall 与 macro-Precision 上分别提升 4.18 与 4.24 个百分点。

我们认为该优势主要的原因有两点：其一，模型通过引入意图辅助任务提供的互补线索，使情感表征能够获得更明确的语用与功能约束，从而减少仅依赖语义表面信息导致的歧义；其二，结构化交互建模能够更充分地聚合对话中的上下文依赖，尤其对 Mastodon 这类跨轮次信息更丰富、上下文关联更强的数据场景

更为有效。总体而言，Mastodon 的结果表明：在联合建模设置下，显式建模任务互补信息与对话结构依赖能够显著提升模型对宏观类别的整体辨识能力。

4.3.2 在 DailyDialog 数据集上的对比

表 2 同时也汇报了 DailyDialog 数据集上的对比结果。与 Mastodon 类似，我们的方法在 macro-Recall 与 macro-Precision 两项指标上均达到当前对比方法中的最优水平，说明模型在“找全”（Recall）与“找准”（Precision）方面具有较好的整体能力。

然而，在 macro-F1 指标上，我们的方法取得 54.54，仍低于部分强基线（例如 Co-GAT 的 61.8 与 DCR-Net 的 59.96）。需要说明的是，macro-F1 并非由 macro-P 与 macro-R 直接计算得到，因此即使 macro-Precision 与 macro-Recall 更高，仍可能出现 macro-F1 相对偏低的情况。这通常意味着：模型在某些类别上存在明显短板，导致这些类别的单类 F1 拉低了宏平均结果。

我们推测该现象与 DailyDialog 的数据特性有关：DailyDialog 存在更明显的类别不均衡与类别间语义相近问题，少数类样本不足容易造成模型在少数类别上的召回或精度波动，从而对 macro-F1 产生更显著的负面影响。尽管我们通过类别加权策略（如 sqrt_inverse_freq）缓解了不均衡问题，但宏平均 F1 仍反映出在极少数类别上的识别仍具有挑战性。

4.3.3 小节

综合两个数据集结果可以看出，我们的方法在 Mastodon 上表现稳定且显著领先，在 DailyDialog 上在 macro-Precision 与 macro-Recall 层面表现突出，但 macro-F1 仍存在提升空间。这表明所提出模型在捕获上下文依赖与任务互补信息方面具有优势，但在更强类别不均衡/更细粒度区分要求的场景下，仍需要进一步提升对少数类与易混淆类别的鲁棒性。

4.4 消融实验

为验证所提出模型中关键组成模块的有效性，我们在保持训练配置一致的前提下开展了消融实验。实验以完整模型为基准，分别构建了移除意图分支（w/o

Intent）、移除交叉注意力引导（w/o Attention）以及将门控替换为线性融合（w/o Gate）三种变体，实验结果如表 3 和表 4 所示。结果显示，完整模型在两个数据集上均取得了最优的综合表现，证明了上下文结构建模与意图引导机制在对话情感识别任务中具有显著的协同互补性。

表 3 Mastodon 数据集消融实验结果

intent	attention	gate	mastodon		
			macro F1	macro R	macor P
✗	✗	✗	62.23	64.55	61.61
☑	✗	☑	61.37	63.03	60.73
☑	☑	✗	67.67	70.46	66.67
☑	✗	✗	62.19	64.26	61.45
☑	☑	☑	**68.76**	**71.48**	**66.94**

表 4 DailyDialog 数据集消融实验结果

intent	attention	gate	DailyDialog		
			macro F1	macro R	macor P
✗	✗	✗	53.34	**56.1**	52.21
☑	✗	☑	52.47	54.55	51.5
☑	☑	✗	52.53	55.83	50.77
☑	✗	✗	52.78	53.06	53.79
☑	☑	☑	**54.4**	53.35	**57.63**

首先，意图信息对情感判别起到了关键的辅助作用。当移除意图分支后，两个数据集上的 Macro-F1 均出现显著下降。这表明对话情感并非仅由话语表层语义决定，而是与对话行为功能（如提问、反驳等）紧密关联。尤其在语义含蓄或强依赖语境的场景中，意图线索能够为情感判别提供额外的功能性约束，有效减少语义歧义。此外，对比“仅保留意图监督但移除交叉注意力”的变体与完整模型可以发现，仅依赖多任务学习带来的隐式参数共享效果有限。这说明性能增益主要来源于交叉注意力机制建立的显式对齐，它使模型能够定向检索关键意图线索，从而更有效地强化情感特征表达。

其次，门控融合机制对于提升模型的鲁棒性至关重要。实验中将门控替换为线性融合导致了整体性能下滑，且在 DailyDialog 上出现了牺牲精度换取召回的现象。这是因为意图与情感的关联强度随语境动态变化：在部分话语中意图线索

可能较弱甚至引入噪声。门控机制通过学习动态权重，能够自适应地调节意图信息的注入强度。它在充分利用有效信息的同时，抑制了不相关意图带来的负迁移风险，从而确保模型在不同对话场景下均能保持稳定的预测能力。

总体而言，本消融实验从不同层面验证了模型设计的合理性。实验结果不仅证实了意图信息作为辅助信号的价值，更强调了“显式引导”与“自适应融合”在跨任务交互中的重要性。这一结论与本文的设计初衷相符，也进一步表明将对话行为层面的意图信息纳入建模，是提升对话情感识别性能的一条有效且鲁棒的途径。

4.5 参数分析

为了全面评估模型在不同配置下的性能表现与鲁棒性，我们将超参数划分为“对优化过程敏感的核心参数”和“与模型结构相关的其他参数”两类进行深入探究。对于意图损失权重、类别权重策略、Dropout 率以及学习率这四个对优化过程极为敏感的参数，实验结果均在 5 个不同随机种子下运行并取平均值，以消除随机性干扰，确保结论的稳定性与可靠性；而对于 GNN 层数、隐藏层维度等模型结构相关参数，则固定随机种子以分析其对模型性能的整体趋势影响。

在 Mastodon 数据集上的实验结果揭示了社交媒体对话场景下的参数敏感性特征。其中，对优化过程敏感的核心参数分析结果如图 2 所示，与模型结构相关的其他参数分析结果如图 3 所示。针对优化过程参数，意图损失权重的选择对模型性能具有显著影响。实验数据显示，随着 $LossWeight(Intent)$ 从 0.1 增加至 1.5，模型性能呈现出典型的先升后降趋势，并在 $LossWeight(Intent)=0.3$时达到最佳 Macro-F1 值 68.64%。这一现象表明，适度的意图监督信号能够有效辅助情感识别，但过大的权重会导致辅助任务主导损失函数，从而干扰主任务的特征学习。在正则化方面，由于 Mastodon 数据集包含大量非正式用语和噪声，模型在 Dropout 设置为 0.3 时泛化能力最强，较低的 Dropout 容易导致过拟合，而过高的设置则会丢失有效信息。此外，模型对学习率较为敏感，在 1e-5 时达到峰值，过大或过小的学习率均会导致性能显著下滑。在类别权重策略上，Balanced 策略表现略优于其他策略。在模型结构参数方面，实验发现浅层的图网络结构更为有效，GNN 层数为 1 时模型表现最佳，随着层数增加至 3 或 4，

性能出现断崖式下跌，这验证了在对话情感任务中，过深的图卷积会导致节点表示趋同的“过平滑”问题。同时，较小的隐藏层维度（256）和中等的 Batch Size（32）能够使模型在保持特征紧凑性的同时获得最优的泛化性能。

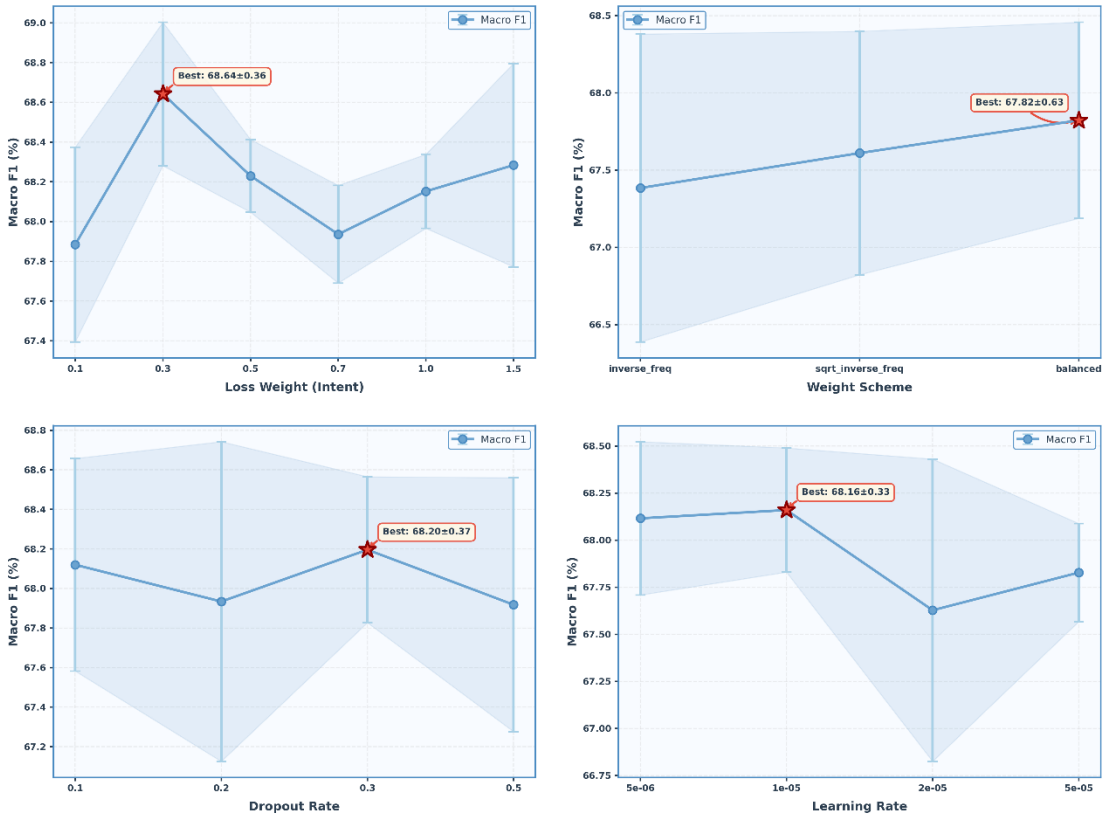

图 2 Mastodon 数据集核心参数分析

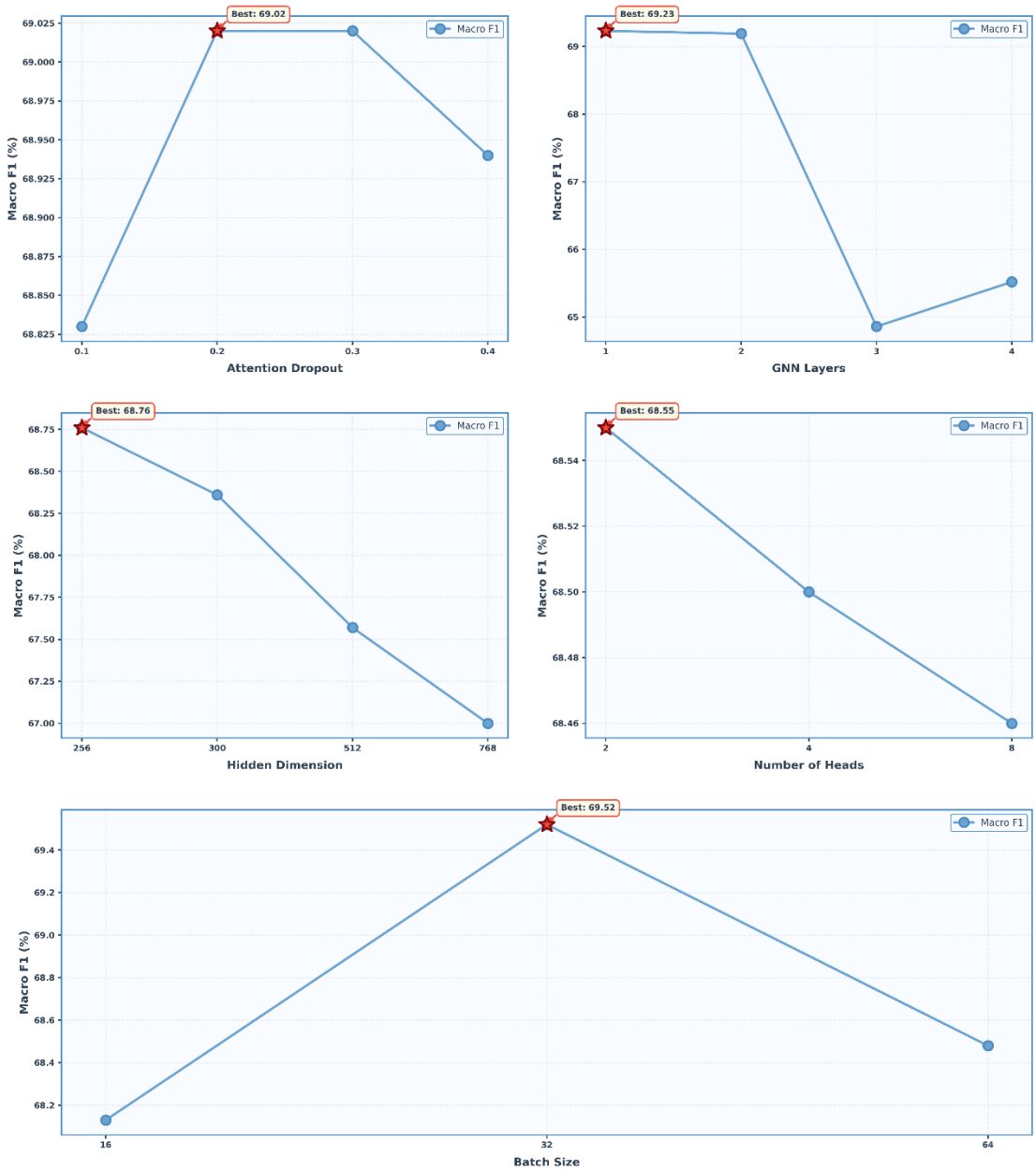

图 3 Mastodon 数据集其他参数分析

对于 DailyDialog 数据集，实验呈现出与 Mastodon 相似但又体现数据特性的趋势。其中，对优化过程敏感的核心参数分析结果如图 4 所示，与模型结构相关的其他参数分析结果如图 5 所示。在优化参数方面，意图损失权重同样在 $LossWeight(Intent)=0.3$ 时达到峰值，这证实了 0.3 是一个跨数据集鲁棒的超参数选择，能够较好地平衡主辅任务。然而，与 Mastodon 不同的是，DailyDialog 的文本更为规范且噪声较少，因此模型偏好较低的正则化强度，在 Dropout 为 0.2 时表现最佳。同时，针对该数据集特定的类别分布，平方根倒数加权策略表

现明显优于其他方案，说明温和的平衡机制更适合此类日常对话数据。在学习率的选择上，该数据集需要更加谨慎的更新步长，在 5e-6 处取得最佳结果。在模型结构方面，DailyDialog 在 GNN 层数为 2 时达到显著峰值，优于层数为 1 的结果，这说明在日常对话中，捕捉二阶依赖即跨越中间话语的上下文对于理解情感流转至关重要，但层数继续增加同样会引发过平滑问题。此外，交叉注意力模块中的 Dropout 设置为 0.3 以及较少的注意力头数（2 Heads）有助于防止模型过度依赖特定的意图-情感对齐模式，从而在有限的数据量下学习到更充分的语义表示。

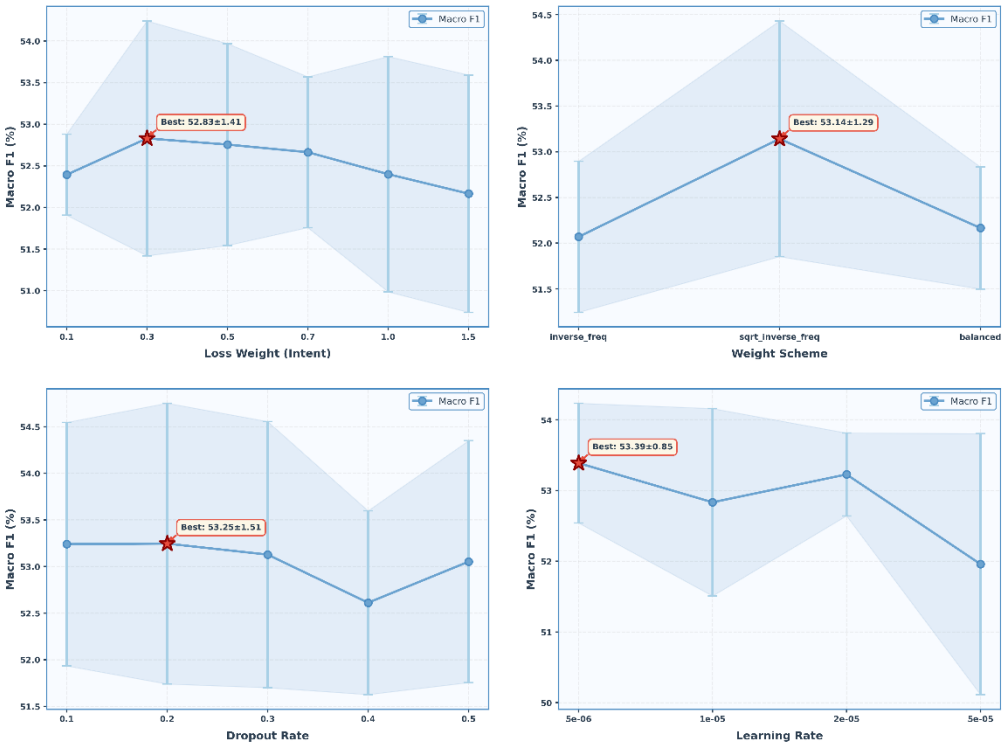

图 4 DailyDilaog 数据集核心参数分析

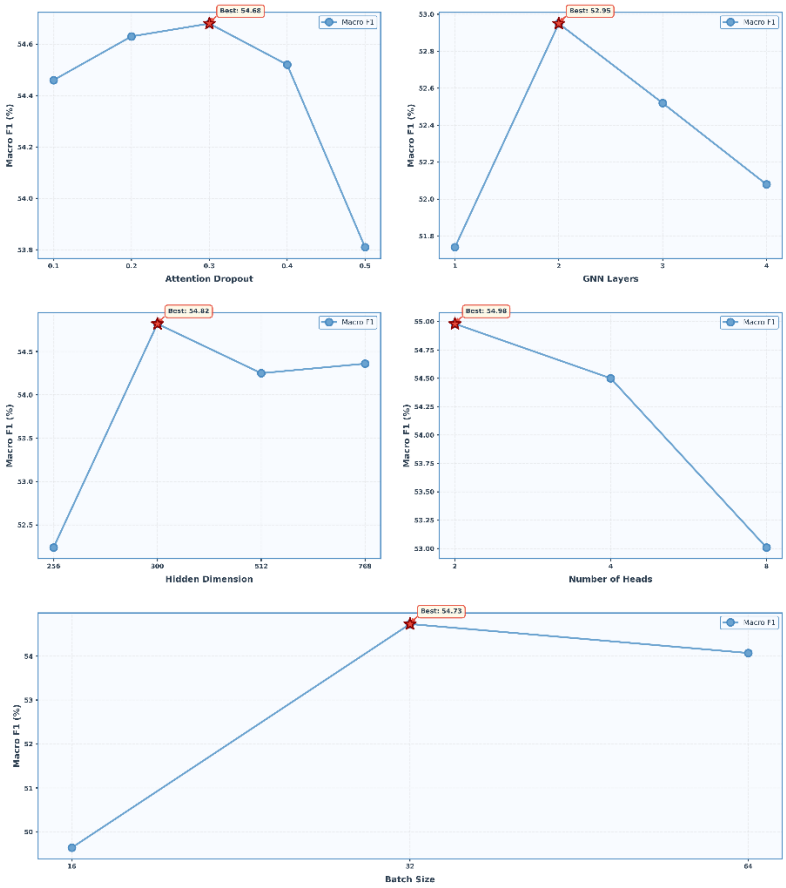

图 5 DailyDilaog 数据集其他参数分析

综合上述两个数据集的分析结果可以得出，意图辅助任务的最佳权重具有较强的跨场景鲁棒性，稳定在 0.3 左右，证明了本文提出的多任务框架的有效性。而正则化参数（如 Dropout）和图结构深度则需根据数据集的噪声水平和对话复杂度进行自适应调整：噪声较大的社交媒体数据需要更强的正则化和更浅层的结构，而规范的日常对话数据则受益于适度的正则化和稍深层的上下文聚合。

5.总结

我们针对对话情感识别任务中上下文依赖捕捉不足以及潜在意图信息未能被有效利用的问题，提出了一种融合意图感知与图神经网络的多任务联合学习框架。该模型以预训练语言模型为基础构建上下文表征，并通过图交互模块，显式地建模了对话中话语间的深层语义关联与情感流转机制。为了进一步增强特征表示的判别性，我们引入意图识别作为辅助任务，通过多任务学习机制挖掘交际意图与情感表达之间的内在相关性，从而在联合优化中实现特征的互补与增强。在

Mastodon 和 DailyDialog 两个具有显著分布差异的基准数据集上的广泛实验，充分验证了所提方法的有效性与优越性。实验结果表明，所提出的框架在核心评价指标上表现优异，取得了与现有主流基线模型相当甚至更优的结果，证明了意图特征对于理解复杂对话情感状态具有重要的补充作用。此外，通过深入的参数敏感性分析与消融实验，我们证实了意图辅助任务权重的跨场景鲁棒性，并揭示了针对不同噪声水平的数据集自适应调整正则化策略与图结构深度的必要性，验证了模型在处理多样化对话场景时的可靠性与泛化能力。