Hierarchical Contrastive Graph Attention Network for Aspect-Based Sentiment Analysis

1st Hao Zhang
School of Computer Science and Engineering
Northeastern University
Shenyang 110819, China
2472115@stu.neu.edu.cn

2nd Baiyou Qiao*
School of Computer Science and Engineering
Northeastern University
Shenyang 110819, China
qiaobaiyou@mail.neu.edu.cn

***Abstract*—Aspect-based sentiment analysis aims to identify the sentiment polarity of specific aspect terms in sentences. Existing graph neural network methods based on dependency trees face two limitations: processing dependency structures at fixed scales without recognizing the distinct roles of syntactic relationships at different granularities, and lacking self-supervised regularization to enhance representation learning. To address these issues, we propose Hierarchical Contrastive Graph Attention Network (HCGAT). The model captures syntactic information at three granularities: word-level graph attention (1-hop) for lexical modifications, phrase-level graph attention (2-hop) for phrasal composition, and sentence-level multi-head attention for global semantics. We design cross-granularity consistency contrastive learning with intra-granularity consistency and inter-granularity discriminability constraints to enhance hierarchical fusion effectiveness. Experiments on three benchmark datasets demonstrate that HCGAT achieves state-of-the-art performance, with particularly significant improvements on Laptop14 which has fewer training samples.**

***Index Terms*—aspect-based sentiment analysis, graph neural network, hierarchical modeling, contrastive learning, dependency tree**

I. Introduction

Aspect-Based Sentiment Analysis (ABSA) aims to identify the sentiment polarity of specific aspect terms in sentences, representing one of the most challenging fine-grained tasks in natural language processing [1]. Unlike sentence-level sentiment analysis, ABSA requires discriminating sentiment polarities for multiple aspect terms within the same sentence. The core challenge lies in accurately establishing semantic associations between aspect terms and their corresponding opinion expressions, particularly when these elements are syntactically distant or when multiple aspects with conflicting sentiments coexist. Early attention-based methods [2], [3] often produce spurious alignments. Recent research has leveraged syntactic dependency trees as explicit structural guidance, and graph neural networks have demonstrated remarkable effectiveness [4]–[6]. Despite these advances, we identify two limitations: **First, existing methods lack principled hierarchical modeling of syntactic structures.** While some work [7], [8] considers multi-hop connections, they typically process different distance ranges through simple masking or concatenation, failing to recognize that syntactic relationships at different granularities have fundamentally distinct semantic functions. Linguistic theory [9] indicates that immediate dependencies encode lexical-level modifications, multi-hop paths capture phrasal composition, while global patterns reflect sentence-level coherence. **Second, multi-granular representations lack consistency constraints.** Existing methods merely use simple concatenation or weighted fusion, failing to ensure that representations of the same aspect at different granularities maintain semantic consistency. Meanwhile, representations of different aspects lack explicit discriminability constraints. To address these limitations, this paper proposes **Hierarchical Contrastive Graph Attention Network (HCGAT)**, with the following innovations: **(1) Linguistically-motivated hierarchical graph encoding.** We explicitly model three granularities: word-level graph attention (1-hop), phrase-level graph attention (2-hop), and sentence-level multi-head attention. These are integrated through a hierarchical fusion module with layer normalization. **(2) Cross-granularity consistency contrastive learning.** Intra-granularity consistency constraint forces representations of the same aspect from different granularities to be close, while inter-granularity discriminability constraint pushes apart representations of different aspects. **(3) Dynamic aspect-aware pooling.** We design an attention-based pooling mechanism that dynamically weights contextual tokens, effectively isolating aspect-specific contexts and preventing sentiment spillover. Experiments on three benchmark datasets demonstrate that HCGAT achieves state-of-the-art performance, with particularly significant improvements on Laptop14 which has the smallest training set. The main contributions of this paper are:

- We propose a linguistically-motivated hierarchical architecture that systematically models syntactic structures at word, phrase, and sentence levels through specialized graph attention mechanisms.
- We design cross-granularity consistency contrastive learning with dual constraints of intra-granularity consistency and inter-granularity discriminability, particularly effective in data-scarce scenarios.
- We design a dynamic aspect-aware pooling mechanism that effectively disentangles sentiment signals in multi-aspect sentences, achieving state-of-the-art results on

* Corresponding author.

three benchmark datasets.

II. Related Work

This section reviews related research in aspect-based sentiment analysis, focusing on the evolution from attention mechanisms to graph neural networks.

A. From Attention to Graph Neural Networks

Early ABSA research relied on attention mechanisms to model aspect-context associations. Representative work includes TD-LSTM [2] for position-aware modeling, ATAE-LSTM [10] combining aspect embeddings with attention, IAN [11] for interactive bidirectional attention, RAM [3] using multi-hop reasoning, and MGAN [12] for multi-granular attention coordination. While effective in capturing relevant contexts, pure semantic attention struggles with complex structures, particularly when aspect and sentiment words are distant or when sentences contain multiple aspects with conflicting polarities.

To address these limitations, recent research has turned to syntactic dependency trees, and graph neural networks operating on dependency structures have become the dominant paradigm. ASGCN [4] first applied GCN on dependency tree topology with aspect-specific masking, and subsequent methods improved upon this with dynamic attention weights and gating mechanisms to handle parsing errors. Recognizing that different dependency types have distinct effects on sentiment propagation, BiGCN [13] separately models lexical and syntactic graphs with gating fusion, R-GAT [5] reshapes dependency trees and learns relation-specific attention weights, and SenticGCN [14] integrates external sentiment knowledge from SenticNet. Observing that path length affects sentiment propagation strength, SSEGCN [7] constructs multi-scale distance masks, DualGCN [15] jointly models syntactic and semantic graphs with distance weighting, and MASGCN [8] designs multi-view attention with adaptive distance matrices. Additionally, ITGCN [16] combines instruction tuning with external knowledge-enhanced dependency graphs, and SSA-GCN [17] utilizes both syntactic and semantic information to construct dual-channel graph structures.

Despite these advances, existing methods predominantly operate at fixed scales—either focusing on direct neighbors or applying uniform processing to predefined distance ranges—lacking systematic architectural designs that recognize the fundamentally different linguistic roles of syntactic relationships at varying granularities. Our work addresses this gap through principled hierarchical modeling with granularity-specific mechanisms.

B. Contrastive Learning in Sentiment Analysis

Contrastive learning learns discriminative representations by maximizing similarity between augmented views of the same sample while minimizing similarity across different samples. Originating in computer vision with SimCLR [18], it has been successfully adapted to NLP through methods like SimCSE [19] and SupCL [20].

In sentiment analysis, contrastive learning applications have become increasingly prevalent. SentiLARE [21] enhances pre-trained models with sentiment knowledge-guided contrastive learning. Scon-ABSA [22] leverages supervised contrastive learning to distinguish aspect-dependent and aspect-invariant sentiment features, first introducing contrastive learning into ABSA tasks. APSCL [23] designs aspect-pair supervised contrastive learning to capture latent sentiment relationships among multiple aspects. DCL-GCN [24] proposes a dual contrastive learning framework that applies contrastive objectives at both semantic encoding and syntax label levels. PRCL-GCN [25] distinguishes between original and knowledge-enhanced dependency graphs through prompted representation joint contrastive learning. CL-GAT [26] combines graph attention networks with supervised contrastive learning through a dual-path approach integrating syntactic and semantic information. Additionally, some work explores data augmentation strategies, such as LWEDA-ACT [27] which utilizes linking words-guided multidimensional emotional data augmentation with adversarial contrastive training.

However, existing work primarily targets sentence-level tasks or applies contrastive learning as an independent module, failing to address how contrastive learning can deeply integrate with multi-granular syntactic modeling. Our work introduces cross-granularity consistency contrastive learning, which enforces representations of the same aspect from different granularities to be consistent through intra-granularity consistency constraints, while enhancing discriminability between different aspects through inter-granularity discriminability constraints, thereby directly serving the effectiveness of hierarchical fusion.

III. Method

A. Problem Definition

Given a sentence $S = \{w_1, w_2, \ldots, w_n\}$ containing n words and an aspect term $a = \{w_i, \ldots, w_j\}$ in the sentence (where $1 \leq i \leq j \leq n$), the ABSA task aims to predict the sentiment polarity $y \in \{\text{positive}, \text{neutral}, \text{negative}\}$ of the aspect term. We obtain the dependency tree T of sentence S through a dependency parser and represent it as an adjacency matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$, where $\mathbf{A}_{ij} = 1$ indicates that words w_i and w_j have a direct dependency relation.

B. Model Overview

As shown in Figure 1, HCGAT takes text sequences with aspect markers and their corresponding dependency tree adjacency matrix $\mathbf{A}$ as input, and outputs the sentiment polarity prediction $\hat{y}$. The model consists of three main modules:

(1) Hierarchical Graph Attention Encoding Module: Contains contextual encoder and hierarchical graph attention encoder, obtaining initial contextual representation $\mathbf{H}^{(0)}$ and three granularity syntactic representations $\mathbf{H}^{(w)}$, $\mathbf{H}^{(p)}$, $\mathbf{H}^{(s)}$.

(2) Hierarchical Fusion Module: Integrates multi-granular features through adaptive gating mechanism, obtaining fused representation $\mathbf{H}^{(\text{fused})}$.

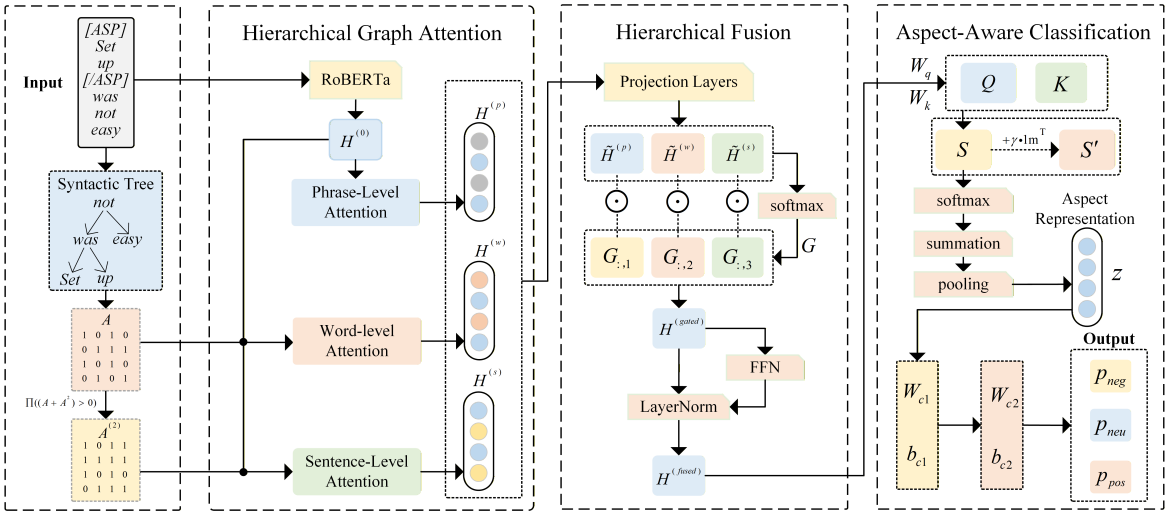

Fig. 1. Overall architecture of the proposed HCGAT model. The framework consists of three main components: (1) Hierarchical Graph Attention Encoding Module that processes input sequences and dependency structures through word-level, phrase-level, and sentence-level encoders; (2) Hierarchical Fusion Module that integrates multi-granular representations via adaptive gating; (3) Aspect-Aware Classification Module that performs sentiment prediction with cross-granularity contrastive learning enhancement.

(3) Aspect-Aware Classification Module: Contains aspect-aware pooling layer and sentiment classifier, extracting aspect-specific representation $\mathbf{z}$ and predicting sentiment polarity.

During training, the cross-granularity consistency contrastive learning module optimizes $\mathcal{L}_{\text{contrast}}$ to enhance the consistency and discriminability of hierarchical representations.

C. Hierarchical Graph Attention Encoding Module

1) Contextual Encoder: The contextual encoder uses pre-trained language model RoBERTa to encode input text. To explicitly mark aspect term positions, we insert special markers [ASP] and [/ASP] before and after aspect terms. After encoding by RoBERTa, we obtain contextual representation matrix $\mathbf{H}^{(0)} \in \mathbb{R}^{n \times d}$, where d is the hidden dimension.

2) Hierarchical Graph Attention Encoder: The hierarchical graph attention encoder captures syntactic information at three granularities: word-level (1-hop) captures direct dependency relations, phrase-level (2-hop) models indirect dependency structures, and sentence-level captures global semantics.

Word-level Graph Attention. The word-level graph attention operates on the original 1-hop adjacency matrix $\mathbf{A}$. For input representation $\mathbf{H}^{(0)}$, the graph attention layer computes attention coefficients between node i and its neighbor node j:

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}_w^\top \left[\mathbf{W}_w \mathbf{h}_i^{(0)} \| \mathbf{W}_w \mathbf{h}_j^{(0)}\right]\right) \tag{1}$$

where $\mathbf{W}_w \in \mathbb{R}^{d \times d}$ is a learnable weight matrix, $\mathbf{a}_w \in \mathbb{R}^{2d}$ is an attention vector, and $\|$ denotes vector concatenation. The attention weights are obtained through softmax normalization:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \tag{2}$$

where $\mathcal{N}_i = \{j \mid \mathbf{A}_{ij} = 1\}$ is the neighbor set of node i. The updated representation is:

$$\mathbf{h}_i^{(w)} = \sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}_w \mathbf{h}_j^{(0)} \tag{3}$$

We stack two word-level GAT layers, obtaining word-level representation $\mathbf{H}^{(w)} \in \mathbb{R}^{n \times d}$.

Phrase-level Graph Attention. The phrase-level graph attention operates on 2-hop adjacency matrices to model indirect dependency relations. The 2-hop adjacency matrix $\mathbf{A}^{(2)}$ is computed as:

$$\mathbf{A}^{(2)} = \mathbb{I}\left((\mathbf{A} + \mathbf{A}^2) > 0\right) \tag{4}$$

where $\mathbb{I}(\cdot)$ is an indicator function that marks both 1-hop and 2-hop reachable node pairs as connected. The computation follows the same procedure as word-level GAT but uses independent parameter matrices $\mathbf{W}_p$ and $\mathbf{a}_p$, obtaining phrase-level representation $\mathbf{H}^{(p)} \in \mathbb{R}^{n \times d}$.

Sentence-level Multi-head Attention. The sentence-level attention captures global semantic information through standard multi-head self-attention:

$$\mathbf{H}^{(s)} = \text{MultiHeadAttn}(\mathbf{H}^{(0)}, \mathbf{H}^{(0)}, \mathbf{H}^{(0)}) \tag{5}$$

where multi-head attention contains h attention heads, obtaining $\mathbf{H}^{(s)} \in \mathbb{R}^{n \times d}$.

D. Hierarchical Fusion Module

The hierarchical fusion module integrates syntactic-semantic information captured at three granularities. First, granularity-specific projection layers align feature distributions:

$$\begin{aligned} \tilde{\mathbf{H}}^{(w)} &= \mathbf{W}_w^{proj} \mathbf{H}^{(w)} + \mathbf{b}_w^{proj} \\ \tilde{\mathbf{H}}^{(p)} &= \mathbf{W}_p^{proj} \mathbf{H}^{(p)} + \mathbf{b}_p^{proj} \\ \tilde{\mathbf{H}}^{(s)} &= \mathbf{W}_s^{proj} \mathbf{H}^{(s)} + \mathbf{b}_s^{proj} \end{aligned} \tag{6}$$

Then, adaptive gating weights for each granularity are computed:

$$\mathbf{G} = \text{softmax}\left(\mathbf{W}_g \tanh\left(\mathbf{W}_c[\tilde{\mathbf{H}}^{(w)} \| \tilde{\mathbf{H}}^{(p)} \| \tilde{\mathbf{H}}^{(s)}]\right)\right) \quad (7)$$

where $\mathbf{W}_c \in \mathbb{R}^{d \times 3d}$, $\mathbf{W}_g \in \mathbb{R}^{3 \times d}$, and $\mathbf{G} \in \mathbb{R}^{n \times 3}$ is the gating weight matrix. The weighted fusion representation is:

$$\mathbf{H}^{(\text{gated})} = \mathbf{G}_{:,1} \odot \tilde{\mathbf{H}}^{(w)} + \mathbf{G}_{:,2} \odot \tilde{\mathbf{H}}^{(p)} + \mathbf{G}_{:,3} \odot \tilde{\mathbf{H}}^{(s)} \quad (8)$$

Finally, a feedforward network with residual connection performs deep transformation:

$$\mathbf{H}^{(\text{fused})} = \text{LayerNorm}\left(\mathbf{H}^{(\text{gated})} + \text{FFN}(\mathbf{H}^{(\text{gated})})\right) \quad (9)$$

E. Aspect-Aware Classification Module

1) Aspect-Aware Pooling: The aspect-aware pooling layer dynamically aggregates context information related to aspect terms. We first construct aspect term mask $\mathbf{m} \in \{0, 1\}^n$, where aspect term positions are 1 and others are 0. Then the attention score matrix is computed:

$$\mathbf{S} = \frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}} \quad (10)$$

where $\mathbf{Q} = \mathbf{W}_q \mathbf{H}^{(\text{fused})}$ and $\mathbf{K} = \mathbf{W}_k \mathbf{H}^{(\text{fused})}$. To enhance attention on aspect term regions, we apply an additive mask:

$$\mathbf{S}' = \mathbf{S} + \gamma \cdot \mathbf{1}\mathbf{m}^\top \quad (11)$$

where γ is a hyperparameter controlling the attention boost degree. The aspect representation vector $\mathbf{z} \in \mathbb{R}^d$ is obtained through softmax normalization, weighted summation, and average pooling over aspect positions.

2) Sentiment Classifier: The aspect representation $\mathbf{z}$ is passed through a two-layer fully connected network to predict sentiment polarity:

$$\hat{\mathbf{y}} = \text{softmax}\left(\mathbf{W}_{c2}\left(\text{ReLU}\left(\mathbf{W}_{c1}\mathbf{z} + \mathbf{b}_{c1}\right)\right) + \mathbf{b}_{c2}\right) \quad (12)$$

where $\mathbf{W}_{c1} \in \mathbb{R}^{(d/2) \times d}$, $\mathbf{W}_{c2} \in \mathbb{R}^{3 \times (d/2)}$, and $\hat{\mathbf{y}} \in \mathbb{R}^3$ is the predicted probability for three sentiment categories.

F. Cross-Granularity Consistency Contrastive Learning

The contrastive learning module enhances the consistency and discriminability of hierarchical representations through self-supervised signals.

Intra-Granularity Consistency Constraint. For the same aspect, its representations from three granularities should be close in the semantic space. We extract aspect-level vectors from each granularity representation using the aspect mask $\mathbf{m}$:

$$\mathbf{z}^{(w)} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n m_i \mathbf{H}_i^{(w)} \quad (13)$$

$$\mathbf{z}^{(p)} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n m_i \mathbf{H}_i^{(p)} \quad (14)$$

$$\mathbf{z}^{(s)} = \frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n m_i \mathbf{H}_i^{(s)} \quad (15)$$

The intra-granularity consistency loss is defined as:

$$\mathcal{L}_{\text{intra}} = 1 - \frac{1}{3}\Big(\cos(\mathbf{z}^{(w)}, \mathbf{z}^{(p)}) + \cos(\mathbf{z}^{(p)}, \mathbf{z}^{(s)}) + \cos(\mathbf{z}^{(w)}, \mathbf{z}^{(s)})\Big) \quad (16)$$

Inter-Granularity Discriminability Constraint. Representations of different aspects should be distant in the semantic space. Given batch size B, we construct an InfoNCE loss:

$$\mathcal{L}_{\text{inter}} = -\frac{1}{B}\sum_{i=1}^B \log \frac{\exp(\text{sim}(\mathbf{z}_i^{(w)}, \mathbf{z}_i^{(p)})/\tau)}{\sum_{j=1}^B \exp(\text{sim}(\mathbf{z}_i^{(w)}, \mathbf{z}_j^{(p)})/\tau)} \quad (17)$$

where τ is a temperature hyperparameter and $\text{sim}(\cdot, \cdot)$ denotes normalized inner product. The total contrastive loss is:

$$\mathcal{L}_{\text{contrast}} = \mathcal{L}_{\text{intra}} + \mu\mathcal{L}_{\text{inter}} \quad (18)$$

G. Training Objective

The main task loss is cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^3 y_c \log \hat{y}_c \quad (19)$$

The total training objective combines classification loss and contrastive learning loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda\mathcal{L}_{\text{contrast}} \quad (20)$$

where λ is a weight hyperparameter balancing the two losses.

IV. Experiments

This section validates the effectiveness of the HCGAT model through extensive experiments. We first introduce experimental settings, then report main experimental results and compare with existing methods, analyze the contribution of each module through ablation experiments, and finally explore the impact of key hyperparameters.

A. Experimental Setup

Datasets: We conduct experiments on three widely used ABSA benchmark datasets: **Restaurant14** and **Laptop14** from SemEval-2014 Task 4 [28], and **Twitter** collected by [29]. Table I shows dataset statistics.

TABLE I
STATISTICAL SUMMARY OF BENCHMARK DATASETS. THE SENTIMENT DISTRIBUTION IS PRESENTED IN THE FORMAT OF "TRAINING SAMPLES / TEST SAMPLES" FOR EACH POLARITY CATEGORY (POSITIVE, NEUTRAL, NEGATIVE).

Dataset	Train	Test	Pos.	Neu.	Neg.
Restaurant14	3608	1119	2164/727	637/196	807/196
Laptop14	2282	632	976/337	455/167	851/128
Twitter	6051	677	1507/172	3016/336	1528/169

Data preprocessing: Stanford CoreNLP 4.5.0 is used for dependency parsing, with [ASP] and [/ASP] markers inserted before and after aspect terms.

Evaluation Metrics: We use **Accuracy** and **Macro-F1** as standard metrics.

Implementation Details. The model is implemented in PyTorch 1.12, with RoBERTa-base-uncased [30] as contextual encoder (hidden dimension d=768). Adam optimizer is used with RoBERTa layer learning rate of 2×10^{-5} and 1×10^{-3} for other layers. Batch size is 16, max sequence length is 128, training for up to 30 epochs with early stopping patience=5. Dropout rate is 0.1 applied after each GAT layer, ReLU activation, and in projection heads. Word-level and phrase-level GAT each stack 2 layers, sentence-level multi-head attention has h=8 heads. For aspect-aware pooling, the attention boost hyperparameter γ=10. Contrastive learning uses temperature τ=0.07 and weight λ of 0.08 (Restaurant14), 0.1 (Laptop14), 0.05 (Twitter). All results are averaged over 3 runs with different random seeds, and improvements over baselines are statistically significant ($p<0.05$ under paired t-test). Experiments are conducted on NVIDIA V100 GPU.

Comparison Methods: We compare with representative methods across different categories. Baseline results are cited from original papers:

Attention-based Methods: ATAE-LSTM [10], IAN [11], RAM [3], MGAN [12].

GNN-based Methods: ASGCN [4], kumaGCN [31], BiGCN [13].

GNN with Pre-trained Models: R-GAT [5], DualGCN [15], SSEGCN [7], DGEDT [6], SenticGCN [14], DRGAT [32], MASGCN [8].

B. Main Results and Analysis

TABLE II
PERFORMANCE COMPARISON ON THREE BENCHMARK DATASETS. BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND SECOND-BEST RESULTS ARE UNDERLINED. Δ REPRESENTS THE PERFORMANCE GAIN OVER THE PREVIOUS STATE-OF-THE-ART METHOD MASGCN. MODELS IN THE THIRD CATEGORY LEVERAGE PRE-TRAINED LANGUAGE MODELS (BERT, ROBERTA, ETC.) AS CONTEXTUAL ENCODERS.

Model	Restaurant14		Laptop14		Twitter	
	Acc	F1	Acc	F1	Acc	F1
Attention-based Methods						
ATAE-LSTM	77.20	—	68.70	—	67.30	—
IAN	78.60	—	72.10	—	—	—
RAM	80.23	70.80	74.49	71.35	69.36	67.30
MGAN	81.25	71.94	75.39	72.47	72.54	70.81
GNN-based Methods						
ASGCN	80.77	72.02	75.55	71.05	72.15	70.40
kumaGCN	81.43	73.64	76.12	72.42	72.45	70.77
BiGCN	81.97	73.48	74.59	71.84	74.16	73.35
GNN with Pre-trained Models						
R-GAT	83.30	76.08	77.42	73.76	75.57	73.82
DualGCN	84.27	78.08	78.48	74.74	75.92	74.29
SSEGCN	84.72	77.51	79.43	76.49	76.51	75.32
DGEDT	86.30	80.00	79.80	75.60	77.90	75.40
SenticGCN	86.92	81.03	82.12	79.05	—	—
DRGAT	86.35	79.92	81.96	78.57	78.84	77.73
MASGCN	87.76	82.56	82.44	79.11	78.58	77.05
HCGAT	**88.56**	**83.79**	**84.65**	**81.59**	**79.18**	**78.40**
Δ vs MASGCN	+0.80	+1.23	+2.21	+2.48	+0.60	+1.35

Table II shows performance comparison of HCGAT with all baseline methods. Our model achieves state-of-the-art performance on all three datasets.

From Table II, HCGAT achieves state-of-the-art performance on all three datasets. Compared with the strongest baselines MASGCN and DRGAT, our model demonstrates consistent improvements across all metrics.

HCGAT outperforms MASGCN by an average of 1.20% in accuracy and 1.69% in F1 score. While MASGCN fuses different distance subgraphs through multi-view attention, our model explicitly captures syntactic information at word-level, phrase-level, and sentence-level granularities through hierarchical modeling. The improvement is most significant on Laptop14, which has the fewest training samples, demonstrating that contrastive learning provides valuable self-supervised signals in data-scarce scenarios.

Compared with DRGAT, which models both dependency and constituency trees, HCGAT achieves average improvements of 1.75% in accuracy and 2.52% in F1. This suggests that hierarchical integration and self-supervised contrastive learning are more effective than flat dual-tree modeling alone. The consistent improvements across diverse datasets—from formal restaurant reviews to informal tweets—validate the robustness of our approach.

C. Ablation Study

To validate the contribution of each component, we conduct comprehensive ablation experiments. Table III presents the results, and we analyze them through the following research questions.

TABLE III
ABLATION STUDY RESULTS ON THREE BENCHMARK DATASETS (ACCURACY/F1). EACH ROW REPRESENTS A MODEL VARIANT WITH SPECIFIC COMPONENTS REMOVED OR MODIFIED TO ASSESS THEIR INDIVIDUAL CONTRIBUTIONS.

Variant	Restaurant14		Laptop14		Twitter	
	Acc	F1	Acc	F1	Acc	F1
HCGAT (Full)	88.56	83.79	84.65	81.59	79.18	78.40
w/o Contrastive	87.09	82.32	82.76	79.70	78.43	77.58
w/o Hierarchical	87.22	82.45	83.01	79.88	78.28	77.52
w/o Phrase-level	87.31	82.51	83.17	80.12	78.31	77.64
w/o Sentence-level	87.49	82.68	83.49	80.38	78.49	77.79
w/o Aspect Pooling	86.98	81.94	82.91	79.82	77.95	77.12
1-hop only	86.63	81.86	82.53	79.47	77.69	76.95
1-hop + 2-hop	87.58	82.94	83.72	80.65	78.61	77.82

RQ1: Does hierarchical modeling improve performance? We validate the effectiveness of hierarchical modeling from two perspectives. First, the "w/o Hierarchical" variant replaces adaptive gating fusion with simple concatenation and linear projection. The results show noticeable performance degradation, indicating that adaptive gating can dynamically adjust the contribution weights of each granularity based on different samples, while simple concatenation fails to capture such sample-level variations. Second, granularity combination

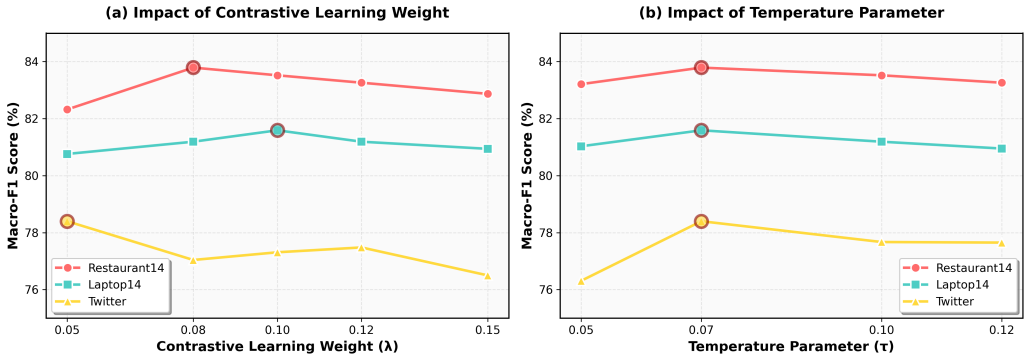

Fig. 2. Hyperparameter sensitivity analysis for contrastive learning components. (a) Effect of contrastive learning weight λ (Eq. 18) on Macro-F1 scores across three datasets. (b) Effect of temperature parameter τ in inter-granularity discriminability constraint (Eq. 16) on Macro-F1 scores.

analysis reveals complementary effects among levels: from using only word-level to progressively adding phrase-level and sentence-level, performance consistently improves. This demonstrates that different granularities capture syntactic-semantic information at different ranges—word-level focuses on direct modification relations, phrase-level extends to indirect dependency structures, and sentence-level supplements global semantic associations—the three work synergistically to achieve complete modeling from local to global.

RQ2: Can contrastive learning enhance aspect-sentiment representations?

Removing contrastive learning leads to performance degradation, with contributions varying notably across datasets: the largest drop occurs on Laptop14 which has the fewest samples, while the smallest drop occurs on Twitter which has the most samples. This phenomenon aligns with expectations—the essence of cross-granularity consistency constraints is to provide additional self-supervised signals by enforcing multi-granular representation alignment. When training data is sufficient, supervised signals alone can learn good representations; however, when data is limited, cross-granularity consistency constraints provide regularization that helps the model learn more consistent and robust hierarchical representations.

RQ3: What are the specific contributions of each granularity level?

Removing phrase-level GAT causes a larger performance drop than removing sentence-level attention. This indicates that in ABSA tasks, indirect dependency relations (e.g., the relationship between an aspect term and the modifier of its modifier) contribute more directly to sentiment judgment. The contribution of sentence-level attention is relatively smaller but still significant, primarily functioning to handle global semantic phenomena such as contrastive structures or adversative relations in sentences.

RQ4: Is aspect-aware pooling necessary for multi-aspect sentences?

Replacing aspect-aware pooling with simple average pooling results in a larger F1 drop than accuracy drop. This is because simple average pooling cannot distinguish the contexts of different aspect terms. When a sentence contains multiple aspect terms with opposite sentiment polarities, sentiment spillover tends to occur—where sentiment signals from one aspect interfere with the classification of another. Aspect-aware pooling effectively isolates the specific context of each aspect term through dynamic focusing mechanism, thereby alleviating this problem.

D. Choice of Pre-trained Encoder

We evaluate three pre-trained encoders with identical model architecture and hyperparameters. Table IV shows F1 scores.

TABLE IV
F1 SCORES (%) WITH DIFFERENT PRE-TRAINED ENCODERS.

Encoder	Restaurant14	Laptop14	Twitter
BERT	83.48	81.27	78.09
DeBERTa	83.57	81.38	78.21
RoBERTa	**83.79**	**81.59**	**78.40**

RoBERTa achieves the best performance, outperforming DeBERTa by 0.22% and BERT by 0.31% on average. However, the small performance gaps indicate that our hierarchical graph attention mechanism can effectively exploit syntactic information regardless of the base encoder, demonstrating the robustness of the proposed architecture. Based on these results, we use RoBERTa in all experiments.

E. Hyperparameter Sensitivity Analysis

This section explores the impact of two key hyperparameters in the contrastive learning module: contrastive learning weight λ and temperature parameter τ. Figure 2 visualizes the sensitivity analysis results.

The contrastive learning weight λ balances the classification loss and contrastive learning loss. As shown in Figure 2(a), the optimal λ reveals an interesting pattern: it is negatively correlated with training data size. Twitter (6051 samples) achieves best performance at λ=0.05, Restaurant14 (3608 samples) at λ=0.08, and Laptop14 (2282 samples) at λ=0.10. This pattern is consistent with our findings in RQ2. For smaller datasets, the model benefits more from self-supervised signals, thus requiring larger λ to fully exploit contrastive

TABLE V
CASE STUDY RESULTS. POS/NEU/NEG DENOTE PREDICTED SENTIMENT POLARITIES. ×=INCORRECT, ✓=CORRECT.

#	Sentence	Aspect	Label	MASGCN	DRGAT	HCGAT
1	Certainly not the best sushi in New York, however, it is always fresh, and the **place** is very clean, sterile.	place	Pos	Pos ✓	Neu ×	Pos ✓
2	While there's a decent **menu**, it shouldn't take ten minutes to get your **drinks** and 45 for a **dessert pizza**.	menu	Pos	Pos ✓	Pos ✓	Pos ✓
		drinks	Neu	Neu ✓	Neg ×	Neu ✓
		dessert pizza	Neu	Neg ×	Neg ×	Neu ✓
3	I complained to the **waiter** and then to the **manager**, but the intensity of rudeness from them just went up.	waiter	Neg	Neg ✓	Neg ✓	Neg ✓
		manager	Neg	Neu ×	Neg ✓	Neg ✓

learning. However, when λ exceeds 0.12, the contrastive loss dominates training and interferes with the main classification task. When λ is too small, the contrastive learning effect becomes insufficient.

The temperature parameter τ controls the sharpness of softmax distribution in contrastive learning. Figure 2(b) shows that τ=0.07 achieves optimal performance on all three datasets, demonstrating good cross-dataset stability. When τ is too low (e.g., 0.05), the softmax distribution becomes overly sharp, causing the model to focus too much on hard negatives and leading to unstable training. When τ is too high ($\geq$0.10), the distribution becomes too smooth, weakening the discriminability between positive and negative pairs. Notably, the model shows robustness within $\tau \in [0.07, 0.10]$, with F1 fluctuations less than 0.5 percentage points across all datasets.

F. Case Study

To qualitatively demonstrate the effectiveness of HCGAT, we select three representative challenging scenarios from the Restaurant14 test set that highlight different advantages of our model. These cases cover contrastive structures, multi-aspect mixed sentiment, and long-distance dependencies—common challenges in ABSA tasks. We compare HCGAT's predictions with two strong baselines MASGCN and DRGAT. Table V presents the comparison results.

The prediction results are compared with MASGCN and DRGAT. Through observation and analysis, we draw the following conclusions:

Case 1 - Contrastive structure: The sentence contains "however" marking a contrast between negative sentiment about sushi and positive sentiment about place. MASGCN correctly identifies the positive sentiment, while DRGAT incorrectly predicts neutral, indicating difficulty in resolving the contrastive structure. HCGAT successfully captures the positive sentiment through hierarchical modeling that effectively distinguishes the two clauses.

Case 2 - Multi-aspect mixed sentiment: The sentence contains three aspects with different polarities. While "menu" has explicit positive modifier "decent", the other two aspects appear in an implicit complaint context. MASGCN correctly handles "menu" and "drinks" but fails on "dessert pizza" due to sentiment spillover. DRGAT only correctly predicts "menu", failing on the other two aspects. HCGAT achieves perfect prediction through aspect-aware pooling that effectively isolates sentiment for each aspect.

Case 3 - Long-distance dependency: The sentiment descriptor "rudeness" appears 11+ tokens away from aspects "waiter" and "manager". DRGAT successfully captures this long-distance dependency for both aspects via dual syntactic trees. MASGCN correctly predicts "waiter" but fails on "manager", predicting neutral instead of negative. HCGAT succeeds through 2-hop adjacency matrices that extend the receptive field to capture distant sentiment-aspect associations.

V. CONCLUSION

In this paper, we propose Hierarchical Contrastive Graph Attention Network (HCGAT), which addresses aspect-based sentiment analysis through multi-granular syntactic modeling and cross-granularity consistency contrastive learning. Experimental results on three benchmark datasets demonstrate that HCGAT achieves state-of-the-art performance, outperforming the strongest baseline by 1.20% in accuracy and 1.69% in F1 on average. Our experimental analysis reveals several valuable findings: (1) Syntactic information at different granularities exhibits complementary effects, with progressive addition of word-level, phrase-level, and sentence-level modeling yielding consistent performance improvements; (2) The contribution of cross-granularity consistency contrastive learning is negatively correlated with data scale, showing more significant regularization effects in data-scarce scenarios; (3) Adaptive gating fusion outperforms simple concatenation, indicating that different samples have varying dependencies on information from each granularity. These findings provide new perspectives for effective utilization of syntactic information in ABSA tasks.

VI. LIMITATIONS AND FUTURE WORK

This study has several limitations. First, the model relies on external dependency parsing tools, and parsing errors may affect performance on informal texts. Second, experiments

are conducted only on English datasets, while sentiment expressions are often closely related to language characteristics and cultural backgrounds, leaving the model's generalization ability on other languages to be verified. Additionally, the current granularity division is predefined and may not adapt to all syntactic structures.Future work will explore leveraging implicit syntactic knowledge from pre-trained language models to reduce dependence on external parsers, and extend the method to multilingual scenarios to verify its cross-lingual applicability.

REFERENCES

[1] M. Wankhade, C. Kulkarni, and A. C. S. Rao, "A survey on aspect base sentiment analysis methods and challenges," *Applied Soft Computing*, vol. 167, p. 112249, 2024.

[2] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective lstms for target-dependent sentiment classification," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers*, 2016, pp. 3298–3307.

[3] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 452–461.

[4] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 4568–4578.

[5] K. Wang, W. Shen, Y. Yang, X. Quan, and R. Wang, "Relational graph attention network for aspect-based sentiment analysis," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 3229–3238.

[6] H. Tang, D. Ji, C. Li, and Q. Zhou, "Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 6578–6588.

[7] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 05, 2020, pp. 8600–8607.

[8] X. Huang, H. Peng, S. Sun, Z. Hao, H. Lin, and S. Wang, "Multi-view attention syntactic enhanced graph convolutional network for aspect-based sentiment analysis," *arXiv preprint arXiv:2501.15968*, 2025.

[9] N. Chomsky, *Syntactic structures*. Walter de Gruyter, 2002.

[10] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606–615.

[11] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017, pp. 4068–4074.

[12] F. Fan, Y. Feng, and D. Zhao, "Multi-grained attention network for aspect-level sentiment classification," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 3433–3442.

[13] M. Zhang and T. Qian, "Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis," in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 3540–3549.

[14] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643, 2022.

[15] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 499–508.

[16] X. Shi, M. Hu, F. Ren, P. Shi, and S. Nakagawa, "Aspect based sentiment analysis with instruction tuning and external knowledge enhanced dependency graph," *Applied Intelligence*, vol. 54, no. 8, pp. 6415–6432, 2024.

[17] J. Chen, H. Fan, and W. Wang, "Syntactic and semantic aware graph convolutional network for aspect-based sentiment analysis," *IEEE Access*, vol. 12, pp. 22 500–22 509, 2024.

[18] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[19] T. Gao, X. Yao, and D. Chen, "Simcse: Simple contrastive learning of sentence embeddings," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 6894–6910.

[20] B. Gunel, J. Du, A. Conneau, and V. Stoyanov, "Supervised contrastive learning for pre-trained language model fine-tuning," *arXiv preprint arXiv:2011.01403*, 2020.

[21] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "Sentilare: Sentiment-aware language representation learning with linguistic knowledge," in *proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 6975–6988.

[22] B. Liang, W. Luo, X. Li, L. Gui, M. Yang, X. Yu, and R. Xu, "Enhancing aspect-based sentiment analysis with supervised contrastive learning," in *Proceedings of the 30th ACM international conference on information & knowledge management*, 2021, pp. 3242–3247.

[23] P. Li, P. Li, and X. Xiao, "Aspect-pair supervised contrastive learning for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 274, p. 110648, 2023.

[24] Y. Huang, A. Dai, S. Cao, Q. Kuang, H. Zhao, and Q. Cai, "A dual contrastive learning-based graph convolutional network with syntax label enhancement for aspect-based sentiment classification," *Frontiers in Physics*, vol. 12, p. 1336795, 2024.

[25] X. Shi, M. Hu, F. Ren, and P. Shi, "Prompted representation joint contrastive learning for aspect-based sentiment analysis," *Knowledge-Based Systems*, vol. 285, p. 111345, 2024.

[26] A. K. Zarandi and S. Mirzaei, "Enhancing aspect-based sentiment analysis through graph attention networks and supervised contrastive learning," *Multimedia Tools and Applications*, vol. 84, no. 25, pp. 30 045–30 080, 2025.

[27] D. Huang, L. Ren, and Z. Li, "Enhancing aspect-based sentiment analysis with linking words-guided emotional augmentation and hybrid learning," *Neurocomputing*, vol. 612, p. 128705, 2025.

[28] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," in *8th International Workshop on Semantic Evaluation August 23-24, 2014.*, 2014, pp. 27–35.

[29] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, 2014, pp. 49–54.

[30] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[31] C. Chen, Z. Teng, and Y. Zhang, "Inducing target-specific latent structures for aspect sentiment classification," in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, 2020, pp. 5596–5607.

[32] L. You, J. Peng, H. Jin, C. Claramunt, H. Zeng, and Z. Zhang, "Drgat: Dual-relational graph attention networks for aspect-based sentiment classification," *Information Sciences*, vol. 668, p. 120531, 2024.