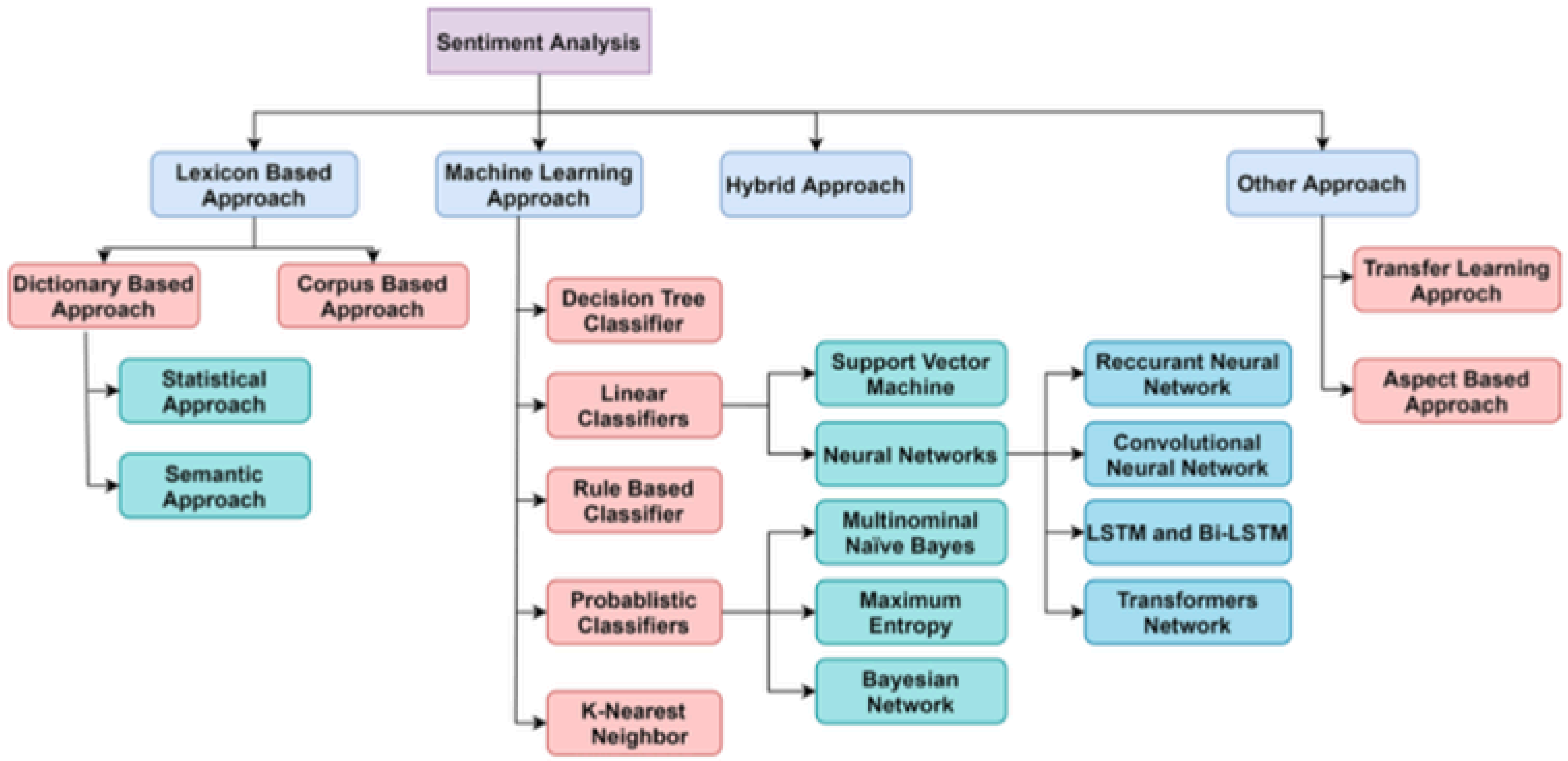
Aspect-Based Sentiment Analysis of Reviews for Requirements Elicitation

Vishal Sudharaka Thenuwara (239364U)
Supervisor: Dr. Nisansa de Silva
Dept. of Computer Science & Engineering, University of Moratuwa
July 2026

Presentation Outline

1. Introduction & Motivation
2. Problem Statement
3. Research Objectives
4. Literature Review & Research Gap
5. Proposed Methodology — DESS
6. Proposed Methodology — SGESS (Span-Level Graph)
7. Experimental Setup
8. Main Results & Comparisons
9. Ablation Study & Design Lessons
10. Dimensional Sentiment Analysis Extension
11. Error Analysis & Limitations
12. Contributions & Conclusion
13. Future Work

Introduction



The Problem:

- "The food was great but the service was terrible" — one sentence, two opinions, two aspects
- Traditional sentiment analysis gives ONE label for this

Evolution:

Document-level → Sentence-level → Aspect-level

ASTE (Aspect Sentiment Triplet Extraction):

- Extracts: (Aspect, Opinion, Sentiment)[23]

- "The price is **reasonable** but the service is **poor**"
→ (price, reasonable, POS)
→ (service, poor, NEG)

- [23] H. Peng et al., "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," *AAAI*, 2020. *(First formalization of ASTE task)*

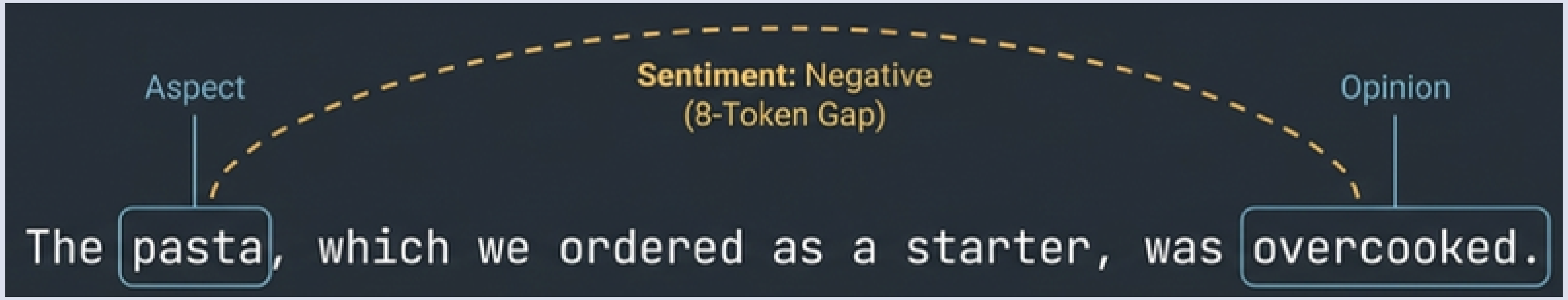
Motivation: Why This Research?

Three weaknesses in current ASTE models:

1. Weak encoder [3]
2. Syntax OR semantics [8]
3. Isolated spans [36]

Our response:

- DeBERTa-v3 → separates content from position
- Dual-channel → syntax + semantics in parallel
- Span-Level Graph → spans talk before prediction



- [3] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL*, 2019.
- [8] P. He et al., "DeBERTa: Decoding-enhanced BERT with disentangled attention," *arXiv:2006.03654*, 2020.
- [36] X. Zhao et al., "Dual Encoder: Exploiting the potential of syntactic and semantic for aspect sentiment triplet extraction," *arXiv:2402.15370*, 2024. (D2E2S)

Problem Statement & Task Definition

Problem Statement

"Customer reviews contain multiple opinions about multiple features in a single sentence."

Task Definition

Task: Given sentence $X = \{x_1, ..., x_n\}$, extract triplets $T = \{(a_k, o_k, s_k)\}$

"The pasta, which we ordered as a starter, was overcooked."

Component	Meaning	Example
a_k	Aspect span (target feature)	"pasta"
o_k	Opinion span (evaluative expression)	"overcooked"
s_k	Sentiment polarity	negative

- [23] H. Peng et al., "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," *AAAI*, 2020
- [26] L. Xu et al., "Learning span-level interactions for aspect sentiment triplet extraction," *arXiv:2107.12214*, 2021. *(Span-ASTE)*

Research Objectives

- R1 : Replace BERT with DeBERTa-v3 for improved disentangled attention and long-range dependency modelling
- R2 : Build a dual-channel architecture integrating syntactic GCN and semantic GCN for richer feature extraction
- R3 : Design the TIN (Token Interaction Network) module for adaptive syntax-semantics feature fusion
- R4 : Enable inter-span communication via span-level graph reasoning before final sentiment classification
- R5 : Document and report negative experimental results to provide reproducible insights for the research community

- [8] P. He et al., "DeBERTa: Decoding-enhanced BERT with disentangled attention," 2020.
- [6] R. Li et al., "Dual graph convolutional networks for aspect-based sentiment analysis," *ACL*, 2021. *(DualGCN)*
- [36] X. Zhao et al., "Dual Encoder (D2E2S)," 2024.
- [26] L. Xu et al., "Learning span-level interactions for aspect sentiment triplet extraction (Span-ASTE)," 2021.

Literature Review: Evolution of ABSA

01 — **Early Methods**
LSTM + Attention sequentials[1,2]

02 — **BERT Era**
BERT-PT, RGAT-BERT, DualGCN-BERT [3,7]

03 — **Advanced Encoders**
RoBERTa & DeBERTa [8]

04 — **Span-Based:**
Span-ASTE, SSJE, SBN [26,35]

05 — **GNN-Based:**
EMC-GCN, D2E2S, STAGE-3D[33]

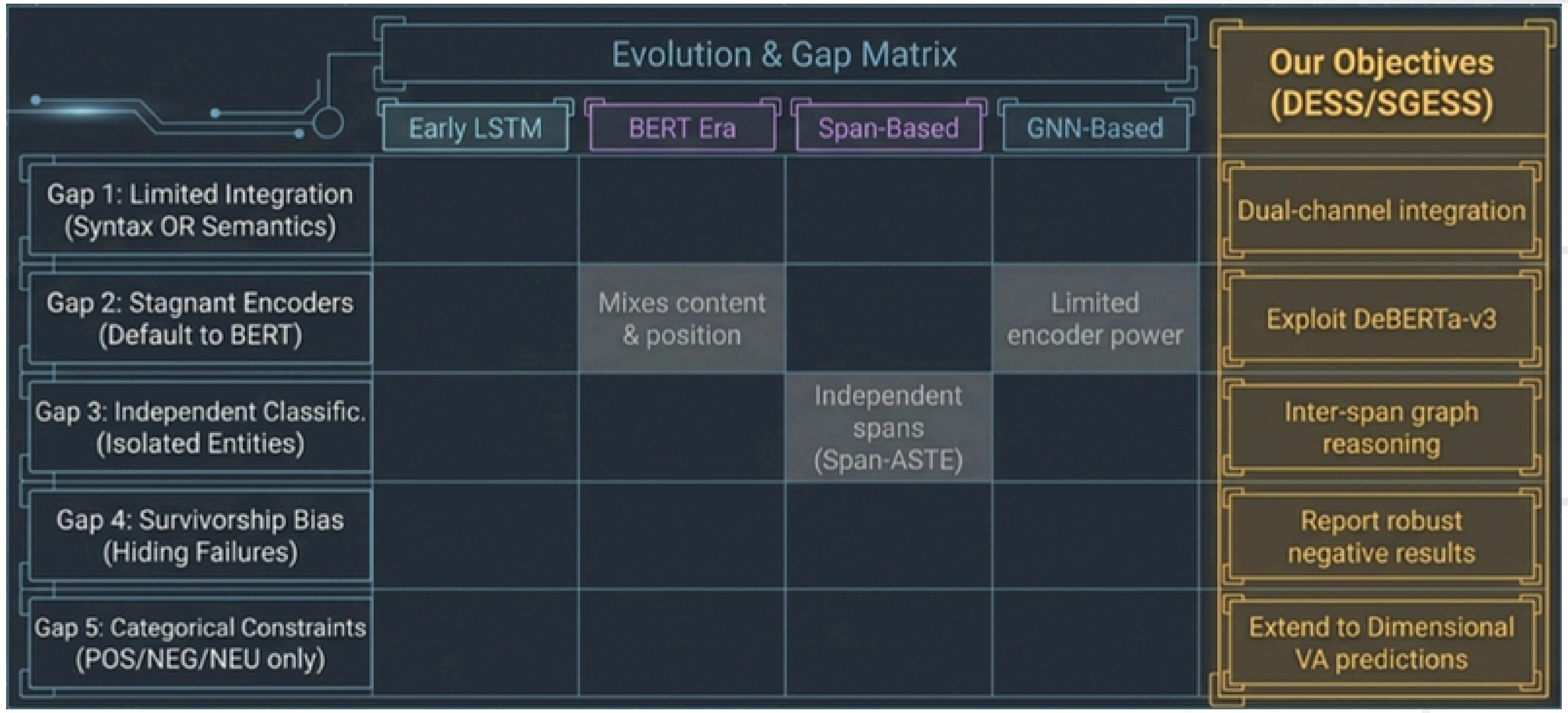
Our Position:
Best encoder (DeBERTa-v3) + Dual GCN (syntax + semantics) + Span-Level Graph Reasoning

First Publication - Fourth Author - Published

Instruct-DeBERTa: A hybrid approach for aspect-based sentiment analysis on textual reviews
..., AVA Malkith, K Isuranda, V Thenuwara... - arXiv preprint arXiv ..., 2024 - arxiv.org

- [1] Y. Wang et al., "Attention-based LSTM for aspect-level sentiment classification," *EMNLP*, 2016. *(ATAE-LSTM)*
- [2] X. Li et al., "Transformation networks for target-oriented sentiment classification," *ACL*, 2018. *(TNet)*
- [3] J. Devlin et al., "BERT," 2019.
- [7] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv:1907.11692*, 2019.
-
- [8] P. He et al., "DeBERTa," 2020.
- [26] L. Xu et al., "Span-ASTE," 2021.
- [33] H. Chen et al., "Enhanced multi-channel graph convolutional network for ASTE," *ACL*, 2022. *(EMC-GCN)*
- [35] S. Liang et al., "STAGE: Span tagging and greedy inference scheme for ASTE," *AAAI*, 2023. *(STAGE-3D)*

Research Gap



1. Few models combine syntax AND semantics in a unified framework[6]
2. Nearly every ASTE method uses BERT — DeBERTa unexplored [8]
3. Literature only reports successes → wasted effort on dead ends
4. Span-based methods: no inter-span communication before sentiment[26]
5. No work adapting ASTE for continuous Valence-Arousal prediction

- [6] R. Li et al., "Dual graph convolutional networks (DualGCN)," *ACL*, 2021.
- [8] P. He et al., "DeBERTa," 2020.
-
- [26] L. Xu et al., "Span-ASTE," 2021.
- [36] X. Zhao et al., "D2E2S," 2024.

Methodology:DESS Architecture Overview

Pipeline Flow:

① **Input Sentence** → DeBERTa-v3 Encoder (disentangled attention)

② **BiLSTM Layers** → Sequential context encoding over token representations

③ **Dual-Channel GCN:**
- Syntactic GCN
- Semantic GCN

④ **TIN Feature Fusion** : merges syntactic & semantic channels

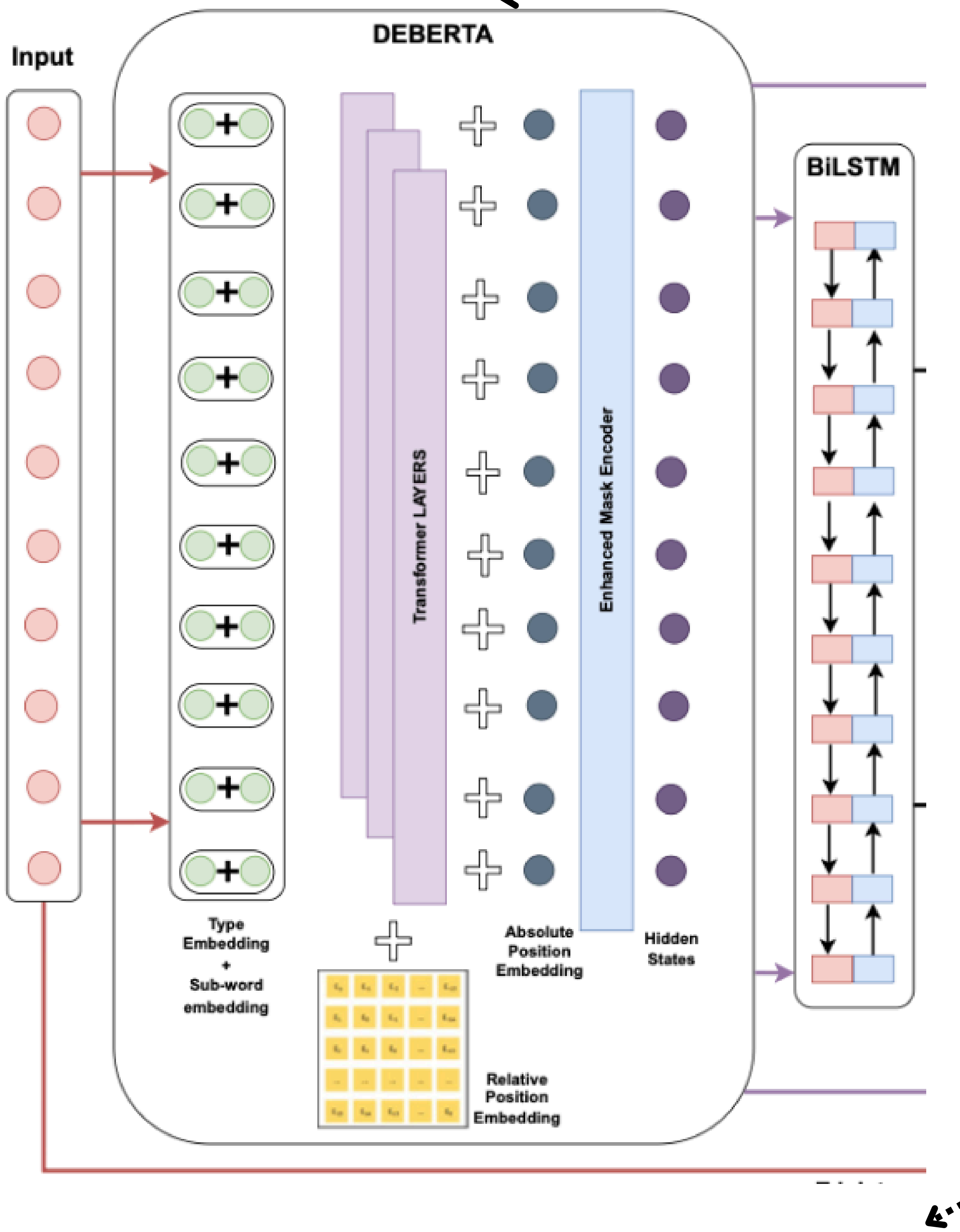
⑤ **Span-Based Classification** → Entity labels + Sentiment polarity per span

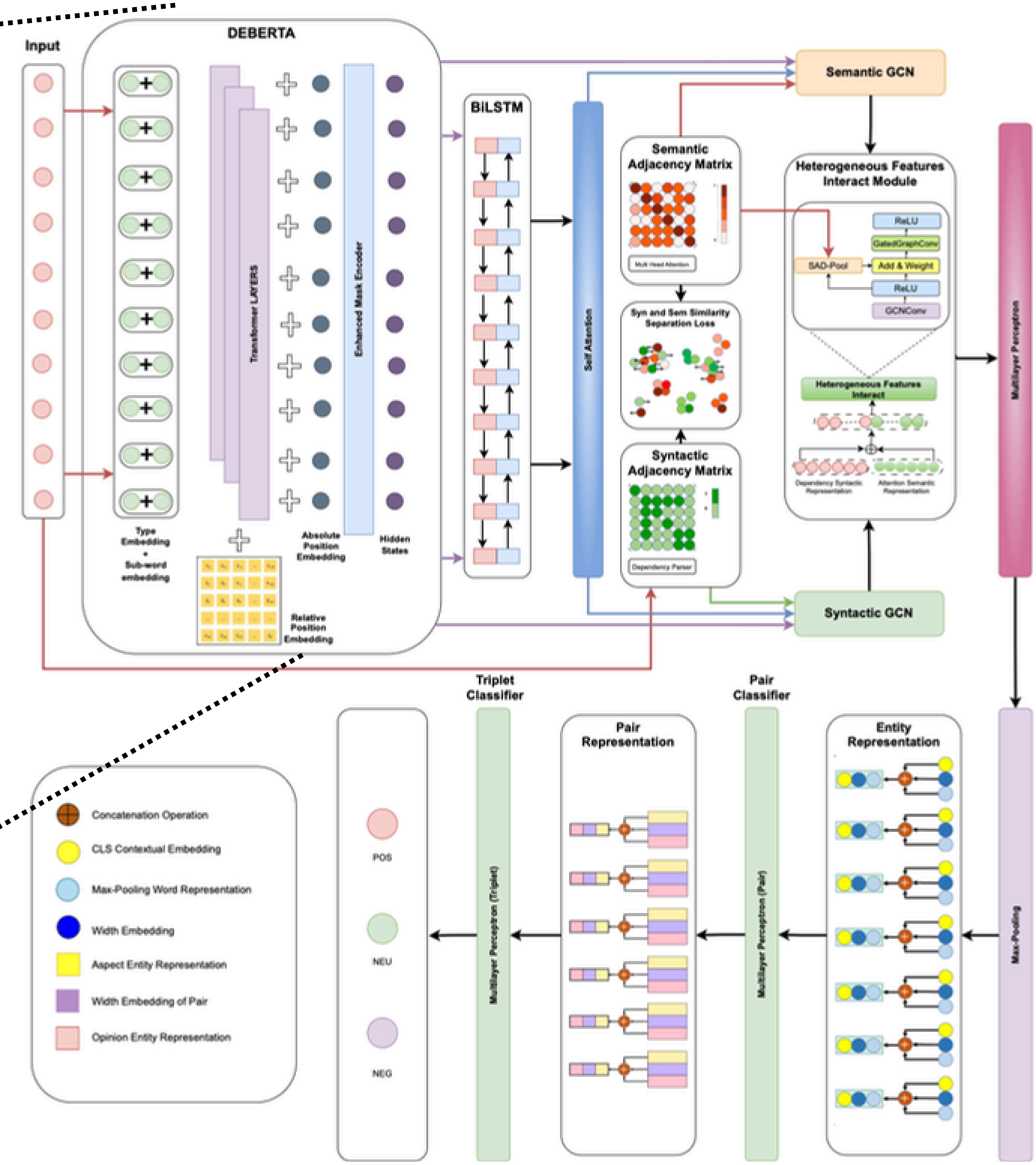
Second Publication - First Author - Published

DESS: DeBERTa Enhanced Syntactic-Semantic Aspect Sentiment Triplet Extraction
V Thenuwara, N de Silva - International Conference on Computational ..., 2025 - Springer

- [8] P. He et al., "DeBERTa," 2020. *(Encoder)*
- [36] X. Zhao et al., "D2E2S," 2024. *(Base dual-channel architecture we extend)*
- [6] R. Li et al., "DualGCN," 2021. *(Dual GCN inspiration)*

Methodology: DESS Architecture Overview



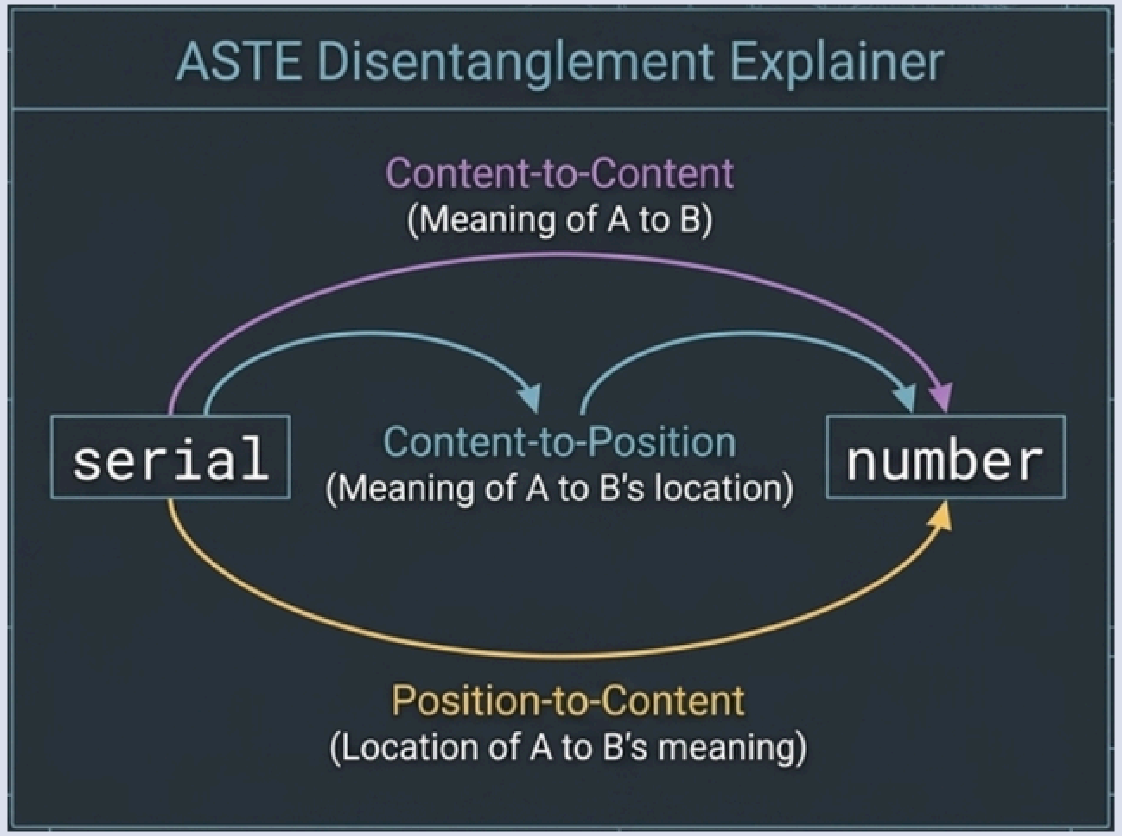


Methodology: DeBERTa Encoder - Why It Matters

DeBERTa introduces disentangled attention with three separate components:

- Content-to-Content
- Content-to-Position
- Position-to-Content

Feature	BERT	DeBERTa-v3
Attention	Single score mixing content + position	Three separate interactions
Position Encoding	Absolute position added to embeddings	Disentangled relative position
Efficacy	General NLU	Long-range dependencies



- [3] J. Devlin et al., "BERT," 2019.
- [8] P. He et al., "DeBERTa: Decoding-enhanced BERT with disentangled attention," 2020.

DESS :Results

Model	14LAP			14RES			15RES			16RES		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+ [22]	30.10	36.90	33.20	39.20	39.80	37.00	41.30	42.10	41.70	41.3	42.1	41.7
RINANTE+ [19]	21.70	18.70	20.10	29.90	30.10	30.00	25.70	22.30	23.90	25.7	22.3	23.9
WhatHowWhy [19]	37.38	50.38	42.87	43.24	48.07	63.66	51.46	57.51	52.32	46.96	64.24	54.21
TOP	57.84	59.33	58.58	54.53	63.59	73.44	68.16	63.30	58.59	63.57	71.98	67.52
Li-unified-R [19]	40.60	43.40	42.30	44.70	51.40	47.80	37.30	54.50	44.30	40.60	44.30	42.30
Peng-two-stage [16]	37.40	50.40	43.90	48.10	57.50	52.30	47.00	64.20	54.20	37.40	50.40	43.90
ChatGPT	35.10	48.90	41.80	41.90	60.60	49.50	44.90	61.70	52.00	35.10	48.90	41.80
JET-BERT [33]	55.40	47.30	51.30	64.50	52.00	57.00	70.40	58.40	64.80	55.40	47.30	51.30
PASTE [17]	55.00	51.60	53.80	58.30	56.70	57.50	65.50	64.40	64.90	55.00	51.60	53.80
JETt [33]	51.48	42.65	46.65	70.20	53.02	60.41	62.14	47.25	53.68	72.12	57.20	63.41
JETo [33]	58.47	43.67	50.00	67.97	60.32	63.92	58.35	51.43	54.67	64.77	61.29	62.98
Unified-BART-ABSA	65.40	65.00	65.20	59.10	59.40	59.30	66.60	68.70	67.70	65.40	65.00	65.20
BMRC [13]	70.60	61.80	66.20	68.50	53.40	60.10	71.20	61.10	66.10	70.60	61.80	66.20
PARAPHRASE	71.10	71.90	71.50	60.40	63.90	62.70	70.10	73.90	72.50	71.10	71.90	71.50
SBSK-ASTE [4]	74.60	71.50	73.10	65.30	63.70	64.50	70.80	72.00	71.40	74.60	71.50	73.10
Span-BiDir [16]	**76.40**	72.40	74.30	69.90	60.40	64.80	71.60	72.60	72.10	76.40	72.40	74.30
Unified	61.41	56.19	58.69	65.52	64.99	65.25	59.14	59.38	59.26	66.60	68.68	67.62
BARTABSA	61.41	56.19	58.69	65.52	64.99	65.25	59.14	59.38	59.26	66.6	68.68	67.62
GAS	65.00	69.50	67.20	56.10	61.80	58.80	66.10	68.70	67.40	65.00	69.50	67.20
GTS-BERT [23]	59.40	51.94	55.42	68.09	69.54	68.81	59.28	57.93	58.60	68.32	66.86	67.58
Double-Encoder [21]	62.12	56.38	59.11	67.95	71.23	69.55	58.55	60.00	59.27	70.65	70.23	70.44
TGA+SFI [21]	65.25	53.79	58.98	71.75	70.52	71.13	62.77	59.79	61.25	68.20	69.26	68.73
EMC-GCN [1]	61.70	56.30	58.80	62.50	62.50	62.50	65.60	71.30	68.40	61.70	56.30	58.80
SCEDD	61.84	60.08	60.95	70.27	73.02	71.62	59.41	62.73	61.03	66.11	71.37	68.64
SA-Transformer [28]	61.28	48.98	54.44	70.76	65.85	68.22	62.82	58.31	60.48	72.01	62.87	67.13
STAGE-3D [12]	71.98	53.86	61.58	78.58	69.58	73.76	73.63	57.90	64.79	76.67	70.12	73.24
Chen-dual-decoder [15]	62.36	60.37	61.35	72.12	73.14	72.62	64.27	60.73	62.45	68.74	71.79	70.23
OTE-MTL [30]	49.20	40.50	44.60	57.90	42.70	48.90	60.30	53.40	56.50	49.20	40.50	44.60
Seq2path [14]	64.57	60.04	62.22	73.28	74.23	73.75	62.62	65.48	64.02	71.59	75.41	73.40
TAGS [15]	65.11	62.20	64.53	77.38	72.86	75.05	70.23	65.73	67.90	76.37	76.85	76.61
BDTF [31]	68.94	55.97	61.74	75.53	73.24	74.35	68.76	63.71	66.12	71.44	73.13	72.27
Sem-Dual Encoder [10]	65.98	58.78	62.17	77.14	75.35	76.23	68.07	66.80	67.43	71.90	76.65	74.20
Dual-MRC [15]	57.39	53.88	55.58	71.55	69.14	70.32	63.78	51.87	57.21	68.60	66.24	67.40
Span-ASTE [24]	63.40	55.80	59.10	62.20	64.50	63.30	69.50	71.20	70.30	63.40	55.80	59.10
SSJE [11]	67.43	54.71	60.41	73.12	71.43	72.26	63.94	66.17	65.05	70.82	72.00	71.38
SBN [2]	65.68	59.88	62.65	76.36	72.43	74.34	69.93	60.41	64.82	71.59	72.57	72.08
ASTE-Base	63.50	64.50	64.00	61.60	65.40	63.00	69.50	75.10	72.00	63.50	64.50	64.00
ASTE-RL [27]	64.80	54.99	59.50	70.60	68.65	69.71	65.45	60.29	62.72	67.21	69.69	68.41
CONTRASTE-Base [16]	72.40	73.20	72.70	62.60	67.20	64.80	72.10	73.90	73.00	72.40	73.20	72.70
CONTRASTE-MTL [16]	73.60	74.40	**74.00**	65.30	66.70	66.10	72.20	**76.30**	74.20	73.60	74.40	74.00
RLI [26]	63.32	57.43	60.96	77.46	71.97	74.34	60.08	70.66	65.41	70.50	74.28	72.34
COM-MRC [29]	75.46	68.91	72.01	62.35	58.16	60.17	68.35	61.24	64.53	71.55	71.59	71.57
D2E2S [32]	67.38	60.31	63.65	75.92	74.36	75.13	70.09	62.11	65.86	77.97	71.77	74.74
D2E2S†	65.70	56.30	60.64	70.09	75.28	72.59	62.24	62.11	62.18	70.99	72.96	71.96
EMC-GCN deberta-v3-base-absa-v1.1†	-	56.03	-	-	68.42	-	-	57.76	-	-	64.66	-
DESS-V3-Base(Ours),[86M]	72.91	63.17	67.69	77.77	76.73	77.24	70.43	68.53	69.46	72.24	**80.72**	76.74
DESS-V3-Large(Ours), [304M]	76.22	62.40	68.63	76.26	**78.06**	77.15	71.07	64.60	67.68	75.48	78.93	77.11
DESS-V2-xxLarge(Ours),[1.5B]	72.88	65.65	69.08	**80.78**	79.20	**79.98**	70.35	75.16	**74.22**	**80.72**	77.33	**77.16**

Methodology: Dual-Channel GCN Architecture

Channel 1: Syntactic GCN

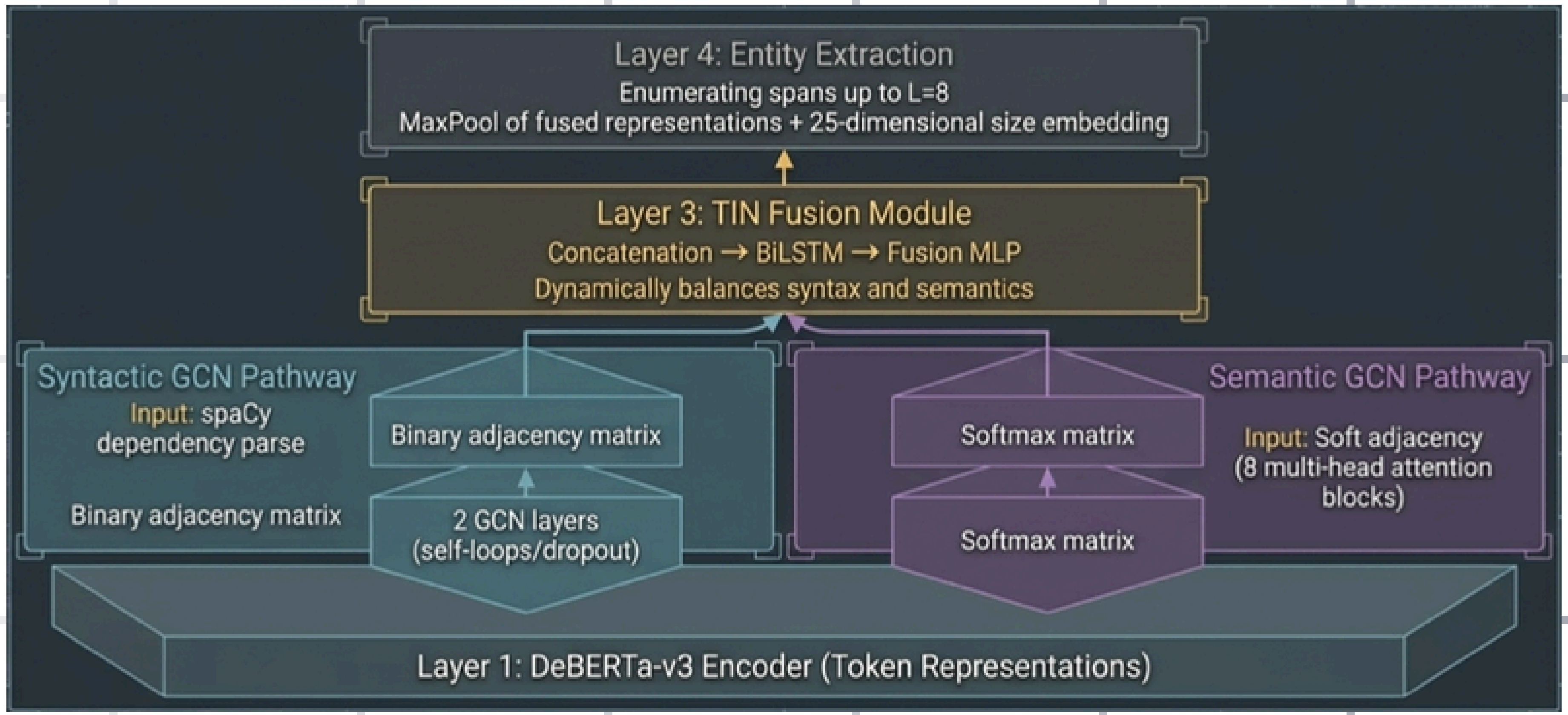
- Input: Dependency parse tree from spaCy
- Captures: grammatical relationships (subject, object, modifiers)

Channel 2: Semantic GCN

- Input: Learned soft adjacency via multi-head attention (8 heads)
- Captures: implicit semantic relationships beyond syntax

TIN Fusion

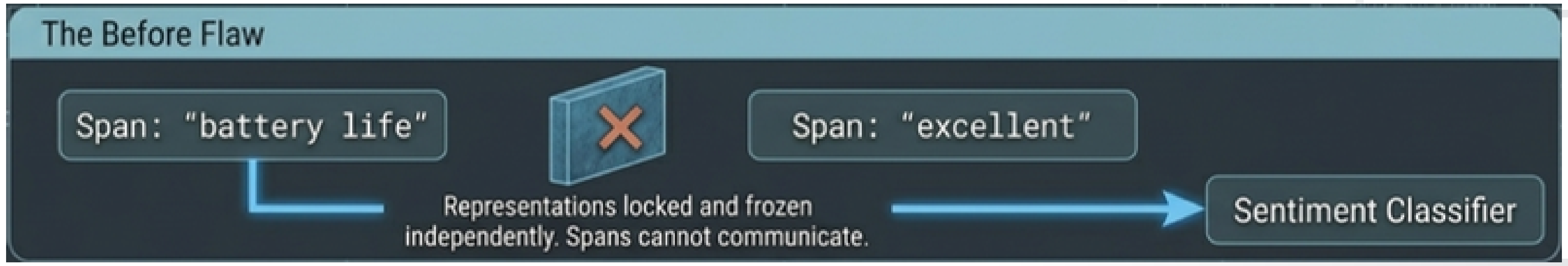
- Residual connections on each channel output
- Concatenation → BiLSTM → Fusion MLP
- Dynamically balances syntactic vs. semantic contributions

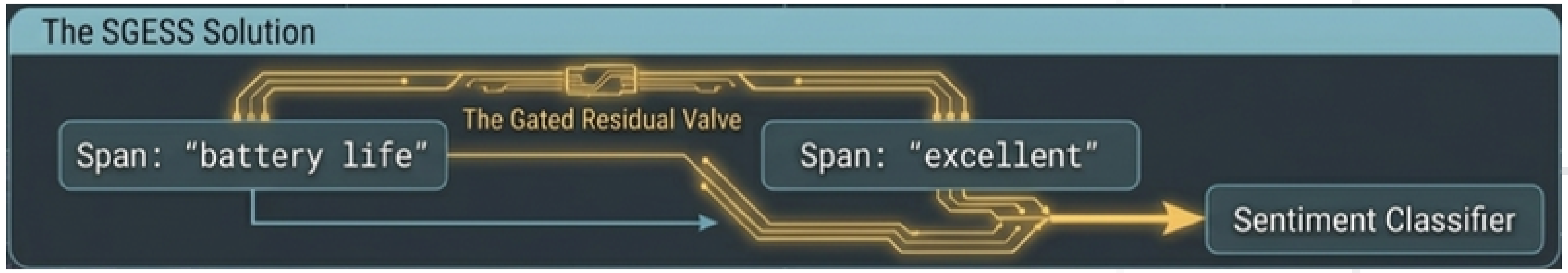


- [6] R. Li et al., "Dual graph convolutional networks for aspect-based sentiment analysis," *ACL*, 2021.
- [9] C. Zhang et al., "Aspect-based sentiment classification with aspect-specific graph convolutional networks," *EMNLP*, 2019. *(ASGCN)*
- [36] X. Zhao et al., "D2E2S," 2024.

Methodology:

Span-Based Entity & Sentiment Classification





Entity Classification:

- Enumerate spans up to L=8 tokens
- Classify → {Aspect, Opinion, None}

Sentiment Classification:

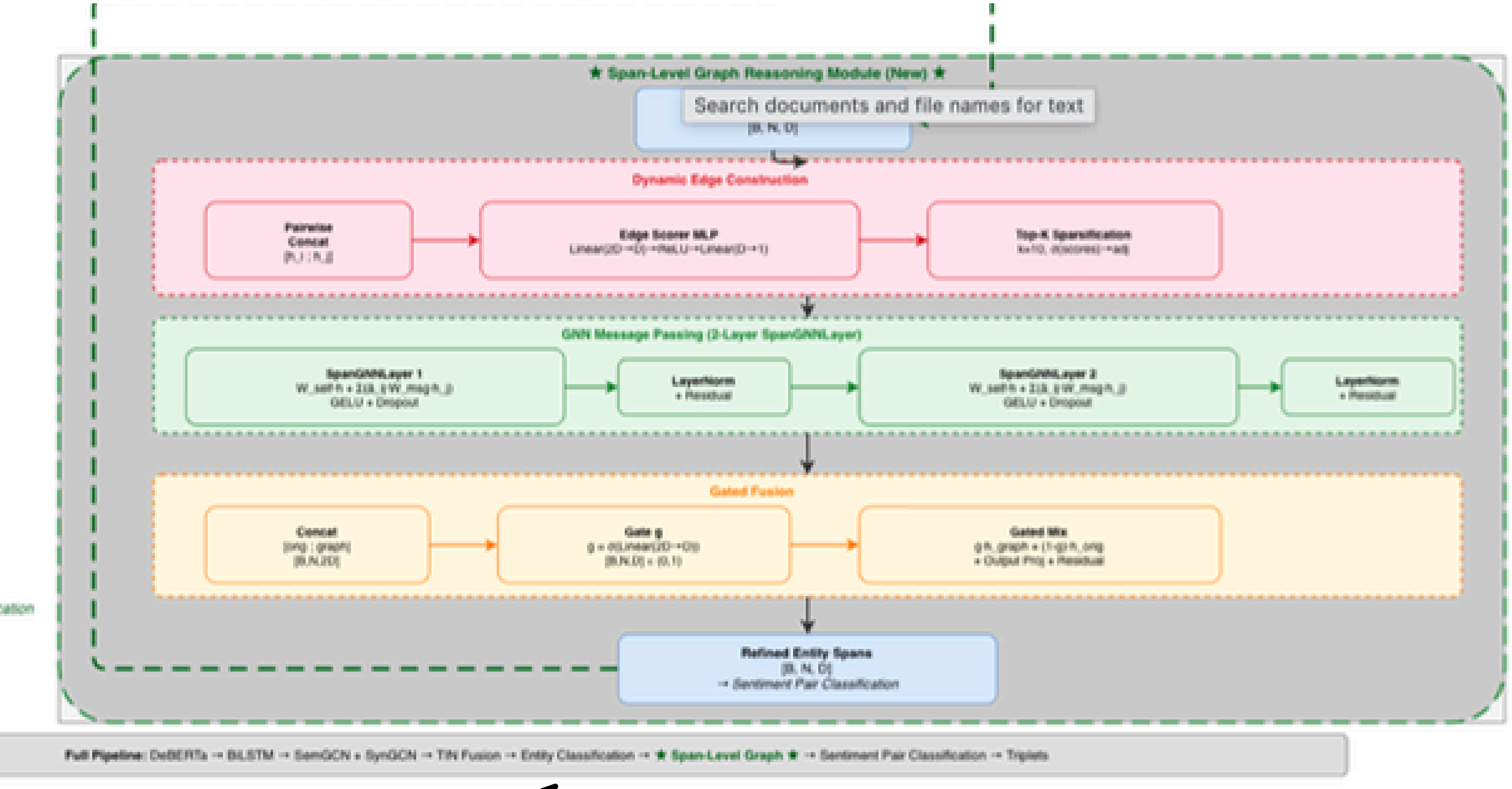
- For each (aspect, opinion) pair: concatenate representations + context + size
- Classify → {Positive, Negative, Neutral}[26]

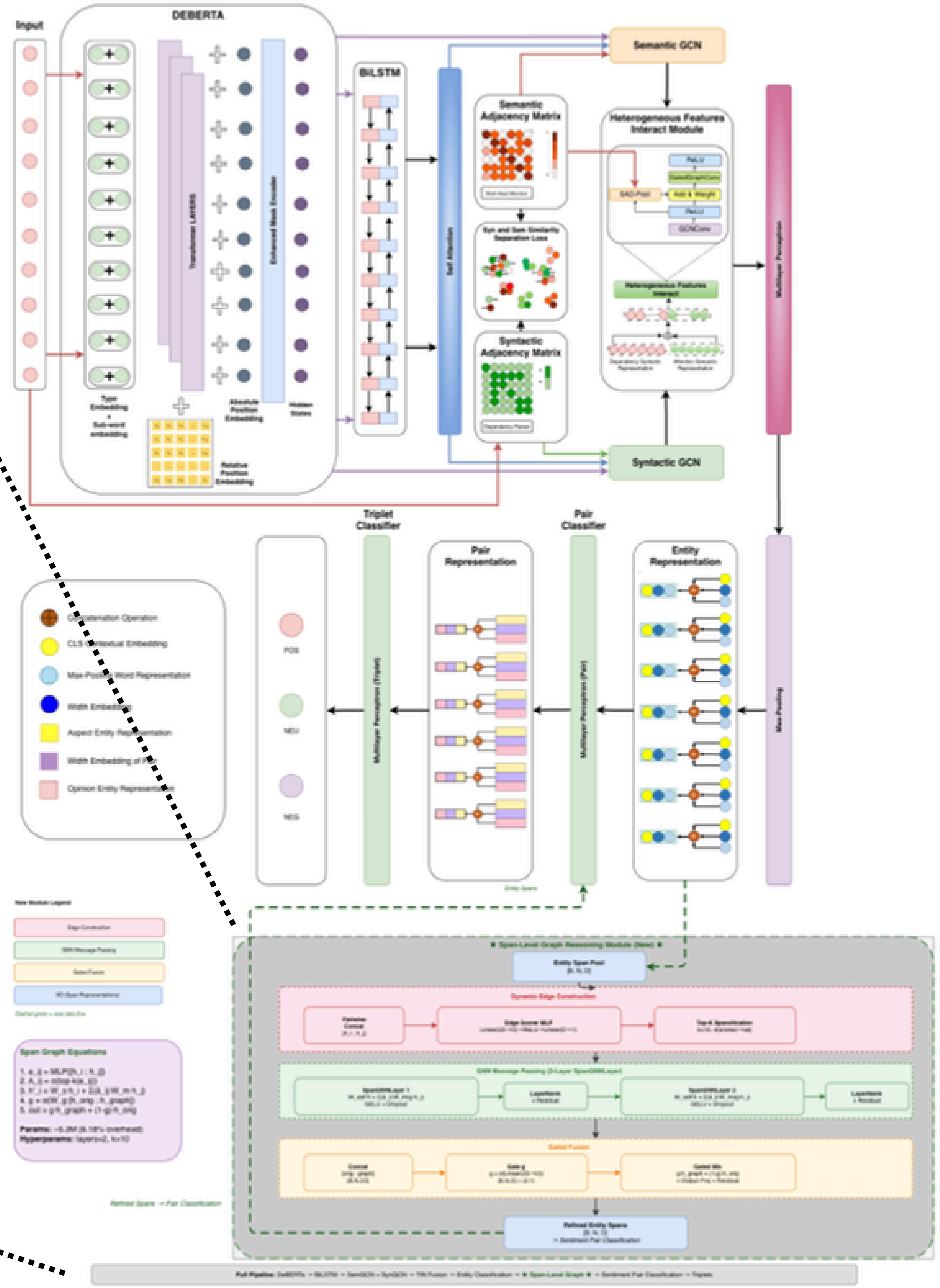
Training Loss :

- $\mathscr{L} = \mathscr{L}$_entity (CE) + $\mathscr{L}$_sentiment (BCE) + 10·$\mathscr{L}$_batch (KL divergence)[36]
- KL term aligns syntactic and semantic adjacency matrices

- [26] L. Xu et al., "Learning span-level interactions for aspect sentiment triplet extraction," 2021. *(Span-ASTE — span enumeration approach)*
- [36] X. Zhao et al., "D2E2S," 2024. *(KL divergence alignment term)*

Methodology:SGESS Architecture Overview





METHODOLOGY

SGESS — SPAN-LEVEL GRAPH REASONING

The Problem with Standard Pipeline:

- Classifies each span independently
- Representations are FROZEN before pairing

Key Design Principles:

- Gated residual ensures model CANNOT degrade below baseline
- Dynamic graph adapts to each sentence
- First span-level graph approach for ASTE

Our Solution :

Pooled Span Representations

↓

Dynamic Edge Construction (learned similarity)

↓

Top-k Sparsification (k=10)

↓

2 Rounds GNN Message Passing (spans share info)

↓

Gated Residual Fusion (safe mixing)

↓

Graph-Refined Spans → Sentiment Classifier

FOURTH PUBLICATION - FIRST AUTHOR - PENDING

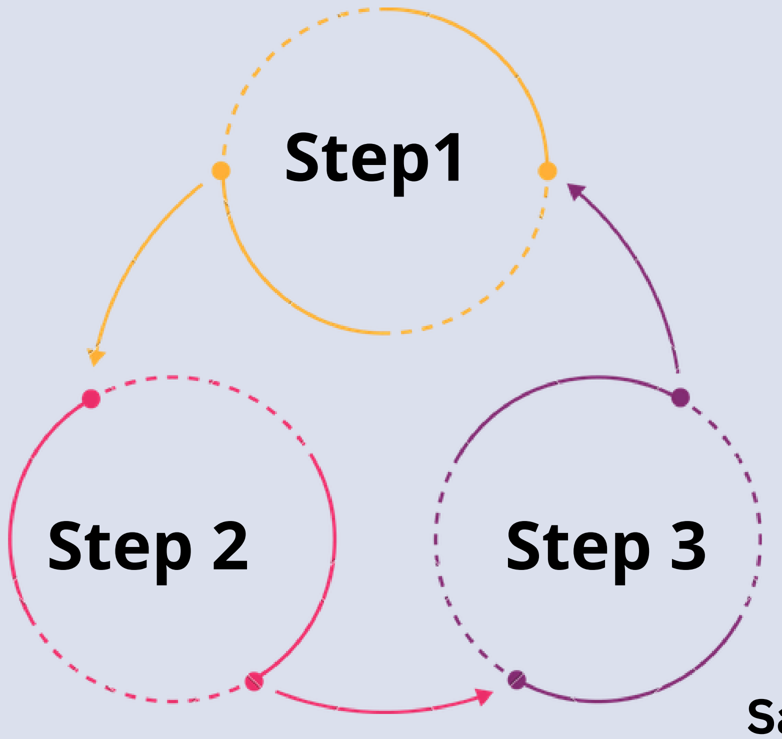
Span-Level Graph Enhanced Syntactic-Semantic Framework for Aspect Sentiment Triplet Extraction

- [26] L. Xu et al., "Span-ASTE," 2021. *(Baseline span-based pipeline we improve upon)*
- [27] Y. Li et al., "A span-sharing joint extraction framework for harvesting aspect sentiment triplets (SSJE)," *KBS*, 2022.
- [29] Y. Chen et al., "A span-level bidirectional network for aspect sentiment triplet extraction (SBN)," 2022.

Methodology: SGESS — Technical Details

- Concatenate every pair → feed through a small MLP → get a relevance score
- sparse graph
- Self-loops removed; padding masked out

Which spans should talk? (Dynamic Edge Construction)



Spans exchange information (GNN Message Passing × 2 rounds)

- Each span updates itself
- Residual + LayerNorm + GELU
- 2 rounds

Safe mixing back

- A learned gate decides
- If graph adds nothing useful → gate closes → model falls back
- Enriched span representations

- [33] H. Chen et al., "Enhanced multi-channel graph convolutional network for ASTE (EMC-GCN)," *ACL*, 2022. *(GNN for ASTE inspiration)*
- [35] S. Liang et al., "STAGE-3D," *AAAI*, 2023. *(Graph-structured ASTE approach)*

Experimental Setup

Datasets:

	Train	Test
14LAP (Laptop):	Train 906	Test 328
14RES (Restaurant):	Train 1266	Test 494
15RES (Restaurant):	Train 754	Test 325
16RES (Restaurant):	Train 857	Test 338

Training Hyper-Parameters:

Epochs:	Optimizer:	Gradient clipping:
120	AdamW	1.0

Learning Rates: 2 x 10^-5 (encoder) | 1 x 10^-4 (other)

Loss Function:
CE (Entity) + BCE (Sentiment) + KL divergence (lambda=10)

Model Scale and Compute Constraints

Model Scale & Compute Constraints

Tier 1:
V3-Base: 86M Params | ~10GB VRAM | 18 min/epoch (P100)

Tier 2:
V3-Large: 304M Params | ~24GB VRAM | 32 min/epoch (A100)

Tier 3:
V2-xxLarge: 1.5B Params | ~38GB VRAM | 65 min/epoch (A100)

- [23] H. Peng et al., "Knowing what, how and why," *AAAI*, 2020. *(ASTE-Data-V2 benchmark origin)*
- [26] L. Xu et al., "Span-ASTE," 2021. *(Dataset splits used)*

Main Results

Model	14LAP			14RES			15RES			16RES		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+ [37]	30.10	36.90	33.20	39.20	39.80	37.00	41.30	42.10	41.70	41.3	42.1	41.7
GTS-BERT [25]	59.40	51.94	55.42	68.09	69.54	68.81	59.28	57.93	58.60	68.32	66.86	67.58
EMC-GCN [33]	61.70	56.30	58.80	62.50	62.50	62.50	65.60	71.30	68.40	61.70	56.30	58.80
Span-ASTE [26]	63.40	55.80	59.10	62.20	64.50	63.30	69.50	71.20	70.30	63.40	55.80	59.10
SSJE [27]	67.43	54.71	60.41	73.12	71.43	72.26	63.94	66.17	65.05	70.82	72.00	71.38
STAGE-3D [35]	71.98	53.86	61.58	78.58	69.58	73.76	73.63	57.90	64.79	76.67	70.12	73.24
Sem-Dual Encoder [38]	65.98	58.78	62.17	77.14	75.35	76.23	68.07	66.80	67.43	71.90	76.65	74.20
Seq2path [31]	64.57	60.04	62.22	73.28	74.23	73.75	62.62	65.48	64.02	71.59	75.41	73.40
TAGS [39]	65.11	62.20	64.53	77.38	72.86	75.05	70.23	65.73	67.90	76.37	76.85	76.61
CONTRASTE-MTL [28]	73.60	74.40	**74.00**	65.30	66.70	66.10	72.20	**76.30**	74.20	73.60	74.40	74.00
D2E2S [36]	67.38	60.31	63.65	75.92	74.36	75.13	70.09	62.11	65.86	77.97	71.77	74.74
D2E2S†	65.70	56.30	60.64	70.09	75.28	72.59	62.24	62.11	62.18	70.99	72.96	71.96
DESS-V3-Base(Ours),[86M]	72.91	63.17	67.69	77.77	76.73	77.24	70.43	68.53	69.46	72.24	**80.72**	76.74
DESS-V3-Large(Ours), [304M]	76.22	62.40	68.63	76.26	**78.06**	77.15	71.07	64.60	67.68	75.48	78.93	77.11
DESS-V2-xxLarge(Ours),[1.5B]	72.88	65.65	69.08	**80.78**	79.20	**79.98**	70.35	75.16	**74.22**	**80.72**	77.33	**77.16**
SGESS-V2-xxLarge(Ours)	71.19	73.24	72.42	**81.73**	**79.78**	**80.75**	**79.23**	75.25	**77.19**	80.37	79.66	**80.01**

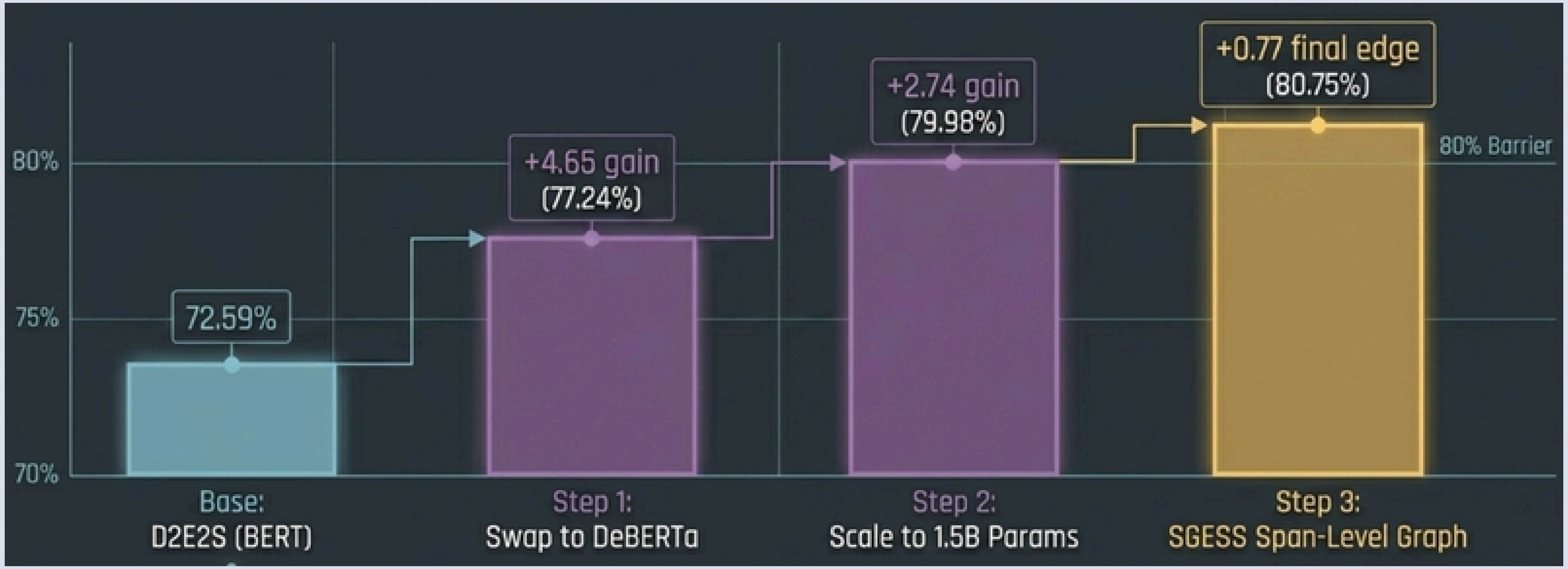
Key Observations:

- Surpasses 40+ prior methods
- Swap: +3–5 F1 points over D2E2S
- Span-Level Graph adds +2.8 to +3.6 F1 on top of DESS

- [38] B. Jiang et al., "A semantically enhanced dual encoder for ASTE," *Neurocomputing*, 2023. *(Sem-Dual Encoder — prev. SOTA 14RES)*
- [39] Y. Mao et al., "A joint training dual-MRC framework for ABSA," *AAAI*, 2021. *(TAGS — prev. SOTA 16RES)*
- [28] R. Mukherjee et al., "CONTRASTE: Supervised contrastive pre-training with aspect-based prompts for ASTE," 2023. *(prev. SOTA 15RES & 14LAP)*

Progressive Improvement Across Variants

F1 Scores on 14RES (progressive improvement):

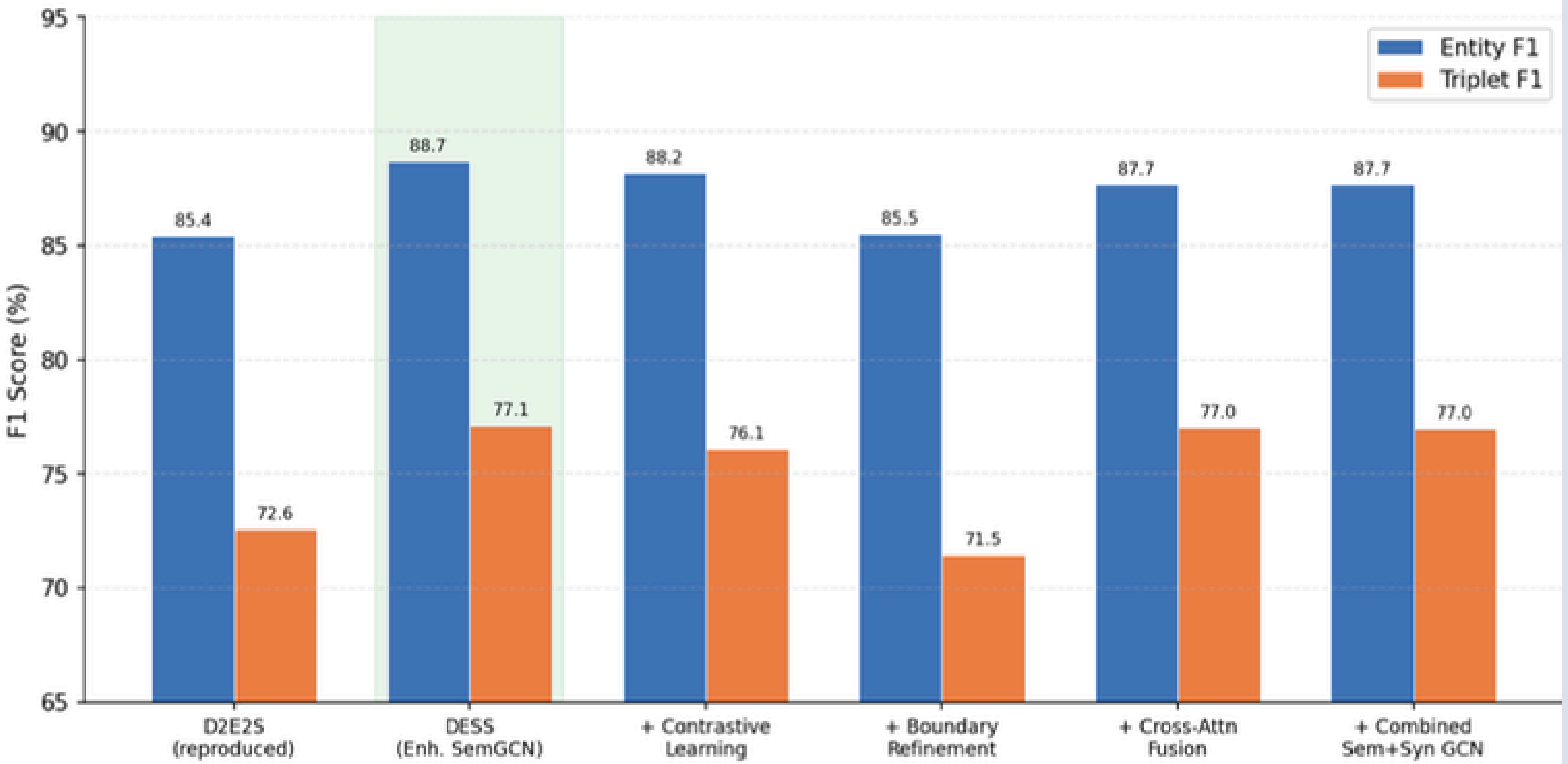


Takeaway:

- Simply replacing BERT → DeBERTa: +4.65 F1 points
- Scaling model size: +2.74 additional points
- Adding Span-Level Graph: +0.77 more points (on top of already-strong baseline)
- Total improvement over baseline: +8.16 F1 points

- [36] X. Zhao et al., "D2E2S," 2024. *(Baseline model we progressively improve)*
- [8] P. He et al., "DeBERTa," 2020. *(Encoder upgrade driving initial gains)*

Ablation Study
What Works and What Doesn't



Configuration	Ent. F1	Trip. F1	Δ
DESS (Enhanced SemGCN)	88.68	77.14	
+ Contrastive Learning	88.19	76.10	−1.04
+ Boundary Refinement	85.53	71.46	−5.68
+ Cross-Attn Fusion	87.66	77.01	−0.13
+ Combined Sem+Syn GCN	87.66	76.99	−0.15
+ Span-Level Graph	**88.90**	**80.75**	+3.61

Why the failures:

- Contrastive Learning - Extra loss competes with entity/sentiment losses
- Boundary Refinement: Massive instability; clashes with DeBERTa's existing positional encoding
- Cross-Attention Fusion: Learned attention converges to near-uniform

- [28] R. Mukherjee et al., "CONTRASTE," 2023. *(Contrastive learning approach for ASTE — inspired our experiment)*
- [36] X. Zhao et al., "D2E2S," 2024. *(Base architecture for ablation)*

Design Principle
Lessons from Negative Results

Key insight:

- With a strong encoder + good GCN, stacking modules causes optimization conflicts not gains.

Pattern of failure:

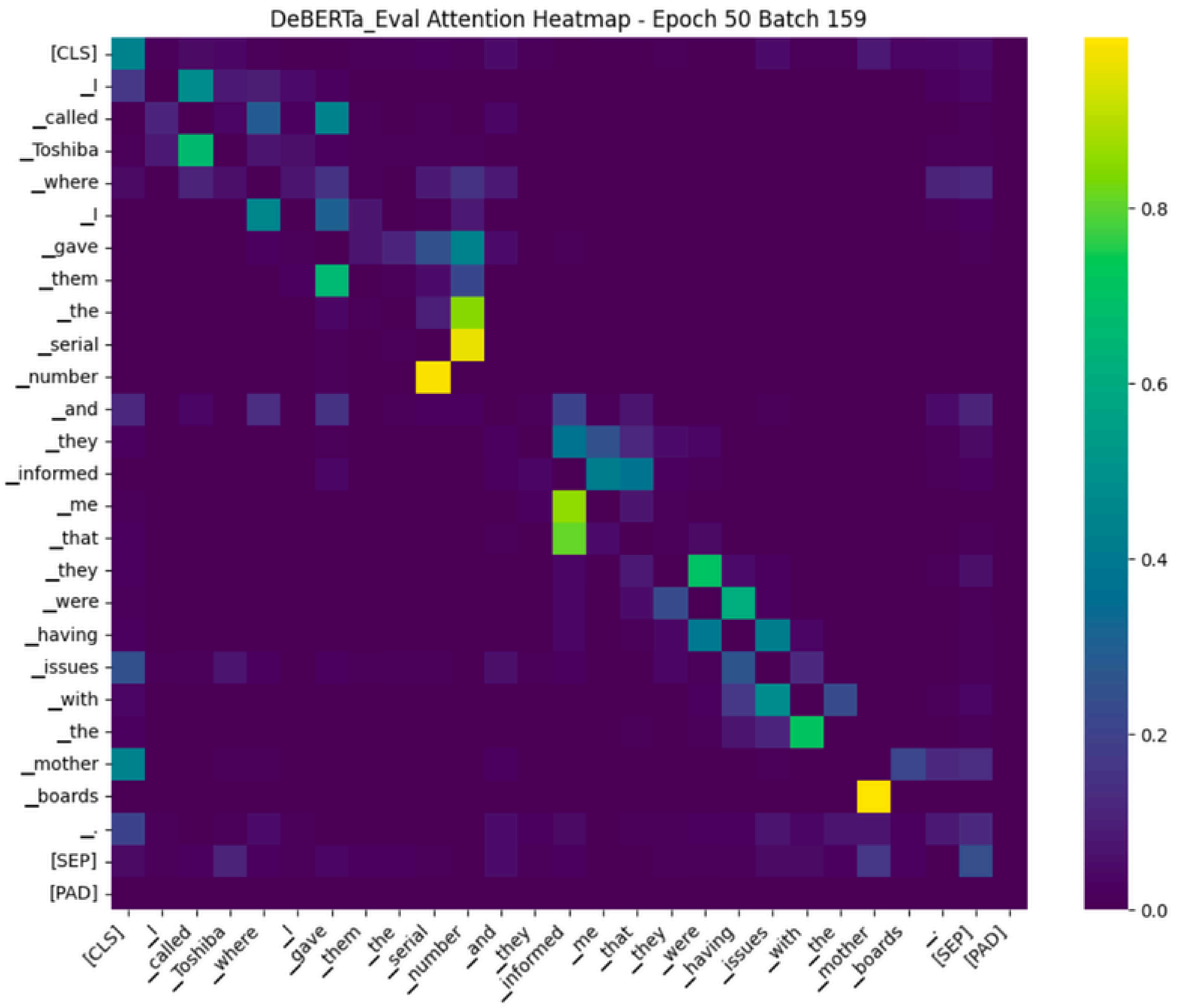
- All 3 failed modules replaced existing components
- Contrastive → modified loss | Boundary → replaced pooling | Cross-Attn → replaced fusion

Pattern of success (Span-Level Graph):

- Added a new stage **between** existing ones
- Does **NOT** touch entity classification or fusion
- Only enriches representations flowing to sentiment classifier
- Gated residual = safety net

Principle: Add **between** stages, not **instead of**. Always keep a residual path.

Attention Visualization



DeBERTa Attention Heatmap Analysis:

- Evaluated at Epoch 50, 14LAP dataset
- Meaningful token pairs:
- "serial" ↔ "number" (entity composition)
- Coherent attention
- Captures dependencies across distance
- Validates: disentangled attention preserves aspect-opinion associations

- [8] P. He et al., "DeBERTa," 2020

Why Span-Level Graph Works

The Core Problem It Solves:

- Standard: Span classified → Pairs formed → Sentiment predicted (representations FROZEN)
- SGESS: Span classified → Graph connects spans → Enriched → Sentiment predicted (spans KNOW each other)

Evidence:

- Low variance: 0.22% std = stable training
- Gated fusion converges to meaningful mixing ratios
- Restaurant (explicit opinions) → bigger gains
- Laptop (implicit/technical) → still +3.34 F1

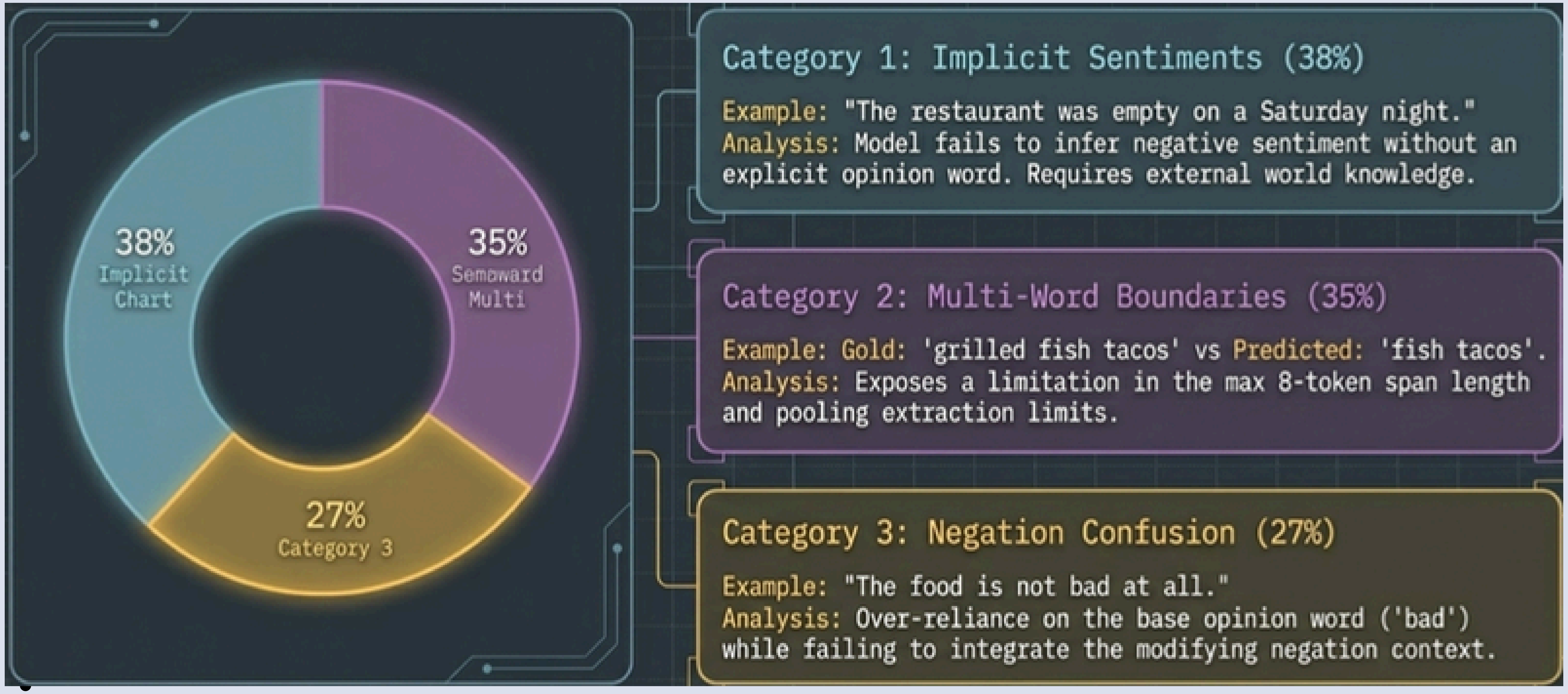
Domain Analysis:

- Restaurant: Explicit opinions ("delicious", "terrible") → easier, big gains from graph
- Laptop: Technical/implicit sentiments → harder but still +3.34 F1

- [26] L. Xu et al., "Span-ASTE," 2021. *(Standard span pipeline we improve)*
- [27] Y. Li et al., "SSJE," *KBS*, 2022. *(Span-sharing approach — still classifies independently)*
- [29] Y. Chen et al., "SBN," 2022. *(Span-level bidirectional — no graph reasoning)*

Error Analysis

100 SAMPLED ERRORS — 3 CATEGORIES:



Takeaways:

- Implicit sentiment = hardest; needs world knowledge
- Boundary: max span length (8) and pooling can improve
- Negation: model focuses on opinion word, misses negation context

Dimensional Sentiment Analysis Extension

Subtask	R^2 (Valence)	R^2 (Arousal)
Subtask 1	0.589	−0.349
Subtask 2	0.659	0.412
Subtask 3	0.681	0.423

Why?

- "acceptable" and "absolutely divine" are both "positive" — but very different
- Dimensional: Valence (pleasantness) × Arousal (intensity), scale 1–9

Adaptation:

- Replace categorical classifier with VA regression head (MLP)

Loss: $\mathscr{L}$_entity + $\mathscr{L}$_VA (MSE) + 10·$\mathscr{L}$_batch

Results (SemEval-2026 Task 3 — Team VYN):

System	Score
Ours (Team VYN)	1.7978
Baseline (Kimi-K2 Thinking)	2.1461
Baseline (Qwen-3 14B)	2.6427

Beat both official baselines (Kimi-K2 and Qwen-3 14B)

Third Publication - First Author - Published

Team VYN at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis [PDF] aclanthology.org
V Thenuwara, W De Mel, N De Silva
Proceedings of the 20th International Workshop on Semantic Evaluation ..., 2026 · aclanthology.org

- SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis (DimABSA) Shared Task.

Contributions Summary

The 5 Contributions

1. Executed the first comprehensive swap of BERT to DeBERTa for ASTE (+3-5 F1).
2. Engineered a robust Dual-Channel (Syntax/Semantics) GCN architecture.
3. Formalized the "Between Stages" design principle through rigorous negative result reporting.
4. Demonstrated framework versatility by dominating the SemEval-2026 Dimensional baselines.
5. Invented Span-Level Graph Reasoning (SGESS) to shatter isolated span classification.

Results: 80.75% (14RES) · 80.01% (16RES) · 77.19% (15RES) · 72.42% (14LAP)

Surpasses 40+ prior methods.

- [8] P. He et al., "DeBERTa," 2020.
- [6] R. Li et al., "DualGCN," 2021.
- [36] X. Zhao et al., "D2E2S," 2024.

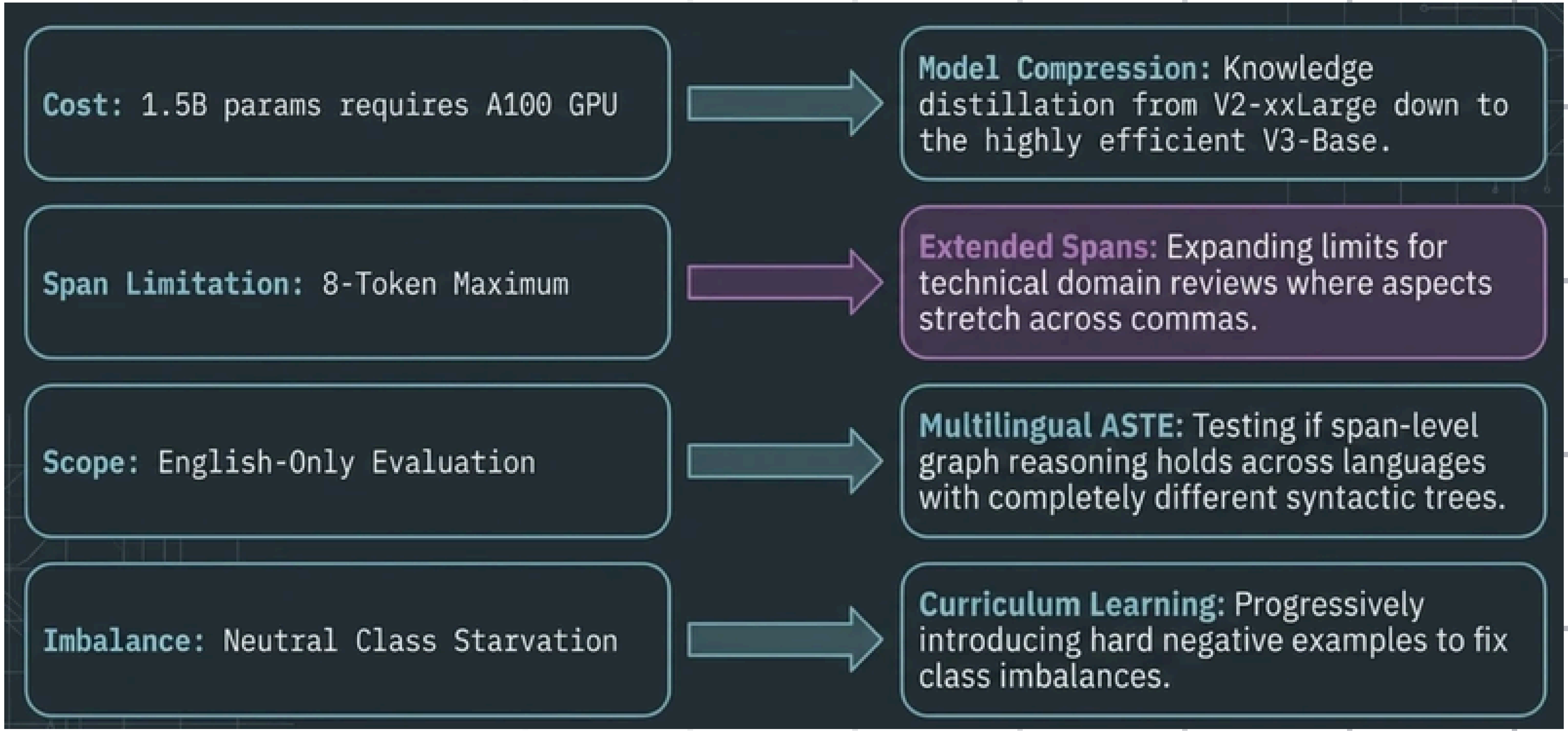
Comparison with Key Baselines

Model	14LAP			14RES			15RES			16RES		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
D2E2S [36]	67.38	60.31	63.65	75.92	74.36	75.13	70.09	62.11	65.86	77.97	71.77	74.74
CONTRASTE-MTL [28]	73.60	74.40	74.00	65.30	66.70	66.10	72.20	76.30	74.20	73.60	74.40	74.00
TAGS [39]	65.11	62.20	64.53	77.38	72.86	75.05	70.23	65.73	67.90	76.37	76.85	76.61
DESS-V3-Base (86M)	72.91	63.17	67.69	77.77	76.73	77.24	70.43	68.53	69.46	72.24	80.72	76.74
DESS-V2-xxLarge (1.5B)	72.88	65.65	69.08	80.78	79.20	79.98	70.35	75.16	74.22	80.72	77.33	77.16
SGESS-V2-xxLarge	71.19	73.24	**72.42**	**81.73**	**79.78**	**80.75**	**79.23**	75.25	**77.19**	80.37	79.66	**80.01**

- SGESS (Ours) 80.75 , 80.01 , 77.19 , 72.42 Margin over nearest competitor: +3.5 to +5.6 F1 points

-- [36] X. Zhao et al., "D2E2S," 2024.
- [28] R. Mukherjee et al., "CONTRASTE," 2023.
- [39] Y. Mao et al., "TAGS," *AAAI*, 2021.
- [35] S. Liang et al., "STAGE-3D," *AAAI*, 2023.
- [26] L. Xu et al., "Span-ASTE," 2021.

Limitations & Future Work



Conclusion

Key Takeaways:

1. **Encoder quality matters most** — DeBERTa's disentangled attention is a direct, significant improvement for ASTE (+3–5 F1 just from the swap)
2. **Dual-channel processing works** — Keeping syntax and semantics separate before fusion produces complementary, balanced representations
3. **Spans should communicate** — The Span-Level Graph Reasoning module is a key missing ingredient; letting spans exchange information before sentiment classification yields consistent +2.8 to +3.6 F1 gains
4. **Negative results are valuable** — Contrastive learning, boundary refinement, and cross-attention fusion all failed; knowing this saves the community significant effort
5. **Design principle** — Add new capabilities between stages, not instead of existing ones; always include a residual fallback path SGESS sets a new state-of-the-art on all four SemEval ASTE benchmarks, surpassing 40+ prior methods.

Thank You

Questions
&
Discussion