

THE ABSTRACT OF DOCTORAL THESIS

I. Summary of the Introduction

- Authors: Nguyễn Thị Cẩm Vân
- **Title:** *Advanced Deep Multimodal Fusion Models for Emotion Recognition in Conversation Under Incomplete and Imbalanced Modalities*
- Field of study of the dissertations: Doctor of Philosophy in Information Systems.
- Major: Information System Code: 9480104
- Name of the training institution: University of Engineering and Technology, VNU.

II. Contents of the Abstract

1. Research objectives and subjects of the dissertation

This dissertation focuses on developing advanced multimodal learning models for two main tasks: (i) multimodal emotion recognition in conversational settings and multimodal sentiment analysis, and (ii) enhancing the robustness of multimodal affective models under practical data conditions such as missing modalities and modality imbalance.

The research objectives include:

- **First**, to develop advanced deep learning models for multimodal emotion recognition in conversation (MERC) that improve information fusion and structured modeling of multi-speaker dialogues, focusing on multimodal fusion strategies at the utterance level and on temporal, speaker-dependent, and conversational structures to enhance recognition performance..
- **Second**, to design robust multimodal learning methods for MERC under incomplete and imbalanced modality conditions, by reconstructing missing features and propagating complementary information across modalities, and by introducing learning strategies that mitigate feature- and modality-level imbalance and prevent the dominance of any single modality.

2. Research methods

The dissertation employs a mixed qualitative-quantitative methodology. Theoretically, it analyzes key models in multimodal learning, graph-based modeling, representation learning, self-supervised learning, and curriculum learning to motivate the proposed architectures. Empirically, it evaluates these architectures on benchmark multimodal emotion and sentiment datasets, with experiments designed to assess overall performance, robustness, and the impact of individual components.

3. Main results and Conclusions

Contributing to the global research trend on multimodal machine learning for conversational emotion recognition, the dissertation accomplishes its objectives through three main contributions:

- **First**, the dissertation proposes structured multimodal fusion models for MERC, including CORECT, a graph-based model that employs a Relational Temporal Graph Convolutional Network (RT-GCN) and a Pairwise Cross-modal Feature Interaction (P-CM) module to jointly capture temporal, speaker-dependent and cross-modal dependencies at the dialogue level, and MultiDAG+CL, which combines a Directed Acyclic Graph (DAG)-based conversational structure with a curriculum learning strategy to better handle emotional shifts and complex dialogues.
- **Second**, the dissertation proposes Mi-CGA, a robust multimodal learning framework for scenarios with incomplete modalities, built around an Incomplete Multimodal Representation (IMR) stage and a Cross-modal Graph Attention Network (CGA-Net) that reconstructs missing modality features and models both intra- and inter-modal relations, enabling models to maintain performance when one or more modalities are partially or completely missing.

- **Finally**, the dissertation introduces Ada2I and SPCL to address modality imbalance and heterogeneous learning difficulty, where Ada2I performs dynamic balancing via Adaptive Feature Weighting (AFW) and Adaptive Modality Weighting (AMW), and SPCL serves as a lightweight plug-and-play module that guides models to learn from easy to hard samples at both utterance and dialogue levels, thereby improving training stability in complex emotional scenarios.

All proposed models are rigorously evaluated through standardized experiments and comparisons with strong baselines, demonstrating significant improvements in performance, reliability and robustness of deep multimodal models for conversational emotion recognition, especially under realistic conditions with incomplete and imbalanced modalities.

List of scientific publications related to the dissertation:

1. **Cam Van Thi Nguyen**, Tuan Mai, Son The, Dang Kieu, and Duc-Trong Le. 2023. “Conversation Understanding using Relational Temporal Graph Neural Networks with Auxiliary Cross-Modality Interaction”. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15154–15167, Singapore. Association for Computational Linguistics <https://doi.org/10.18653/v1/2023.emnlp-main.937> (Scopus, CORE Rank A* Conference)
2. **Cam-Van Thi Nguyen**, Cao-Bach Nguyen, Duc-Trong Le, and Quang-Thuy Ha. 2024. “Curriculum Learning Meets Directed Acyclic Graph for Multimodal Emotion Recognition”. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4259–4265, Torino, Italia. ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.380/> (Scopus, CORE Rank B Conference).
3. **Cam-Van Thi Nguyen**, The-Son Le, Anh-Tuan Mai, and Duc-Trong Le. 2024. “Ada2I: Enhancing Modality Balance for Multimodal Conversational Emotion Recognition”. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*. Association for Computing Machinery, New York, NY, USA, 9330–9339. <https://doi.org/10.1145/3664647.3681648> (Scopus, CORE Rank A* Conference)
4. **Cam-Van Thi Nguyen**, Hai-Dang Kieu, Quang-Thuy Ha, Xuan-Hieu Phan, Duc-Trong Le. “Mi-CGA: Cross-modal Graph Attention Network for robust emotion recognition in the presence of incomplete modalities”, *Neurocomputing*, Volume 623, 2025, 129342, ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2025.129342> (SCIE Q1 Journal, Impact Factor: 6.5)
5. **Phuong-Anh Nguyen**, The-Son Le, Duc-Trong Le, **Cam-Van Thi Nguyen***. 2026. “Leveraging Self-Paced Curriculum Learning for Enhanced Modality Balance in Multimodal Conversational Emotion Recognition.” *Neural Computing and Applications* <https://doi.org/10.1007/s00521-026-12160-6> (Scopus Q1 Journal, in press)

Hanoi, / /2026

PHD CANDIDATE

SUPERVISOR

CONFIRMATION FROM THE TRAINING UNIVERSITY