ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

HỘI NGHỊ KHOA HỌC TRẺ VNU-HCMUS YSC 2025
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM · LẦN II - NĂM 2025

MULTI-VIEW VISION TRANSFORMERS FOR ROBUST MUSHROOM SPECIES CLASSIFICATION

Quan Hoang Ngoc*, Truong Trong-Dai Le*, Long Hoang Huu Nguyen, Anh-Thu Tran-Van

Computer Science Faculty, University of Information Technology (UIT), VNU-HCMC
Faculty of Information Technology, HCMC University of Science, VNU-HCMC

INTRODUCTION

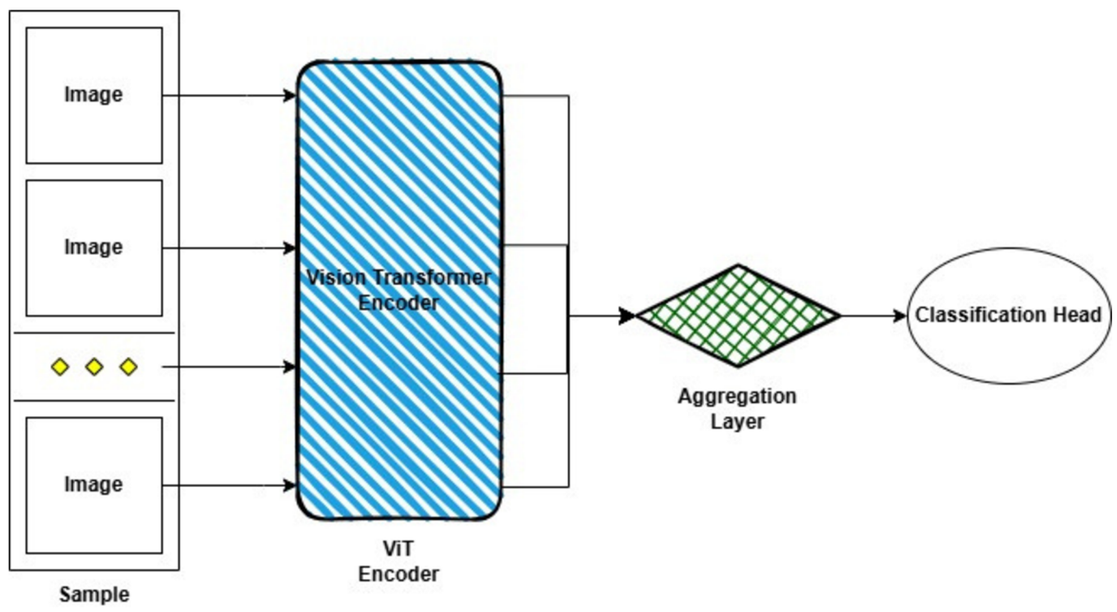
Mushroom species classification from images poses a considerable challenge due to visual variations among samples of the same species (**intra-class variation**) caused by differences in viewing angles, lighting conditions, and individual specimen characteristics of images. To address this issue, we propose a novel **Multi-View Vision Transformer (Multi-View ViT)** framework that enhances classification robustness and generalization by effectively capturing the diversity within species.

Multi-View Vision Transformer

Leverages multiple images from the same species as a single input sample (each capturing distinct views).

- Features are extracted independently using a Vision Transformer Encoder for each image.
- Features are aggregated (typically via mean pooling) to form a unified multi-view embedding.
- This aggregated representation aims for a comprehensive understanding, improving resilience to intra-class variations and augmenting input data.

Multi-View ViT Diagram



Multi-view learning unlocks better generalization for species with visual variations.

BENCHMARK

Evaluated on the **Olympiad in AI at HCMC 2025** dataset, which includes 1,200 labeled training images, 32x32 pixels, evenly distributed. Our method outperformed baseline models.

RESULT

Semi-final round	
Accuracy F1-score	97.33% (Top 1) 97.29%
Final round	
Overall Accuracy	98.89%

CONCLUSION

- Multi-view aggregation effectively reduces intra-class variation, improving accuracy and robustness.
- Promising for other vision tasks with high intra-class variability.

Keywrods:
Mushroom Image Classification; Deep Learning; Vision Transformer (ViT); Multi-View Learning; Intra-Class Variation; Feature Aggregation.

Contact info:
22521178@gm.uit.edu.vn