# PROPOSAL: ĐIỂM DANH & NHẬN DIỆN CẢM XÚC ĐA LUỒNG (PART 1)

**Tên đề tài (Tiếng Việt):** Nghiên cứu và Phát triển Hệ thống Thị giác máy tính Đa nhiệm (Multi-task Learning) cho bài toán Định danh và Phân tích Cảm xúc Đám đông trong Lớp học thông minh.

**Tên đề tài (Tiếng Anh - Suggested):** *Real-time Multi-Face Analysis in Crowded Classrooms: A Unified Framework for Robust Face Recognition and Affective Computing under Unconstrained Conditions.*

## 1. Đặt vấn đề (Problem Statement)

Bài toán "lớp học thông minh" hiện nay thường tách rời hai tác vụ: (1) Điểm danh (Face Recognition) và (2) Đánh giá thái độ (Emotion Recognition). Việc chạy 2 mô hình riêng biệt gây tốn tài nguyên và độ trễ lớn.

- **Thách thức "In-the-wild":**
    1. **Biến thiên tỉ lệ (Scale Variation):** Sinh viên ngồi bàn đầu có khuôn mặt $200 \times 200$ pixel, nhưng sinh viên bàn cuối chỉ còn $20 \times 20$ pixel. Các mô hình CNN thông thường sẽ trượt (miss) các khuôn mặt nhỏ này.
    2. **Che khuất (Occlusion):** Trong lớp học đông, sinh viên thường bị che khuất bởi người ngồi trước hoặc laptop.
    3. **Gán nhãn sai (Identity-Emotion Mismatch):** Khi tracking nhiều người, hệ thống dễ bị nhầm lẫn: "Cảm xúc của sinh viên A nhưng lại gán cho sinh viên B".
- **Câu hỏi nghiên cứu (Research Question):** Làm thế nào để thiết kế một kiến trúc mạng nơ-ron chia sẻ đặc trưng (Feature Sharing) để giải quyết đồng thời bài toán Nhận diện và Cảm xúc với độ trễ thấp (< 50ms) trên các khuôn mặt có độ phân giải thấp (Low-resolution)?

## 2. Mục tiêu nghiên cứu (Research Objectives)

1. **Giải quyết bài toán "Tiny Face":** Tối ưu hóa mô hình phát hiện khuôn mặt để bắt được các đối tượng nhỏ ở xa (Small Object Detection) mà không làm tăng số lượng tham số quá lớn.
2. **Đề xuất cơ chế Attention cho Cảm xúc:** Phát triển module chú ý (Attention Module) để tập trung vào các vùng quan trọng (mắt, miệng) nhằm phân loại cảm xúc chính xác ngay cả khi khuôn mặt bị che khuất một phần.
3. **Xây dựng Pipeline định danh liên tục:** Kết hợp thuật toán Tracking (như ByteTrack) để duy trì định danh (ID) của sinh viên trong suốt buổi học, tránh việc điểm danh lặp lại hoặc mất dấu.

## 3. Tổng quan tài liệu & Khoảng trống nghiên cứu (Literature Review & Gap Analysis)

Đề tài cần biện luận dựa trên các công trình SOTA (State-of-the-art):

### 3.1. Phát hiện khuôn mặt (Face Detection):

- **RetinaFace (2020):** Là chuẩn vàng với kiến trúc Feature Pyramid Network (FPN), xử lý tốt các khuôn mặt đa tỉ lệ. Tuy nhiên, bản ResNet-50 quá nặng.
- **YOLOv8-Face (2023):** Sự cải tiến của dòng YOLO cho khuôn mặt. Đây là hướng đi mới: tốc độ cực nhanh và độ chính xác tiệm cận RetinaFace [1].

### 3.2. Nhận diện định danh (Face Recognition):

- **ArcFace (2019):** Sử dụng *Additive Angular Margin Loss* để tối ưu hóa khoảng cách giữa các lớp nhân vật. Đây vẫn là nền tảng vững chắc nhất.
- **AdaFace (CVPR 2022):** Đây là bước đột phá cần trích dẫn. AdaFace giải quyết vấn đề **Low Quality Face** (ảnh mờ, thiếu sáng) bằng cách điều chỉnh hàm Loss dựa trên chất lượng ảnh (Quality Adaptive Margin). Đây là chìa khóa cho sinh viên ngồi bàn cuối [2].

### 3.3. Nhận diện cảm xúc (Face Emotion Recognition - FER):

- Các nghiên cứu cũ dùng VGG/ResNet thường thất bại khi gặp ảnh nhiễu.
- **POSTER++ (ICCV 2023):** Kết hợp Vision Transformer và CNN để bắt cả chi tiết cục bộ (mắt/miệng) và toàn cục. Đây là SOTA hiện tại cho FER [3].

### 3.4. Khoảng trống nghiên cứu (The Gap):

- Hầu hết các nghiên cứu hiện tại xử lý các tác vụ một cách *tuần tự* (Sequential): Detect -> Crop -> Recognize ID -> Recognize Emotion. Cách này rất chậm.
- **Đề xuất:** Xây dựng mô hình **Multi-task Learning (MTL)**, nơi một Backbone (thân mạng) duy nhất trích xuất đặc trưng chung, sau đó rẽ nhánh (Branching) ra 2 đầu ra: ID và Emotion. Điều này giúp giảm 40-50% chi phí tính toán.

## 4. Tài liệu tham khảo minh chứng (References)

*Đề tài cần đọc và trích dẫn các bài báo chủ chốt này (Giai đoạn 2022-2024):*

1. **Về Face Detection hiện đại:**
    - *Paper:* "YOLOv8-Face: A Real-Time Face Detector based on YOLOv8" (arXiv preprint, 2023).
    - *Paper:* "LFFD: A Light and Fast Face Detector for Edge Devices" (CVPR 2019 - tuy cũ nhưng vẫn là chuẩn mực cho lightweight).
2. **Về Face Recognition chất lượng thấp (Low-res):**

- *Paper:* "AdaFace: Quality Adaptive Margin for Face Recognition" (CVPR 2022). *Bài báo quan trọng nhất cho đề tài này.*
- *Paper:* "MagFace: A Universal Representation for Face Recognition and Quality Assessment" (CVPR 2021).

3. **Về Emotion Recognition (SOTA):**
   - *Paper:* "POSTER++: A Cleaner and Stronger Facial Expression Recognition Network" (ICCV 2023).
   - *Paper:* "Affective Processes: Interaction of Emotion and Attention in a Multi-Task Framework" (IEEE Transactions on Affective Computing, 2024).

### Chain of Thought cho phần tiếp theo:

Ở **Phần 2**, tôi sẽ đi sâu vào **Phương pháp nghiên cứu (Methodology)**. Tôi sẽ thiết kế một pipeline cụ thể:

- **Input:** Video stream từ Camera IP trong lớp học.
- **Detector:** Sử dụng **YOLOv8-Face** (vì cần real-time).
- **Tracker:** Sử dụng **ByteTrack** (để giữ ID khi sinh viên quay mặt đi chỗ khác).
- **Recognizer:** Một mạng CNN nhẹ (MobileFaceNet) được huấn luyện lại với hàm loss **ArcFace**.
- **Classifier:** Mạng phân lớp cảm xúc.
- Tôi sẽ giải thích cách ghép nối các khối này để không bị "nghẽn cổ chai" (bottleneck).

## 5. Phương pháp nghiên cứu & Kiến trúc đề xuất (Methodology)

Đề tài đề xuất kiến trúc **End-to-End Multi-task Tracking & Analysis Pipeline**. Hệ thống không xử lý từng ảnh rời rạc mà xử lý theo dòng video (Video Stream) để tận dụng thông tin ngữ cảnh.

### *Module 1: Phát hiện & Theo vết đa đối tượng (Detection & Tracking)*

Thay vì chỉ detect khuôn mặt ở từng frame (gây giật lag và mất ID khi sinh viên cúi đầu), ta sử dụng cơ chế Tracking.

- **Detector:** Sử dụng **SCRFD** (Sample and Computation Redistribution for Face Detection).
  - *Lý do:* SCRFD (CVPR 2022) hiện là SOTA về cân bằng tốc độ/chính xác, tốt hơn RetinaFace ở các khuôn mặt nhỏ (tiny faces).
- **Tracker:** Sử dụng **ByteTrack**.

- *Đóng góp khoa học:* ByteTrack tận dụng cả các bounding box có độ tin cậy thấp (low score) - thường là các khuôn mặt bị mờ hoặc che khuất - để duy trì ID liên tục. Điều này cực kỳ quan trọng trong lớp học khi sinh viên thường xuyên cúi xuống viết bài.

*Module 2: Đánh giá chất lượng khuôn mặt (Face Quality Assessment - FQA)*

Đây là bộ lọc thông minh ("Gatekeeper").

- **Vấn đề:** Các frame bị mờ (motion blur) khi sinh viên quay đầu sẽ làm giảm độ chính xác nhận diện.
- **Giải pháp:** Tích hợp thuật toán **MagFace**.
  - Nếu điểm chất lượng $Q < threshold$ , hệ thống sẽ bỏ qua frame đó, không đưa vào mô hình nhận diện để tiết kiệm tài nguyên, nhưng vẫn giữ ID từ Tracker.

*Module 3: Mạng Đa nhiệm Chia sẻ Đặc trưng (Multi-task Feature Extraction)*

Thay vì chạy 2 mạng riêng biệt, ta thiết kế một mạng **ResNet-50 (hoặc MobileFaceNet cho Edge Device)** làm backbone chung.

- **Cấu trúc rẽ nhánh (Branching Architecture):**
  - **Shared Layers:** Trích xuất các đặc trưng hình học cơ bản (mắt, mũi, miệng).
  - **Branch 1 (Identity):** Sử dụng **AdaFace Loss** để tối ưu hóa việc phân biệt danh tính (Class Separation).

$$L_{AdaFace} = -\log \frac{e^{s\left(\cos\theta_{y_i} - m(Q)\right)}}{e^{s\left(\cos\theta_{y_i} - m(Q)\right)} + \sum_{j \neq y_i} e^{s\cos\theta_j}}$$

    *(Trong đó $m(Q)$ là biên độ thay đổi tùy theo chất lượng ảnh $Q$ - Đây là điểm mới).*
  - **Branch 2 (Emotion):** Sử dụng **DAN (Deep Attention Network)**. Cơ chế Attention giúp mạng tự động "zoom" vào vùng mắt và khóe miệng để phân loại cảm xúc, bỏ qua phần nền nhiễu.

## 6. Kế hoạch thực nghiệm (Implementation Plan)

### *6.1. Chiến lược dữ liệu (Data Strategy)*

Đề tài không thể tự thu thập hàng triệu ảnh để train từ đầu. Phải dùng chiến lược **Transfer Learning**:

1. **Pre-training (Giai đoạn 1):** Huấn luyện backbone trên tập dữ liệu khổng lồ **MS-Celeb-1M** (cho nhận diện) và **FER-2013/AffectNet** (cho cảm xúc).
2. **Fine-tuning (Giai đoạn 2 - Quan trọng):**

- Thu thập dữ liệu thực tế tại phòng học (khoảng 10 video, mỗi video 15 phút).
- Sử dụng kỹ thuật **Few-shot Learning**: Chỉ cần 5-10 ảnh khuôn mặt của mỗi sinh viên để "đăng ký" (enroll) vào hệ thống.

### *6.2. Metrics & Benchmarking*

- **Độ chính xác nhận diện (Identification Rate):** Đo bằng Rank-1 Accuracy và TAR@FAR (True Accept Rate tại False Accept Rate thấp, ví dụ 1e-3). *Yêu cầu: >98% trong điều kiện lý tưởng, >90% khi bị che khuất.*
- **Độ chính xác cảm xúc:** Sử dụng Confusion Matrix. Đặc biệt chú ý tỉ lệ nhầm lẫn giữa "Neutral" (Bình thường) và "Bored" (Chán) / "Sad" (Buồn).
- **Hiệu năng hệ thống (System Performance):**
  - Đo FPS (Frames Per Second) trên GPU (như NVIDIA RTX 3060) và Edge Device (Jetson Nano/Xavier).
  - Mục tiêu: Xử lý >= 25 FPS với 30 khuôn mặt đồng thời.

## 7. Tài nguyên kỹ thuật & Mã nguồn (Resources)

Đây là bộ công cụ "chuẩn công nghiệp" để sinh viên bắt đầu:

**A. GitHub Repositories (Nền tảng):**

1. **InsightFace (Must-have):** github.com/deepinsight/insightface
   - Đây là thư viện toàn diện nhất hiện nay. Chứa sẵn **SCRFD** (Detect), **ArcFace/AdaFace** (Recognition). Sinh viên nên xây dựng dự án dựa trên thư viện này.
2. **ByteTrack:** github.com/ifzhang/ByteTrack
   - SOTA về tracking. Code rất sạch và dễ tích hợp với detector khác.
3. **AffectNet / DAN:** github.com/yaoying/DAN
   - Mạng Deep Attention Network cho nhận diện cảm xúc, hoạt động rất tốt trên các bộ dữ liệu lớn.

**B. Dataset để tải về:**

- **AffectNet:** Bộ dữ liệu cảm xúc lớn nhất và "tự nhiên" nhất (không phải diễn xuất trong phòng lab).
- **WIDER Face:** Dùng để test khả năng detect khuôn mặt nhỏ/đám đông.

Lời khuyên:

1. **Vấn đề Riêng tư (Privacy Ethics):** Đây là "tử huyệt" của các đề tài camera lớp học. Trong đề tài, ĐỀ TÀI **bắt buộc** phải có một mục nói về bảo mật.
   - *Giải pháp kỹ thuật:* Hệ thống không lưu video gốc. Chỉ lưu **Vector đặc trưng (Embedding Vectors)** (chuỗi số hóa). Không thể khôi phục lại mặt người từ vector này -> Đảm bảo tính riêng tư (GDPR compliance).

2. **Xử lý "Unknown" (Người lạ):** Hệ thống sẽ làm gì nếu có người lạ vào lớp?
    - ĐỀ TÀI cần thiết lập ngưỡng (threshold) khoảng cách Cosine. Nếu khoảng cách > Threshold -> Gán nhãn "Unknown" thay vì cố gán ghép vào một sinh viên nào đó.
3. **Tập trung vào "Góc nghiêng" (Profile Face):** Trong lớp, sinh viên hay quay ngang nói chuyện. Hầu hết model chết ở góc > 45 độ.
    - *Tip:* Khi Fine-tuning, hãy Augment (tăng cường) dữ liệu bằng cách xoay ảnh, thêm nhiễu để mô hình học được tính bất biến (invariance).