

Đề tài về **Head Pose Estimation (HPE)** có giá trị lõi (contribution) phải nằm ở:

1. **Tính thực tiễn (Robustness):** Chạy được trên thiết bị yếu (Edge device/Webcam thường) với độ trễ thấp.
2. **Mô hình hóa hành vi (Behavior Modeling):** Từ góc quay đầu (Head Pose) phải suy ra được trạng thái “Tập trung” (Attention State) theo chuỗi thời gian, chứ không phải chỉ là lát cắt từng khung hình.

Dưới đây là **Phần 1: Tổng quan, Tính cấp thiết và Tổng quan tài liệu (Literature Review)**. Phần này đóng vai trò “bán” ý tưởng của bạn cho hội đồng, chứng minh rằng đề tài này đáng để làm.

PROPOSAL: HỆ THỐNG ĐÁNH GIÁ MỨC ĐỘ TẬP TRUNG (PART 1)

Tên đề tài (Tiếng Việt): Nghiên cứu và Phát triển Framework Đa phương thức Đánh giá Mức độ Tập trung Thời gian thực trong E-Learning sử dụng Học sâu và Ước lượng Tư thế Đầu (Deep Head Pose Estimation).

Tên đề tài (Tiếng Anh - Suggested): *Real-time Multimodal Student Engagement Assessment Framework in E-Learning based on Lightweight Deep Head Pose Estimation and Temporal Attention Modeling.*

1. Đặt vấn đề (Problem Statement)

Sự bùng nổ của E-learning sau đại dịch Covid-19 đã đặt ra một thách thức lớn: “**Sự vắng mặt trong hiện diện**” (Absent presence). Người học đăng nhập vào hệ thống nhưng tâm trí không tập trung vào bài giảng.

- **Thực trạng:** Các phương pháp giám sát truyền thống (click chuột, thời gian online) không phản ánh đúng sự tập trung nhận thức (cognitive engagement).
- **Hạn chế công nghệ hiện tại:**
 - Các hệ thống dùng Eye-tracking chuyên dụng thì quá đắt đỏ và xâm lấn.
 - Các giải pháp dùng Webcam hiện tại (chỉ detect khuôn mặt) dễ bị đánh lừa và hoạt động kém khi thiếu sáng hoặc góc quay nghiêng lớn (extreme poses).
- **Câu hỏi nghiên cứu (Research Question):** Làm thế nào để ước lượng chính xác mức độ tập trung của người học chỉ thông qua camera đơn (monocular camera) chất lượng thấp, trong điều kiện môi trường không kiểm soát (in-the-wild), với chi phí tính toán tối thiểu?

2. Mục tiêu nghiên cứu (Research Objectives)

Đề tài không chỉ dừng lại ở việc “tính góc quay đầu”, mà tập trung vào các mục tiêu sau để đạt hàm lượng khoa học cao:

- Xây dựng mô hình Lightweight HPE:** Phát triển mạng nơ-ron tích chập nhẹ (dựa trên MobileNetV3 hoặc GhostNet) để ước lượng 3 góc Euler (Pitch, Yaw, Roll) với độ sai số $< 5^\circ$ trên tập dữ liệu in-the-wild (như 300W-LP, AFLW2000).
- Mô hình hóa chuỗi thời gian (Temporal Modeling):** Tích hợp cơ chế **Self-Attention (Transformer)** hoặc **LSTM** để phân tích sự thay đổi tư thế đầu theo thời gian. *Lý do:* Một cái gật đầu nhẹ có thể là “đồng ý/tập trung”, nhưng gục đầu kéo dài là “ngủ gật/mất tập trung”. Ảnh tĩnh không thấy được điều này.
- Đề xuất chỉ số “Engagement Score”:** Xây dựng công thức toán học định lượng mức độ tập trung tổng hợp từ hướng nhìn (Gaze vector) và tư thế đầu.

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

Đây là phần quan trọng nhất để chứng minh tính mới của đề tài.

3.1. Các phương pháp hình học truyền thống (Geometric Methods):

- Trước đây, các nghiên cứu sử dụng thuật toán **PnP (Perspective-n-Point)** để tính toán tư thế đầu từ các điểm đặc trưng (facial landmarks).
- Nhược điểm:* Phụ thuộc hoàn toàn vào độ chính xác của việc phát hiện landmark. Nếu landmark bị che khuất (occlusion) hoặc lệch, kết quả góc quay sẽ sai lệch lớn [1].

3.2. Các phương pháp Học sâu (Deep Learning Approaches):

- Các mô hình SOTA (State-of-the-art) gần đây như **Hopenet (ResNet-50)** [2] hay **FSA-Net** đã đạt độ chính xác cao bằng cách hồi quy trực tiếp góc quay từ ảnh mà không cần landmark trung gian.
- Gần đây nhất (2023-2024), xu hướng chuyển sang **6DRepNet** [3] (biểu diễn 6D rotation matrix) để giải quyết vấn đề “Gimbal lock” (khóa trục) của góc Euler truyền thống.

3.3. Khoảng trống nghiên cứu (The Research Gap): Dù độ chính xác của việc đoán góc (Head Pose) đã rất cao, nhưng **sự liên kết giữa Head Pose và “Sự tập trung” (Engagement)** vẫn còn lỏng lẻo trong các nghiên cứu gần đây:

- Hầu hết các mô hình chạy theo hướng *Frame-based* (xử lý từng ảnh rời rạc), bỏ qua thông tin ngữ cảnh thời gian (*Temporal Context*).
- Các mô hình chính xác (như ResNet-50) quá nặng để chạy trên trình duyệt web của sinh viên, trong khi các mô hình nhẹ (MobileNet) thường hy sinh độ chính xác quá nhiều.

- => **Đề xuất của đề tài:** Sử dụng kiến trúc **Teacher-Student Learning** (Knowledge Distillation) để nén mô hình lớn vào mô hình nhỏ, kết hợp với module phân tích chuỗi thời gian để đánh giá sự tập trung ổn định.

4. Tài liệu tham khảo & Minh chứng (References)

Dưới đây là danh sách các bài báo uy tín và mới nhất (2022-2024) mà bạn nên trích dẫn để tăng độ tin cậy:

1. Về phương pháp Lightweight HPE (Mới nhất 2024):

- *Paper:* “Online Learning State Evaluation Method Based on Face Detection and Head Pose Estimation” (MDPI Sensors, 2024). *Bài này đề xuất mạng DB-Net nhẹ cho di động, rất sát với đề tài.*
- *Paper:* “A Novel Student Engagement Analysis of Real Classroom Teaching Using Unified Body Orientation Estimation” (MDPI Sensors, 2025). *Cập nhật mới nhất về việc kết hợp hướng cơ thể.*

2. Về kỹ thuật AI nâng cao (6DoF & Representation):

- *Paper:* “6DoF Head Pose Estimation through Explicit Bidirectional Interaction with Face Geometry” (ECCV 2024). *Dùng để tham khảo phần toán học nâng cao.*
- *Paper:* “Rank Pose: Learning Generalised Fish-Eye Head Pose Estimation from Multi-View Images” (CVPR 2023).

3. Về ứng dụng trong Giáo dục (EdTech):

- *Paper:* “Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement” (Springer, 2023).

5. Tài nguyên học tập & GitHub (Resources)

Đây là các kho tài liệu và mã nguồn mở tốt nhất để nghiên cứu sinh bắt đầu ngay lập tức:

• Sách nền tảng (Must Read):

- *Computer Vision: Algorithms and Applications (2nd Edition, 2022)* - Richard Szeliski. (Đặc biệt chương 12 về Face & Head).
- *Deep Learning for Computer Vision with Python* - Adrian Rosebrock (Phần thực hành).

• GitHub Repositories (Code cơ sở để phát triển):

1. **6DRepNet (SOTA hiện nay):** github.com/thohemp/6DRepNet (Mô hình rất tốt để tham khảo cách biểu diễn góc quay 6D).
2. **Lightweight Head Pose:** github.com/Shaw-git/Lightweight-Head-Pose-Estimation (Code này dùng MobileNet, rất phù hợp để tối ưu hóa cho web/mobile).

3. **MediaPipe (Google):** google.github.io/mediapipe/ (Cực mạnh về real-time landmark, có thể dùng làm baseline để so sánh).