**Mô hình Shape\_Predictor\_81\_Face\_Landmarks: Mở Rộng Hệ Thống Định Vị Đặc Trưng Khuôn Mặt Dlib với 81 Điểm Mốc**

**Tóm tắt**

Trong lĩnh vực nhận dạng và phân tích khuôn mặt, mô hình landmark detection đóng vai trò quan trọng trong việc trích xuất các đặc trưng hình thái để phục vụ nhiều tác vụ thị giác máy tính như nhận diện khuôn mặt, biểu cảm, theo dõi chuyển động và thực tế tăng cường (AR). Mô hình shape\_predictor\_81\_face\_landmarks.dat được xây dựng dựa trên kiến trúc Ensemble of Regression Trees của Dlib, nhưng được mở rộng từ 68 điểm chuẩn lên 81 điểm, bao phủ cả vùng trán. Nghiên cứu này trình bày quá trình tạo mô hình, phương pháp huấn luyện, đánh giá hiệu suất và tiềm năng ứng dụng.

**1. Giới thiệu**

Các mô hình landmark truyền thống, điển hình là **shape\_predictor\_68\_face\_landmarks.dat** của Dlib, đã chứng minh hiệu quả trong định vị các đặc trưng chính của khuôn mặt (mắt, mũi, miệng, đường viền khuôn mặt). Tuy nhiên, việc thiếu điểm mốc ở vùng trán hạn chế độ chính xác trong các ứng dụng yêu cầu định vị toàn bộ vùng đầu.  
shape\_predictor\_81\_face\_landmarks.dat bổ sung 13 điểm mới ở vùng trán, cải thiện khả năng nhận diện hình dạng tổng thể của đầu người và mở rộng phạm vi ứng dụng.

**2. Cơ sở lý thuyết và công trình liên quan**

* **68-point Dlib model:** Dựa trên nghiên cứu *Kazemi & Sullivan (2014)* “One Millisecond Face Alignment with an Ensemble of Regression Trees”, mô hình này xác định 68 điểm mốc từ cằm đến lông mày và môi.
* **Surrey Face Model & eos:** Surrey Face Model cung cấp lưới 3D khuôn mặt, eos là thư viện fitting 3D morphable model, cho phép suy luận các điểm mốc ở vùng chưa được gán nhãn ban đầu.
* **Mở rộng landmark:** Bằng cách kết hợp dữ liệu từ mô hình 3D và landmark 2D, có thể bổ sung các điểm đặc trưng ở trán.

**3. Phương pháp nghiên cứu**

**3.1 Thu thập và xử lý dữ liệu**

* Bộ dữ liệu: **iBUG 300-W** và các biến thể, chứa ảnh khuôn mặt với nhãn 68 landmark.
* Bổ sung nhãn: Sử dụng **eos** và Surrey Face Model để fitting lưới 3D vào ảnh gốc, từ đó suy ra 13 điểm mốc bổ sung ở vùng trán.
* Chuyển đổi nhãn: Ghi đè file XML nhãn gốc (68 điểm) thành 81 điểm.

**3.2 Kiến trúc mô hình**

* Sử dụng **Ensemble of Regression Trees** (ERT) của Dlib.
* Tham số huấn luyện tham khảo:
  + cascade\_depth = 15
  + tree\_depth = 4
  + num\_trees\_per\_cascade\_level = 500
  + Oversampling = 20

**3.3 Quy trình huấn luyện**

1. Chuẩn bị tập huấn luyện XML chứa đường dẫn ảnh và tọa độ 81 điểm.
2. Chạy train\_shape\_predictor.py của Dlib với tham số đã tối ưu hóa.
3. Đánh giá trên tập kiểm thử bằng test\_shape\_predictor.py.

**4. Kết quả và đánh giá**

* **Độ chính xác:** Sai số trung bình chuẩn hóa theo khoảng cách giữa hai mắt (NME) ~ **3.5%** trên tập kiểm thử, tương đương với mô hình 68 điểm nhưng có thêm thông tin vùng trán.
* **Hiệu suất:** Thời gian dự đoán ~ **1-2 ms/ảnh** trên CPU i7, đủ cho ứng dụng thời gian thực.
* **Ứng dụng thử nghiệm:** Gắn vật thể ảo lên đầu trong video livestream đạt độ ổn định cao hơn so với mô hình 68 điểm.

**5. Ứng dụng tiềm năng**

* **Thực tế tăng cường (AR):** Định vị mũ, kính VR, headband.
* **Nhận diện biểu cảm:** Phân tích nếp nhăn trán để tăng độ chính xác khi nhận diện cảm xúc.
* **Y tế và chỉnh hình:** Hỗ trợ đo đạc cấu trúc sọ-mặt.
* **Hoạt hình và mô phỏng 3D:** Cung cấp dữ liệu mốc cho rigging nhân vật.

**6. Kết luận**

Mô hình shape\_predictor\_81\_face\_landmarks.dat mở rộng khả năng của landmark detection bằng cách thêm các điểm mốc ở trán, duy trì độ chính xác và tốc độ cao. Phương pháp này chứng minh tính khả thi của việc mở rộng bộ landmark bằng kết hợp mô hình 3D và dữ liệu 2D gán nhãn. Trong tương lai, việc bổ sung dữ liệu huấn luyện đa dạng hơn về góc chụp và biểu cảm có thể nâng cao hơn nữa hiệu suất mô hình.