

Xử lý Dữ liệu lớn với  
kỹ thuật Cá nhân hóa  
DynamicID

Nhóm 1 - 64HTTT2

# TỔNG QUAN

- DynamicID là một khung làm việc không cần tinh chỉnh [tuning-free framework] được đề xuất để tạo sinh hình ảnh người cá nhân hóa
- Mục tiêu chính của DynamicID là tạo ra các hình ảnh duy trì được danh tính nhất quán từ các hình ảnh được tham chiếu đến của 1 người cụ thể đồng thời kết hợp liền mạch các thông tin từ mô tả văn bản [text prompts] để cho phép chỉnh sửa thuộc tính khuôn mặt một cách linh hoạt



# TỔNG QUAN

Mặc dù hiện nay có nhiều phương pháp có thể tạo sinh hình ảnh cá nhân đã đạt đến khả năng bảo toàn danh tính với độ trung thực nhưng nó vẫn có 2 vấn đề hạn chế cần phải giải quyết :

- Hạn chế trong việc tạo hình ảnh đa danh tính.
- Khả năng chỉnh sửa thuộc tính khuôn mặt không đầy đủ.

→ DynamicID được tạo ra để giải quyết 2 bài toán trên

- Tạo hình ảnh cá nhân hóa đạt được khả năng đa danh tính.
- Chỉnh sửa khuôn mặt linh hoạt .

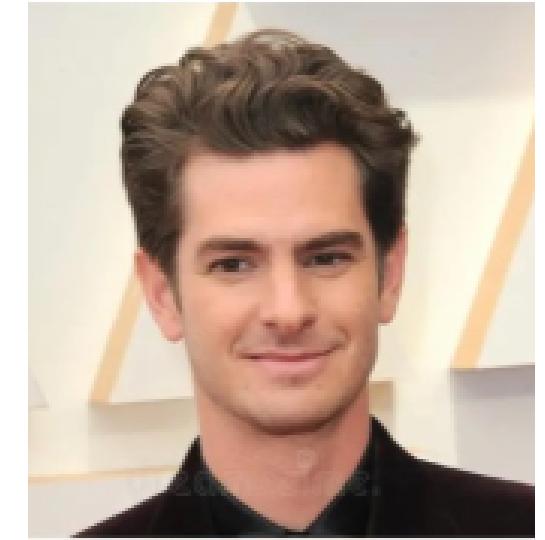
# TỔNG QUAN

DynamicID đạt được mục tiêu thích ứng zero-shot từ bối cảnh đơn danh tính sang đa danh tính thông qua 3 thay đổi chính:

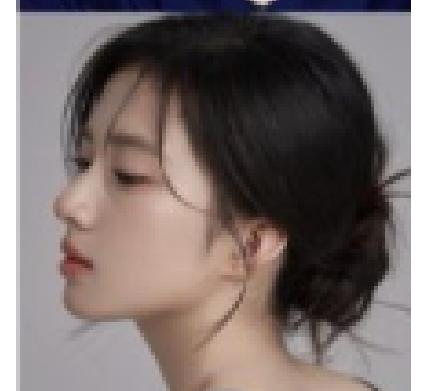
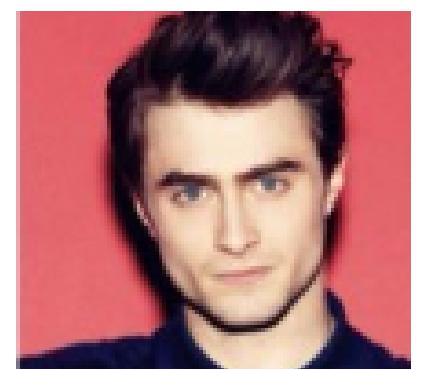
- Semantic activated attention - SAA [Cơ chế kích hoạt theo ngữ nghĩa]
- Identity-Motion Reconfigurator (IMR) [Bộ Định hình lại Chuyển động-Danh tính]
- Mô hình Đào tạo Tách rời Tác vụ [Task-Decoupled Training Paradigm]

# INPUT

1. Ảnh Tham chiếu Danh tính: chỉ cần một ảnh tham chiếu duy nhất cho mỗi danh tính, bất kể là kịch bản đơn danh tính hay đa danh tính.
2. Điều kiện Văn bản [Textual Prompts]
  - a. Text Prompts: Văn bản mô tả bối cảnh, hành động, phụ kiện, trang phục, và phong cách của chủ thể trong ảnh .
  - b. Face Prompts: Các văn bản cụ thể dùng để mô tả biểu cảm và hướng khuôn mặt.



Prompt: "sad drink coffee"



Prompt: "two people at war, both facing camera"

# OUTPUT

Đầu ra của DynamicID là hình ảnh con người cá nhân hóa đã được tạo sinh bao gồm các đặc điểm sau:

- Cá nhân hóa đa danh tính
- Độ trung thực Danh tính Cao
- Chỉnh sửa Khuôn mặt Linh hoạt
- Hình ảnh chất lượng cao



# DATASET

Tên dataset: Dataset-for-dynamicid

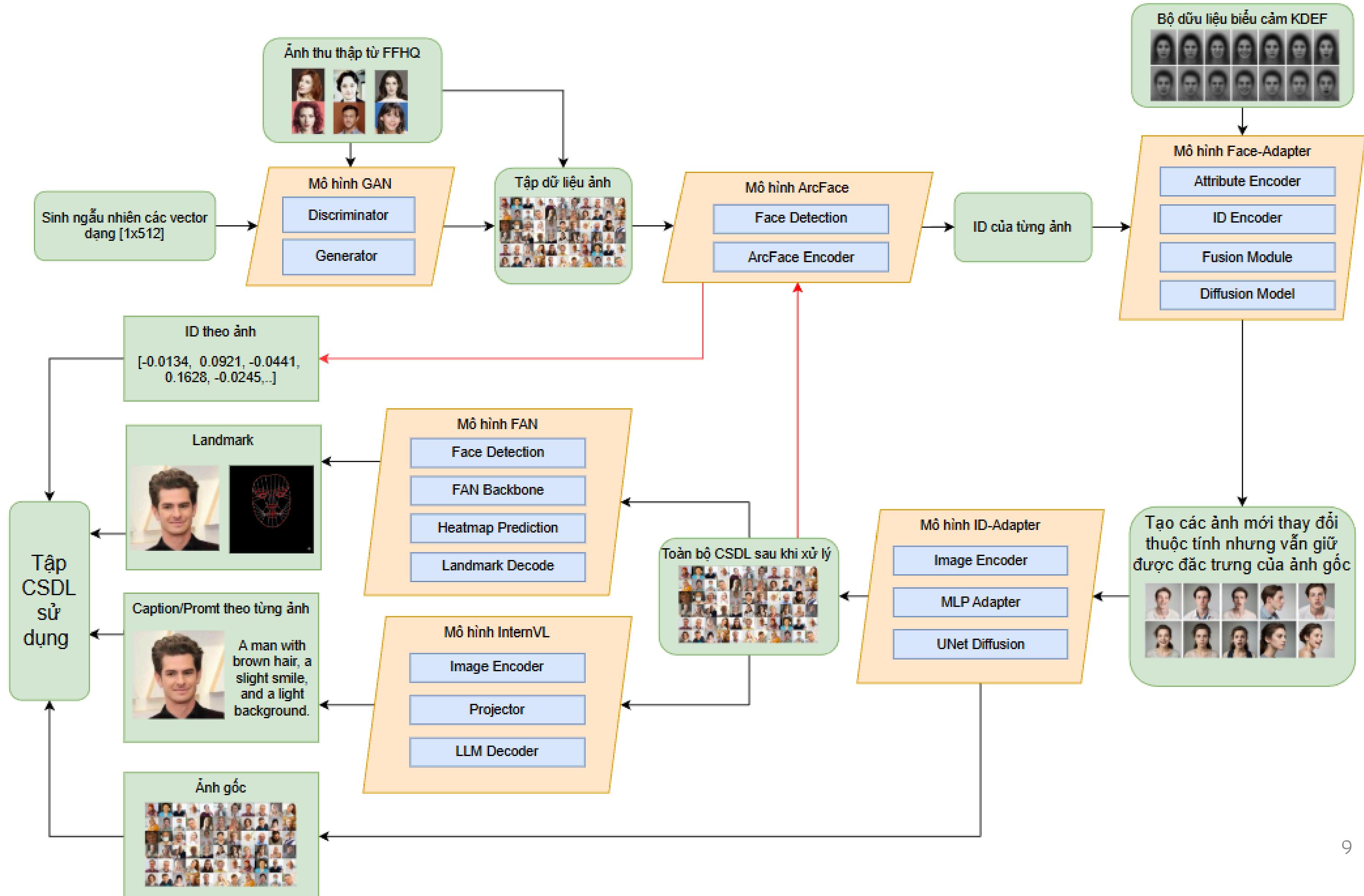
Quy mô: 7000 bức ảnh gương mặt với sự đa dạng về độ tuổi, sắc tộc, bối cảnh hình ảnh, phù hợp thị giác máy tính và học sâu, đặc biệt là những tác vụ liên quan đến việc tạo sinh và xử lý khuôn mặt người chất lượng cao.

# DATASET

Dựa vào đầu vào của 2 tập train SAA và IMR trong project có các yêu cầu về xử lý dữ liệu đầu vào:

- Resize ảnh về 512 x 512.
- Lọc các ảnh lỗi.
- Detect lanmark
- ArcFace embedding
- Sinh Caption/Prompt mô tả của từng ảnh

# QUÁ TRÌNH XỬ LÍ DỮ LIỆU



# DATASET

Vấn đề đặt ra:

- Dữ liệu lớn 7000 ảnh với dung lượng là 10.18 Gb
  - Các thao tác xử lý dữ liệu lặp đi lặp lại [Đọc ảnh, resize ảnh, tính landmark, tính embedding]
- Nên sử dụng pyspark để xử lý phân tán dữ liệu.

# MÔ HÌNH XỬ LÝ PYSPARK

Input

Split

Worker Node

Aggregate & Output

Bộ dữ liệu  
ảnh gồm  
7000 ảnh.  
Vd: 1.jpg;  
2.png;...

1000 ảnh

1000 ảnh

.....

1000 ảnh

Worker 1

Mô hình FAN (tạo Lanmark)  
Mô hình InternVL (tạo Promt mô tả)  
Mô hình Arc\_Face (tạo ID cho ảnh)

Worker 2

Mô hình FAN (tạo Lanmark)  
Mô hình InternVL (tạo Promt mô tả)  
Mô hình Arc\_Face (tạo ID cho ảnh)

Worker 3  
...

Worker 4

Mô hình FAN (tạo Lanmark)  
Mô hình InternVL (tạo Promt mô tả)  
Mô hình Arc\_Face (tạo ID cho ảnh)

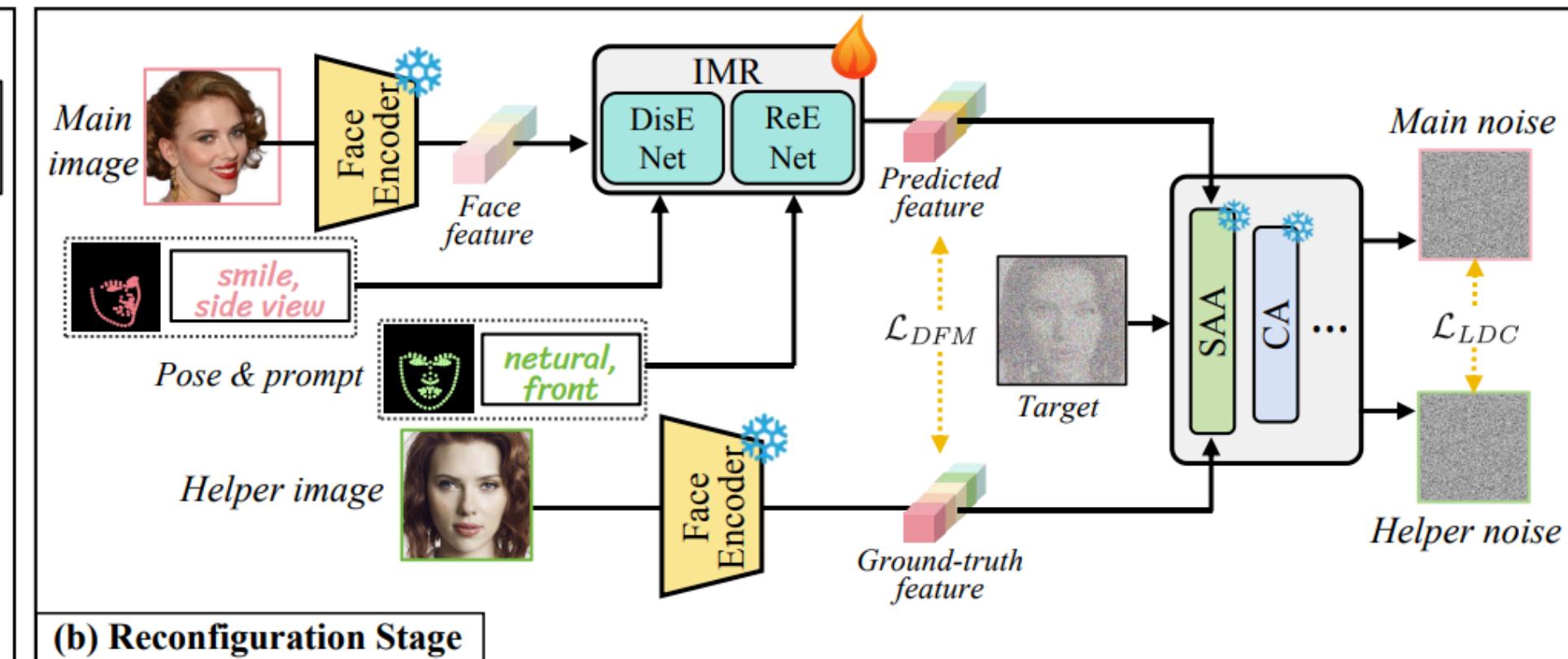
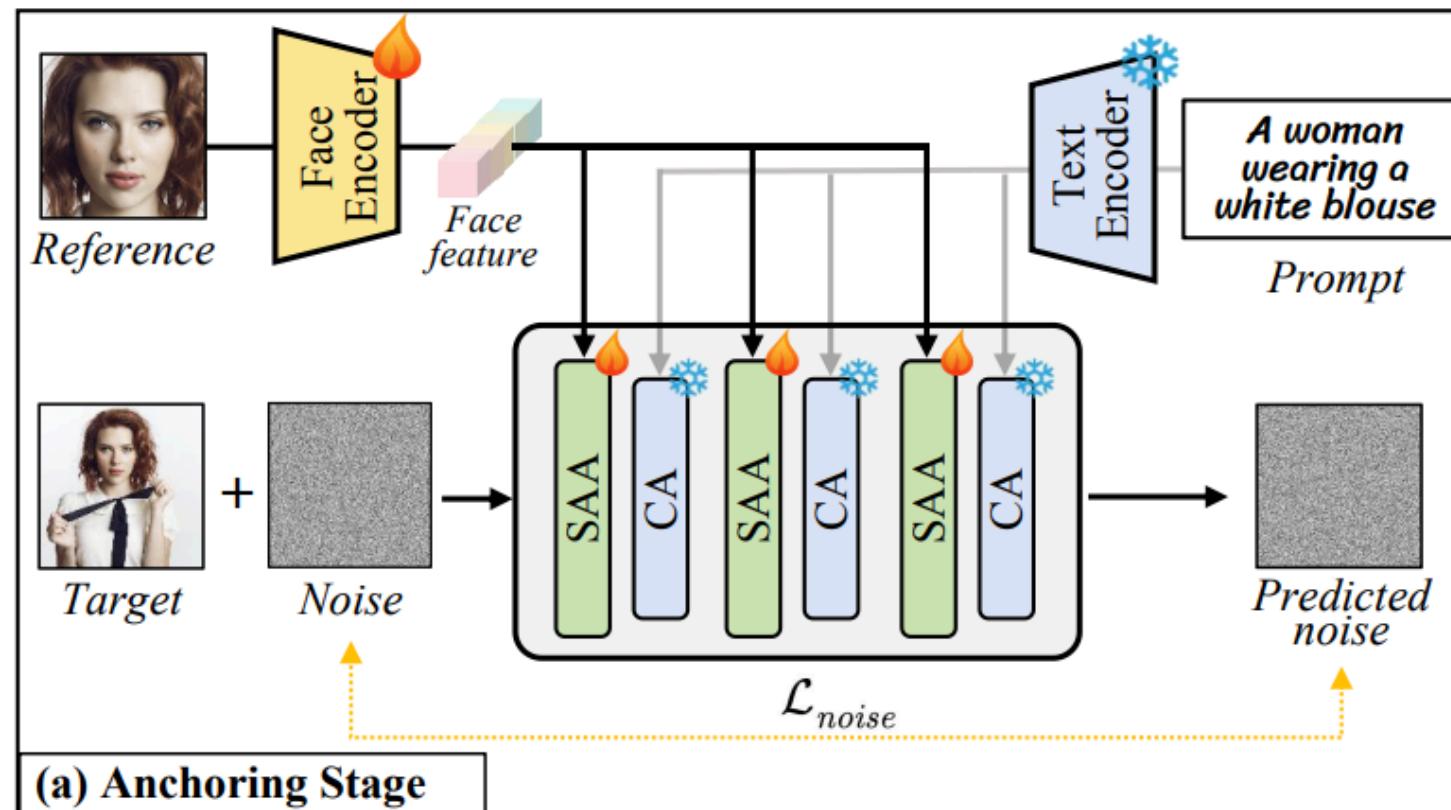
Tổng hợp kết quả  
theo từng ảnh

VD: file ảnh 1 gồm:

- 1.jpg (ảnh gốc)
- 1.png (Lanmark)
- 1.text (Promt)
- 1.pt (ID ảnh)

# LUỒNG TRIỀN KHAI

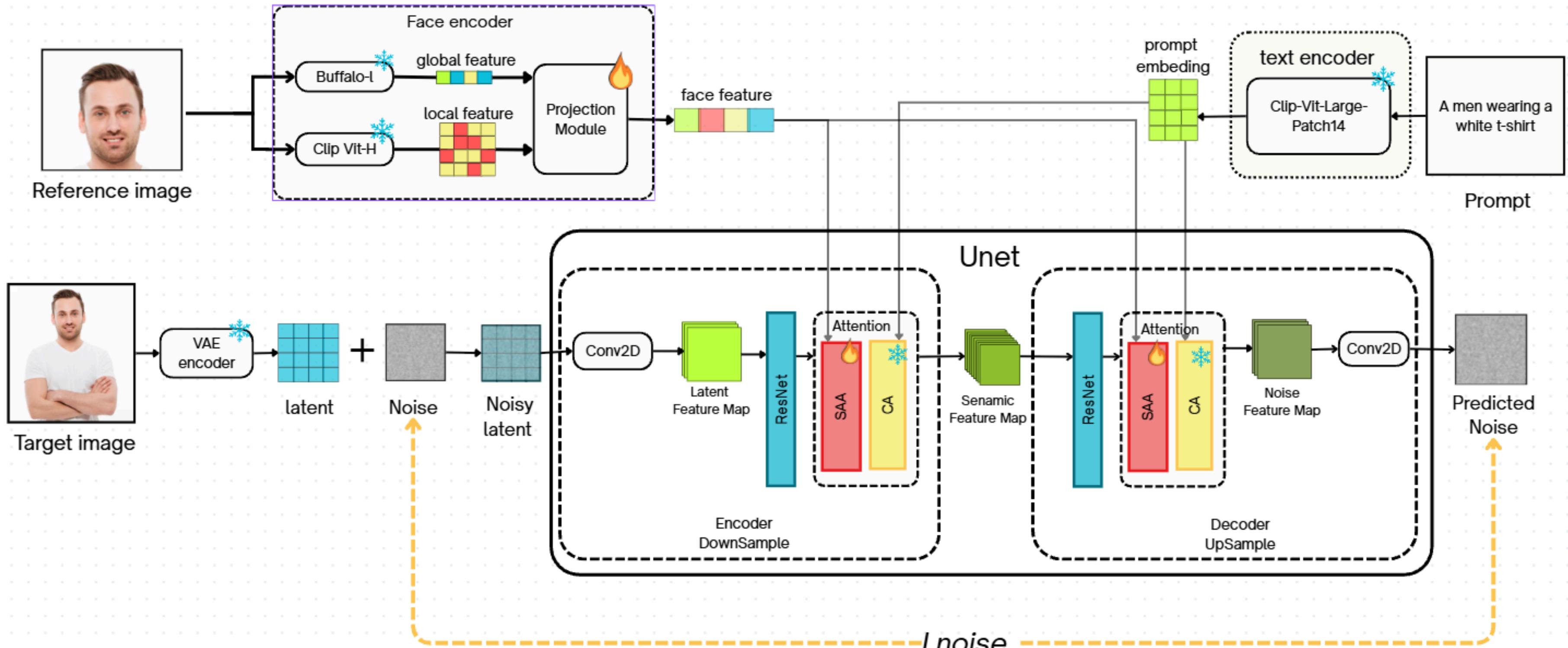
Để khắc phục nhu cầu về các bộ dữ liệu chuyên biệt quy mô lớn, DynamicID áp dụng mô hình đào tạo tách rời tác vụ [task-decoupled training paradigm], được chia thành hai giai đoạn chiến lược



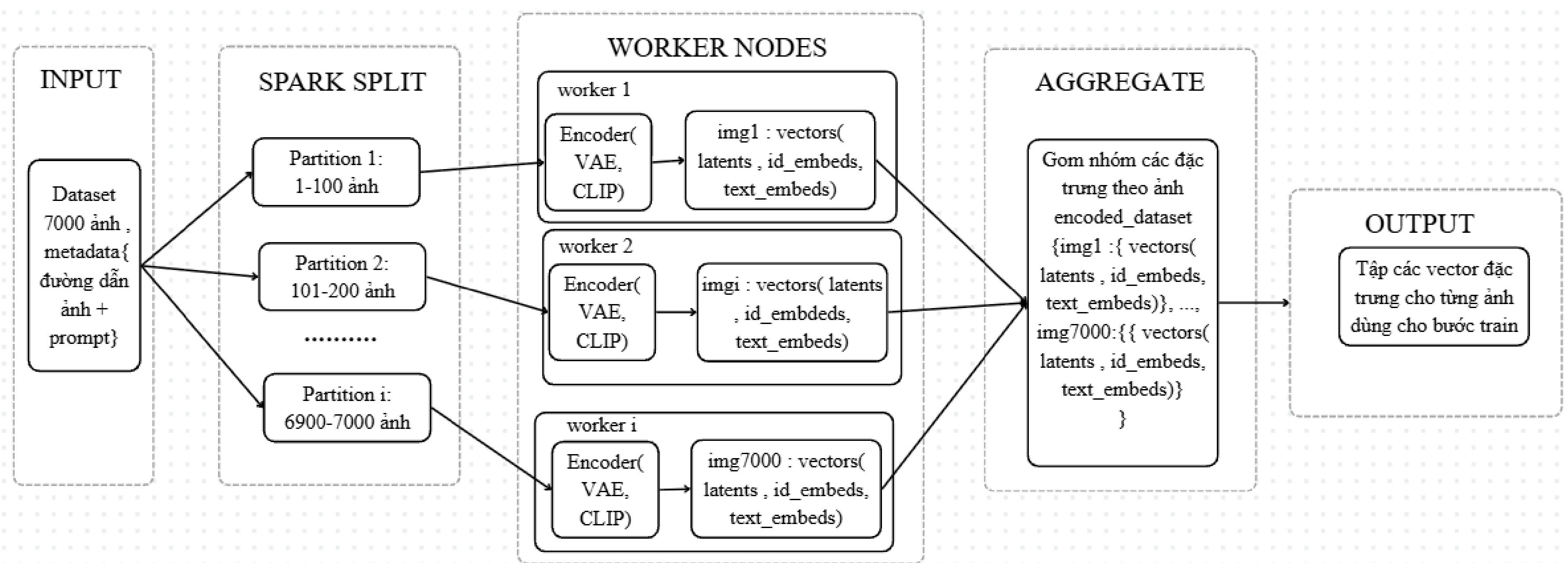
Giai đoạn neo danh tính

Giai đoạn tái cấu hình

# ANCHORING STAGE

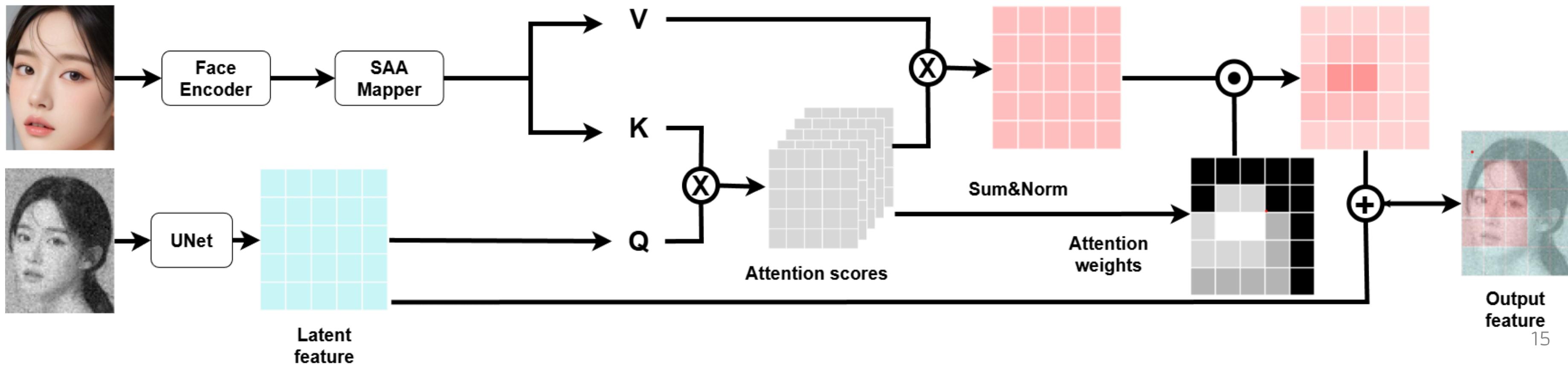


# ANCHORING STAGE



# MÔ HÌNH SAA

- Semantic-Activated Attention [SAA] là một cơ chế tiêm đặc trưng sáng tạo, sử dụng cỗng kích hoạt cấp độ truy vấn.
- Nó điều chỉnh động mức độ kích hoạt của các truy vấn tiềm ẩn hình ảnh dựa trên mức độ liên quan ngữ nghĩa của chúng đối với các khóa đặc trưng khuôn mặt.
- Mục đích là giảm thiểu sự gián đoạn đối với hành vi của mô hình nền tảng khi tiêm đặc trưng danh tính.



# RECONFIGURATION STAGE

Đây là giai đoạn tiếp theo, tập trung chuyên sâu vào khả năng chỉnh sửa khuôn mặt

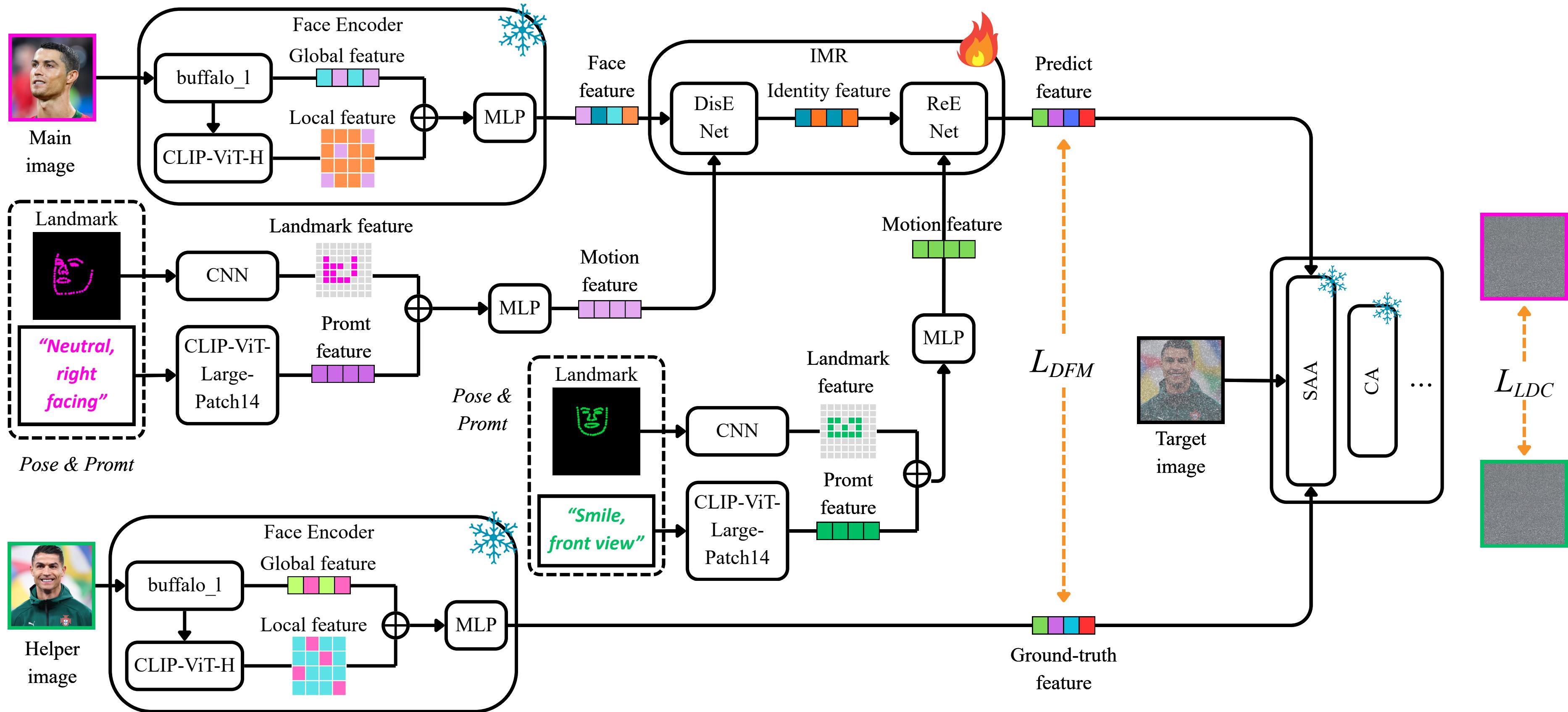
- Mục tiêu: Chuyên môn hóa việc đào tạo Bộ Định hình lại Chuyển động-Danh tính [IMR] cho khả năng chỉnh sửa khuôn mặt linh hoạt và chi tiết.
- Thành phần được Huấn luyện: Sau khi SAA và bộ mã hóa khuôn mặt đã được tối ưu hóa trong Giai đoạn Neo, các thành phần này được đóng băng, và trọng tâm chuyển sang đào tạo IMR.

# MÔ HÌNH IMR

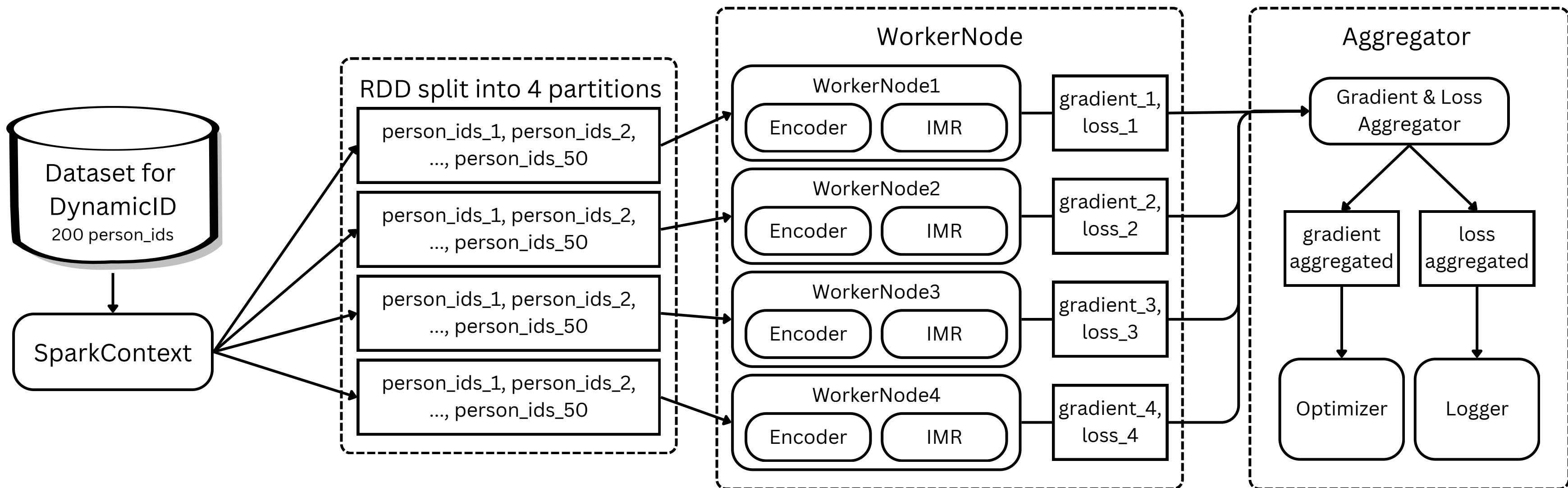
Identity-Motion Reconfigurator [IMR]:

- IMR là một bộ biến đổi đặc trưng có thể huấn luyện [trainable feature transformer].
- Nó áp dụng thao tác trong không gian đặc trưng để tách rời [disentangle] và định hình lại [reconfigure] các đặc trưng chuyển động [motion features, như biểu cảm và định hướng] và đặc trưng danh tính [ID features, như cấu trúc và vân da].
- IMR hoạt động như một trung gian, nhận đầu ra của bộ mã hóa khuôn mặt, sửa đổi đặc trưng và cung cấp đầu ra đã sửa đổi này cho SAA, từ đó hỗ trợ khả năng chỉnh sửa khuôn mặt linh hoạt và độ trung thực cao.

# RECONFIGURATION STAGE



# MÔ HÌNH XỬ LÝ PYSPARK



# KẾT LUẬN

Ưu điểm của DynamicID:

- Khả năng Cá nhân hóa Đa Danh tính Xuất sắc
- Khả năng Chỉnh sửa Khuôn mặt Linh hoạt và Chi tiết
- Hiệu quả Huấn luyện và Giảm Sự Phụ thuộc vào Dữ liệu

Nhược điểm của DynamicID:

- Giới hạn về mô hình nền tảng
- Giới hạn Hiệu suất
- Vấn đề Đạo đức và Xã hội

# HƯỚNG PHÁT TRIỂN

Ngắn hạn:

- Sử dụng các framework phân tán phù hợp hơn chẳng hạn apache spark thay vì Map reduce
- Triển khai hoàn chỉnh mô hình
- Mở rộng tập dữ liệu phù hợp hơn.

Dài hạn:

- Tối ưu hoá mô hình, nghiên cứu các phương pháp hoàn chỉnh hơn
- Xây dựng hệ thống tự động thu thập dữ liệu, train, đánh giá, triển khai

THANKS FOR  
LISTENING!