

# Report White Balance

Report day: 5/03/2025

Name: Nguyễn Thanh Hòa

Team: 2

Team leader: Hòa

Team member: Nguyễn Thanh Hòa, Nguyễn Đình Vũ Hải, Nguyễn Quý Toàn



## Task title 1: Implement InstructPix2Pix for White Balance

### ▼ 1. Purpose

Triển khai mô hình `instruct-pix2pix` để thực hiện cân bằng sáng.

Đảm bảo mô hình nhận ảnh tối làm đầu vào + câu prompt (yêu cầu) và tạo ra ảnh sáng hơn làm đầu ra.

### ▼ 2. Action

Tải mô hình `instruct-pix2pix` từ Hugging Face.

Triển khai lên FastAPI + docker:

- Thiết lập FastAPI để tạo API suy luận.
- Cấu hình Docker để đóng gói FastAPI server.

### ▼ 3. Result

- Đã tạo ứng dụng FastAPI để nhận ảnh đầu vào và trả về ảnh đã xử lý cùng với câu prompt.
- Đã đóng gói FastAPI server bằng Docker

### ▼ 4. Upcomming

- Kiểm thử hiệu suất API với các đầu vào ảnh khác nhau.
- Đánh giá chất lượng ảnh đầu ra.
- Tối ưu tốc độ



## Task title 2: Quantize (fp16, int8, int4)

### ▼ 1. Purpose

Thực hiện ba **Quantize**: fp16, int8 và int4.

Giảm kích thước mô hình và cải thiện hiệu suất

Đảm bảo khả năng tương thích với FastAPI và Docker.

### ▼ 2. Action

Chuyển đổi các thành phần của mô hình (text\_encoder, unet, vae) sang định dạng ONNX.

- Gặp vấn đề với model Unet khi chuyển sang định dạng Onnx

Áp dụng **Quantize**:

- Áp dụng **Quantize** fp16 cho mô hình ONNX.
- Thử nghiệm **Quantize** int8 bằng ONNX Runtime.
- Nghiên cứu khả năng áp dụng **Quantize** int4.

### ▼ 3. Result

Gặp một số mất mát nhỏ về độ chính xác, cần tinh chỉnh thêm

Cần kiểm thử thêm để đánh giá

### ▼ 4. Upcomming

Cân bằng giữa độ chính xác và hiệu suất.

Đo lường sự khác biệt về hiệu suất giữa các mức