

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS2205 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS2205.FEB2025

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



ỨNG DỤNG MÔ HÌNH ĐỐI KHÁNG ĐỂ BẢO VỆ DANH TÍNH TRÊN ẢNH

Lê Minh Tài - 240202026

Tóm tắt



- Link Github:
<https://github.com/taileuit/ppnckh>
- Link YouTube video:
<https://youtu.be/R-hkp3pw6pk>

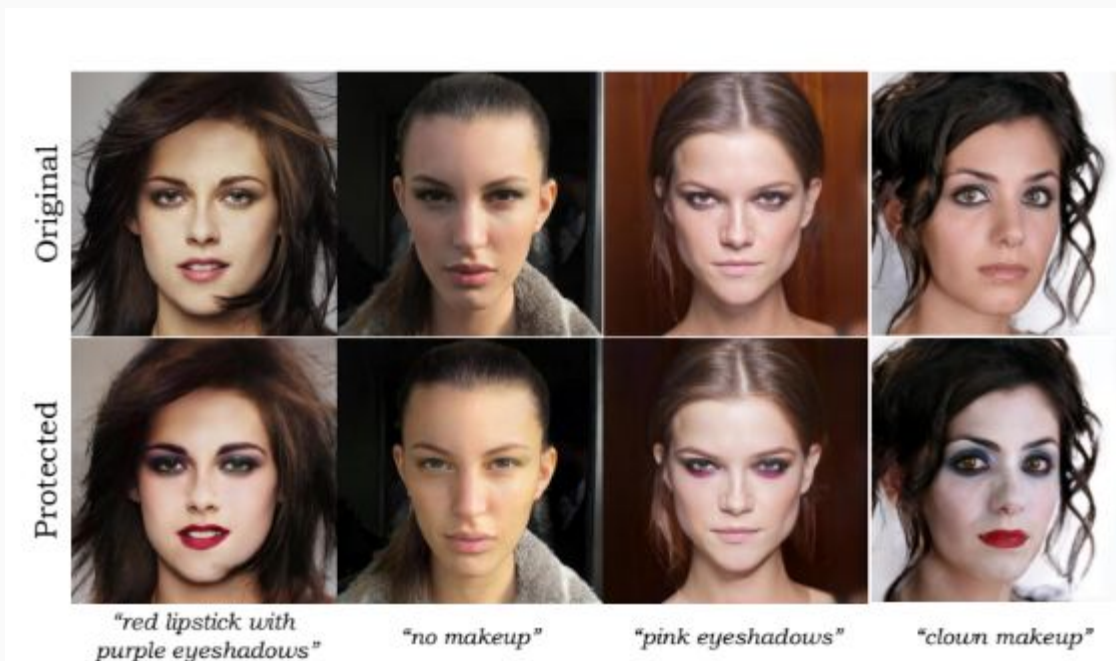
Giới thiệu

🤖 AI nhận diện khuôn mặt ngày càng phổ biến, xâm phạm quyền riêng tư.




🛡️ Người dùng mạng xã hội có thể bị nhận diện trái phép từ ảnh chân dung/selfie.

🆔 Giải pháp mới: CLIP2Protect (CVPR 2023)

→ Tạo ảnh “trang điểm” bằng mô tả văn bản, đánh lừa AI nhưng vẫn thẩm mỹ.



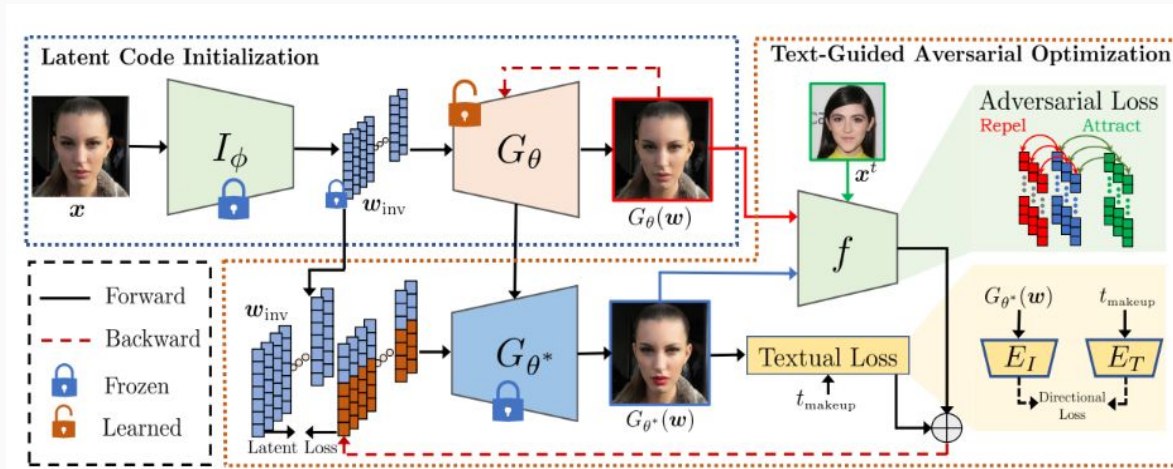
Mục tiêu

-  Phân tích cơ chế kỹ thuật của CLIP2Protect.
-  Cài đặt, thử nghiệm mô hình trên dữ liệu ảnh chân dung.
-  Đánh giá hiệu quả và khả năng ứng dụng thực tế (mạng xã hội).

Nội dung và Phương pháp

Phân tích lý thuyết mô hình CLIP2Protect:

- Kết hợp giữa StyleGAN2 và CLIP
- Pipeline gồm:
 - Khởi tạo latent từ ảnh gốc
 - Tối ưu ảnh theo mô tả văn bản
- So sánh với TIP-IM, AMT-GAN, AdvMakeup



Nội dung và Phương pháp

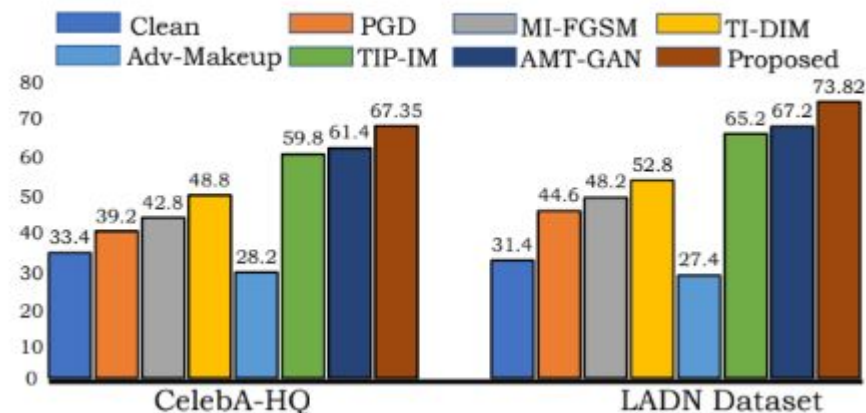
Triển khai thử nghiệm thực tế:

- Cài đặt từ GitHub (Colab/local GPU)
- Dữ liệu: CelebA-HQ
- Input: ảnh + mô tả (ex: “son đỏ”, “mắt khói”)
- Đánh giá:
 - Tỷ lệ nhận diện đúng trước/sau
 - FID Score

Kết quả dự kiến


Kết quả kỹ thuật:


- Tỷ lệ nhận diện AI giảm 30–50%
- $FID < 50 \rightarrow$ ảnh vẫn giữ thẩm mỹ




Kết quả dự kiến

Ứng dụng thực tiễn:

 Bảo vệ danh tính trước khi đăng ảnh đại diện/selfie

 Giao diện đơn giản → phổ biến cho người dùng cá nhân

 Mở rộng: Web app với mô tả ảnh → tạo ảnh bảo vệ

Privacy Protection

Choose an image:

Enter a description:

Original Image

Altered Image

So sánh với các phương pháp khác

Bảng so sánh:

	Adv-Makeup [71]	TIP-IM [70]	AMT-GAN [22]	Ours
Natural outputs	Yes	Partially	Partially	Yes
Black box	Yes	Yes	Yes	Yes
Verification	Yes	No	Yes	Yes
Identification	No	Yes	No	Yes
Unrestricted	Yes	No	Yes	Yes
Text guided	No	No	No	Yes

📌 CLIP2Protect (Ours) là phương pháp duy nhất hỗ trợ điều khiển bằng văn bản, hoạt động tốt trong môi trường black-box, đồng thời cho phép người dùng cá nhân hóa ảnh đầu ra theo ý muốn.

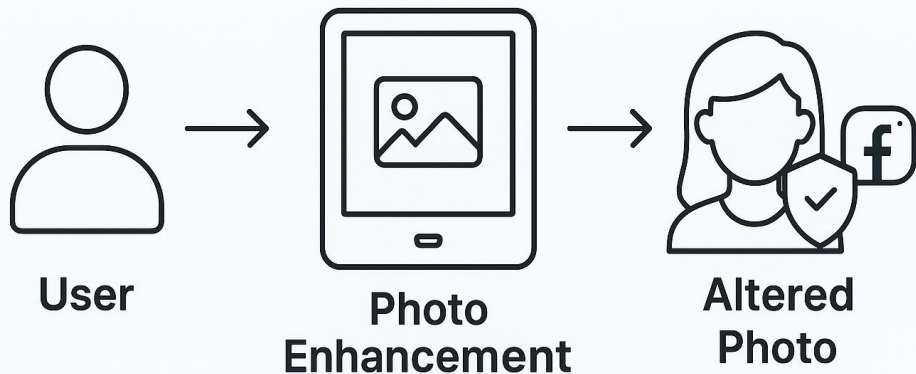
Hạn chế & Đề xuất

Một số hạn chế:

- Cần GPU để chạy → khó phổ biến
- Giao diện người dùng chưa thân thiện

Đề xuất cải tiến:

- Xây dựng web app bằng Streamlit hoặc Flask
- Khảo sát trải nghiệm người dùng
- Tích hợp trước bước đăng ảnh lên mạng xã hội



Tài liệu tham khảo

[1] Fahad Shamshad, Muzammal Naseer, Karthik Nandakumar.

Clip2Protect: Protecting Facial Privacy Using Text-Guided Makeup via Adversarial Latent Search.

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 20595–20605.

[2] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila.

Analyzing and Improving the Image Quality of StyleGAN.

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8110–8119.

[3] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever.

Learning Transferable Visual Models From Natural Language Supervision.

arXiv preprint arXiv:2103.00020, 2021.

[4] Shengshan Hu, Xiaogeng Liu, Yechao Zhang, Minghui Li, Leo Yu Zhang, Hai Jin, Libing Wu.

Protecting Facial Privacy: Generating Adversarial Identity Masks via Style-Robust Makeup Transfer.