TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ KỸ THUẬT TP. HỒ CHÍ MINH
HCM·UTE
HCMC University of Technology and Engineering

# SECURE AND REVERSIBLE FACE DE-IDENTIFICATION WITH FORMAT-PRESERVING ENCRYPTION

**Authors: HeeHwan Kim, Sungjune Park, Daeseon Choi**


Seminar Report

Presenter: Võ Lê Phúc Hậu

Source: IEEE Access (Volume 13) – July 2025

# NỘI DUNG TRÌNH BÀY

1. Background
2. Problem
3. Motivation
4. Existing Attention Limitations
5. Proposed Method
6. Results
7. Security and Complexity Analysis
8. Insight & Conclusion

# 1. Background

- **Face de-identification:** quá trình làm cho khuôn mặt trong ảnh không dễ liên kết với danh tính gốc nhưng vẫn giữ các thông tin khác như pose, emotion, light, context
- **Reversible de-identification:** ảnh đã ẩn danh vẫn có thể khôi phục về danh tính gốc khi có đủ điều kiện uỷ quyền
- **Identify embedding/identify vector:** biểu diễn số học của danh tính khuôn mặt
- **Format-preserving encryption (FPE):** phương pháp mã hoá bảo toàn định dạng

# 2. Problem: Facial Data Security

- Sự bùng nổ của kỹ thuật số và giao dịch trực tuyến dẫn đến việc thu thập và lưu trữ lượng lớn dữ liệu sinh trắc học
- Những hình ảnh lưu trữ trong hệ thống giám sát, CSDL hoặc nền tảng xác thực thường hay bị tấn công, nếu bị rò rỉ có thể dẫn đến:
  - Đánh cắp danh tính
  - Lừa đảo, giả mạo
  - Deepfake

# 3. Motivation: Mối quan hệ giữa quyền riêng tư và khả năng khôi phục dữ liệu

- Các quy định pháp lý như GDPR yêu cầu bảo vệ dữ liệu nghiêm ngặt. Tuy nhiên, trong nhiều trường hợp cần khôi phục lại danh tính từ dữ liệu đã ẩn danh

→ Nhu cầu cấp thiết là phát triển một công nghệ có khả năng:

- Khử định danh mạnh mẽ để ngăn chặn truy cập trái phép
- Cho phép khôi phục dữ liệu một cách chính xác
- Duy trì được chất lượng hình ảnh và các đặc trưng để phục vụ cho các tác vụ khác

*→ Cần phát triển công nghệ khử định danh và cho phép khôi phục dữ liệu*

# 4. Existing Attention Limitations

Những hạn chế chính của các phương pháp hiện tại:

- **Kỹ thuật truyền thống (Blur, Pixelation, Noise):** dễ thực hiện nhưng làm giảm chất lượng ảnh và có khả năng bị phá vỡ bởi các mô hình hiện đại.
- **Kỹ thuật dựa trên AI (GANs):** Tạo ra hình ảnh chân thực nhưng không thể khôi phục lại
- **Kỹ thuật khôi phục dựa trên mật khẩu:** thường dựa trên các khoá ngắn/yếu → dễ bị brute-force và thiếu tính bền vững về cấu trúc

# 5. Proposed Method

Nghiên cứu đề xuất một **mô hình tích hợp Format-Preserving Encryption – FPE) vào các mạng Face Swapping**

- ✓ Mục tiêu: cải thiện độ bảo mật thông qua mật mã mà vẫn giữ nguyên định dạng dữ liệu
- ✓ Mô hình tích hợp: Áp dụng cho FaceShifter và SimSwap
- ✓ Cốt lõi: Sử dụng thuật toán FF3 để mã hoá identify vector mà không làm thay đổi định dạng, cấu trúc số và số chiều của dữ liệu

# 5. Proposed Method

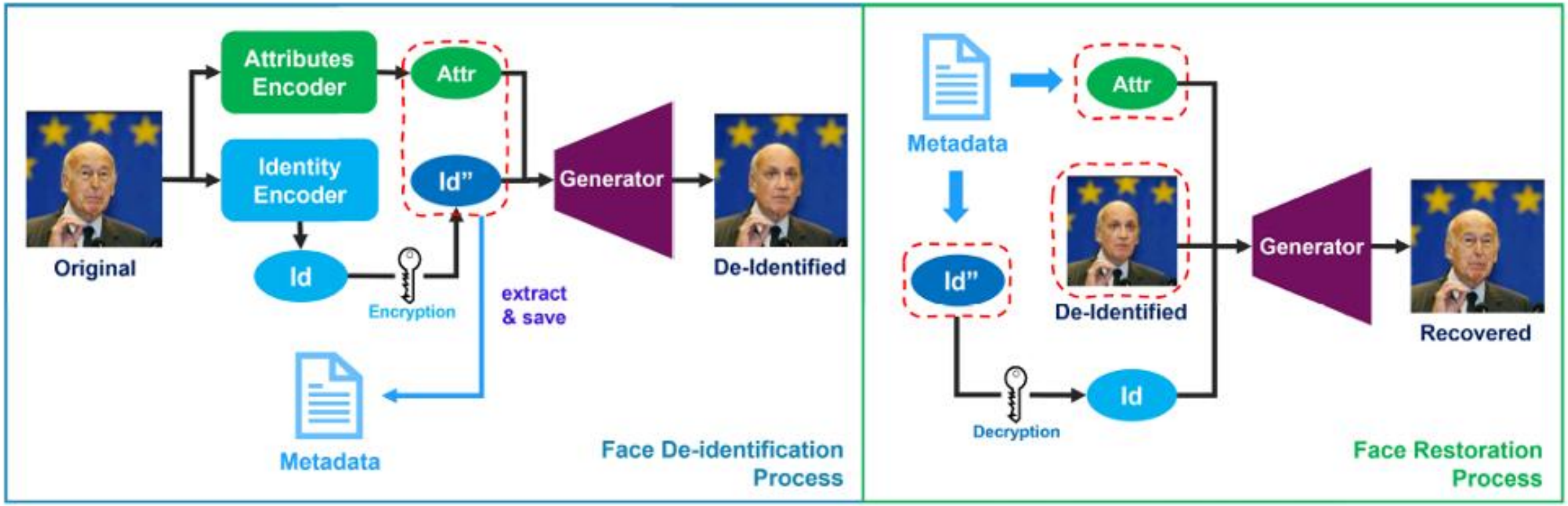



**FIGURE 1.** Framework of face de-identification process and restoration process.

# 5. Proposed Method

- FaceShifter (AEI-NET) là một dual-stage framework được thiết kế để hoán đổi khuôn mặt độ phân giải cao
  - Trích xuất các đặc điểm danh tính từ ảnh nguồn bằng bộ mã hoá Arc Face:

$$z_{id} = ArcFace(X_s) \quad (1)$$

  - Sau đó, các thuộc tính như biểu cảm, tư thế, ánh sáng sẽ được extract bằng Multi-Level Attribute (MLA) dựa trên kiến trúc U-Net

$$MLA(X_t) = z_{att} \quad (2)$$

$$z_{att} = \{z_{att}^1, z_{att}^2, \ldots, z_{att}^n\} \quad (3)$$

# 5. Proposed Method

- FaceShifter (AEI-NET) là một dual-stage framework được thiết kế để hoán đổi khuôn mặt độ phân giải cao

$$z_{id} = ArcFace(X_s) \quad (1)$$

$$MLA(X_t) = z_{att} \quad (2)$$

$$z_{att} = \{z_{att}^1, z_{att}^2, \ldots, z_{att}^n\} \quad (3)$$

  - Các đặc điểm kết hợp lại với nhau bởi bộ sinh AAD và hình ảnh gốc để tạo ra hình ảnh mới:

$$G(X_s, z_{att}, z_{id}) = X_{deid} \quad (4)$$

# 5. Proposed Method

- SimSwap là face swapping framework có độ chân thực cao, cho phép biến đổi danh tính một cách tổng quát.
  - Trích xuất các đặc điểm danh tính từ ảnh nguồn bằng bộ mã hoá Arc Face:

$$z_{id} = ArcFace(X_s) \qquad (5)$$

  - Sau đó, hình ảnh đích được mã hoá thành bản đồ đặc điểm chứa thông tin thuộc tính (vector danh tính):

$$F_t = Encoder(X_t) \qquad (6)$$

# 5. Proposed Method

- SimSwap là face swapping framework có độ chân thực cao, cho phép biến đổi danh tính một cách tổng quát.

$$z_{id} = ArcFace(X_s) \quad (5)$$

$$F_t = Encoder(X_t) \quad (6)$$

  - Các đặc điểm kết hợp lại với nhau và hình ảnh gốc để tạo ra hình ảnh mới:

$$G(X_s, F_t, z_{id}) = X_{deid} \quad (7)$$

# 5. Proposed Method

- Format-Preserving Encryption: là một phương pháp mã hoá bảo toàn định dạng
- Nghiên cứu sử dụng thuật toán FF3 (Feistel Finite Field) – một phương pháp FPE phổ biến:

$$str \leftarrow ExtDigits(z_{id}^{(i)})$$ Ex. $z_{id}$=0.0234589 → "234589"

$$enc \leftarrow FF3_{encrypt}(str, key)$$ Ex. "234589" → "594932"

$$z_{id}^{\prime(i)} \leftarrow (-1) \times enc \times 10^{-7}$$ Ex. "594932" → 594932 x $10^{-7}$ → 0.0594932

# 5. Proposed Method

**Algorithm 1** De-Identification Algorithm

**Require:** $X_s$ (source image), $X_t$ (target image), $key$

**Ensure:** $X_{deid}$ (de-identified image), metadata: $[z'_{id}, z_{att}]$

$z_{id} \leftarrow ArcFace(X_s)$

$z_{att} \leftarrow MLA(X_t)$ or $Encoder(X_t)$

**for** each $z_{id}^{(i)}$ in $z_{id}$ **do**

$str \leftarrow ExtDigits(z_{id}^{(i)})$

$enc \leftarrow FF3_{encrypt}(str, key)$

$z_{id}^{\prime(i)} \leftarrow (-1) \times enc \times 10^{-7}$

**end for**

$X_{deid} \leftarrow G(X_s, z_{att}, z'_{id})$

**/* Store metadata for restoration */**

Store $z'_{id}$ and $z_{att}$ securely as metadata

**return** $X_{deid}$, metadata $[z'_{id}, z_{att}]$

# 5. Proposed Method

**Algorithm 2** Restoration Algorithm

**Require:** $X_{deid}$ (de-identified image), metadata $[z'_{id}, z_{att}]$, $key$

**Ensure:** $X_{restored}$ (restored image)

**Load metadata for restoration */**

Retrieve encrypted identity vector $z'_{id}$ and attributes $z_{att}$ from metadata

**for** each $z_{id}^{\prime(i)}$ in $z'_{id}$ **do**

$str \leftarrow ExtDigits(z_{id}^{\prime(i)})$

$dec \leftarrow FF3_{decrypt}(str, key)$

$z_{id}^{\prime\prime(i)} \leftarrow (-1) \times dec \times 10^{-7}$

**end for**

$X_{restored} \leftarrow G(X_{\text{deid}}, z_{att}, z''_{id})$

**return** $X_{restored}$

# 6. Results

## De-identification and restoration accuracy of FaceShifter & SimSwap

| Dataset | Verification Model | De-Identified | Restored | De-Identified | Restored |
|---|---|---|---|---|---|
| LFW | ArcFace | 99.38% | 99.99% | 99.04% | 99.98% |
| | FaceNet512 | 99.81% | 99.87% | 99.95% | 99.56% |
| | VGG-Face | 99.85% | 99.99% | 91.91% | 99.98% |
| FFHQ | ArcFace | 99.18% | 99.99% | 99.10% | 99.89% |
| | FaceNet512 | 99.99% | 99.81% | 99.95% | 97.49% |
| | VGG-Face | 99.05% | 99.99% | 90.96% | 99.86% |
| VGGFace2-HQ | ArcFace | 96.87% | 98.96% | 98.41% | 99.85% |
| | FaceNet512 | 99.92% | 97.53% | 99.98% | 97.71% |
| | VGG-Face | 97.91% | 99.08% | 92.75% | 99.90% |
| CelebA-HQ | ArcFace | 96.53% | 100.00% | 97.84% | 99.91% |
| | FaceNet512 | 98.22% | 99.73% | 99.95% | 98.51% |
| | VGG-Face | 97.02% | 99.99% | 92.51% | 99.98% |
| | | FaceShifter | | SimSwap | |

# 6. Results

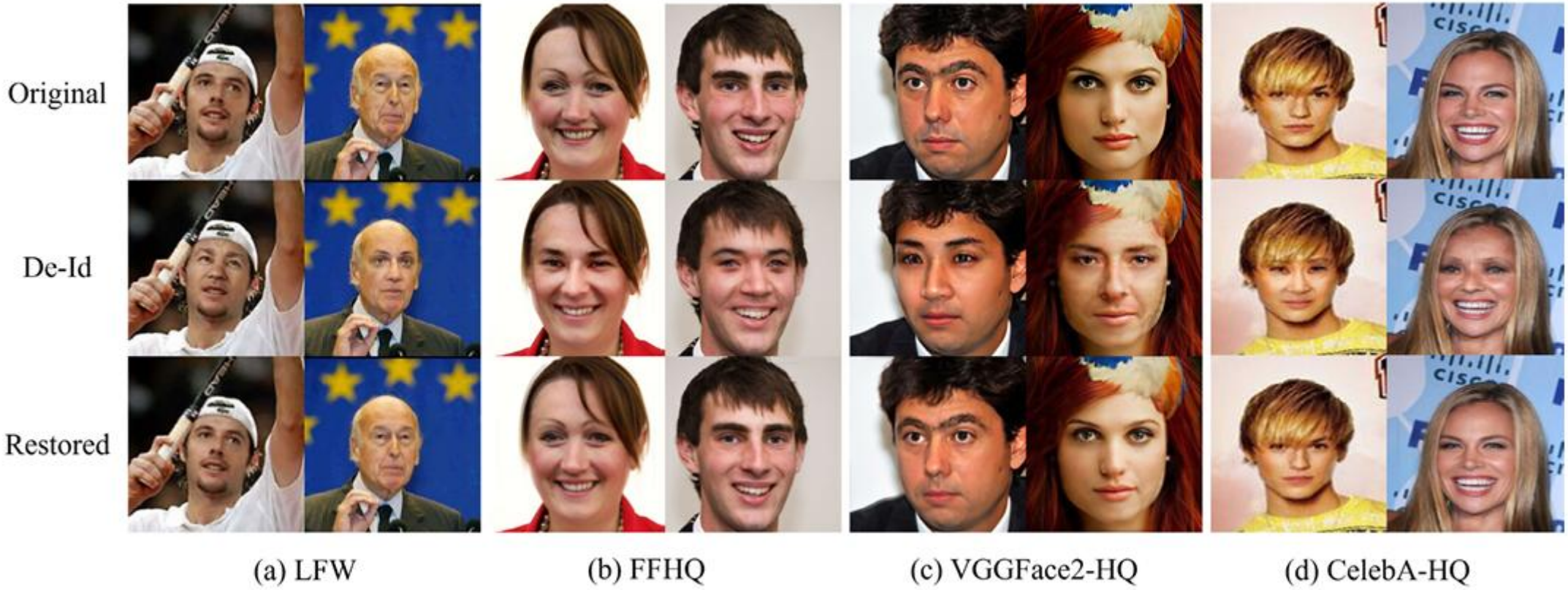



**FIGURE 2.** **Visual results of the de-identification and restoration process using FaceShifter across four datasets: (a) LFW, (b) FFHQ, (c) VGGFace2-HQ, and (d) CelebA-HQ. Each row shows: original, de-identified, and restored images.**

# 6. Results

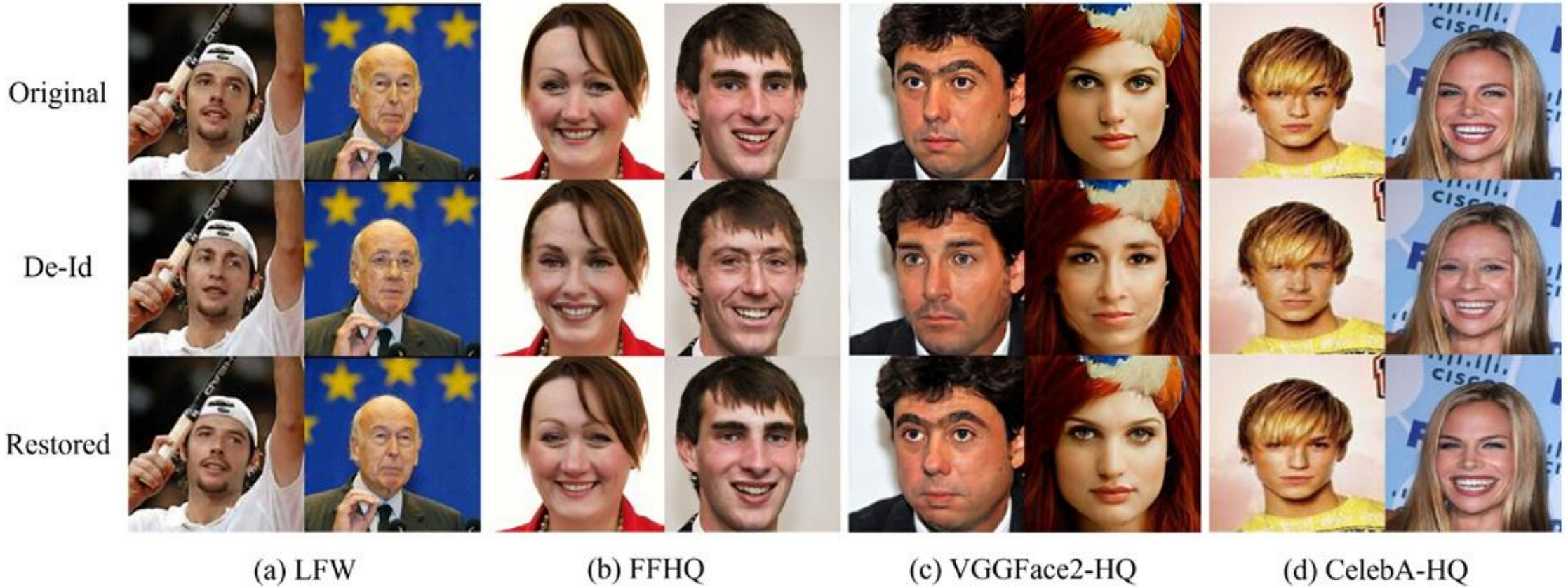



**FIGURE 3.** **Visual results of the de-identification and restoration process using SimSwap. The datasets and image arrangement are consistent with those shown in Figure 2.**

# 6. Results

## Quantitative evaluation of image quality for de-identified and restored outputs

| Dataset | Metric | De-Identified | Restored | De-Identified | Restored |
|---|---|---|---|---|---|
| LFW | SSIM ↑ | 0.9486 | 0.9657 | 0.9514 | 0.9579 |
| | FID ↓ | 1.0787 | 0.6668 | 0.8298 | 0.6452 |
| | LPIPS ↓ | 0.0240 | 0.0187 | 0.0208 | 0.0170 |
| | PSNR ↑ | 30.4010 | 34.3245 | 31.9974 | 33.5091 |
| | BRISQUE ↓ | -0.2347 / 31.7490 | +1.8156 / 33.7993 | +1.3454 / 33.3291 | +1.6945 / 33.6782 |
| FFHQ | SSIM ↑ | 0.8978 | 0.9056 | 0.8341 | 0.8190 |
| | FID ↓ | 6.7071 | 4.7227 | 19.7097 | 15.8504 |
| | LPIPS ↓ | 0.0519 | 0.0565 | 0.0797 | 0.0773 |
| | PSNR ↑ | 25.7045 | 28.6898 | 26.1322 | 26.0516 |
| | BRISQUE ↓ | +3.5324 / 6.6942 | +9.7293 / 12.8911 | +11.8354 / 14.9972 | +12.6911 / 15.8529 |
| VGGFace2-HQ | SSIM ↑ | 0.8721 | 0.8472 | 0.7312 | 0.6719 |
| | FID ↓ | 6.2875 | 3.8112 | 19.9047 | 16.7163 |
| | LPIPS ↓ | 0.0651 | 0.0851 | 0.1309 | 0.1285 |
| | PSNR ↑ | 23.9561 | 24.6397 | 22.3970 | 21.0113 |
| | BRISQUE ↓ | -0.1860 / 14.2166 | +4.6634 / 19.0660 | +10.1827 / 24.5853 | +10.9031 / 25.3057 |
| CelebA-HQ | SSIM ↑ | 0.9399 | 0.9543 | 0.9219 | 0.9248 |
| | FID ↓ | 2.1414 | 2.2271 | 3.8333 | 2.8213 |
| | LPIPS ↓ | 0.0277 | 0.0307 | 0.0293 | 0.0291 |
| | PSNR ↑ | 28.8664 | 31.8741 | 29.4248 | 30.1624 |
| | BRISQUE ↓ | +0.4876 / 15.1296 | +3.7176 / 18.3596 | +0.8797 / 15.5217 | +2.0768 / 16.7188 |

# 7. Security & Complexity Analysis

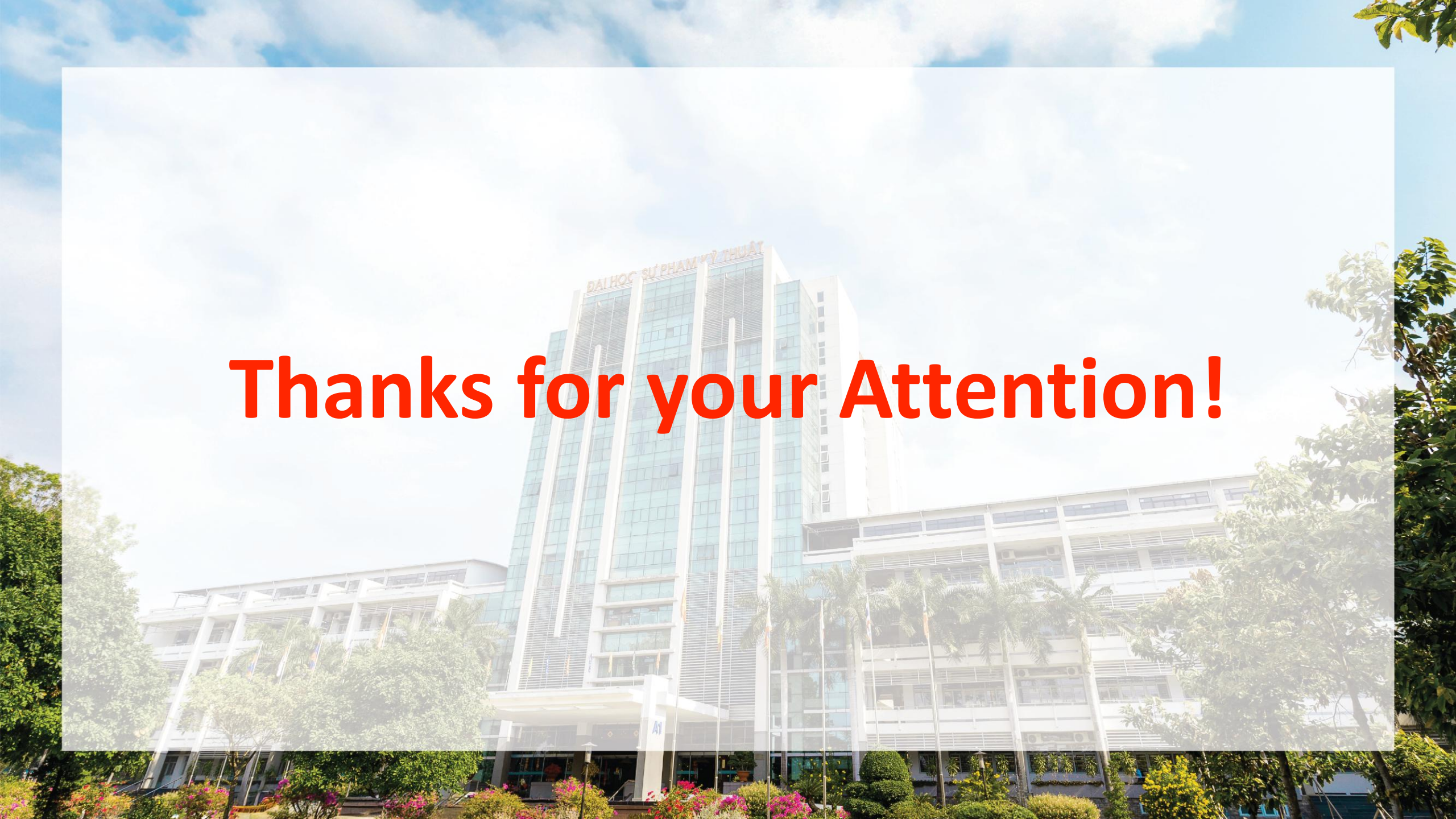
- **Chống tấn công:** việc sử dụng FPE có khả năng chống lại các cuộc tấn công so với mật khẩu thông thường
- **An toàn dữ liệu:** Ngay cả khi vector nhúng bị rò rỉ, danh tính gốc vẫn được bảo vệ nếu như không có khoá giải mã
- **Độ phức tạp:** cần tài nguyên tính toán lớn

# 8. Insight & Conclusion

## Các kết quả thực nghiệm cho thấy:

1. Phương pháp giải quyết thành công mâu thuẫn giữa bảo mật quyền riêng tư và khả năng phục hồi dữ liệu.
2. Việc tích hợp mật mã và quá trình biểu diễn đặc điểm là một hướng đi có thể phát triển

➔ Nghiên cứu đã giải quyết các vấn đề liên quan đến bảo mật dữ liệu sinh trắc học mà không làm mất đi giá trị sử dụng của dữ liệu




# Thanks for your Attention!