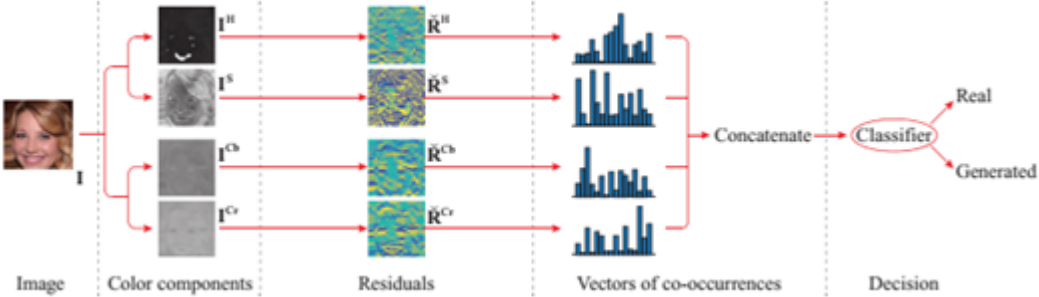


## MẪU BÁO CÁO CỦA MỖI HV

<b>Họ và tên (IN HOA)</b>	NGUYỄN THÀNH DUY (CH2001027) HOÀNG HẢI NAM (CH2002039) HUỖNH ĐỨC TÂM (CH2002044)
<b>Ảnh</b>	
<b>Số buổi vắng</b>	0
<b>Bonus</b>	30
<b>Tên đề tài (VN)</b>	NHẬN DIỆN ẢNH TẠO TỪ MẠNG SÂU DÙNG CHÊNH LỆCH THÀNH PHẦN MÀU SẮC
<b>Tên đề tài (EN)</b>	Identification of deep network generated images using disparities in color components
<b>Giới thiệu</b>	<ul style="list-style-type: none"><li>● <b>Bài toán:</b> Trong những năm gần đây, các mô hình chỉnh sửa hình ảnh phát triển rất nhanh [1-2]. Trước đây, mô hình chỉnh sửa hình ảnh chỉ tạo được các ảnh có cấu trúc đơn giản và hình ảnh sẽ khác xa so với thực tế. Do đó không khó phân biệt đâu là ảnh thật, đâu là ảnh được chỉnh sửa bằng mắt thường. Tuy nhiên, với sự phát triển của những kiến trúc mạng sâu hiện đại, đặc biệt là GANs (generative adversarial networks) [1], chất lượng của ảnh được tạo ra được cải thiện [3-4], và rất khó để có thể phân biệt bức ảnh được tạo ra từ mô hình mạng sâu (DNG) bằng mắt thường.</li><li>● <b>Lý do chọn đề tài:</b> Mặc dù việc sử dụng các mô hình chỉnh sửa hình ảnh giúp thuận lợi cho việc xử lý ảnh [5-6], tuy nhiên nó cũng gây ra nhiều rủi ro nghiêm trọng về bảo mật. Việc sử dụng ảnh giả hoặc video đã qua chỉnh sửa để tạo ra các tin tức giả thất thiệt, các khuôn mặt được tạo có thể được đăng</li></ul>

	<p>trên mạng xã hội để làm giả thông tin cá nhân hoặc được sử dụng để tấn công hệ thống xác thực sinh trắc học. Gần đây, cả phương tiện truyền thông [7] và cộng đồng nghiên cứu [8] đã thể hiện mối quan tâm lớn về tác động tiêu cực của hình ảnh DNG, và một số chính phủ [9] thậm chí đã sửa đổi luật để ngăn chặn việc chia sẻ ác ý nội dung phương tiện giả mạo do phần mềm máy học tạo ra như DeepFake. Để xác định tính xác thực của hình ảnh và tránh các vấn đề an ninh tiềm ẩn, điều quan trọng là phải xác định hình ảnh DNG.</p> <ul style="list-style-type: none"> <li>● Input: hình của một người</li> <li>● Output: xác định ảnh đó là ảnh thật hay ảnh DNG</li> </ul> 
Mục tiêu	<ul style="list-style-type: none"> <li>● Xác định ảnh thật và ảnh DNG với tỉ lệ chính xác cao hơn các mô hình hiện hành.</li> </ul>
Nội dung và phương pháp thực hiện	<p><b>Nội dung</b></p> <p>Trong phạm vi của đề cương này, các nội dung sẽ được thực hiện sẽ là: Tiến hành phân tích cách thức tạo nên các hiện vật về khía cạnh thành phần màu sắc của ảnh DNG, điều tra sự khác biệt giữa ảnh DNG và ảnh thật trong một số không gian màu. Sau đó, xây dựng một bộ tính năng để ghi lại các đặc tính của hình ảnh DNG để xác định chúng. Cuối cùng, là xây dựng một số kịch bản và đánh giá hiệu suất nhận dạng ảnh DNG.</p> <p><b>Phương pháp thực hiện</b></p> <ul style="list-style-type: none"> <li>● Phân tích hình ảnh từ góc độ màu sắc <ul style="list-style-type: none"> <li>○ Tìm hiểu cách thức tạo ra hình ảnh DNG <p>Tìm hiểu cách các hiện vật trong một hình ảnh DNG được tạo ra từ GAN về góc độ màu sắc. Xem xét các nguyên lý bố trí màu sắc ở các pixel kế nhau trong cùng sự vật. So sánh hình ảnh DNG với hình ảnh được tạo ra từ máy chụp để từ đó tìm ra điểm khác biệt, làm chìa khóa để nhận dạng nguồn gốc hình ảnh. 10</p> </li> <li>○ Nhận biết các thành phần màu sắc trong ảnh DNG</li> </ul> </li> </ul>

Theo các nghiên cứu, GAN thường tạo ra hình ảnh trong không gian RGB, chúng có xu hướng tuân theo các thuộc tính của hình ảnh thực trong không gian RGB, trong khi ít chú ý hơn đến các thuộc tính trong không gian màu khác. Do đó, sự khác biệt giữa hình ảnh DNG và hình ảnh thực là không rõ ràng trong không gian màu RGB, tuy nhiên chúng có thể rõ ràng hơn trong các không gian màu khác. Chúng ta phân tích hình ảnh DNG trong ba không gian màu khác nhau là RGB, HSV và YCbCr và tìm kiếm thành phần màu khác biệt để xác định hình ảnh DNG.

Với các hình ảnh thứ  $I$  trong tập dữ liệu, tính hệ số tương quan giữa các pixel liên kề trong mỗi thành phần màu  $I^c$  của nó ( $c \in \{R, G, B, H, S, V, Y, Cb, Cr\}$ ):

$$r_i^c = \frac{\sum_{j=1}^m \sum_{k=1}^{n-1} (I_{j,k}^c - \bar{I}^c)(I_{j,k+1}^c - \bar{I}^c)}{\sqrt{\sum_{j=1}^m \sum_{k=1}^{n-1} (I_{j,k}^c - \bar{I}^c)^2 \sum_{j=1}^m \sum_{k=1}^{n-1} (I_{j,k+1}^c - \bar{I}^c)^2}},$$

Với  $\bar{I}^c$  là giá trị trung bình của  $I^c$ , và  $m$  và  $n$  là chiều cao và chiều rộng của hình ảnh.

Đối với một tập hợp các ảnh DNG, chúng tôi tính toán  $r_i^c$  cho mỗi ảnh và xây dựng biểu đồ của  $r_i^c$  là  $\mathbb{H}_{DNG}^c$ . Làm tương tự với tập ảnh thực là biểu đồ  $\mathbb{H}_{Real}^c$ .

$$d_{\chi^2}(\mathbb{H}_{DNG}^c, \mathbb{H}_{Real}^c) = \frac{1}{2} \sum_x \frac{(\mathbb{H}_{DNG}^c(x) - \mathbb{H}_{Real}^c(x))^2}{\mathbb{H}_{DNG}^c(x) + \mathbb{H}_{Real}^c(x)},$$

Với  $x$  là chỉ số bin.  $d_{\chi^2}(\mathbb{H}_{DNG}^c, \mathbb{H}_{Real}^c)$  có thể dùng làm chỉ số nhận biết.

- Phân tích độ chênh lệch trong miền dư vi phân bậc nhất

Phần này sẽ loại bỏ nội dung hình ảnh bằng tính năng lọc thông cao và sau đó nghiên cứu sự chênh lệch về phần dư hình ảnh. Lấy phần dư hình ảnh bằng cách áp dụng toán tử vi phân bậc nhất.

$$\mathbf{R}_{j,k}^c = \mathbf{I}_{j,k}^c - \mathbf{I}_{j,k+1}^c, \quad c \in \{R, G, B, H, S, V, Y, Cb, Cr\}$$

Với  $I^c$  là thành phần thứ  $c$  của ảnh  $I$  và  $R^c$  là phần dư tương ứng.

- Trích xuất các thuộc tính từ thành phần màu sắc

Dựa trên các phân tích trên, chúng ta sẽ trích xuất các thuộc tính từ các thành phần màu sắc trong miền còn lại của hình ảnh. Trong giai đoạn trích xuất đối tượng, trước tiên chúng ta tính toán phần dư của ảnh bằng toán tử vi phân bậc

	<p>nhất (phần trên) và xử lý trước chúng bằng phương pháp cắt bớt. Sau đó, hợp nhất các ma trận cộng hưởng [10] như một loại mô tả đặc trưng để tạo thành tập hợp đặc trưng.</p> <ul style="list-style-type: none"> <li>○ Cắt bớt phần dư của hình ảnh <p>Khi đã thu được phần dư ảnh chúng ta cần xử lý trước phần dư trước khi tính toán ma trận đồng xuất hiện. Để giảm số lượng các giá trị phân biệt, các ảnh dư <math>\mathbf{R}^c (c \in \{H, S, Cb, Cr\})</math> được cắt bớt như sau:</p> <math display="block">\check{\mathbf{R}}^c(x, y) = \begin{cases} \tau, &amp; \mathbf{R}^c(x, y) \geq \tau, \\ \mathbf{R}^c(x, y), &amp; -\tau &lt; \mathbf{R}^c(x, y) &lt; \tau \\ -\tau, &amp; \mathbf{R}^c(x, y) \leq -\tau, \end{cases}</math> <p>Sau khi cắt bớt, các ảnh dư thu được <math>\check{\mathbf{R}}^c</math> chỉ chứa các giá trị nguyên trong phạm vi <math>[-\tau, \tau]</math>. Sau đó, chúng được sử dụng để tính toán các ma trận đồng xuất hiện.</p> </li> <li>○ Trích xuất các thuộc tính cùng xuất hiện <p>Tổng cộng có 4 ma trận đồng xuất hiện được tính từ <math>\check{\mathbf{R}}^H, \check{\mathbf{R}}^S, \check{\mathbf{R}}^{Cb}</math>, và <math>\check{\mathbf{R}}^{Cr}</math>. Ma trận đồng xuất hiện của một mảng 2-D <math>\mathbf{V}</math> được tính như sau:</p> <math display="block">\mathbf{C}(\theta_1, \theta_2, \dots, \theta_d) = \frac{1}{n} \sum_{x,y} \mathbb{1} \left( \begin{aligned} &amp;\mathbf{V}(x, y) = \theta_1, \\ &amp;\mathbf{V}(x + \Delta x, y + \Delta y) = \theta_2, \dots, \\ &amp;\mathbf{V}(x + (d - 1)\Delta x, y + (d - 1)\Delta y) = \theta_d \end{aligned} \right)</math> </li> <li>○ Tiến hành thực nghiệm <p>Theo các nội dung đã phân tích ở trên, chúng ta sẽ tiến hành như các mô tả để thu về bộ dữ liệu đồng thời đánh giá kết quả đạt được so với các phương pháp trước đó đã thực hiện.</p> </li> <li>● Thực hiện các thí nghiệm với tập dữ liệu theo các trường hợp giả định. <ul style="list-style-type: none"> <li>○ Trường hợp dữ liệu training phù hợp</li> <li>○ Trường hợp dữ liệu training không phù hợp</li> <li>○ Trường hợp không biết mô hình</li> </ul> </li> </ul>
<b>Kết quả dự kiến</b>	<ul style="list-style-type: none"> <li>● So sánh giữa các phương pháp: <ul style="list-style-type: none"> <li>- Bảng kết quả so sánh hiệu suất (%) cho dữ liệu training phù hợp.</li> <li>- Bảng kết quả so sánh hiệu suất (%) dữ liệu training không phù hợp. <ul style="list-style-type: none"> <li>❖ Nguồn hình ảnh không phù hợp (cùng một loại ngữ nghĩa).</li> <li>❖ Nguồn hình ảnh không phù hợp (các loại ngữ nghĩa khác nhau).</li> <li>❖ Các mô hình GAN không phù hợp.</li> </ul> </li> <li>- Bảng kết quả hiệu suất (%) trường hợp không biết mô hình.</li> </ul> </li> </ul>

- *Bộ dữ liệu:*

- Bộ dữ liệu ảnh thực: 5 bộ dữ liệu hình ảnh thực với hai loại nội dung ngữ nghĩa( ví dụ : khuôn mặt và phòng ngủ) và các độ phân giải khác nhau đã được sử dụng trong các thí nghiệm.
- Bộ dữ liệu hình ảnh DNG: hình ảnh DNG được tạo bởi 6 các loại GAN bao gồm: DCGAN, WGAN-GP, ProGAN, StyGAN, BigGAN và CocoGAN.
  - ❖ Đối với DCGAN và WGAN-GP: tạo ra hình ảnh LR DNG với kích thước 128x128.
  - ❖ Đối với ProGAN, StyGAN, BigGAN và CocoGAN: tạo ra hình ảnh khuôn mặt 1024x1024 và / hoặc hình ảnh phòng ngủ 256x256.

Dataset	Category	Content	Resolution	Quantity	Note
$\mathcal{R}_{F-LR}$	Real	Face	$128 \times 128$	200,000	Selected from CelebA
$\mathcal{R}_{FI-LR}$	Real	Face	$128 \times 128$	10,000	Selected from LFW
$\mathcal{R}_{B-LR}$	Real	Bedroom	$128 \times 128$	200,000	Selected from LSUN bedroom
$\mathcal{R}_{F-HR}$	Real	Face	$1024 \times 1024$	100,000	Combination of CelebA-HQ and FFHQ
$\mathcal{R}_{B-HR}$	Real	Bedroom	$256 \times 256$	100,000	Selected from LSUN bedroom
$\mathcal{G}_{F-LR}^*$	Generated	Face	$128 \times 128$	200,000	GANs trained with CelebA
$\mathcal{G}_{FI-LR}^*$	Generated	Face	$128 \times 128$	10,000	GANs trained with LFW
$\mathcal{G}_{B-LR}^*$	Generated	Bedroom	$128 \times 128$	200,000	GANs trained with LSUN bedroom
$\mathcal{G}_{F-HR}^*$	Generated	Face	$1024 \times 1024$	100,000	GANs trained with CelebA-HQ or FFHQ
$\mathcal{G}_{B-HR}^*$	Generated	Bedroom	$256 \times 256$	100,000	GANs trained with bedroom images

† The asterisk \* denotes the type of GAN. \*  $\in$  {DCGAN, WGAN-GP} for LR datasets, and \*  $\in$  {ProGAN, StyGAN, BigGAN, CocoGAN} for HR bedroom datasets, and \*  $\in$  {ProGAN, StyGAN} for HR face datasets.

*Bảng các bộ dữ liệu trong thực nghiệm*

**Tài liệu tham khảo**

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative adversarial nets”, in *Proceedings of the Conference Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.
- [2] A. van den Oord, N. Kalchbrenner, K. Kavukcuoglu, “Pixel recurrent neural networks”, in *Proceedings of the International Conference Machine Learning (ICML)*, pp. 1747–1756, 2016.
- [3] T. Karras, T. Aila, S. Laine, J. Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation”, in *Proceedings of the International Conference Learning Representations (ICLR)*, 2018.
- [4] T. Karras, S. Laine, T. Aila, “A style-based generator architecture for generative adversarial networks”, 2018, *arXiv:1812.04948*. [Online]. Available: <https://arxiv.org/abs/1812.04948>
- [5] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., “Photo-realistic single image super-resolution using a generative adversarial network”, in *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, 2017.

- |  |  |
|--|--|
|  | <ul style="list-style-type: none"><li>• [6] S. Iizuka, E. Simo-Serra, H. Ishikawa, “Globally and locally consistent image completion”, <i>ACM Trans. Graphics</i> 36 (4), 107:1–107:14, 2017.</li><li>• [7] J. Snow, “AI could set us back 100 years when it comes to how we consume news”, 2017. [Online]. Available: <a href="https://www.technologyreview.com/s/609358">https://www.technologyreview.com/s/609358</a></li><li>• [8] W. Knight, “The us military is funding an effort to catch deepfakes and other AI trickery”, 2018. [Online]. Available: <a href="https://www.technologyreview.com/s/611146">https://www.technologyreview.com/s/611146</a></li><li>• [9] Virginia bans ‘deepfakes’ and ‘deepnudes’ pornography, 2019. [Online]. Available: <a href="https://www.bbc.com/news/technology-48839758">https://www.bbc.com/news/technology-48839758</a></li><li>• [10] R.M. Haralick, K. Shanmugam, I. Dinstein, “Textural features for image classification”, <i>IEEE Trans. Syst. Man, Cybern. SMC</i> 3 (6), pp. 610–621, 1973.</li></ul> |
|--|--|