

# **BÁO CÁO CUỐI KỲ**

**Môn học**

**CS2205.CH1501 -**

**PHƯƠNG PHÁP LUẬN NCKH**

**Giảng viên**

**PGS.TS. LÊ ĐÌNH DUY**

**Thời gian**

**03/2021 - 06/2021**

----- *Trang này cố tình để trống* -----

## HƯỚNG DẪN


### ***Yêu cầu:***

- *Bước 1: Chọn File/Make a copy để tạo ra một file theo template mẫu [https://docs.google.com/document/d/1pu86lH6STGaVk2JH70n3jWx8qt9Eue\\_imVTQhg3s/](https://docs.google.com/document/d/1pu86lH6STGaVk2JH70n3jWx8qt9Eue_imVTQhg3s/). Đặt tên tập tin này là: CS2205.CH1501.RM.FinalReport.MSHV*
- *Bước 2: Điền các thông tin về đề cương đề tài vào file GDocs trên. Tối đa 6 trang.*
- *Bước 3: Copy toàn bộ nội dung đề cương đề tài và Paste vào cuối tập tin này (tránh ghi đè lên nội dung của HV khác).*
- *Bước 4: Nộp bài (Turn in) theo yêu cầu trên Classroom. Chọn Add or Create và chọn Link đến file Google ở trên. Lưu ý đặt quyền Anyone with the link - Viewer. Trong phần Private Comment, cung cấp thông tin của github repos, thông tin các thành viên của nhóm và các ghi chú khác nếu có. Lưu một phiên bản pdf của đề cương trên github repos*

### ***Lưu ý:***


- *Việc tuân thủ các hướng dẫn, các yêu cầu theo mẫu là **bắt buộc** và được đánh giá trong điểm tổng kết của đồ án môn học.*
- ***Deadline: 25/07/2021***


----- *Trang này cố tình để trống* -----

<b>Họ và tên (IN HOA)</b>	TRẦN TIÊU VY
<b>Ảnh</b>	
<b>Số buổi vắng</b>	
<b>Bonus</b>	Ghi số lần comment trên Google Classroom
<b>Tên đề tài (VN)</b>	PHÁT HIỆN NGƯỜI KHÔNG ĐEO KHẨU TRANG - KHÔNG NÊN QUÁ DÀI, NHƯNG CŨNG KHÔNG NÊN QUÁ NGẮN (QUÁ RỘNG)
<b>Tên đề tài (EN)</b>	OPTIONAL - KHÔNG BẮT BUỘC
<b>Giới thiệu</b>	<p><i>Hướng dẫn:</i></p> <ul style="list-style-type: none"> <li>• Bài toán/vấn đề mà đề tài muốn giải quyết</li> <li>• Lí do chọn đề tài, khả năng ứng dụng thực tế, tính thời sự</li> <li>• Mô tả input và output, nên có hình minh hoạ</li> </ul>
<b>Mục tiêu</b>	<ul style="list-style-type: none"> <li>• Trong vòng 3 ý</li> <li>• Lưu ý viết sao cho có thể đánh giá/lượng hoá được như thế nào là đạt được mục tiêu</li> </ul>

<b>Nội dung và phương pháp thực hiện</b>	<ul style="list-style-type: none"> <li>• <i>Viết chi tiết các nội dung và phương pháp để đạt mục tiêu</i></li> <li>• <i>Lưu ý Mục tiêu → Nội dung → Phương pháp phải có kết nối với nhau.</i></li> </ul>
<b>Kết quả dự kiến</b>	<ul style="list-style-type: none"> <li>• <i>Phần mềm ứng dụng</i></li> <li>• <i>Thuật toán,</i></li> <li>• <i>So sánh giữa các phương pháp</i></li> <li>• <i>Bộ dữ liệu, etc</i></li> </ul>
<b>Tài liệu tham khảo</b>	<ul style="list-style-type: none"> <li>• <i>Theo định dạng DBLP</i></li> <li>• <i>Điền sai format sẽ bị trừ điểm</i></li> </ul>

- *Các bài nộp bắt đầu từ trang 7.*
- *Lưu ý Paste vào cuối file để tránh ảnh hưởng đến các bài nộp trước đó.*

<b>Họ và tên (IN HOA)</b>	ĐỖ THỊ THÙY LINH CH2002037
<b>Ảnh</b>	
<b>Số buổi vắng</b>	0
<b>Bonus</b>	6
<b>Tên đề tài (VN)</b>	ĐỘ PHÂN GIẢI HÌNH ẢNH SIÊU CAO THÔNG QUA TINH CHỈNH LẬP LẠI.
<b>Tên đề tài (EN)</b>	IMAGE SUPER - RESOLUTION VIA ITERATIVE REFINEMENT
<b>Giới thiệu</b>	<ul style="list-style-type: none"> <li>● <b>Bài toán/vấn đề mà đề tài muốn giải quyết:</b> Tạo ra hình ảnh có độ phân giải cao phù hợp với hình ảnh đầu vào có độ phân giải thấp chỉ bằng hạt nhân của bức ảnh.</li> <li>● <b>Lí do chọn đề tài, khả năng ứng dụng thực tế, tính thời sự:</b> Siêu độ phân giải hình ảnh là một thách thức vì nhiều hình ảnh đầu ra có thể nhất quán với một hình ảnh đầu vào duy nhất và sự phân bố có điều kiện của hình ảnh đầu ra được cung cấp đầu vào thường không phù hợp tốt với các phân phối tham số đơn giản. Do đó, trong khi các phương pháp dựa trên hồi quy đơn giản với các mạng tích tụ chuyển tiếp có thể hoạt động để có độ phân giải siêu cao tại tỷ lệ phóng đại thấp, chúng thường thiếu độ trung thực cao - chi tiết cần thiết cho tỷ lệ phóng đại cao.</li> </ul>

	<p>Các phương pháp tiếp cận thường mắc phải nhiều hạn chế khác nhau; ví dụ: các mô hình tự động phục hồi rất tốn kém cho việc tạo hình ảnh có độ phân giải cao, NF và VAE thường mang lại chất lượng mẫu dưới mức tối ưu, và GAN yêu cầu các thủ thuật chính quy hóa và tối ưu hóa được thiết kế cẩn thận để khắc phục sự không ổn định của tối ưu hóa.</p> <p>Một cách tiếp cận mới được đề nghị là SR3 (Siêu phân giải thông qua Sàng lọc lặp lại) để tạo hình ảnh có độ trung thực cao một cách vô điều kiện trên cả khuôn mặt và hình ảnh tự nhiên.</p> <ul style="list-style-type: none"><li>● <i>Mô tả input và output, nên có hình minh họa</i></li></ul> <p>Input là hình ảnh với độ phân giải thấp.</p> <p>Output là hình ảnh có độ phân giải cao, đầy đủ chi tiết được hiển thị rõ nét.</p> <div><div>Input</div><div>SR3 output</div><div>Reference</div><div></div></div>
Mục tiêu	<ul style="list-style-type: none"><li>● Hiểu rõ nguyên lý thực hiện của mô hình SR3 trong việc xử lý và nâng cao độ phân giải của hình ảnh.</li><li>● Ứng dụng mô hình SR3 trong việc phục hồi ảnh cũ có độ phân giải thấp.</li><li>● Tìm hiểu sự khác nhau ở kết quả thu được giữa mô hình SR3 với các mô hình khác trong thúc đẩy hiệu suất mô hình khuếch tán lên mức cao nhất đối với những tiêu chuẩn nâng cấp độ phân giải hình ảnh.</li></ul>



## Nội dung và phương pháp thực hiện

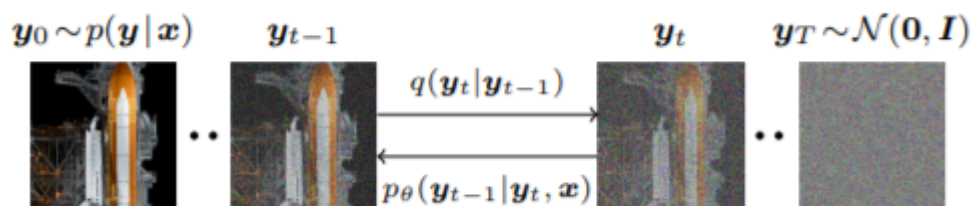
### ❖ Nội dung:

- SR3 được lấy cảm hứng từ nghiên cứu gần đây về các mô hình tạo gen sâu và các phương pháp tiếp cận dựa trên học tập gần đây đối với độ phân giải siêu cao. Mô hình tự động phục hồi (AR) có thể mô hình hóa khả năng ghi nhớ ký dữ liệu chính xác, nắm bắt các bản phân phối phong phú. Tuy nhiên, việc tạo điểm ảnh tuần tự của chúng rất tốn kém, hạn chế ứng dụng cho các hình ảnh có độ phân giải thấp. Chuẩn hóa các luồng cải thiện khi lấy mẫu tốc độ trong khi mô hình hóa khả năng dữ liệu chính xác, nhưng cần các phép biến đổi tham số hóa khả nghịch với một định thức Jacobian có thể kiểm soát được giới hạn khả năng biểu đạt của chúng.
- Lấy mẫu động lực học Langevin từ các hàm điểm đã học gần đây đã được chứng minh là có hiệu quả để tạo ra hình ảnh vô điều kiện chất lượng cao. Những mô hình này cũng đã được khái quát thành thời gian liên tục. Các mô hình so khớp và khuếch tán điểm số giảm thiểu cũng đã được tìm thấy thành công trong việc tạo hình và tổng hợp giọng nói.

### ❖ Phương pháp thực hiện:

#### Sử dụng mô hình khuếch tán phát âm có điều kiện

- Quá trình khuếch tán thuận  $q$  (từ trái sang phải) dần dần thêm nhiễu Gaussian vào hình ảnh mục tiêu. Suy luận ngược lại quá trình suy luận  $p$  (từ phải sang trái) lặp đi lặp lại biểu thị hình ảnh đích được điều chỉnh trên hình ảnh nguồn  $x$ .



- Sự phân bố của các hình ảnh trung gian trong suy luận chuỗi được định nghĩa theo quy trình khuếch tán thuận điều đó dần dần thêm tiếng ồn Gaussian vào tín hiệu thông qua một Chuỗi Markov, ký hiệu là  $q(y_t | y_{t-1})$ . Mục tiêu của mô hình là đảo ngược quá trình khuếch tán Gaussian bằng cách khôi phục lặp đi lặp lại tín hiệu từ nhiễu thông qua chuỗi Markov ngược điều kiện trên  $x$ .
- Để giúp đảo ngược quá trình khuếch tán, chúng tôi tận dụng thông tin bổ sung bên dưới dạng hình ảnh nguồn  $x$  và tối ưu hóa mô hình làm giảm thần kinh  $\theta$  như nhập hình ảnh nguồn  $x$  và một hình ảnh

mục tiêu nhiều nhằm mục đích khôi phục hình ảnh mục tiêu không ồn ào  $y_0$ .

**Algorithm 1** Training a denoising model  $f_\theta$

```

1: repeat
2:    $(x, y_0) \sim p(x, y)$ 
3:    $\gamma \sim p(\gamma)$ 
4:    $\epsilon \sim \mathcal{N}(0, I)$ 
5:   Take a gradient descent step on
      $\nabla_\theta \|f_\theta(x, \sqrt{\gamma}y_0 + \sqrt{1-\gamma}\epsilon, \gamma) - y_0\|_p^p$ 
6: until converged
  
```

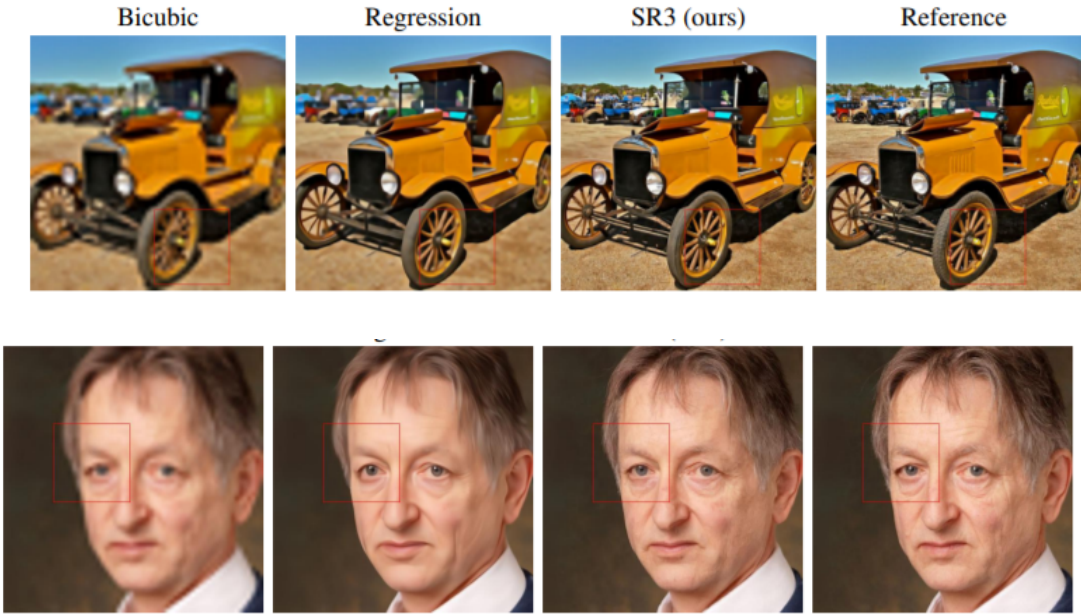
**Algorithm 2** Inference in  $T$  iterative refinement steps

```

1:  $y_T \sim \mathcal{N}(0, I)$ 
2: for  $t = T, \dots, 1$  do
3:    $z \sim \mathcal{N}(0, I)$  if  $t > 1$ , else  $z = 0$ 
4:    $y_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( y_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_\theta(x, y_t, \gamma_t) \right) + \sqrt{1-\alpha_t} z$ 
5: end for
6: return  $y_0$ 
  
```

❖ **Thực nghiệm**

- Đánh giá hiệu quả của mô hình SR3 trong siêu phân giải trên khuôn mặt, hình ảnh tự nhiên và hình ảnh tổng hợp thu được từ một mô hình tạo gen có độ phân giải thấp.



**Kết quả dự kiến**

*So sánh giữa các phương pháp:*

- Tính nhất quán: là một thước đo về tính nhất quán của các đầu ra siêu phân giải, việc tính toán MSE giữa các đầu ra lấy mẫu xuống và các đầu vào có độ phân giải thấp cho thấy SR3 đạt được khả năng đánh lỗi tính nhất quán tốt nhất.

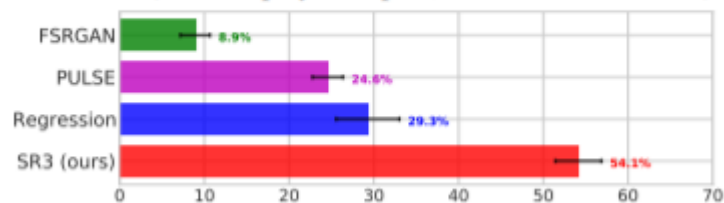
Metric	PULSE [28]	FSRGAN [7]	Regression	SR3
PSNR $\uparrow$	16.88	23.01	<b>23.96</b>	23.04
SSIM $\uparrow$	0.44	0.62	<b>0.69</b>	0.65
Consistency $\downarrow$	161.1	33.8	2.71	<b>2.68</b>

- Phân loại chính xác: phân loại đối tượng trên hình ảnh có độ phân giải thấp. SR3 vượt trội hơn các phương pháp hiện có với tỷ lệ chênh lệch lớn về lỗi phân loại top 1 và top 5, chứng tỏ chất lượng cảm nhận cao của đầu ra SR3.

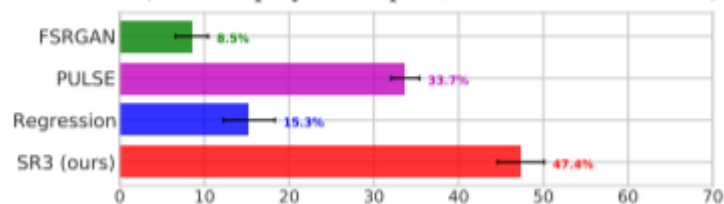
Method	Top-1 Error	Top-5 Error
Baseline	0.252	0.080
DRCN [22]	0.477	0.242
FSRCNN [13]	0.437	0.196
PsyCo [35]	0.454	0.224
ENet-E [44]	0.449	0.214
RCAN [64]	0.393	0.167
Regression	0.383	0.173
SR3	<b>0.317</b>	<b>0.120</b>

- Tỷ lệ đánh lừa của SR3 là gần đến 50%, chỉ ra rằng SR3 tạo ra hình ảnh chân thực và trung thực với các đầu vào có độ phân giải thấp. Tỷ lệ đánh lừa đối với FSRCNN và PULSE trong Task-1 là thấp hơn đường cơ sở hồi quy và SR3.

Fool rates (3 sec display w/ inputs,  $16 \times 16 \rightarrow 128 \times 128$ )



Fool rates (3 sec display w/o inputs,  $16 \times 16 \rightarrow 128 \times 128$ )



#### ❖ Bộ dữ liệu:

Tạo khuôn mặt vô điều kiện và các mô hình tạo ImageNet có điều kiện lớp bằng cách sử dụng DDPM trên cùng bộ dữ liệu đã thảo luận ở trên. Cho tập huấn và thử nghiệm, chúng tôi sử dụng các hình ảnh có độ phân giải thấp được lấy mẫu xuống bằng cách sử dụng phép nội suy hai chiều với tính năng khử răng cưa. Đối với ImageNet, chúng tôi loại bỏ các hình ảnh có hình ảnh có cạnh ngắn hơn độ phân giải mục tiêu

<b>Tài liệu tham khảo</b>	<p>[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Image Super-resolution via Progressive Cascading Residual Network. In CVPR, 2018.</p> <p>[2] Martin Arjovsky, Soumith Chintala, and Leon Bottou. Wasserstein GAN. In arXiv, 2017.</p> <p>[3] David Berthelot, Peyman Milanfar, and Ian Goodfellow. Creating high resolution images with a latent adversarial generator. arXiv preprint 2003.02365, 2020.</p> <p>[4] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096, 2018.</p> <p>[5] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning Gradient Fields for Shape Generation. In ECCV, 2020.</p> <p>[6] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, and William Chan. WaveGrad: Estimating Gradients for Waveform Generation. In ICLR, 2021.</p> <p>[7] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages 2492–2501, 2018.</p> <p>[8] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. In ICCV, 2017.</p> <p>[9] Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. In NIPS, 2015.</p> <p>[10] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. arXiv:1605.08803, 2016</p>
---------------------------	---