
BÁO CÁO MÔ HÌNH SINH REALNVP

Ngày 29 tháng 11 năm 2020

Giáo viên hướng dẫn: PSG.TS Thân Quang Khoát

Sinh viên thực hiện: Nguyễn Lý Thái Dương, Nguyễn Văn Chức

GIỚI THIỆU

Mô hình RealNVP được xuất phát từ ý tưởng của Flow Model trong đó chúng ta muốn tìm một xác suất $p_{\theta}(x)$ với biến x liên tục để sinh ra ảnh. Mục đích của mô hình này nhằm tìm được một phương pháp để train dữ liệu, với mỗi giá trị dữ liệu x mới thì cần đánh giá được giá trị $p_{\theta}(x)$. Từ việc biết được phân phối $p_{\theta}(x)$, chúng ta có thể tiến hành lấy mẫu và sinh ra được dữ liệu mới. Hay nói theo cách khác, chúng ta cần hahir đi tìm một mô hình biến ản có ý nghĩa.

Trên thực tế việc tìm ra $p_{\theta}(x)$ gặp rất nhiều khó khăn. Do đó một ý tưởng nảy ra trong mô hình Flow model là tìm cách biến đổi x thành một biến z bằng một song ánh

$$x = \{x\}$$

mà trong đó chúng ta đã biết được phân phối của biến z này. Cụ thể trong mô hình RealNVP, chúng ta lấy :

$$z \sim \mathcal{N}(0, 1)$$

XÂY DỰNG MÔ HÌNH

0.0.1 Mô hình Flow Model

Vì hàm $z = \{x\}$ là một hàm song ánh, việc xem xét x và z là tương đương nhau. Như vậy điểm mấu chốt là đi tìm hàm $f()$. Để thỏa mãn tính chất song ánh và thỏa mãn cho x và z là các biến tuân theo các phân phối liên tục thì $\{f\}$ phải là các hàm đơn biến và khả nghịch.

Trong số các hàm tác dụng lên một biến tuân theo phân phối, hàm CDF (Cumulative Density Function) luôn cho phép biến đổi một phân phối bất kỳ về một phân phối Uniform. Do đó, mọi phân phối có thể biến được cho nhau.

Mô hình RealNVP

Dựa trên ý tưởng của mô hình Flow, mô hình RealNVP áp dụng cho biến nhiều chiều x và train theo hướng tối ưu hóa theo Negative Log likelihood:

$$\operatorname{argmin}_{\theta} \mathbb{E}_x [-\log p_{\theta}(x)] \quad (1)$$

Theo phép đổi biến từ x sang z chúng ta có:

$$p_{\theta}(x) = p(f_{\theta}(x)) \left| \det \frac{\partial f_{\theta}(x)}{\partial x} \right| \quad (2)$$

Thay vào (1) ta có:

$$\operatorname{argmin}_{\theta} \mathbb{E}_x [-\log p_{\theta}(x)] = \mathbb{E}_x \left[-\log p_{\theta}(z) - \log \det \left| \frac{\partial f_{\theta}(x)}{\partial x} \right| \right] \quad (3)$$

Chúng ta có thể nhận thấy một chuỗi các song ánh thì cũng tương đương với một song ánh. Do đó trong thuật toán RealNVP sử dụng một chuỗi các biến đổi song ánh. Do lý do đơn giản và thuận tiện cho việc tính toán tác giả sử dụng biến đổi song ánh có dạng

affine. Xét chuỗi biến đổi từ x sang z :

$$x \longrightarrow f_1 \longrightarrow f_2 \longrightarrow \cdots \longrightarrow f_k \longrightarrow z \quad (4)$$

hay:

$$z = f_k \circ \cdots \circ f_1 (x) \quad (5)$$

Trong đó mỗi biến đổi affine $f_i(x)$ có dạng:

$$f(x) = A^{-1} (x + b) \quad (6)$$

Như vậy việc tính toán định thức trong công thức (3) chính là định thức của A . Xét biến đổi Affine cụ thể mà ta chia đầu vào của mỗi layer Affine này thành hai phần bằng nhau $x_{1:d/2}, x_{d/2+1,d}$. Khi đó biến đổi có dạng:

$$z_{1:d/2} = x_{1:d/2} \quad (7)$$

$$z_{d/2+1,d} = x_{d/2+1,d} s_{\theta} (x_{1:d/2}) + t_{\theta} (x_{1:d/2}) \quad (8)$$

Khi đó ma trận định thức có dạng:

$$\begin{bmatrix} \mathbb{I} & 0 \\ \frac{\partial z_{d/2+1,d}}{\partial x_{1:d/2}} & diag(s_{\theta}(x_{1:d/2})) \end{bmatrix}$$

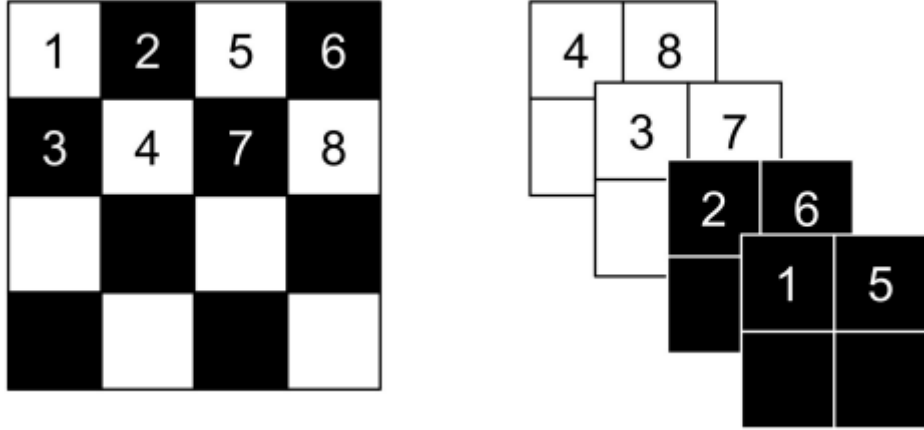
Với mục đích model có thể học được đa dạng sự tương quan giữa các thành phần của bức

ảnh, mô hình RealNVP cho phép sử dụng các mask thay thế cho affine liên tục như mô tả trên.

$$z = b \odot x + (1 - b) \odot (x \odot \exp s(b \odot x) + t(b \odot x)) \quad (9)$$

$$z = b \odot x + (1 - b) \odot (x \odot \exp s(b \odot x) + t(b \odot x))$$

Trong đó b là mask của mô hình. Trên thực tế, mô hình sử dụng hai loại mô hình là checkerboard và channelwise.



Hình 1: Hai loại mask

Trong đó để chuyển từ bức ảnh to kích thước $w * h * c$ về bức ảnh kích thước $\frac{w}{2} * \frac{h}{2} * 4c$ bằng cách chia bức ảnh ban đầu thành các bức ảnh nhỏ kích thước 2*2. Chia các phần tử tương ứng của mỗi bức ảnh này vào các channel nhỏ hơn. Để minh họa rõ hơn về cách thức phân chia có thể theo dõi ảnh trên.

Kiến trúc mạng

- Checkerboard Affine type 1
- Actnorm
- Checkerboard Affine type 0
- Actnorm
- Checkerboard Affine type 1
- Actnorm
- Checkerboard Affine type 0
- Squeeze
- Channelwise Affine type 0
- Actnorm
- Channelwise Affine type 1
- Actnorm
- Channelwise Affine type 0
- Unsqueeze
- Checkerboard Affine type 1
- Actnorm
- Checkerboard Affine type 0
- Actnorm
- Checkerboard Affine type 1

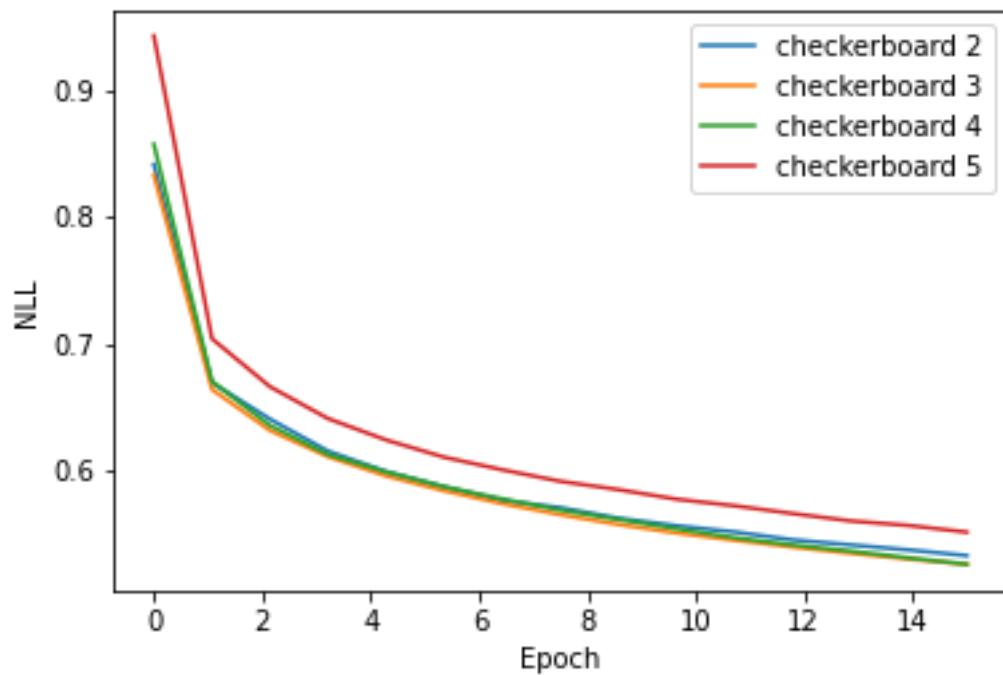
CÀI ĐẶT- KIỂM THỬ

Dữ liệu

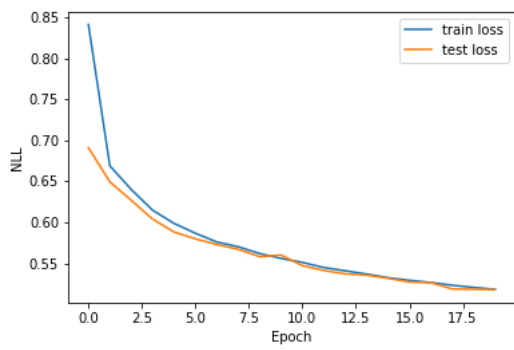
Dữ liệu sử dụng cho cài đặt và kiểm thử là bộ CELEB trong đó, lượng ảnh dùng để train và test có kích thước 10000 và 6000. Mỗi ảnh là ảnh màu kích thước 32 x 32.

0.0.2 Ảnh hưởng số lượng Checker board

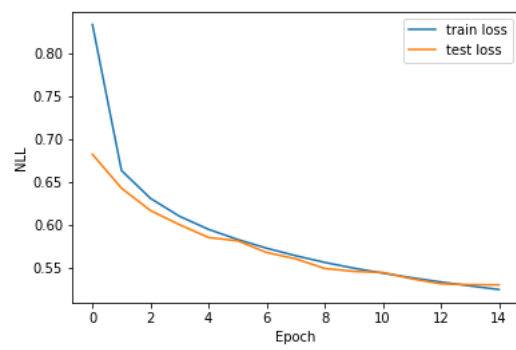
So sánh số lượng các checker board của khối checker board ban đầu.



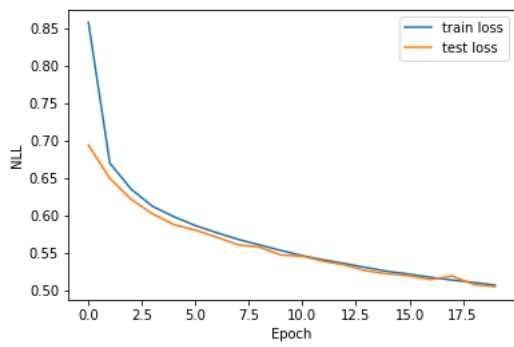
Hình 2: So sánh số lượng checkerboard



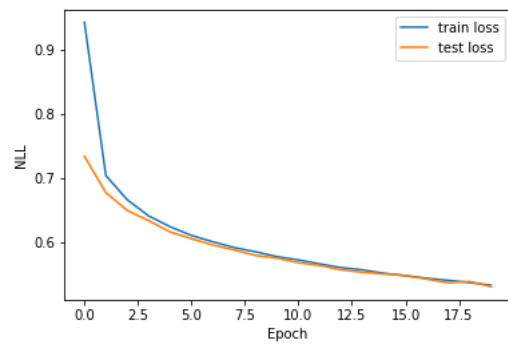
Hình 3: 2 checkerboard



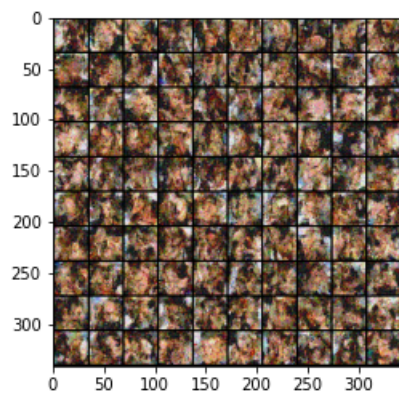
Hình 4: 3 checkerboard



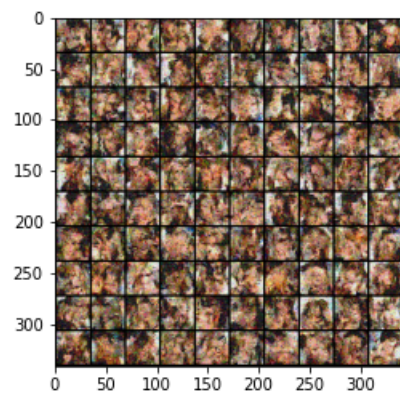
Hình 5: 4 checkerboard



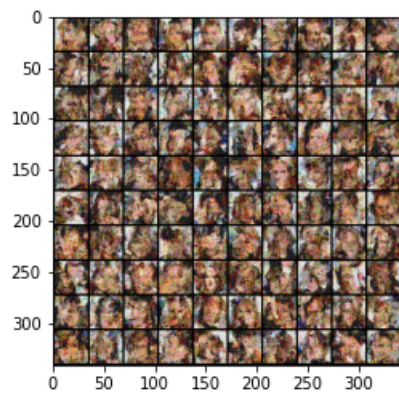
Hình 6: 5 checkerboard



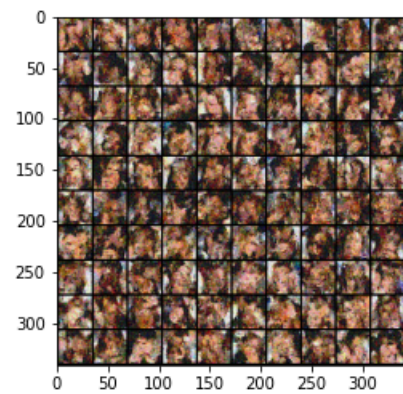
Hình 7: 2 checkerboard



Hình 8: 3 checkerboard

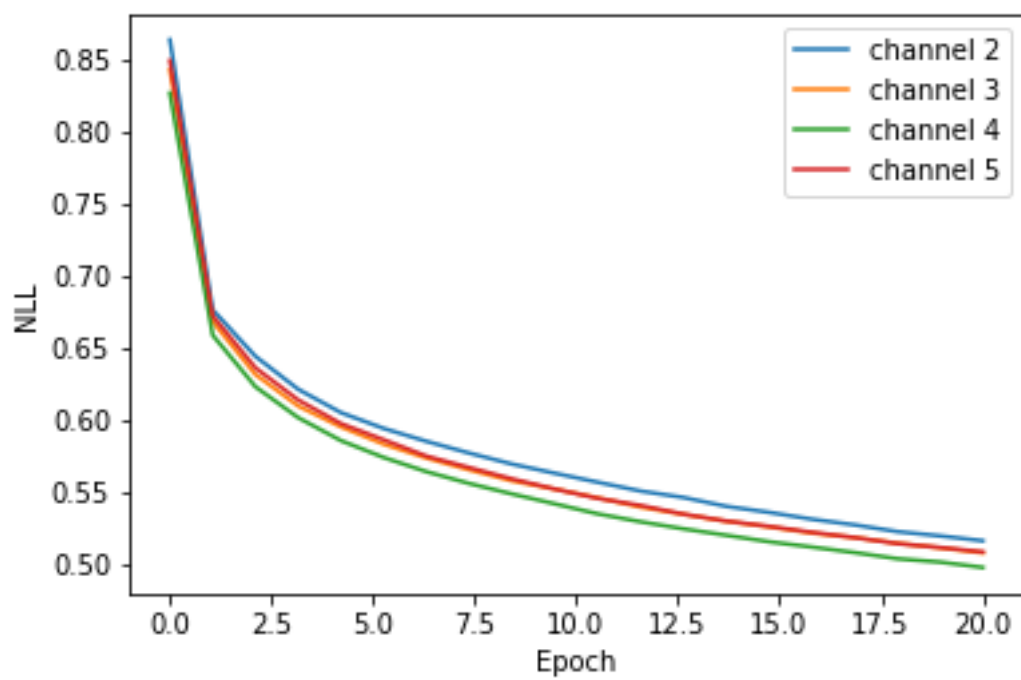


Hình 9: 4 checkerboard

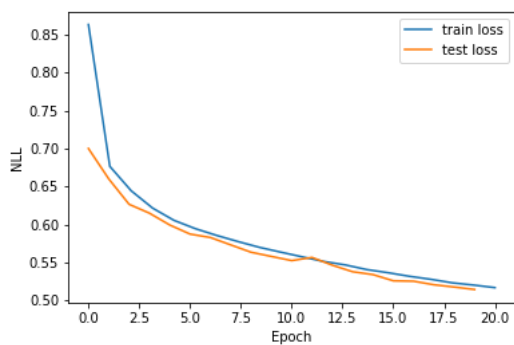


Hình 10: 5 checkerboard

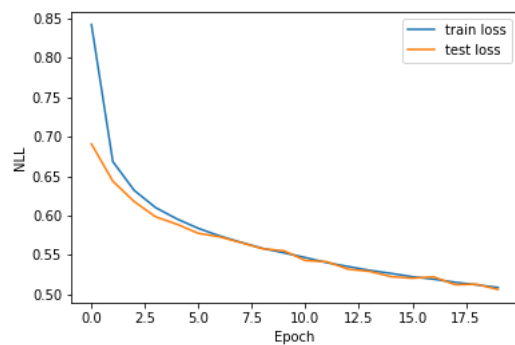
Ảnh hưởng số lượng channel wise



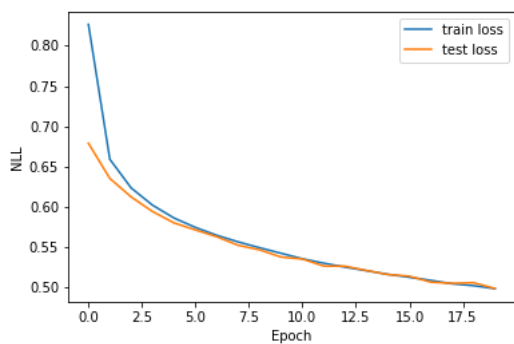
Hình 11: So sánh số lượng channelwise



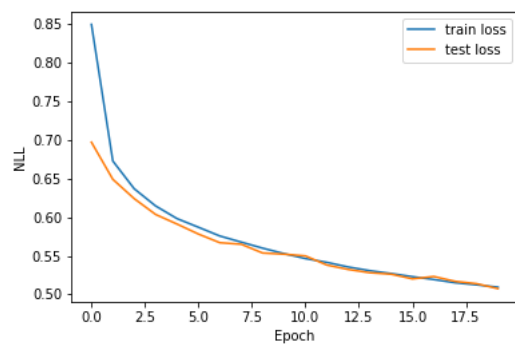
Hình 12: 2 channel



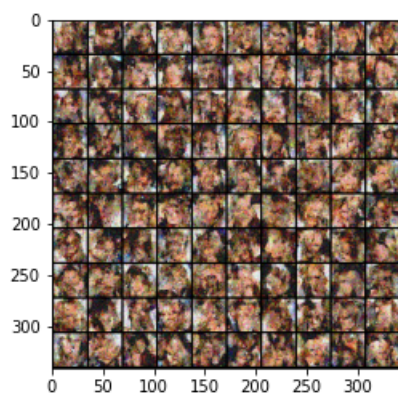
Hình 13: 3 channel



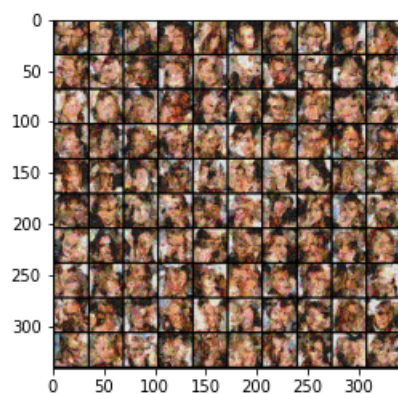
Hình 14: 4 channel



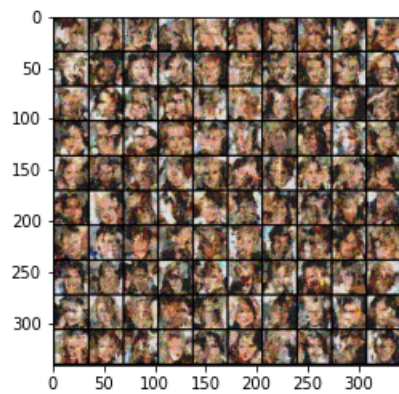
Hình 15: 5 channel



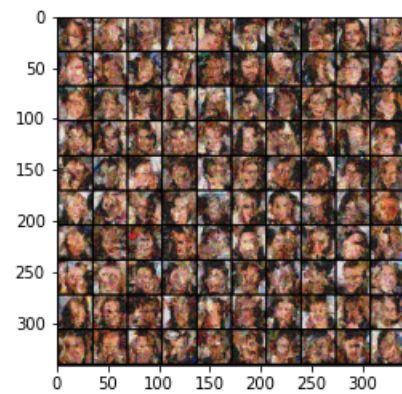
Hình 16: 2 channel



Hình 17: 3 channel



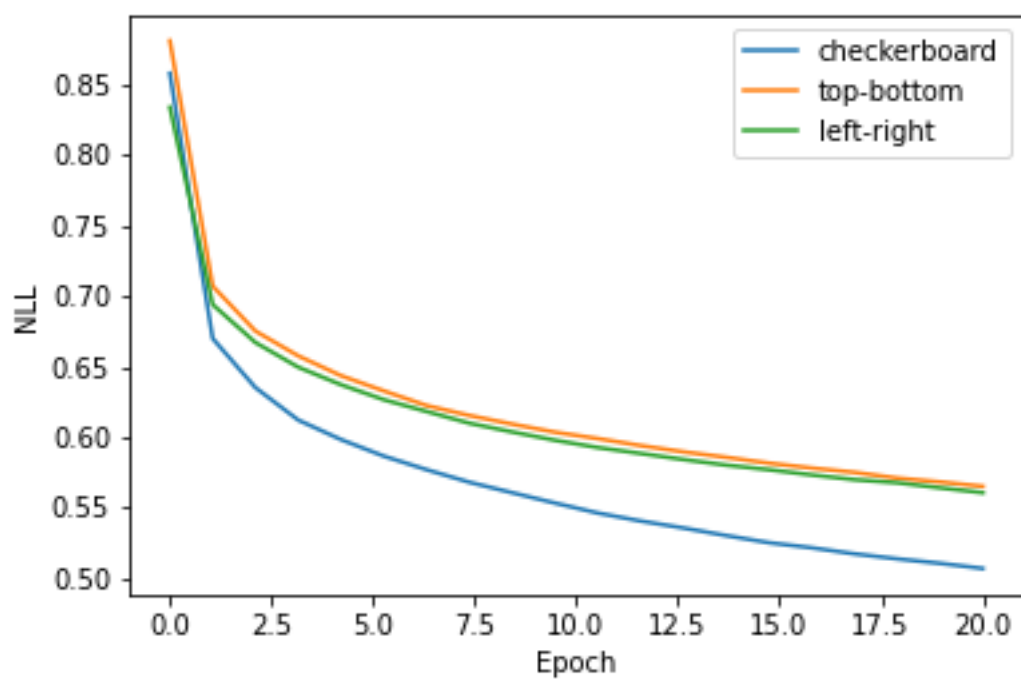
Hình 18: 4 channel



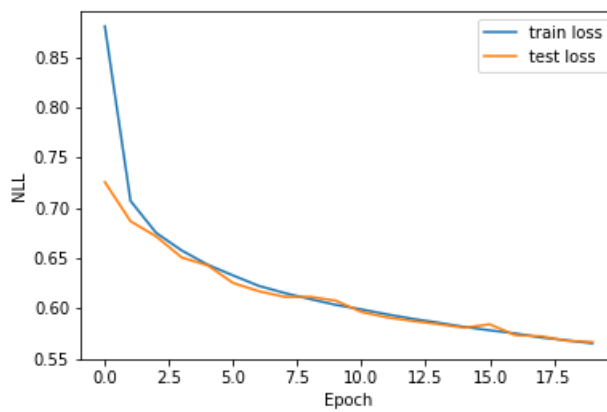
Hình 19: 5 channel

Ảnh hưởng của các loại mask khác nhau trong lớp affine

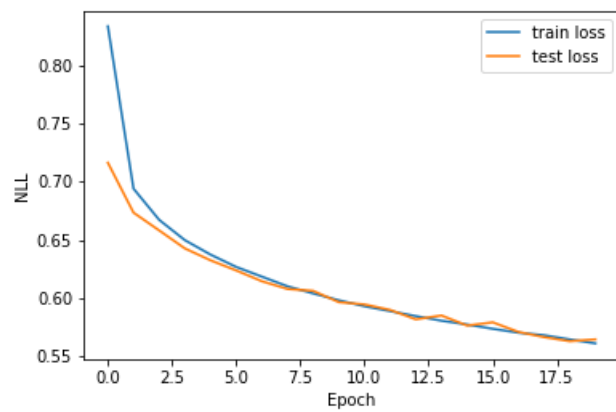
Trong phần này nhóm tiến hành thí nghiệm với hai mask là các cặp nửa top-bottom và left-right. Kết quả train cho hai loại mask mới này:



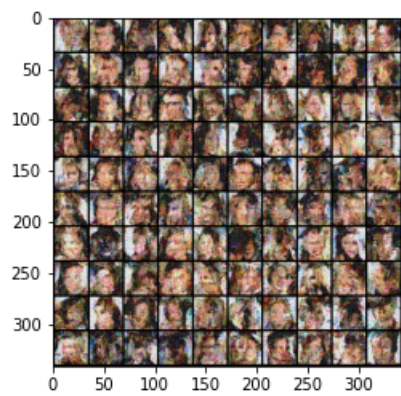
Hình 20: So sánh các loại mask



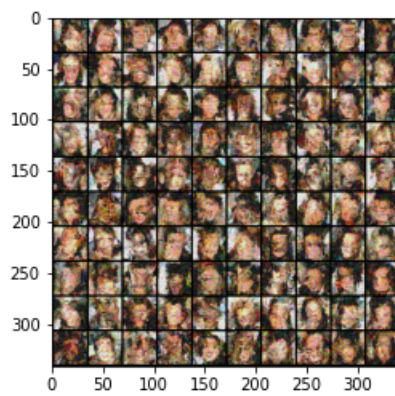
Hình 21: top-bottom mask



Hình 22: left- right mask



Hình 23: top-bottom mask



Hình 24: left-right mask

Kết quả có thể nhận thấy rằng dùng mask có dạng checkerboard tốt hơn so với dùng các loại mask khác do bắt nguồn từ lý thuyết các ô màu đen được suy diễn từ các ô màu trắng. Do đó, mask có dạng checkerboard đảm bảo mối tương quan giữa các pixel của bức ảnh cao hơn so với hai loại mask còn lại.

KẾT LUẬN

Trong bài báo cáo này, nhóm đã trình bày những khái niệm và lý thuyết cơ bản cũng như cấu trúc của mô hình RealNVP. Từ những kết quả dễ thấy trong việc thay đổi các tham số trong mô hình, cho thấy lý thuyết đã được kiểm chứng. Bên cạnh đó, các thông số về lựa chọn cấu trúc mô hình của nhóm tác giả đặt ra ban đầu đã được tối ưu. Tuy nhiên, về mặt trực quan, chất lượng ảnh sinh ra kể cả từ những mô hình tốt nhất thì kết quả vẫn không được tốt.