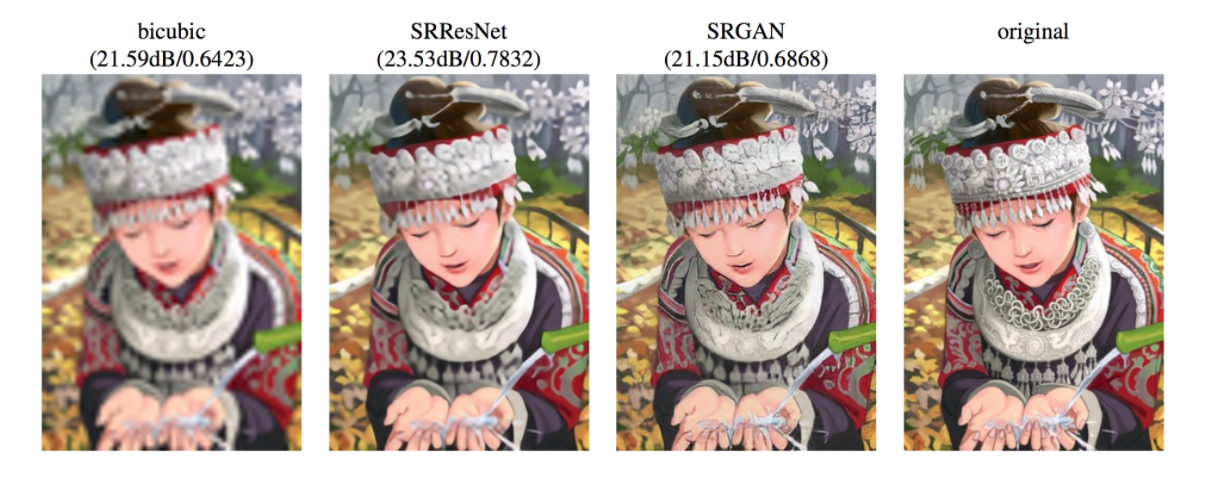
**Bài toán siêu phân giải**

**(Super-resolution)**

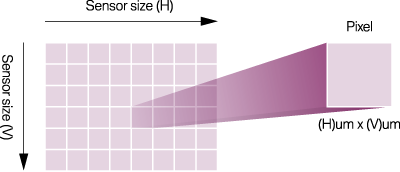


Giới thiệu

*Siêu phân giải là cụm từ dành cho những bức ảnh có độ phân giải cao và chất lượng ảnh cao hơn so với ảnh gốc được đưa ra. Bài toán siêu phân giải là bài toán chuyển từ ảnh có độ phân giải thấp sang ảnh có độ phân giải cao sao cho chất lượng ảnh được cải thiện hơn so với bức ảnh gốc. Với sự phát triển của trí tuệ nhân tạo, ngày càng xuất hiện nhiều phương pháp cũng như mô hình trí tuệ nhân tạo để giải quyết bài toán này.*

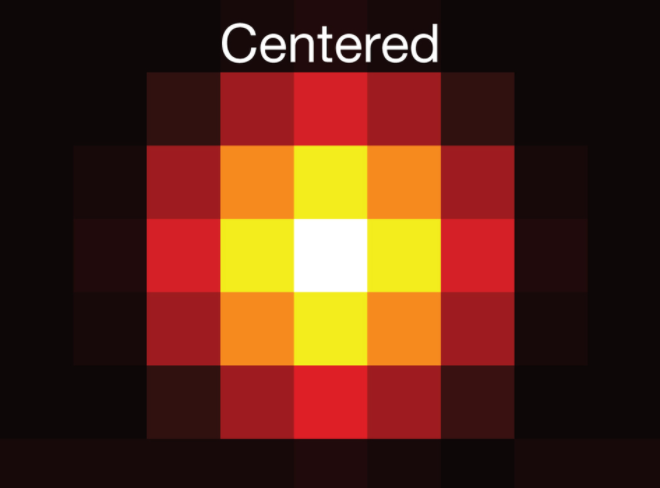
# Ảnh số và siêu phân giải.

Ảnh số (Digital image) là ảnh được biểu diễn trên các thiết bị điện tử hiện nay (máy tính, điện thoại, máy ảnh kỹ thuật số), những bức ảnh được biểu diễn dưới dạng Frame hay một ma trận hay chiều, mỗi giá trị trên ma trận thể hiện thông tin về màu sắc tại vị trí trên ma trận đó.



* **Độ phân giải (Resolution)**: là kích cỡ ma trận điểm ảnh biểu diễn cho ảnh (đơn vị pixel, megapixel).
* **Kênh màu (Channel)**: là số lượng ma trận biểu diễn cho ảnh (thông thường là 1 đối với ảnh xám, 3 đối với ảnh màu).
* **Điểm ảnh (Pixel)**: các giá trị trong ma trận thể hiện màu sắc của ảnh (giá trị này thông thường là 0 đến 255).
* **Kích thước ảnh (Size)** : là kích thước vật lý của bức ảnh (inch, cm, mm), ngoài ra của được dùng để nói lên kích thước vật lý của từng điểm ảnh.
* **ppi** (pixels per inch) là mật độ thông tin mà các màn hình có thể thu nhận trên mỗi inch.
* **dpi** - (dots per inch) là số điểm trên diện tích 01 inch vuông được tính trong in ấn.

*Thông thường một bức ảnh có chất lượng cao phụ thuộc khá nhiều vào máy ảnh của thiết bị chụp ảnh cũng như thuật toán định vị* ***(Texture)*** *các giá trị điểm ảnh cho bức ảnh đó,* ***các thông tin thể hiện cho bức ảnh sẽ bị giới hạn bởi độ phân giải máy chụp ảnh (hay nói cách khác máy chụp ở độ phân giải càng cao ảnh càng chất lượng).***



**Các phép toán nội suy:**

**….**

Bài toán siêu phân giải hay cải thiện (**enhance**) chất lượng ảnh khá phổ biến trong lĩnh vực xử lý ảnh số và thị giác máy tính, tăng cường độ phân giải của ảnh cũng như tăng những thông tin nhận được từ bức ảnh có độ phân giải thấp. Bài toán mang tính ứng dụng khá cao trong cuộc sống như cải thiện chất lượng ảnh:

* Các thiết bị chụp ảnh có độ phân giải thấp.
* Những đối tượng trong ảnh quá nhỏ so với kích thước của bức ảnh dẫn đến việc thể hiện thông tin của các đối tượng bị hạn chế.
* Ảnh bị nhiễu cũng như nhòe do môi trường điền kiện.
* Ứng dụng sử dụng đòi hỏi ảnh có độ phân giải cao.
* Cải thiện ảnh phục vụ cho các bài toán khác (nhận diện).

Tuy nhiên bài toán siêu phân giải là một bài toán khó, hầu như không có thuật toán thủ công nào có thể xử lý bài toán một cách tốt nhất. Với sự phát triển của trí tuệ nhân tạo và sự ra đời của Neural Network đã cho ra những phương pháp giải quyết bài toán một cách có hiệu quả.

# Convolutional neural network vs Super-resolution.

## Super-resolution Convolutional neural network (SRCNN).

SRCNN được giới thiệu vào năm 2014 bởi Chao Dong và các đồng nghiệp, mô hình đưa ra áp dụng dựa trên mạng Neural tích chập bao gồm 3 phần:

* **Patch extraction & representation**: Rút trích các đặc trưng ảnh (Low-resolution).
* **Non-linear mapping:** Mapping phi tuyến giữa 2 tầng tích chập (mapping vector này sang vector khác không làm thay đổi chiều).
* **Recontruction:** Tái cấu trúc ảnh từ danh sách các Fearture map đã được mapping.

Dữ liệu ảnh bao gồm ảnh gốc (Ground-truth) và ảnh đã được giảm độ phân giải bằng các phép toán nội suy (Bicubic, Gaussian).

Mô hình theo hướng Pre-upsampling (Ảnh sẽ được tăng độ phân giải trước khi đưa vào mô hình).

Input image: LR up-sampled (ISR).

Conv Layer 1: Patch extraction.

* N1 filter of size C \* f1 \* f1.
* Activation function: ReLU.
* Parametes: C \* f1 \* f1 \* N1 and N1  bias.

Conv Layer 2: Non-linear Mapping.

* N2 filter of size N1 \* f2 \* f2.
* Activation function: ReLU.
* Parametes: N1 \* f2 \* f2 \* N2 and N2  bias.

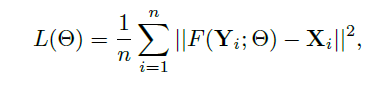
Conv Layer 3: Patch extraction.

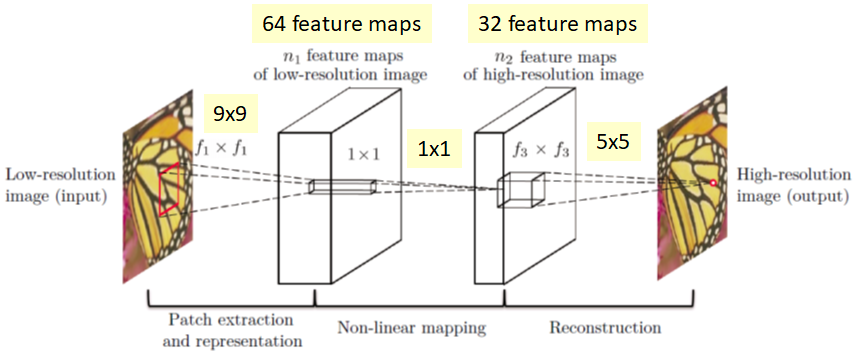
* 1 filter of size N2 \* f2 \* f2.
* Activation function: Identity.
* Parametes: N2 \* f3 \* f3 \* C and C bias.

ReLU activation: F(x) = max(0, x).

Identity: F(x) = x.

Loss Function: Mean squared error (MSE) between the N reconstructed HR images and the N original true HR images.





Kiến trúc đề xuất

*Total parameters : 8032 params*

## Fast Super-Resolution Convolutional Neural Networks (FSRCNN).

Được đồng tác giả SRCNN giới thiệu nhằm mục đích cải thiện tốc độ trainning của SRCNN và kết quả mô hình đưa ra.

Input Image: LR single channel.

Conv. Layer 1: Feature extraction

* 56 filters of size 1 x 5 x 5.
* Activation function: PReLU.
* Parameters: 1 x 5 x 5 x 56 = 1400 weights and 56 biases.

Conv. Layer 2: Shrinking

* 12 filters of size 56 x 1 x 1.
* Activation function: PReLU.
* Parameters: 56 x 1 x 1 x 12 = 672 weights and 12 biases.

Conv. Layers 3–6: Mapping

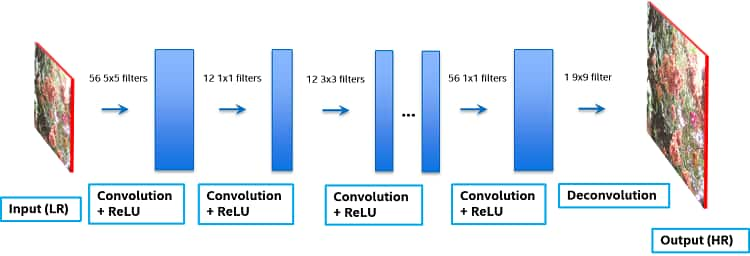
* 4 x 12 filters of size 12 x 3 x 3.
* Activation function: PReLU.
* Parameters: 4 x 12 x 3 x 3 x 12 = 5184 weights and 48 biases.

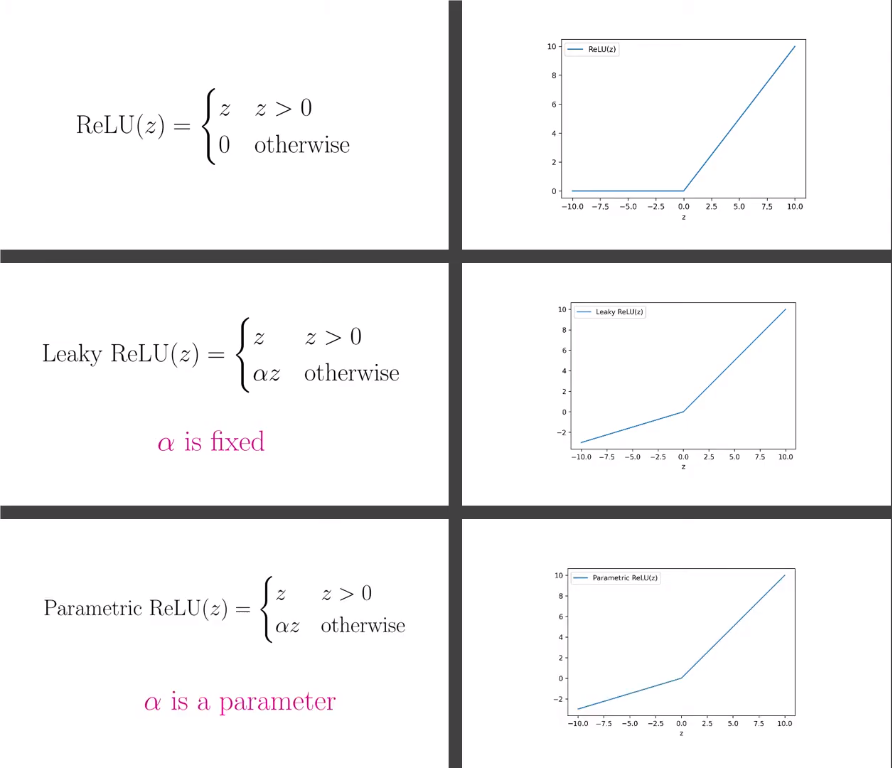
Conv. Layer 7: Expanding

* 56 filters of size 12 x 1 x 1.
* Activation function: PreLU.
* Output: 12 feature maps.
* Parameters: 12 x 1 x 1 x 56 = 672 weights and 56 biases.

DeConv Layer 8: Deconvolution

* One filter of size 56 x 9 x 9.
* Activation function: PReLU.
* Parameters: 56 x 9 x 9 x 1 = 4536 weights and 1 bias.

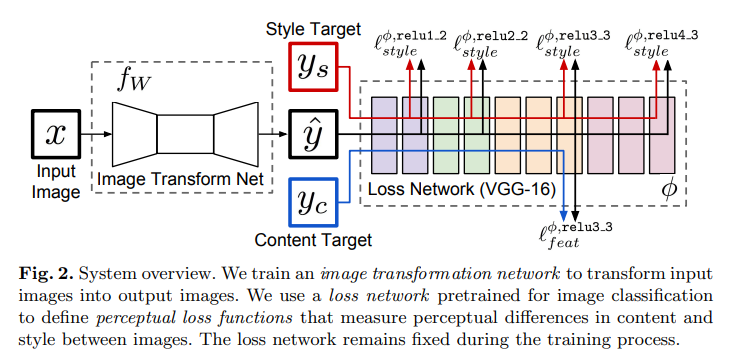
Total parameters: 12644 (plus a very small number of parameters in PReLU layers)



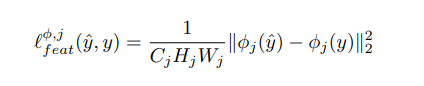
## Perceptual Loss vs Subpixel Convolution layer.

### Perceltual Loss (Feature Loss).

Sử dụng MSE làm hàm Loss trực tiếp giữa giá trị ảnh gốc đối với ảnh đầu ra của mô hình gặp nhiều nhược điểm, vì thực chất MSE chỉ đo cường độ khác biệt giữa giá trị thực so với giá trị đầu ra chứ không đó lường cấu trúc cũng như đặc trưng của ảnh.

Bằng việc tận dụng các mô hình máy học đã được trainning sẵn để phục vụ cho việc tính toán hàm Loss, những mô hình này được gọi là Loss model.

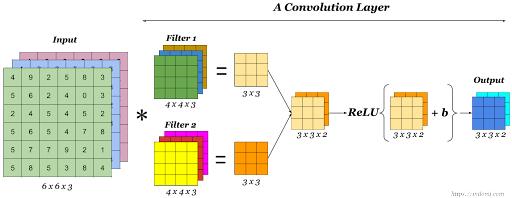
Đối với bài toán **Super-resolution** thì **Perceptual loss** **function** được áp dụng đó là Feature Loss (Content Loss):

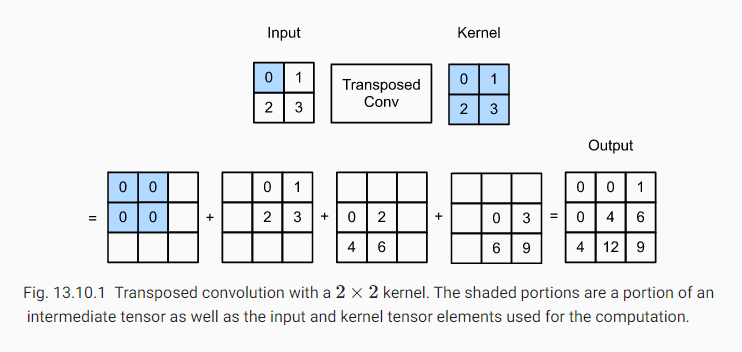


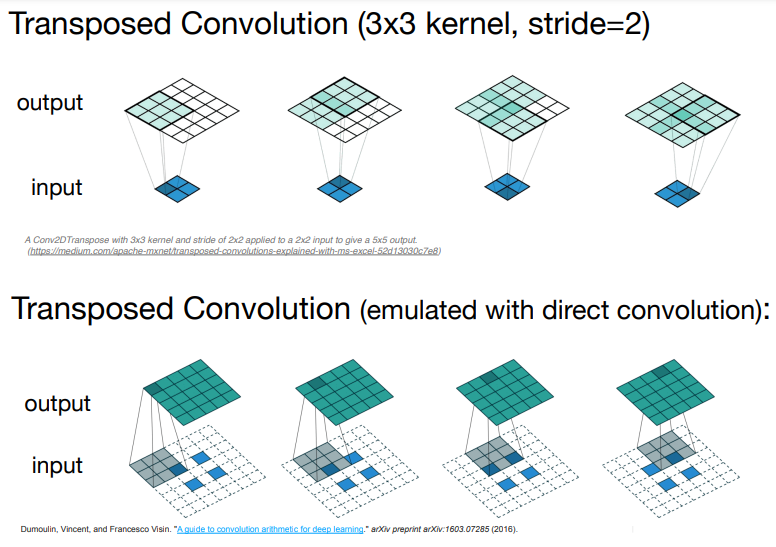
Trong đó *φj*(y)là Fearture Map của layer thứ j trong mô hình *φ,* trong đó y là input của mô hình.

Hàm dùng khoảng cách Euclidean (MSE) để so sánh giữa 2 giá trị *y*, .

### Sub-pixel Convolutional Neural network (ESPCN).

**Convolution layer**

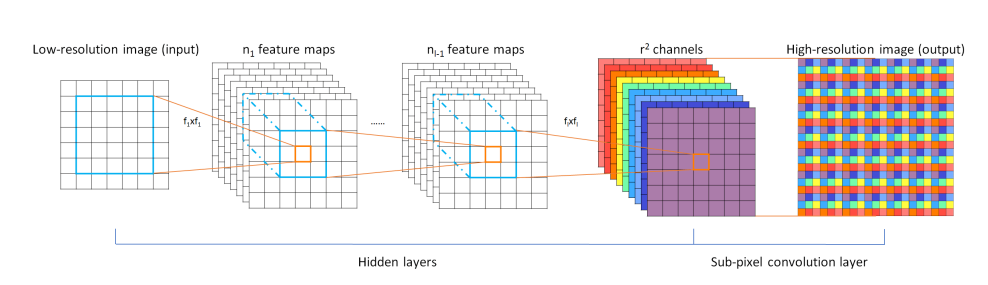
**Transposed Convolution layer (Fractionally Strided Convolution)**

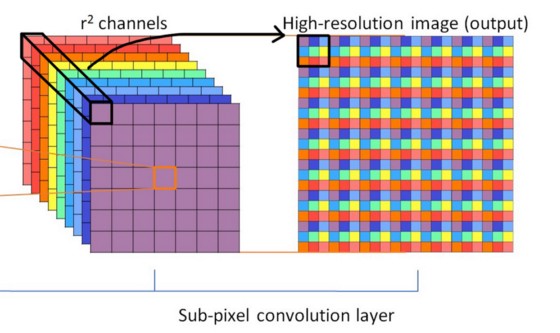
**Về mặt kỹ thật, Transposed Convolution là một phép tính tích chập thông thường với việc biến đổi đầu bằng cách chèn các số 0 vào input, do vậy còn được gọi với cái tên Fractionally Stride Convolution**

**Sub-pixel Convolutional layer**

Đối với các bài toán Super-resolution cần có những upsampled layer (tăng kích cỡ feature map), thông thường Transposed convolution được sử dụng trong bài toán này, tuy nhiên về mặt kỹ thuật thì cách tính layer này giống như convolution layer với input được biến đổi, điều này dẫn đến kết quả không tốt đối với bài toán.

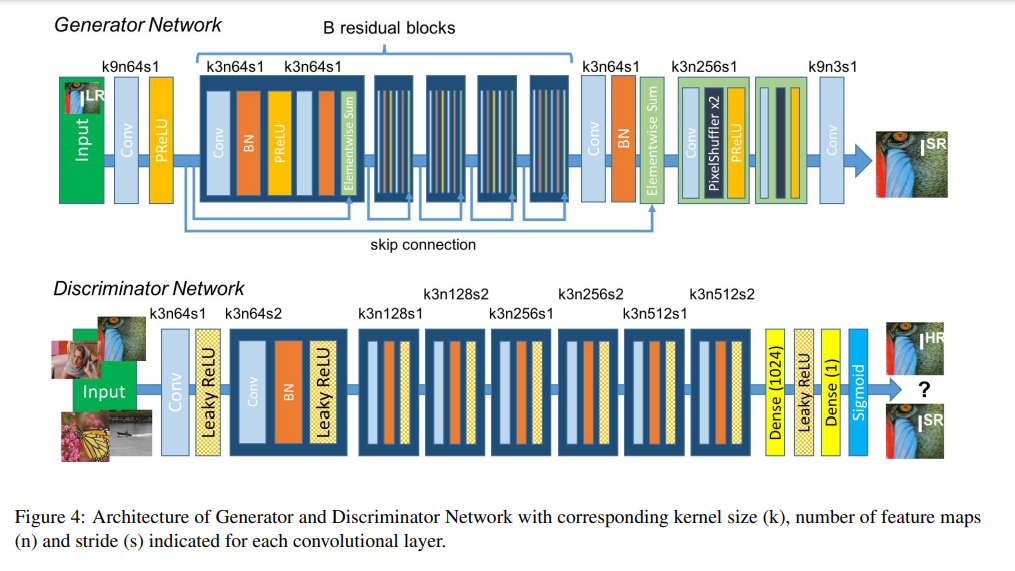
Sub-pixel được thay thế Transposed convolution layer trong các bài toán Super resolution sau này.



r là hệ số scale(x2, x3, …), số feature map của output sẽ bằng

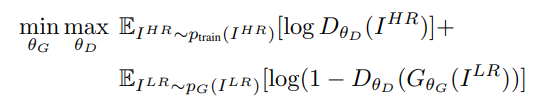
# Super-resolution Generative Adversarial Network (SRGAN).

SRGAN được Christian Ledig, Lucas Theis và các đồng nghiệp giới thiệu vào năm 2017 bằng việc ứng dụng mạng GAN vào bài toán Super resolution, bài báo giới thiệu 2 mô hình giải quyến bài toán là SRGAN và SRResNet, trong đó SRResNet là phần generator của SRGAN.

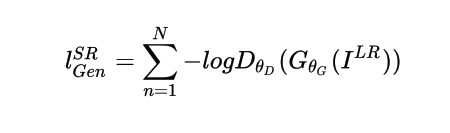


Đối việc trainning SRGAN thì mục tiêu trainning là Perceptual loss, nên hàm loss chính sẽ là tổng của 2 dạng loss:

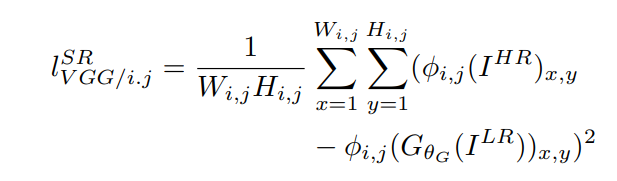
* **Adversarial Loss:**

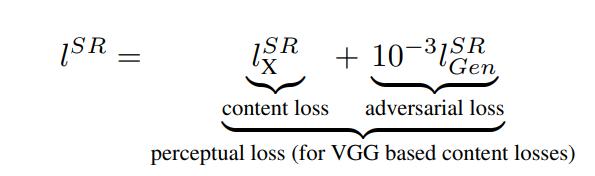


Generator loss



* **Content Loss:**

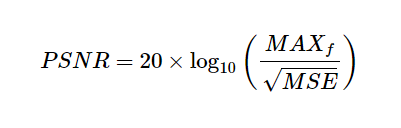


Perceptual loss cho Generator, trong quá trình trainning Generator cố gắng học dựa trên 2 tiêu chí là đánh lừa Discriminator và tạo ảnh siêu phân giải sao cho nét gần như ảnh gốc.

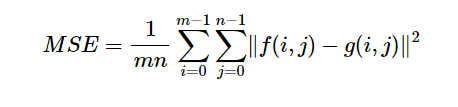
# Đánh giá các mô hình siêu phân giải.

Để đánh giá các mô hình siêu phân giải ta cần một độ đo (metrics) sao cho việc đánh giá ảnh giữa các mô hình là như nhau, để đánh giá thường có 2 metrics được sử dụng là:

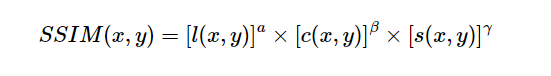
* **Peak signal-to-noise ratio (PSNR)**: Đây là độ đo thường dùng để so sánh độ giống nhau giữa 2 ảnh, độ đo dựa vào hàm MSE.



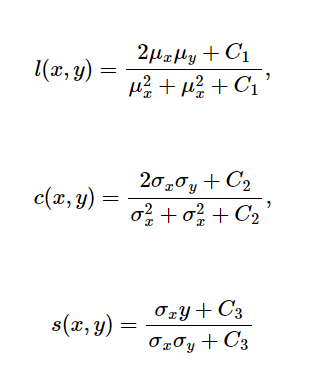
Trong đó:



* **The Structural Similiarity Index (SSIM):** Là độ đo mới khắc phục nhược điểm của PSNR trong, độ đó kiểm tra sự tương đồng giữa 2 bức ảnh dựa trên 3 yếu tố: luminance (l), contrast(c) và tructure(s).



Trong đó



Trong đó σ là độ lệch, μ là kỳ vọng và C là hằng số.

# Backpropagation on Convolutional layer.