Overview of Research on Image Super-Resolution

Trần Thị Vân Anh
Computer Vision Team of AithingsLab

16-2-2024

Table of Contents

Abstract

Tăng cường hình ảnh là một trong những kỹ thuật quan trọng nhất của xử lý ảnh, phản ánh đặc trưng bản chất của lĩnh vực này. Tùy vào mục đích mỗi bài toán khác nhau mà ta sử dụng phương pháp khác nhau như tăng cường ảnh trong lĩnh vực y tế, ảnh vệ tinh ... .Tăng cường hình ảnh được hiểu đơn giản như sau : thu nhỏ ảnh tạo ảnh LR, phóng to ảnh tạo ảnh HR Vấn đề đặt ra cần tái tạo ảnh HR từ 1 hoặc nhiều ảnh LR. Ảnh được biểu diễn dưới dạng ma trận số ví dụ ta có kích thước 3*3 => tạo ảnh 12*12 Các giá trị pixel (ptu trong ma trận) mới đượ tạo ra bằng cách nào ? Lúc này các kỹ thuật tăng độ phân giải ảnh đã ra đời , mỗi kỹ thuật có nhiều cách thực hiện khác nhau. Nhìn chung ta có thể chia giai đoạn phát triển của nó như sau :
xử lý ảnh truyền thống: nội suy -> mạng tích chập sử dụng học sâu -> transformer

Keywords: SISR,MISR,SR,…

Overview of Research on Image Super-Resolution

Trong quá trình tái tạo ảnh chất lượng cao (HR) từ ảnh có độ phân giải thấp (LR), các giá trị pixel mới trong ảnh HR có thể được tạo ra thông qua các phương pháp như:

1. **Interpolation (Nội suy)**: Sử dụng các phương pháp nội suy như nội suy tuyến tính, nội suy đa giác, hoặc nội suy spline để ước lượng giá trị pixel mới dựa trên các giá trị pixel có sẵn trong ảnh LR.
2. **Deep Learning-based Approaches (Tiếp cận dựa trên học sâu)**: Sử dụng các mô hình máy học, đặc biệt là các mạng nơ-ron tích chập (CNNs), để học cách biến đổi ảnh LR thành ảnh HR. Các mô hình này thường được huấn luyện trên các cặp ảnh LR và HR để học các biến đổi phức tạp từ đầu vào đến đầu ra.
3. **Transformer-based Approaches (Tiếp cận dựa trên Transformer)**: Các kiến trúc dựa trên Transformer, như Swin Transformer, cũng đã được sử dụng để tăng cường độ phân giải ảnh. Chúng thường sử dụng cấu trúc shifted windows và self-attention để hiểu và biểu diễn thông tin không gian trong hình ảnh, giúp tạo ra các phiên bản ảnh HR chất lượng cao.

1.Phương pháp tăng cường ảnh truyền thống

một vài các phương pháp truyền thống phổ biến để tăng cường hình ảnh :

- giảm nhiễu hình ảnh phơi sáng nhiều lần
- khử mờ từng khung hình
- sub-pixel image localization:

+ ngoại suy:

***Ngoại suy tuyến tính**: cách sử dụng phương pháp hồi quy tuyến tính hoặc phương pháp của bình phương tối thiểu để xác định các hệ số tương ứng với một hàm tuyến tính, và sau đó sử dụng hàm này để dự đoán giá trị cho các điểm mới.

***Ngoại suy đa giác**: mối quan hệ giữa các điểm dữ liệu được xấp xỉ bằng các đa giác. Các phương pháp ngoại suy đa giác bao gồm đa giác Lagrange, đa giác Newton và đa giác spline. Các đa giác này được sử dụng để tạo ra các đa giác "khớp" với dữ liệu và sau đó được sử dụng để dự đoán giá trị cho các điểm mới.

***Ngoại suy cục bộ (Local interpolation)**: giả định rằng mối quan hệ giữa các điểm dữ liệu chỉ phụ thuộc vào một phần của không gian. Do đó, ta có thể sử dụng thông tin từ các điểm lân cận để dự đoán giá trị cho một điểm mới. Các phương pháp ngoại suy cục bộ bao gồm kỹ thuật nội suy và kỹ thuật bên ngoài suy.

***Ngoại suy dựa trên học máy (Machine learning-based extrapolation)**: sử dụng các mô hình học máy để học mối quan hệ giữa các điểm dữ liệu và dùng mô hình đã học được để dự đoán giá trị cho các điểm mới.

+ nội suy :

*** Nearest-neighbor interpolation (nội suy láng giềng gần nhất)**: là phương pháp đơn giản nhất, các pixel trong ảnh sẽ dùng giá trị của pixel trong ảnh gần nó nhất.

*Bilinear interpolation (nội suy song tuyến): phương pháp này sẽ nội suy giá trị của một pixel bằng cách tính trung bình có trọng số 4 (2x2) pixel lân cận.

***Bicubic interpolation (nội suy song khối)**: tương tự như bilinear interpolation nhưng với 16 (4x4) pixel lân cận.

***Lanczos interpolation (nội suy Lanczos)**: sử dụng thuật toán tính trung bình giá trị pixel bằng hàm sin.

**điều kiện tất cả as đến từ 1 nguồn duy nhất

=>nhược điểm : ảnh output khá mờ , bị răng cưa , k thể restore lại được các chi tiết trong ảnh gốc .

2.Phân loại SR

Có hai loại chính của super resolution:

1. **Single Image Super Resolution (SISR)**: Trong SISR, một mô hình được huấn luyện để tăng cường độ phân giải của một hình ảnh đơn lẻ. Mô hình này sẽ học cách chuyển đổi các hình ảnh có độ phân giải thấp thành các hình ảnh có độ phân giải cao hơn bằng cách học các mối quan hệ và cấu trúc trong dữ liệu hình ảnh.
2. **Multi-Image Super Resolution (MISR)**: Trong MISR, nhiều hình ảnh có độ phân giải thấp được sử dụng để tạo ra một hình ảnh có độ phân giải cao hơn. Các kỹ thuật như việc kết hợp thông tin từ nhiều hình ảnh, tự động chụp ảnh từ các góc độ khác nhau, hoặc sử dụng dữ liệu từ các cảm biến khác nhau có thể được áp dụng để nâng cao độ phân giải.

=> learn & deep learning , transformer

3.Learn in SR

- tìm hiểu mối tương quan giữa LR và HR bằng việc train 1 lượng lớn data
- sử dụng phương pháp mã hóa và học sâu : lấy mẫu ở tần số cao để dự đoán HR , SC(spare coding) chọn ảnh HR và LR tạo từ điển -> tính toán hệ số thưa (sparse coefficient) -> tạo SR - Timofte's Anchored Neighborhood Regression (ANR): kết hợp từ điển thưa và hồi quy khu vực theo chốt
- ưu điểm : khắc phục nhược điểm phương pháp nhúng , có tính mở rộng mạnh mẽ
- nhược điểm : bị hạn chế bởi model framework, k nhạy cảm với dữ liệu nhiễu , hình ảnh tái tạo và độ phức tạp tính toán phụ thuộc vào tử điển , thời gian tính toán dài và hiệu suất tính toán thấp , từ điển k đầy đủ => ảnh tạo ra chất lượng thấp

4.Deep learning in SR

Kỹ thuật này có thể làm tăng nội dung thông tin của hình ảnh, nhưng không có gì đảm bảo rằng các tính năng được nâng cấp tồn tại trong hình ảnh gốc và không nên sử dụng

các công cụ nâng cấp tích chập sâu trong các ứng dụng phân tích có đầu vào không rõ ràng. Những phương pháp này có thể gây ảo giác cho các đặc điểm hình ảnh, khiến chúng không an toàn khi sử dụng trong y tế .

4.1. chỉ số đánh giá hiệu suất hình ảnh

- PSNR (Peak Signal-to-Noise Ratio): Tỉ lệ tín hiệu-tới-nhiễu cực đại
- SSIM(Structural Similarity Index): chỉ số tương đồng cấu trúc, tương đồng cấu trúc

=>Giá trị càng cao của cả hai chỉ số này, thì giá trị pixel của kết quả tái tạo càng gần với tiêu chuẩn.

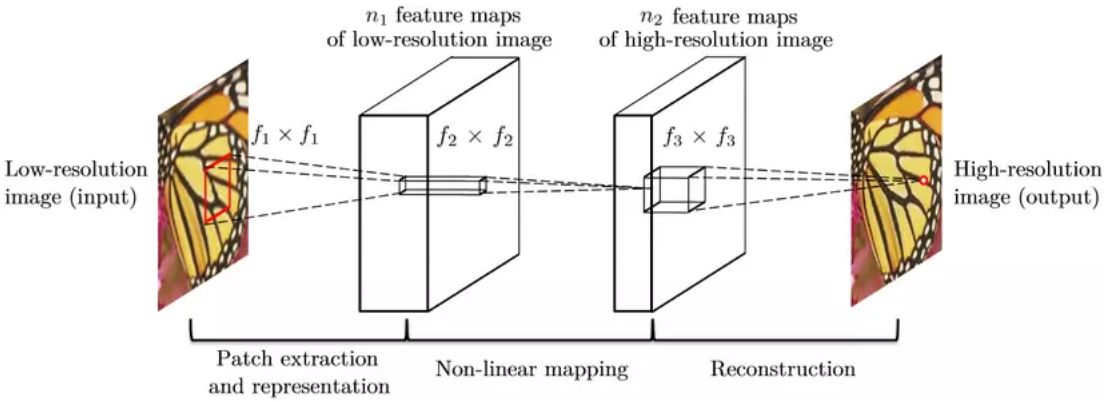
- MSE(Mean Squared Error): hàm lỗi

4.2. các phương pháp

a.SRCNN

-sử dụng lớp tích chập kích thước 9*9,1*1,5*5 => output kích thước 64,32

-MSE ->PSNR cao hơn => giảm thời gian tính toán và tăng độ chính xác



b. FSRCNN

- cải thiện của SRCNN : tăng tốc độ hàm tính toán
- những thay đổi :

+ sử dụng nội suy lấy mẫu tăng

+thay đổi chiều thuộc tính, hạt nhân tích chập nhỏ hơn , nhiều lớp ánh xạ để cải thiện thời gian thực

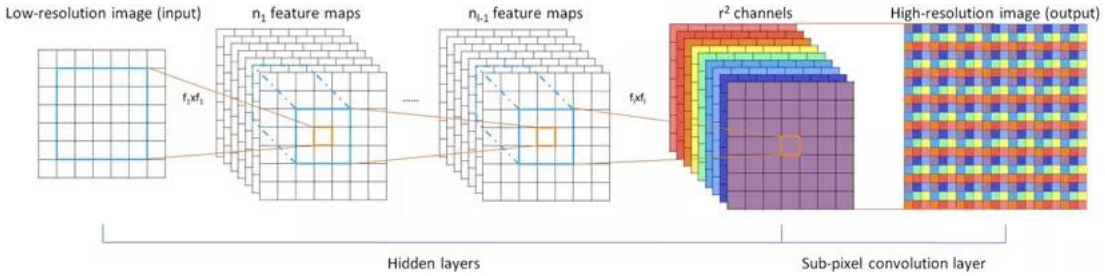
+ các lớp tích chập có thể chia ra =>cải thiện thời gian và chất lượng hình ảnh

c. ESPCNN

-cải tiến của SRRCNN

- chức năng nội suy quá trình tăng kích thước ảnh được tích hợp ngầm trong lớp tích chập trước và có thể được học tự động

=> các thuộc tính có thể được trích xuất trực tiếp trên kích thước của LR ảnh gốc => realtime & accurancy



-cần để ý phần chọn hệ số upscale r

d.VDSR

- cải thiện của SRCNN: sử dụng lượng lớn các lớp

=> độ chính xác ấn tượng

- đề xuất rằng LR và HR rất giống nhau, điều này có nghĩa là thông tin tần số thấp được mang bởi LR tương tự với thông tin tần số thấp của HR

-> gây tốn thời gian thực tế

=>chỉ cần học các phần sót lại của phần tần số cao giữa HR và LR

=> ý tưởng cộng dồn input LR(dùng nội suy để chuyển về size mong muốn) và residual error để mạng học

- cải tiến

+ mạng có 20 lớp

+sử dụng residual learning -> giảm tính toán, tăng tốc độ trung bình

+ adaptive gradient cropping -> tránh vấn đề bùng nổ độ dốc khi huấn luyện mạng noron

+các hình ảnh có độ phóng đại khác nhau đưa vào mạng huấn luyện

e. DRCN

-áp dụng RNN và residual learning cho ảnh SR

- model gồm có

+ mạng nhúng -> trích xuất đặc trưng

+ mạng học nền (Inference Network) sử dụng cấu trúc đệ quy để tăng cường việc truyền thông tin giữa các tầng và thiết lập kết nối ngữ cảnh-> khôi phục các chi tiết tần số cao trong hình ảnh

+ Mạng khôi phục (Reconstruction Network) nhận đầu vào là kết quả tích chập của mỗi tầng và hình ảnh đầu vào->kết hợp chúng để tạo ra kết quả khôi phục.

- ưu điểm giảm tham số của mạng , tránh bùng nổ gradient, sai số đầu ra mỗi tầng đệ quy và tổng sai số đầu ra lấy làm hàm mục tiêu khi huấn luyện

f.DRRN

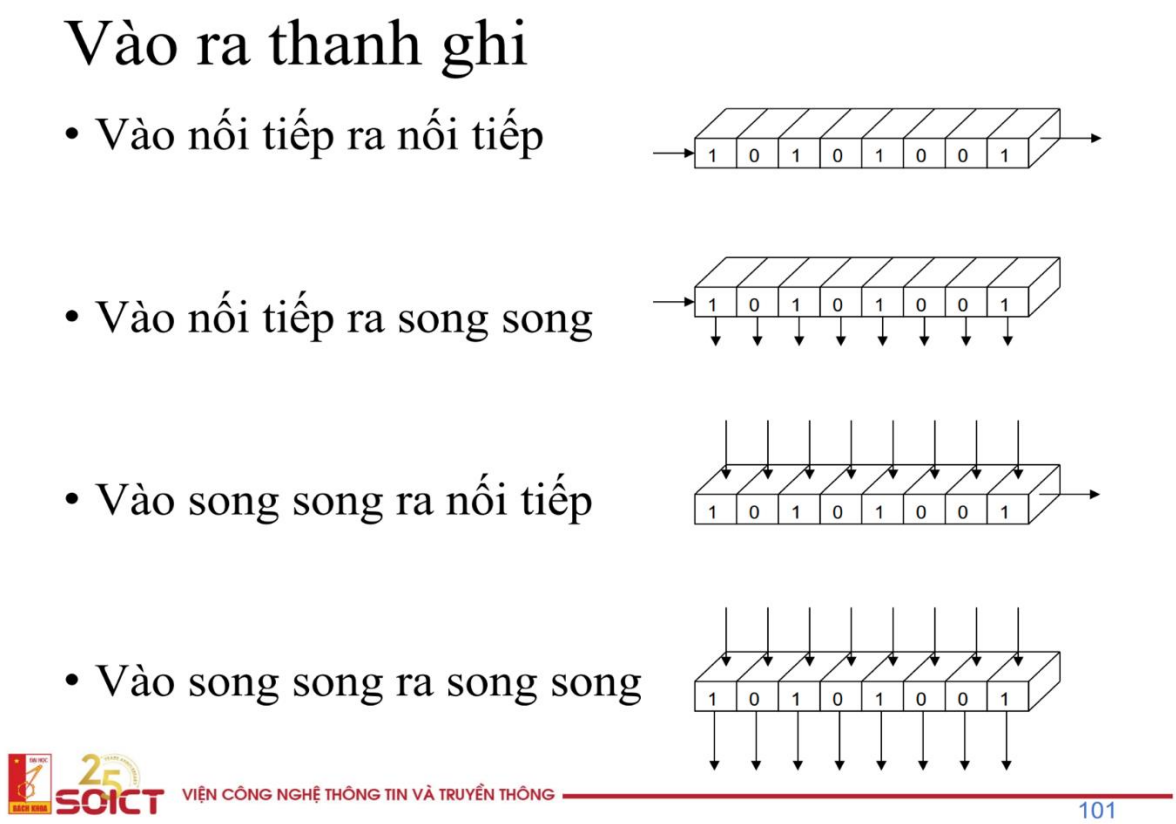
= local residual learning+ global residual learning + recursive learning của nhiều trọng số ở chế độ nhiều đường dẫn (1 recursive block and 25 residualunits with a 52-layer network structure)

-> việc điều chỉnh các cấu trúc ResNet hiện tại và các cấu trúc khác, và sử dụng một cấu trúc mạng sâu hơn để cải thiện kết qủa

***ResNet (Residual Network) là một kiến trúc mạng nơ-ron sâu được thiết kế để giải quyết vấn đề giảm mất thông tin (vanishing gradient problem) khi huấn luyện các mô hình nơ-ron sâu.

5.Transformer

Bây giờ ta sẽ có 1 cách nhìn khác về việc lấy thông tin và truyền thông tin của học sâu ở thanh ghi dịch môn điện tử cntt khi thiết kế mạng tích chập



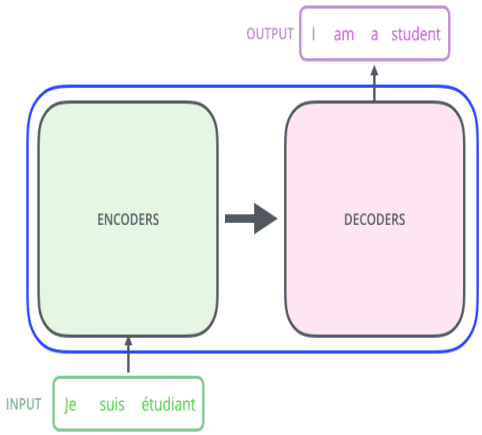
- CNN : thông tin và nối tiếp- ra nối tiếp

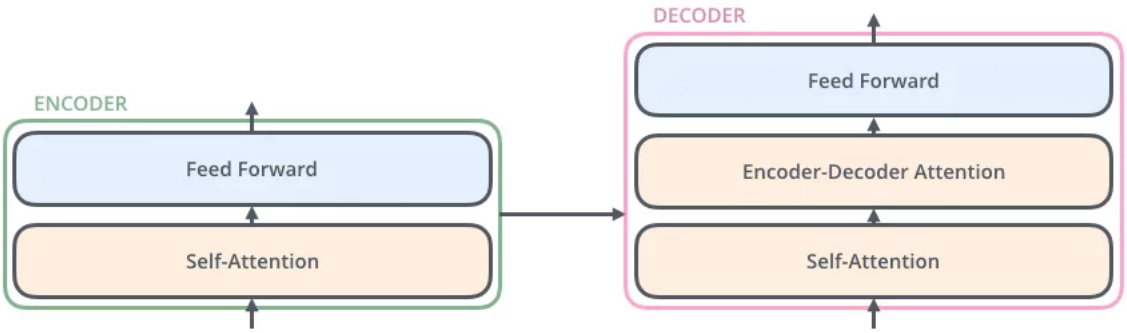
Các mạng CNN có thể dễ dàng được thực hiện song song ở một tầng nhưng không có khả năng nắm bắt các phụ thuộc chuỗi có độ dài biến thiên.

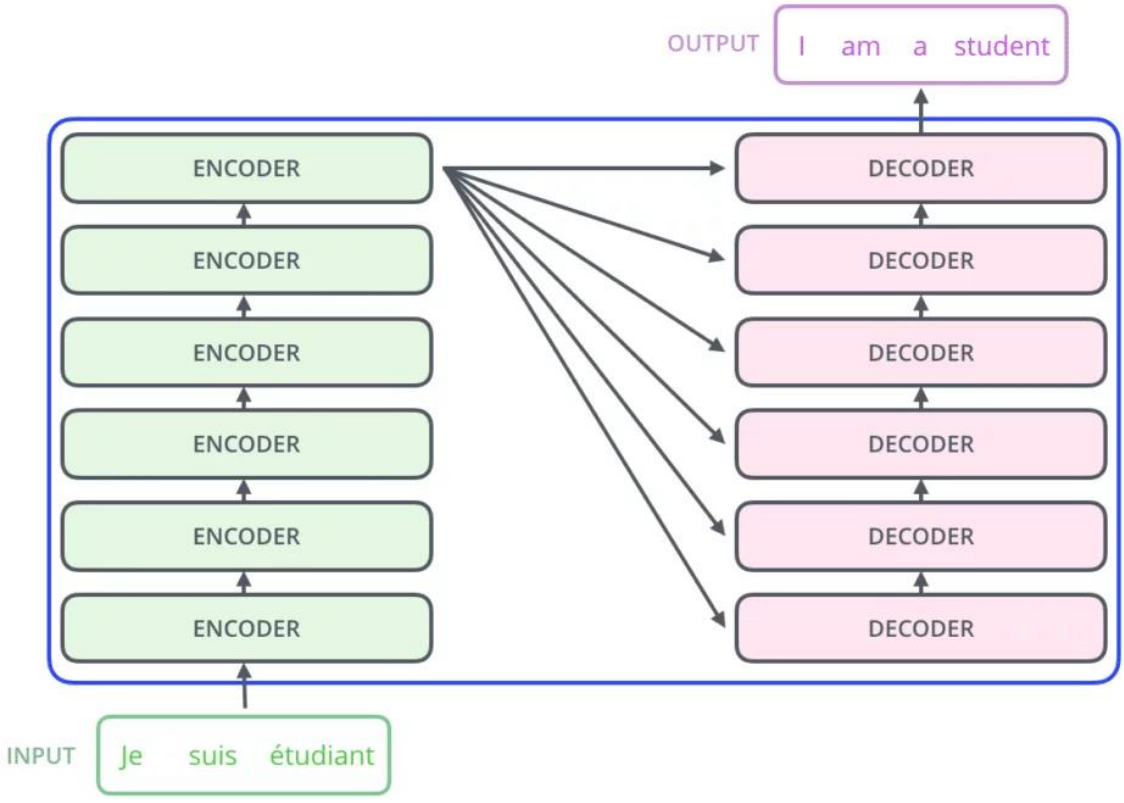
- RNN :có nhiều loại thông tin vào song song - ra nối tiếp, vào nối tiếp - ra song song ...

Các mạng RNN có khả năng nắm bắt các thông tin cách xa nhau trong chuỗi có độ dài biến thiên, nhưng không thể thực hiện song song trong một chuỗi.

=>transformer: tận dụng ưu điểm của cả 2 mạng này


OUTPUT
I am a student
ENCODERS
DECODERS
INPUT
Je suis étudiant


ENCODER
Feed Forward
Self-Attention
DECODER
Feed Forward
Encoder-Decoder Attention
Self-Attention


OUTPUT
I am a student
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
DECODER
DECODER
DECODER
DECODER
DECODER
DECODER
INPUT
Je suis étudiant

encoder: vào nối tiếp - ra nối tiếp

decoder: vào song song - ra nối tiếp

->Transformer có những ưu điểm quan trọng như khả năng xử lý song song hiệu quả, khả năng lấy thông tin từ xa và khả năng mở rộng dễ dàng

5.1. TR-MISR

- deep learning không thể tận dụng tốt MISR

=> end-to-end framework: TR-MISR

(end to end learning: một mô hình học dự đoán trực tiếp từ dữ liệu đầu vào đến đầu ra mà không cần các bước xử lý trung gian)

- gồm 3 phần:

encoder (residual blocks), transformer(fusion module), decoder (subpixel convolution.)

5.2. **S**hifted **win**dows **Transformer** (Swin Transformer)

-điểm mới:

+ Kiến trúc phân cấp: hierarchical Transformer

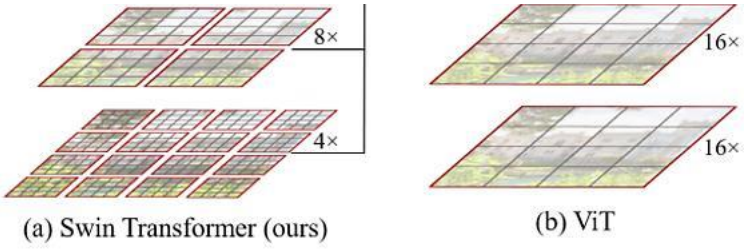


Figure 1. (a) The proposed Swin Transformer builds hierarchical feature maps by merging image patches (shown in gray) in deeper layers and has linear computation complexity to input image size due to computation of self-attention only within each local window (shown in red). It can thus serve as a general-purpose back

+ Self-attention cục bộ trong local window

+Shifted window

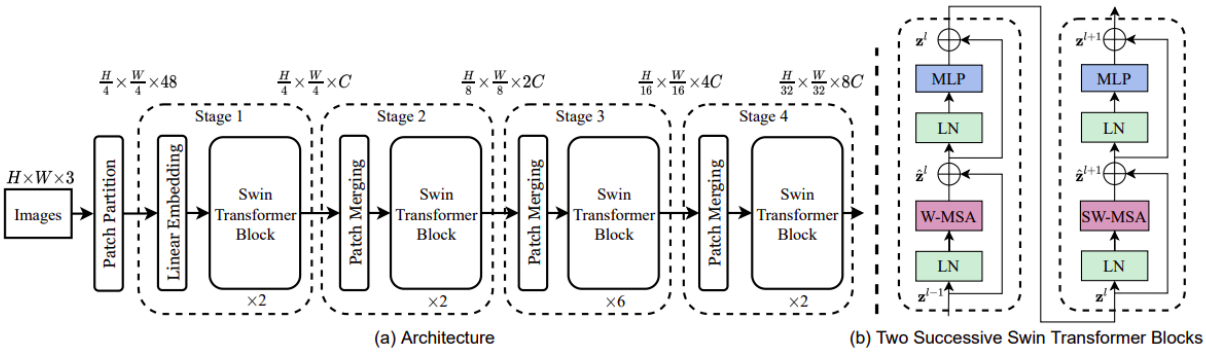


Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.

References

1.Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network

https://arxiv.org/pdf/1609.05158.pdf

2. Overview of Research on Image Super-ResolutionReconstruction

https://sci-hub.se/https://ieeexplore.ieee.org/document/9404113

3. TR-MISR: Multiimage Super-Resolution Based on Feature Fusion With Transformers

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9684717

4. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows

https://arxiv.org/pdf/2103.14030.pdf