BỘ GIÁO DỤC VÀ ĐÀO TẠO BỘ QUỐC PHÒNG

HỌC VIỆN KỸ THUẬT QUÂN SỰ

VŨ ANH TÚ

NGHIÊN CỨU MÔ HÌNH MẠNG SRGAN TRONG NÂNG CAO ĐỘ PHÂN GIẢI ẢNH VÀ ỨNG DỤNG

LUẬN VĂN THẠC SĨ

Chuyên ngành: Hệ thống thông tin

Hà Nội – Năm 2019

BỘ GIÁO DỤC VÀ ĐÀO TẠO BỘ QUỐC PHÒNG

HỌC VIỆN KỸ THUẬT QUÂN SỰ

VŨ ANH TÚ

NGHIÊN CỨU MÔ HÌNH MẠNG SRGAN TRONG NÂNG CAO ĐỘ PHÂN GIẢI ẢNH VÀ ỨNG DỤNG

Chuyên ngành: Hệ thống thông tin

Mã số: 8 48 01 04

CÁN BỘ HƯỚNG DẪN KHOA HỌC

Cán bộ hướng dẫn chính: TS. NGUYỄN VĂN GIANG

Hà Nội - Năm 2019

CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI

HỌC VIỆN KỸ THUẬT QUÂN SỰ

Cán bộ chấm phản biện 1: .....

Cán bộ chấm phản biện 2: .....

Luận văn thạc sĩ được bảo vệ tại:

HỘI ĐỒNG CHẤM LUẬN VĂN THẠC SĨ

HỌC VIỆN KỸ THUẬT QUÂN SỰ

Ngày … tháng … năm 2019

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN XÁC NHẬN CHỈNH SỬA LUẬN VĂN THẠC SĨ

Họ và tên tác giả luận văn: Vũ Anh Tú

Đề tài luận văn: Nghiên cứu mạng SRGAN trong nâng cao độ phân giải ảnh và ứng dụng

Chuyên ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Cán bộ hướng dẫn: TS.Nguyễn Văn Giang

Tác giả, cán bộ hướng dẫn khoa học và Hội đồng chấm luận văn xác nhận tác giả đã sửa chữa, bổ sung luận văn theo biên bản họp Hội đồng ngày .....với các nội dung như sau:

.....

.....

.....

.....

Ngày ..... tháng .... năm 2019

Cán bộ hướng dẫn	**Tác giả luận văn**
(Ký và ghi rõ họ tên)	*(Ký và ghi rõ họ tên)*
TS. Nguyễn Văn Giang	Vũ Anh Tú

CHỦ TỊCH HOẶC THƯ KÝ HỘI ĐỒNG

(Ký và ghi rõ họ tên)

Tôi xin cam đoan:

Những kết quả nghiên cứu được trình bày trong luận văn là hoàn toàn trung thực, của tôi, không vi phạm bất cứ điều gì trong luật sở hữu trí tuệ và pháp luật Việt Nam. Nếu sai tôi hoàn toàn chịu trách nhiệm trước pháp luật.

TÁC GIẢ LUẬN VĂN

(Ký và ghi rõ họ tên)

Vũ Anh Tú

MỤC LỤC

Trang

TÓM TẮT LUẬN VĂN

Họ và tên học viên: Vũ Anh Tú

Chuyên ngành: Hệ thống thông tin Khóa: K29A

Cán bộ hướng dẫn: TS. NGUYỄN VĂN GIANG

Tên đề tài: Nghiên cứu mô hình mạng SRGAN trong nâng cao độ phân giải ảnh và ứng dụng

Tóm tắt luận văn:

Trong luận văn tác giả trình bày mạng học sâu SRGANs, là mạng đối nghịch tạo sinh dành cho siêu phân giải đơn ảnh với chỉ số phóng lớn là x4. Trong mạng học sâu SRGANs tác giả đề xuất bài toàn được tối ưu bằng cách tối ưu hàm mất mát cảm quan (Perceptual loss) bao gồm hàm mất mát phân biệt và hàm mất mát nội dung. Mạng phân biệt được huấn luyện để phân biệt ảnh siêu phân giải và ảnh phân giải cao. Bên cạnh đó mạng tạo sinh đặc biệt tính toán hàm mất mát qua chỉ số tương đồng cảm quan thay vì tương đồng điểm ảnh, điều này làm cải thiện đáng kể chi tiết ảnh. Ngoài ra mạng SRGANs sử dụng các khối dư (Residual Block) cho mạng sâu hơn làm tăng hiệu quả khôi phục ảnh với ảnh bị giảm mẫu lớn (heavily downsampled image)

DANH MỤC CÁC KÍ HIỆU, VIẾT TẮT

STT	Viết tắt	Viết đầy đủ	Dịch nghĩa
1.	CNN	Convolution Neural Network	Mạng nơ-ron tích chập
2.	GANs	Generative Adversarial Network	Mạng đối nghịch tạo sinh
3.	HR	High Resolution	Độ phân giải cao
4.	LR	Low Resolution	Độ phân giải thấp
5.	MAP	Maximum a Posteriori estimation	Ước lượng cực đại hóa hậu nghiệm
6.	MLE	Maximum Likelihood estimation	Ước lượng cực đại hóa xác suất đồng thời
7.	MSE	Mean Squared Error	Sai số toàn phương trung bình
8.	ReLU	Rectified Linear Unit	Đơn vị tuyến tính chỉnh lưu
9.	PReLU	Parameterized Rectified Linear Unit	Tham số hóa đơn vị tuyến tính chỉnh lưu
10.	PSNR	Peak Signal-to-Noise Ratio	Tỉ số tín hiệu cực đại trên nhiễu
11.	SR	Super Resolution	Siêu độ phân giải
12.	SRGANs	Super resolution Generative Adversarial Network	Mạng siêu phân giải đối nghịch tạo sinh
13.	SSIM	Structural Similarity	Chỉ số tương đồng cấu trúc

DANH MỤC HÌNH VẼ

Trang

MỞ ĐẦU

Nhu cầu thị yếu của con người về cảm nhận độ nét ngày càng cao, điều đó được hiện hữu trong cả nhu cầu hàng ngày của con người cũng như mục đích nghiên cứu. Khi hình ảnh có độ nét tốt hơn thì việc thu nhận thông tin sẽ được đầy đủ và chân thực hơn. Trong ngành học máy, ảnh phân giải cao, nhiều chi tiết sẽ là bước tiền xử lí dữ liệu trước khi đưa vào các mô hình huấn luyện nhằm tăng độ chính xác. Về mặt nghiên cứu ở các ngành khác như y tế, quân sự, công nghiệp… thì đều cần ảnh có độ phân giải cao để tăng tính chính xác trong các nghiệp vụ để hỗ trợ việc ra quyết định. Do đó nhu cầu ứng dụng thực tế của siêu phân giải ảnh là rất lớn, đây cũng là hướng nghiên cứu được chú trọng trong nhiều năm trước.

Nhìn chung các nghiên cứu về siêu phân giải ảnh đã đạt được nhiều bước tiến trong cả độ chính xác và tốc độ nhờ với sự phát triển nhanh chóng của mạng học sâu (Deep Learning), từ đó siêu phân giải đơn ảnh sử dụng kĩ thuật học máy được quan tâm nhiều hơn cũng như có nhiều hướng nghiên cứu hơn. Sử dụng mạng đối nghịch tạo sinh GANs là một hướng nghiên cứu rất mới, hoàn toàn khác các kỹ thuật tiếp cận trước đây nhằm tạo ra ảnh siêu phân giải chân thực nhất với cảm nhận của con người.

Các nghiên cứu của luận văn có thể được ứng dụng thực tế cho các hệ thống máy ảnh số phân giải cao, cũng như là bước tiền xử lí dữ liệu tin cậy cho các hệ thống trong ngành thị giác máy khác.

Đề tài luận văn thạc sĩ: Nghiên cứu mô hình mạng SRGAN trong nâng cao độ phân giải ảnh và ứng dụng

Bố cục được chia thành 4 chương:

Chương 1: Tổng quan về siêu phân giải ảnh

Chương 2: Các phương pháp siêu phân giải ảnh cổ điển

Chương 3: Tổng quan về kĩ thuật học sâu và mạng SRGANs

Chương 4: Cài đặt và thử nghiệm

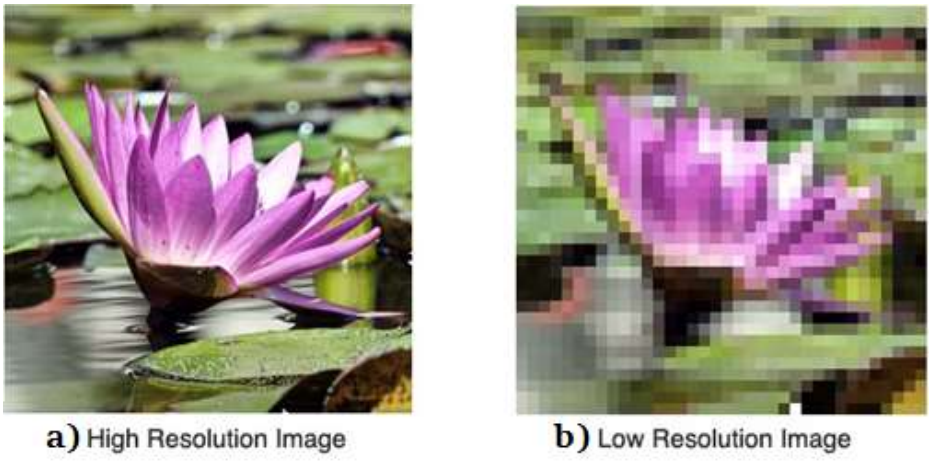
Trong đó Chương 1 trình bày tổng quan về bài toán nâng cao độ phân giải bao gồm khái niệm, cơ sở lý thuyết và các yếu tố ảnh hưởng đến chất lượng ảnh từ camera. Chương 2 trình bày khái quát về một số phương pháp nâng cao độ phân giải ảnh cổ điển, bao gồm siêu phân giải ảnh đơn khung và đa khung. Chương 3 tổng quan về mạng học sâu và kiến trúc mạng SRGANs áp dụng trong siêu phân giải ảnh. Cuối cùng chương 4 trình bày phương pháp cài đặt, thông tin về bộ dữ liệu, phương pháp đánh giá cũng như kết quả thu đươc của thuật toán

Để có thể hoàn thành tốt luận văn, tôi vô cùng biết ơn thầy giáo hướng dẫn của tôi, Tiến sĩ Nguyễn Văn Giang - bộ môn Hệ thống thông tin - Khoa Công nghệ thông tin, với lời khuyên, đã tận tình hướng dẫn, chỉ bảo tôi. Tôi xin cảm ơn khoa Công nghệ thông tin và cán bộ đơn vị đã tạo mọi điều kiện cho tôi.

Chương 1. TỔNG QUAN VỀ SIÊU PHÂN GIẢI ẢNH

1.1. Khái niệm về độ phân giải ảnh

Độ phân giải là một khái niệm dễ gây nhầm lẫn vì nó được sử dụng với nhiều ý nghĩa rất khác nhau. Trong trường hợp của một màn hình thì độ phân giải chính là số điểm ảnh mà nó có thể hiển thị. Ví dụ, một màn hình có độ phân giải là 768x768, tức là nó sẽ có 768x768 điểm ảnh.



Hình 1.1: Ảnh độ phân giải cao và ảnh độ phân giải thấp

Tuy nhiên, trong trường hợp của một ảnh số thì độ phân giải cần được hiểu theo một ý nghĩa khác. Ví dụ như ở Hình 1.1, hai ảnh a và b đều có cùng một kích thước là 128x128 pixel, và nó có cùng độ phân giải? Nhận xét này rõ ràng không hợp lý. Ở khái niệm độ phân giải lại mang ý nghĩ như là khả năng phân biệt các chi tiết trong một ảnh và như thế, hiển nhiên độ phân giải của ảnh a phải lớn hơn ảnh b, bởi ta có thể nhận biết các chi tiết ở ảnh a tốt hơn ảnh b.

Từ hai ví dụ trên, ta có thể định nghĩa rằng độ phân giải của một ảnh là khả năng phân biệt rõ ràng các chi tiết nhỏ nhất trong bức ảnh. Các nhà nghiên cứu trong lĩnh vực xử lý ảnh số và thị giác máy tính thường sử dụng khái niệm độ phân giải ảnh trong một số tình huống sau [1]:

- *Độ phân giải trong không gian (Spatial Resolution)*: là độ phân giải được đánh giá bằng số lượng điểm ảnh trong một Inch (PPI). Đây chính là khả năng biểu diễn thông tin của bức ảnh trên một đơn vị diện tích, theo đó trên cùng 1 diện tích, ảnh nào có mật độ điểm ảnh lớn hơn thì nó sẽ chứa nhiều thông tin hơn và do đó sẽ có độ phân giải tốt hơn;

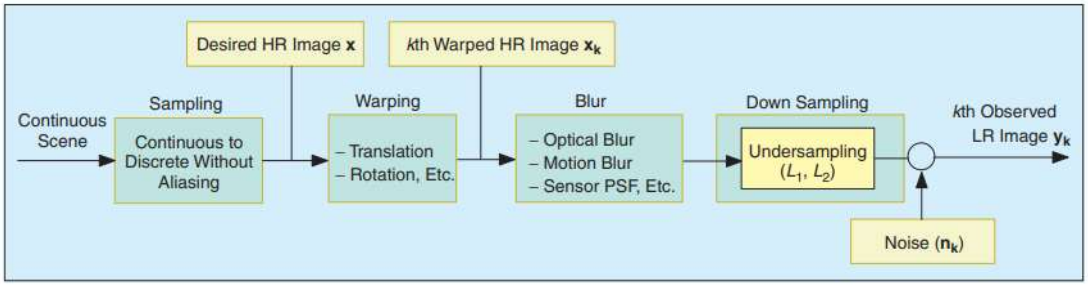
- *Độ phân giải về độ sáng (Brightness Resolution)*: là mức độ phân biệt các cường độ ánh sáng khác nhau ở một pixel bất kì. Khái niệm liên quan đến số mức lượng tử của năng lương ánh sáng mà một cảm biến quang trong Camera thu nhận được. Thông thường, Brightness Resolution của một ảnh đơn sắc (ảnh xám) là 256 tương ứng với 8 bit lượng tử. Và đối với ảnh màu là 24 bit. Khái niệm này thường được biết đến như là độ sâu màu của một ảnh;

- *Độ phân giải về thời gian (Temporal Resolution)*: là số lượng khung hình được ghi nhận trong thời gian một giây. Điều này liên quan đến khả năng phân biệt chuyển động của các đối tượng theo thời gian. Thông thường, đối với một video thì chỉ số này phải ít nhất là 24 frame/s để mắt người không thể cảm nhận được sự chuyển đổi giữa các khung ảnh.

Trong phạm vi luận văn này, khái niệm độ phân giải được sử dụng là độ phân giải trong không gian. Và quá trình chuyển đổi từ một ảnh có độ phân giải thấp thành ảnh có độ phân giải cao gọi là siêu phân giải ảnh (Super Resolution).

1.2. Mô hình thu nhận ảnh

Để thu nhận hình ảnh cần 2 thành phần chính là linh kiện nhạy với phổ năng lượng điện từ trường : thành phần thứ nhất tạo tín hiệu điện ở đầu ra tỷ lệ với mức năng lượng mà bộ cảm biến (camera), thứ hai là bộ số hóa. Một ảnh g(x,y) ghi được từ camera là ảnh liên tục tạo nên mặt phẳng 2 chiều, ảnh cần chuyển sang dạng thích hợp để được xử lý bằng máy tính. Phương pháp biến đổi một ảnh (hay một hàm) liên tục trong không gian cũng như theo giá trị thành dạng số rời rạc được gọi là số hóa ảnh. [2] [3]



Hình 1.2: Mô hình thu nhận ảnh từ ảnh độ phân giải cao bị giảm chất lượng

Việc biến đổi gồm 2 bước:

- Bước 1: đo các giá trị trên khoảng không gian gọi là lấy mẫu;
- Bước 2: ánh xạ cường độ (hoặc giá trị) đo được thành một số hữu hạn các mức rời rạc gọi là lượng tử hóa.

Hình ảnh ban đầu sau khi đi qua cảm biến của camera biến đổi ảnh từ hàm liên tục trong không gian thành hàm rời rạc, sau đó được hệ thống xử lý phù hợp với các tham số của hệ thống thành hình ảnh có thể xử lý bằng máy tính. Các dạng định dạng thường được sử dụng như ảnh đen trắng (với định dạng IMG) , ảnh đa cấp xám cho tới ảnh màu (BMP, GIF,JPEG…) đều tuân thủ cấu trúc chung: mào đầu tệp, dữ liệu nén, bảng màu. Trong quá trình thu nhận ảnh dưới sự tác động khách quan của môi trường cũng cấu hình camera và các yếu tố chủ quan hình ảnh thu được không được như hình ảnh ban đầu, làm ảnh hưởng đến chất lượng ảnh thu nhận được.

1.3. Các yếu tố làm giảm độ phân giải ảnh

Một bối cảnh thực $X(x,y)$ được ghi nhận bởi một camera sẽ chịu ảnh hưởng bởi chuyển động tương đối giữa ống kính của camera và bối cảnh, sự chuyển động của các đối tượng trong ảnh và độ rung của thiết bị trong quá trình chụp. Tiếp đó, dưới tác động của các nhiễu động trong không khí $H^{atm}(x,y)$, cường độ ánh sáng đi đến cảm biến trong quá trình chụp sẽ thay đổi dẫn đến hiện tượng nhòe ảnh. Bên cạnh dưới tác động của Point Spread

Function (PSF) ở ống kính $H^{cam}(\mathrm{x},\mathrm{y})$, các điểm ảnh cũng bi nhòe đi. Tiếp đó, quá trình rời rạc hóa tín hiệu ở các điểm cảm biến CCD cũng là một nguyên nhân tạo ra ảnh chất lượng thấp $Y(\mathrm{m},\mathrm{n})$. Ta có thể mô hình hóa quá trình này bằng biểu thức toán học như sau:

$$Y(\mathrm{m},\mathrm{n}) = \mathrm{D}[\mathrm{H}^{cam}(\mathrm{x},\mathrm{y}) \otimes F(\mathrm{H}^{atm} \otimes X(\mathrm{x},\mathrm{y}))] + \mathrm{V}(\mathrm{m},\mathrm{n}) \quad (1.1)$$

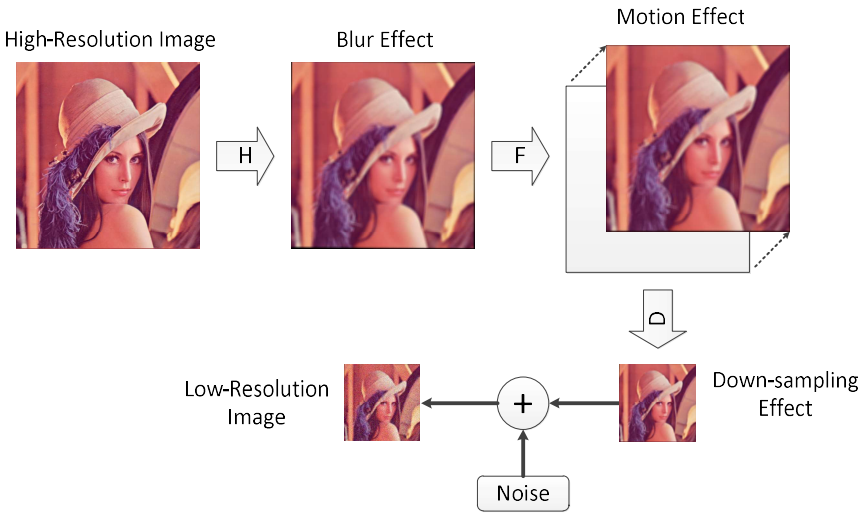
Trong đó $\otimes$ là tích chập, F là hàm chuyển động, D là quá trình giảm mẫu và V(m,n) là ảnh hưởng của các nhiễu cộng. Trong nghiên cứu của mình, Farsiu và cộng sự đã chuyển đổi biểu thức (1.1) thành dạng tích các ma trận để sử dụng cho các ảnh số như sau:

$$Y_k = D_k \otimes H_k \otimes F_k \otimes X \otimes V_k \quad (1.2)$$

Với:

- X: ảnh có độ phân giải cao;
- Y_k : ảnh có độ phân giải thấp thứ k;
- F_k : ảnh hưởng của sự chuyển động;
- $H_k = H_k^{atm} * H_k^{cam}$: toán tử biểu diễn ảnh hưởng của SPF cũng như nhiễu động của môi trường khiến ảnh bị mờ;
- D_k: ảnh hưởng của quá trình giảm mẫu;
- V_k : nhiễu cộng.

Để thuận tiện cho quá trình khảo sát giải thuật, từ công thức (1.2), tác giả đã xây dựng một mô hình tạo ra các ảnh có độ phân giải thấp từ một ảnh ban đầu có độ phân giải cao. Theo đó, toán tử lọc Gaussian Filter được sử dụng để mô phỏng ảnh hưởng của H_k, và mô hình nhiễu ngẫu nhiên được sử dụng để mô phỏng ảnh hưởng của V_k. Kết quả ngõ ra của hệ thống được trình bày ở Hình 1.3 bên dưới:



Hình 1.3: Mô hình hóa quá trình thu nhận ảnh độ phân giải thấp

Qua đó, từ hình ảnh HR ban đầu chịu tác động của nhiều yếu tố từ việc như giảm mật độ điểm ảnh, làm mờ ảnh hay ảnh hưởng bởi nhiễu làm hình ảnh thu được không còn được chất lượng như hình ảnh ban đầu. Các yếu tố ảnh hưởng đến chất lượng hình ảnh như:

Thứ nhất, chất lượng của camera với độ phân giải thấp, độ phân giải của camera là thước đo chất lượng hình ảnh camera thu được . Mỗi hình ảnh được tạo ra từ vô số điểm ảnh, do đó độ phân giải càng cao thì hình ảnh thu được càng sắc nét ngược lại càng thấp thì hình ảnh càng thấp .Vì vậy, hình ảnh thu được từ những camera này sẽ làm giảm mật độ điểm ảnh của hình ảnh xuống sẽ không còn như hình ảnh ban đầu. Các yếu tố bao gồm: bao gồm Ống kính, cảm biến hình ảnh, bộ xử lý tín hiệu hình ảnh (đại diện bởi các tính năng như tự phơi sáng, tự cân bằng trắng, tự động lấy nét, chống ngược sáng WDR, chống nhiễu kỹ thuật số DNR, độ nét). Mỗi ống kính có một độ phân giải quang học, độ chính xác của các ống kính thiết kế để nhận các tia sáng đến chính xác điểm đích ở trên bảng cảm biến. Cảm biến hình ảnh được sử dụng để tận dụng tối đa lượng ánh sáng đến cảm biến bằng cách thay đổi cấu trúc cảm biến.

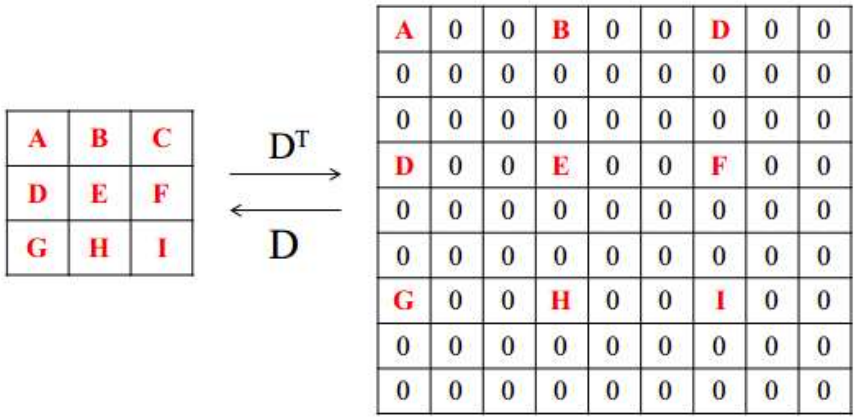
Thứ hai, độ rung lắc khi chụp ảnh làm bức ảnh giảm chất lượng so với hình ảnh ban đầu. Có 2 trường hợp tạo sự rung lắc:

- Đối tượng chụp chuyển động: Khi chụp một đối tượng đang chuyển động, ảnh rất dễ bị nhòe mờ đó có thể do tốc độ chụp quá chậm.
- Máy ảnh bị rụng lắc: Tương tự như khi chụp đối tượng chuyển động, khi máy ảnh bị rung lắc hoặc do tốc độ chụp cũng có thể khiến hình ảnh không được sắc nét.

Chương 2. CÁC PHƯƠNG PHÁP SIÊU PHÂN GIẢI ẢNH CỔ ĐIỂN

2.1. Siêu phân giải ảnh sử dụng phương pháp nội suy

Quá trình giảm mẫu là một trong những nguyên nhân gây ra tạo ra ảnh chất lượng thấp. Cụ thể, như ví dụ ở Hình 2.1, khi tiến hành giảm mẫu từ ma trận 9x9 xuống còn 3x3 (D) thì một lượng thông tin bị mất đi. Và khi đó độ phân giải của ảnh giảm xuống. Tuy nhiên quá trình tăng mẫu (D^T) từ một ảnh có độ phân giải thấp lại không cải thiện chất lượng ảnh, bởi vì lượng thông tin bị mất đi sẽ được thay thê bằng các giá trị 0, và đây là nguyên nhân tạo ra hình có độ phân giải thấp mà tác giả đã trình bày ở chương trước.



Hình 2.1: Quá trình tăng/giảm mẫu

Ưu điểm của phương pháp nội suy là độ phức tạp thấp và có thể thực hiện trong thời gian thực. Nhược điểm của phương pháp là chỉ áp dụng khử mờ khi các ảnh làm mờ giống nhau và thuật toán chưa tối ưu. [2]

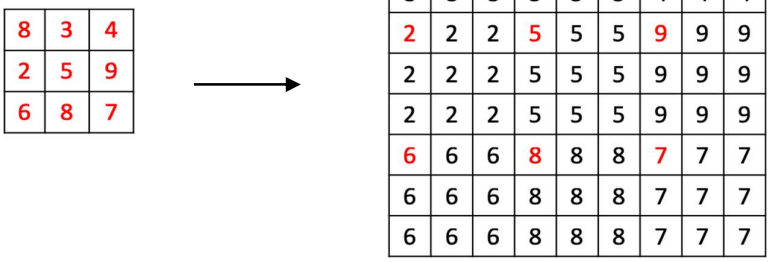
Quá trình nội suy là một trong những hoạt động cơ bản trong xử lý ảnh. Chất lượng hình ảnh phụ thuộc rất nhiều vào kỹ thuật nội suy được sử dụng. Các kỹ thuật nội suy được chia thành hai loại, kỹ thuật nội suy xác định (deterministic) và thống kê (statistical). Sự khác biệt là các kỹ thuật nội suy xác định giả định một biến đổi nhất định giữa các điểm mẫu, chẳng hạn như

tuyến tính trong trường hợp nội suy tuyến tính. Phương pháp nội suy thống kê xấp xỉ tín hiệu bằng cách giảm thiểu sai số ước tính. Quá trình xấp xỉ này có thể dẫn đến các giá trị mẫu ban đầu không được sao chép. Vì các phương thức thống kê có tính toán không hiệu quả, trong bài viết này chỉ các kỹ thuật xác định mới sẽ được thảo luận.

Một số kỹ thuật nội suy đã được phát triển và thường được sử dụng. Trong có có các phương pháp được sử dụng phổ biến nhất là các kỹ thuật nội suy nearest neighbor, linear và Bicubic. Ít phổ biến hơn là các phương thức nội suy đa thức và Lagrange.

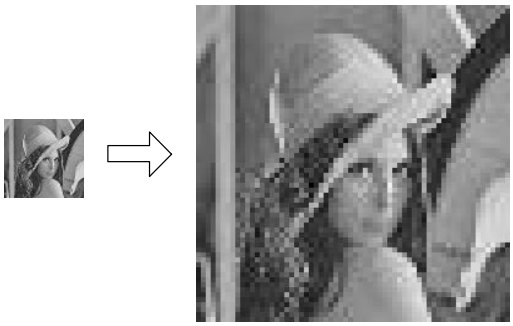
a. Nearest Neighbor Interpolation

Đây là phương pháp đơn giản nhất, các điểm ảnh bị khuyết sẽ được điền bằng các giá trị điểm ảnh ngay trước đó, ví dụ Hình 2.2 là ta thực hiện tăng mẫu ảnh có kích thước từ 3x3 lên 8x8 bằng phương pháp nội suy Nearest Neighbor. [4]



Hình 2.2: Tăng mẫu kích thước từ 3x3 lên 9x9 bằng Nearest Neighbor

Áp dụng giải thuật cho hình ảnh thử nghiệm ta có kết quả như sau:

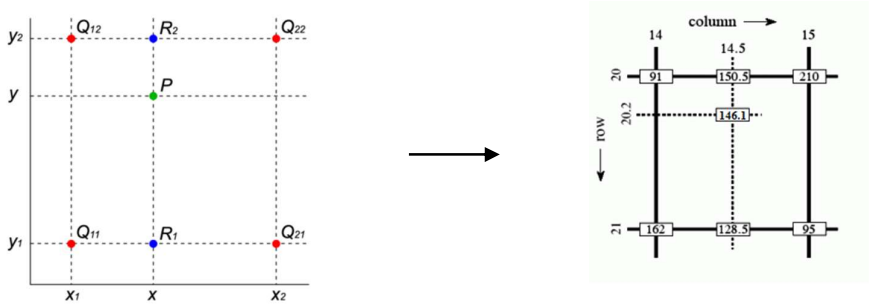


Hình 2.3. Kết quả nội suy sử dụng phương pháp Nearest Neighbor

Sự tích chập trong miền không gian với hàm h(x) tương đương với miền tần số của hàm sinc. Vì các tùy biến, hàm sinc làm cho bộ lọc thông thấp kém. Kỹ thuật này đạt được độ phóng đại bằng cách nhân bản pixel, và rút gọn bằng cách lấy mẫu điểm thưa thớt. Đối với những thay đổi quy mô lớn, nội suy lân cận gần nhất tạo ra hình ảnh với răng cưa.

b. Bilinear Interpolation

Đây là phương pháp mở rộng của phương pháp nội suy tuyến tính được sử dụng cho một hàm hai biến $f(x,y)$ trên không gian 2 chiều, ý tưởng chính của phương pháp này là nội suy theo một hướng rồi tiếp tục nội suy theo hướng còn lại. [4]



Hình 2.4. Mô phỏng một quá trình nội suy sử dụng Bilinear

Xét ví dụ ở Hình 2.4, ở đây ta đã xác định được giá trị tại các điểm $Q_{11}(x_1,y_1)$, $Q_{12}(x_1,y_2)$, $Q_{21}(x_2,y_1)$, $Q_{22}(x_2,y_2)$ và cần tính giá trị tại điểm P, ta có 2 bước để thực hiện nội suy:

- Bước 1: Nội suy tuyến tính các giá trị R_1 và R_2 trên trục x

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x_1 - x}{x_1} f(Q_{21}) \quad \text{với } R_1 = (x, y_1)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x_1 - x}{x_1} f(Q_{22}) \quad \text{với } R_2 = (x, y_2)$$

- Bước 2: Từ 2 điểm R_1và R_2 thực hiện thêm một phép nội suy tuyến tính để tìm ra P

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y_1 - y}{y_2 - y_1} f(R_2)$$

Áp dụng thuật toán trên vào ma trận 3x3 như ở Hình 2.2 ta được ma trận mới như sau:

8	3	4
2	5	9
6	8	7

→

8	8	6	5	3	3	4	4	4
8	8	8	3	3	3	4	4	4
6	6	5	4	4	4	5	6	6
4	4	4	4	4	5	6	7	7
2	2	3	4	5	6	7	9	9
3	3	4	5	6	7	8	8	8
5	5	5	6	7	7	7	8	8
6	6	7	7	8	8	7	7	7
6	6	7	7	8	8	7	7	7

Hình 2.5: Tăng mẫu kích thước từ 3x3 lên 9x9 bằng Bilinear

Áp dụng giải thuật cho hình ảnh thử nghiệm ta có kết quả như sau:

Hình 2.6: Kết quả nội suy sử dụng phương pháp Bilinear

Nội suy Bilinear thể hiện sự đơn giản trong giải thuật bằng việc cho rằng quan hệ giá trị mức xám giữa các điểm ảnh là tuyến tính. Do đó với các vùng ảnh có tần số thấp thì nội suy cho kết quả tốt, nhưng với những vùng chi tiết ảnh có tần số cao, độ biến thiên giá trị mức xám cao hay quan hệ giá trị mức xám giữa các điểm ảnh là phi tuyến, thì kết quả nội suy kém.

c. Bicubic Interpolation:

Một phép nội suy được sử dụng rất phổ biến cho hiệu quả tốt nhất trong các phép nội suy ở miền không gian là Bicubic. Để cải thiện chất lượng hơn so với Bilinear, nội suy Cubic thể hiện mối quan hệ giữa mức xám giữa các điểm ảnh là dạng đa thức bậc 3.

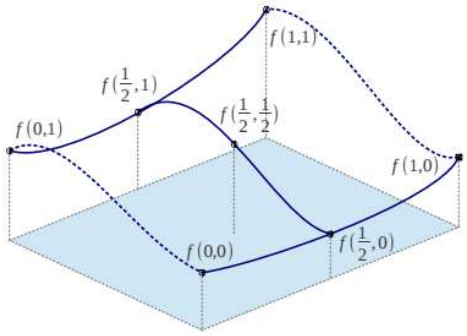
Trong phép nội suy này, một điểm được suy ra từ thông tin của 4 điểm xung quanh nó. Hình 2.7 cung cấp một ví dụ cho phép nội suy bicubic, theo đó: [4]

- $f(\frac{1}{2},0)$ xác định bởi $f(0,0)$, $f(1,0)$, $\partial_x f(0,0)$, $\partial_x f(1,0)$
- $f(\frac{1}{2},1)$ xác định bởi $f(0,1)$, $f(1,1)$, $\partial_x f(0,1)$, $\partial_x f(1,1)$
- $\partial_y f(\frac{1}{2},0)$ xác định bởi $\partial_y f(0,0)$, $\partial_y f(1,0)$, $\partial_{xy} f(0,0)$, $\partial_{xy} f(1,0)$
- $\partial_y f(\frac{1}{2},1)$ xác định bởi $\partial_y f(0,1)$, $\partial_y f(1,1)$, $\partial_{xy} f(0,1)$, $\partial_{xy} f(1,1)$

- $\partial_{xy} f(\frac{1}{2},\frac{1}{2})$ xác định bởi $f(\frac{1}{2},0)$, $f(\frac{1}{2},1)$, $\partial_y f(\frac{1}{2},0)$, $\partial_y f(\frac{1}{2},1)$
- $f(\frac{1}{2},0)$ xác định bởi $f(0,0)$, $f(1,0)$, $\partial_y f(0,0)$, $\partial_y f(1,0)$
- $f(\frac{1}{2},1)$ xác định bởi $f(0,1)$, $f(1,1)$, $\partial_x f(0,1)$, $\partial_x f(1,1)$
- $\partial_y f(\frac{1}{2},0)$ xác định bởi $\partial_y f(0,0)$, $\partial_y f(1,0)$, $\partial_{xy} f(0,0)$, $\partial_{xy} f(1,0)$
- $\partial_y f(\frac{1}{2},1)$ xác định bởi $f(0,0)$, $f(1,0)$, $\partial_x f(0,0)$, $\partial_x f(1,0)$

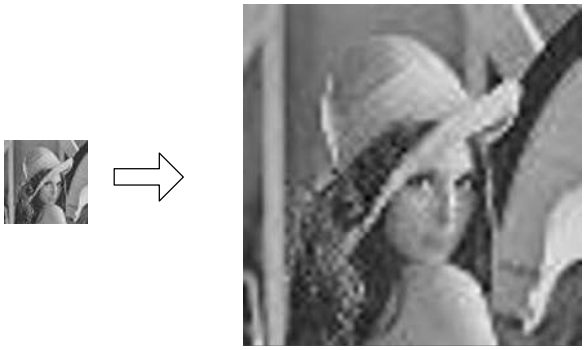
Trong đó:

- $f(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij} x^i y^j$
- $\partial_x f(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} i a_{ij} x^{i-1} y^j$
- $\partial_y f(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} j a_{ij} x^i y^{j-1}$
- $\partial_{xy} f(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} ij a_{ij} x^{i-1} y^{j-1}$



Hình 2.7: Ví dụ phép nội suy bicubic

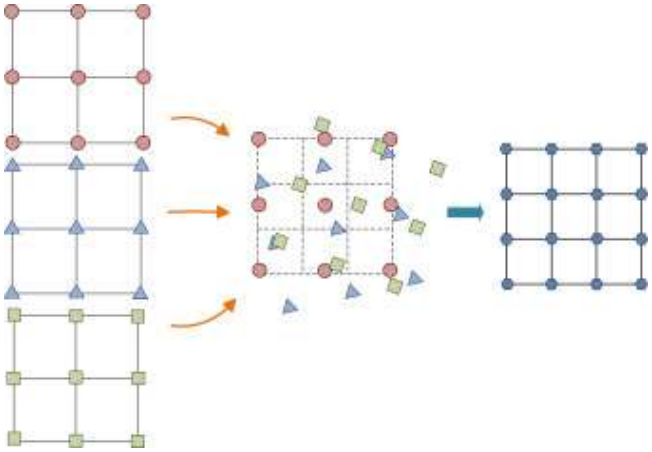
Khi sử dụng phép nội suy này để phóng đại ảnh ví dụ ở Hình 2.3 ta được kết quả ảnh siêu phân giải (Hình 2.8) mượt hơn so với hai phép nội suy Bilinear và Nearest.



Hình 2.8: Kết quả phép biến đổi bicubic

2.2. Siêu phân giải ảnh đa khung hình

Phương pháp này lợi dụng chính sự rung động của camera khi chụp, gây ra sự xê dịch các khung hình. Như minh hoạ ở Hình 2.9, chính sự xê dịch này vô hình chung tạo ra thông tin bị thiếu hụt, không được lấy mẫu ở ảnh phân giải thấp, từ đó để tạo ra ảnh phân giải cao HR. Do được bổ xung thông tin thiếu hụt nên kỹ thuật siêu phân giải ảnh đa khung cho kết quả tốt hơn rõ rệt so với các kỹ thuật siêu phân giải nội suy đơn khung. [5]



Hình 2.9: Quá trình siêu phân giải ảnh đang khung hình

Nhìn chung các phương pháp siêu phân giải ảnh đa khung đều thực hiện hai bước chính là quá trình quá trình xác nhận ảnh (Registration) và qúa trình khôi phục ảnh (Reconstruction).

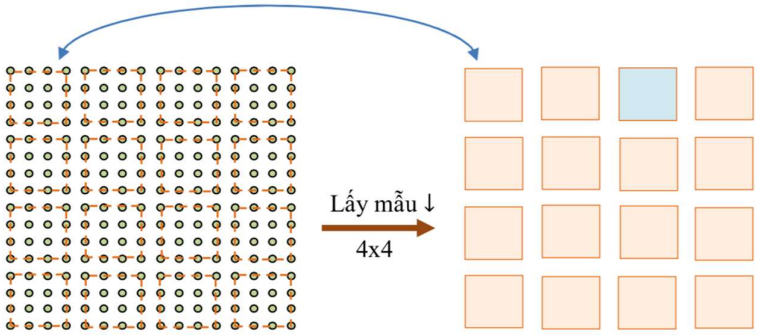
- *Xác nhận ảnh*: Xác định các thông số dịch và góc xoay (ϕ) giữa các khung tham khảo so với khung chính xét trên chung một hệ trục toạ độ;

- *Khôi phục ảnh*: dựa vào tập hợp các thông số chuyển (x, y, ϕ) các ảnh phân giải thấp LR sẽ được sắp xếp lại trên một cùng một hệ trục toạ độ, sau đó sử dụng kỹ thuật nội suy không gian để khôi phục ảnh phân giải cao HR.

2.2.1. Siêu phân giải ảnh dựa trên mô hình thống kê

Qua việc áp dụng lý thuyết Baysian để xây dựng mối quan hệ tổng quát giữa thành phần ẩn (là các điểm ảnh của ảnh phân giải cao HR) và các biến trạng thái ảnh phân giải thấp LR đầu vào. Chúng bao gồm lõi mờ của ảnh, chuyển động của các pixel điểm ảnh và nhiễu.

Mặt khác ta có thể mô hình hóa tổng quan hệ thống thu nhận ảnh của camera. Về mặt vật lý, giá trị mức xám của mỗi pixel độ phân giải thấp thu được là trung bình cộng giá trị mức xám của các pixel độ phân giải cao trong nội vùng của nó. Kết quả, với cảnh thực khi được chụp, thì ảnh thu được bao giờ cũng có độ phân giải thấp và luôn bị mờ đi so với ảnh thực tế. Chuỗi ảnh HR gốc là f_{HRi}. Ta gọi U là toán hạng lấy mẫu không gian của camera, K là lõi mờ của camera, w_i là nhiễu nội của hệ thống camera và $s = s(x, y)$ biến tọa độ không gian ảnh 2 chiều.



Hình 2.10: Quá trình lấy mẫu xuống

Ta gọi frame thu được thứ i là f_{LRi}, là ma trận của các pixel điểm ảnh 2 chiều. Mô hình toán học đơn giản của hệ thống thu nhận ảnh video, cho frame thứ i, được thể hiện như sau: [2] [5]

$$f_{LRi}(\mathrm{s}) = UKf_{HRi}(\mathrm{s}) + \mathrm{w}_i \tag{2.2.1}$$

Trong đó U là toán hạng lấy mẫu không gian ảnh.

Trong thực tế, một cách tổng quát luôn có sự chuyển động của camera và cảnh hay các chi tiết, đối tượng trong cảnh. Sự chuyển động này có thể là sự chuyển dịch theo phương ngang, phương thẳng đứng và xoay một cách tùy ý giữa các frame.

Ta gọi Δs_i là thông số dịch chuyển theo phương ngang và phương thẳng đứng, với toán hạng dịch là $F_{\Delta s_i}$. Và θ_i là thông số góc xoay với toán hạng R_{θ_i} của khung f_{HRi} so với frame gốc f_{HR1}, vậy ta có phương trình quan hệ giữa chuỗi khung HR f_{HRi} và khung gốc f_{HR1} là :

$$f_{HRi} = f_{HR1}(\mathrm{R}_{\theta_i}(\mathrm{s} + \Delta \mathrm{s}_i)) \tag{2.2.2}$$

Từ phương trình 2.2.1 và 2.2.2 ta thu được mối quan hệ giữa ảnh phân giải thấp và khung ảnh phân giải cao ban đầu là:

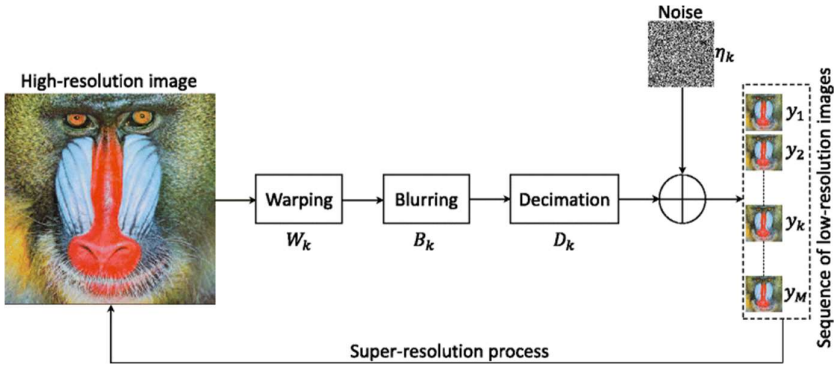
$$f_{LRi}(\mathrm{s}) = UKf_{HR1}(\mathrm{R}_{\theta_i}(\mathrm{s} + \Delta \mathrm{s}_i)) + \mathrm{w}_i \tag{2.2.3}$$

Do đó, ta thấy rằng siêu phân giải ảnh đa khung là giải bài toán ngược của pt (2.2.3), tìm ảnh gốc HR, f_{HR1}, từ tập các ảnh video LR thu được . Vậy có thể nói, siêu phân giải ảnh là hình thức sử dụng các thuật toán bằng phần mềm, khôi phục tạo ra ảnh HR từ ảnh ngõ vào video LR với mục tiêu gia tăng độ rõ nét của chi tiết ảnh.

2.2.2. Mô hình Bayes cho ước lượng chuyển động

Theo lý thuyết xác suất, ta có tập biến quan sát được là chuỗi các frame ảnh LR đầu vào $\{f_{LRi}\}$, với i = {1..N} . Mặt khác từ pt (2.2.3), ta có tập các thông số cần ước lượng là: [6]

- f_{HR1} khung ảnh gốc phân giải cao;
- $\{\Delta s_i\}$ thông số ước lượng chuyển dịch phẳng theo phương ngang và phương thẳng đứng của frame thứ i so với frame hiện tại;
- $\{\theta_i\}$ thông số ước lượng xoay của frame thứ i so với frame hiện tại;
- $\{w_i\}$ thông số ước lượng nhiễu của frame thứ i.



Hình 2.11: Quá trình mô hình suy giảm chất lượng ảnh từ khung ảnh gốc

Trong MAP, ta biết trước một giá thiết được gọi là thông tin tiền nghiệm của tham số θ. Từ giả thiết này ta có thể suy ra được các khoảng giá trị và phân bố của tham số. Ta có công thức ước lượng xác suất đồng thời như sau:

$$\theta = \arg\max p\left(\theta \mid x_1 \ldots x_n\right) \tag{2.2.4}$$

Thông thường, hàm tối ưu trong khó xác định dạng một cách trực tiếp. Chúng ta thường biết điều ngược lại, tức nếu biết tham số, ta có thể tính được

hàm mật độ xác suất của dữ liệu. Vì vậy, để giải bài toán MAP, ta thường sử dụng quy tắc Bayes.

Theo công thức xác suất Bayessian ta có:

$$p(\mathrm{y}\,|\,\mathrm{x}) = \frac{p(\mathrm{x}\,|\,\mathrm{y})\,\mathrm{p}(\mathrm{y})}{\sum_{y} p(\mathrm{x},\mathrm{y})} \tag{2.2.5}$$

Bài toán MAP công thức 2.2.4 thường được biến đổi thành:

$$\theta = \arg\max\left[\Pi p(\mathrm{x}_i\,|\,\theta)\,\mathrm{p}(\theta)\right] \tag{2.2.6}$$

Áp dụng Baysian MAP vào bài toán siêu phân giải đa khung, công việc xác nhận ảnh (registration) trở thành việc ước lượng độ dịch chuyển của các khung ảnh phân giải thấp LR, sao cho khi ghép chúng lại ta được sự khớp nhất giữa các ảnh. Khi đó các vùng thông tin bị thiếu sẽ được bổ xung bằng nhiều ảnh LR.

Ta có xác xuất f_{LRi} với f_{HR1}, Δs_i, w_i, θ_i là phân bố xác suất tương đồng, các thông số này là độc lập với nhau. Vậy việc ước lượng các tham số này quy về:

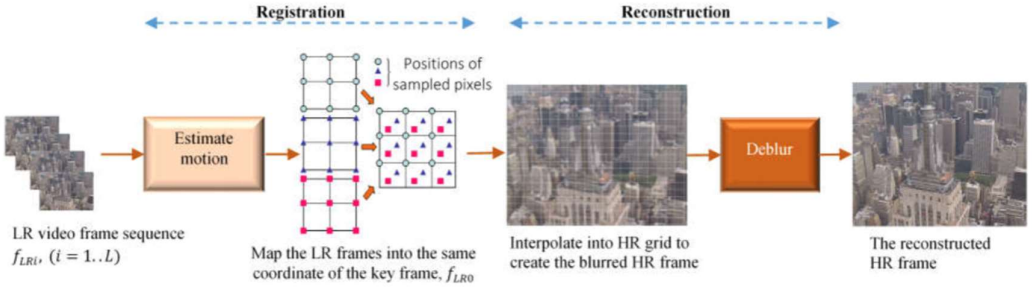
$$\left\{f^*_{HR1}, \left\{\Delta \mathrm{s}^*_i\right\}, \left\{w^*_i\right\}, \{\theta_i\}\,|\,\{f_{LRi}\}\right\} = \arg\max p\left(f_{HRi}, \{\Delta \mathrm{s}_i\},\ \{w_i\},\ \{\theta_i\}\,|\,\{f_{LRi}\}\right)$$

Để tìm hàm phân bố xác suất của các thông số ước lượng Δ, và , ta dựa vào quan điểm các thông số được ước lượng tối ưu: khi ảnh HR được khôi phục trơn nhẵn (smoothing); vector chuyển động và nhiễu của các pixel điểm ảnh cũng tương đối smoothing.

2.2.3. Mô hình cho hệ thống siêu phân giải ảnh

Mô hình cho hệ thống video siêu phân giải bao gồm hai giai đoạn xử lý chính, ước lượng chuyển động và khôi phục. Từ chuỗi các frame LR ngõ vào, ta xác nhận để tìm chuyển động giữa frame tham khảo f_{LRi}, và frame chính f_{HR1}. Quá trình xác nhận sẽ thực hiện ước lượng chuyển động của từng block các pixel điểm ảnh. Tiếp đến, các block pixel của các frame LR ngõ vào được sắp xếp trên cùng một hệ trục tọa độ của frame chính.

Sau đó ta dùng nội suy Bicubic để khôi phục frame HR. Frame này là chính là ảnh HR được khôi phục của frame HR mờ, Sau cùng, giải mờ cho frame HR mờ ta được frame HR chính cần được khôi phục.



Hình 2.12: Mô hình siêu phân giải ảnh đa khung

a. Ước lượng (xác nhận) chuyển động

Quá trình ước lượng chuyển động thực hiện qua hai bước chính, ước lượng xoay và ước lượng dịch.

Trong ước lượng dịch có hai bước, ước lượng dịch thô và ước lượng dịch tinh. Ước lượng dịch thô xác định chuyển động trong phạm vi số nguyên lần pixel. Ước lượng dịch tinh xác định trong phạm vi nhỏ hơn một pixel.

Như đã được nói đến ở phần trước, để ảnh video thu được có chất lượng, các hệ thống thu hình thường di chuyển chậm. Chúng được gắn cố định trên các hệ trượt, tịnh tiến với góc quay trong phạm vi nhỏ, từ -2 độ đến

2 độ. Độ chính xác của phép ước lượng xoay gia tăng theo từng lớp, 1 độ cho lớp M, 0.1 độ cho lớp N. Vậy độ chính xác cho phép ước lượng xoay là ±0.1 độ. Giá trị này đủ để ảnh HR được khôi phục chính xác.

b. Khôi phục ảnh HR

Giải thuật khôi phục ảnh HR cũng giống như trong phần ước lượng được chia làm hai phần là phần giải mờ và phần khôi phục tăng chất lượng ảnh.

Đầu tiên, ta loại bỏ các pixel nhiễu (hay suy biến) từ các khối điểm ảnh của các chuỗi ảnh LR đầu vào. Chuyển động bất đồng bộ của các chi tiết ảnh trong gây ra sự suy biến tại những vùng chi tiết này của ảnh HR được khôi phục. Những pixel suy biến trong được xác định như sau:

$$\left(|\, \mathrm{B}_{LRi} - \mathrm{UKR}_{\theta_i} \mathrm{F}_{\Delta si} \mathrm{B}_{HR1} \,|\right)_{xy} > Threshold$$

Dựa trên các thông số được ước lượng, các frame LR sau khi đã được loại bỏ các điểm ảnh bị suy biến sẽ được sắp xếp trên cùng hệ trục tọa độ với khung ảnh chính. Tiếp theo, ta dùng nội suy Bicubic để khôi phục frame HR frame tại module U, khung HR được khôi phục chính là ảnh mờ của frame chính HR f_{HR1}. Tiếp đến, frame mờ HR được giải mờ tại module. Kết quả tạo ra frame HR xấp xỉ với ảnh thực của frame chính HR. Chúng ta sử dụng kỹ thuật bộ lọc Wiener để giải mờ ảnh.

Để tăng chất lượng của ảnh HR được khôi phục, ta sử dụng vòng lặp. Điều kiện để kiểm tra vòng lặp hội tụ được xác định. Do ta chỉ có một biến đầu vào và ra là $\{f_{LRi}\}$, nên ta có điều kiện tối ưu để vòng lặp hội tụ là:

$$\left\{f_{HR1}^{*}\right\} = \arg\min\left\{|\, \Delta f_{HR1} \,|\right\}$$

Sau đó, frame chính LR được gán trở lại với frame HR mờ và tiếp tục lặp vòng từ bước để gia tăng độ chính xác của thông tin chi tiết của frame chính HR được khôi phục. Để giới hạn thời gian xử lý, số vòng lặp được giới hạn không quá 4 lần. Đây là giá trị lựa chọn tốt nhất qua các thực nghiệm.

Chương 3. TỔNG QUAN VỀ KĨ THUẬT HỌC SÂU VÀ MẠNG SRGANs

Chương 3 trình bày các khái niệm cơ bản về học sâu, là một nhánh của ngành máy học được nghiên cứu rất nhiều trong nhiều lĩnh vực hiện nay.

Kết cấu Chương 3 gồm có 3 phần là: Tổng quan về học máy và các khái khái niệm cơ bản về học máy, một số khái niệm liên quan đến kĩ thuật học sâu, siêu phân giải ảnh sử dụng mạng học sâu. Trong chương này, tác giả đi sâu vào mô hình CNN và mạng học sâu SRGANs.

3.1. Tổng quan về kỹ thuật học máy

Tom Mitchell, giáo sư nổi tiếng của Đại học Carnegie Mellon University - CMU định nghĩa cụ thể và chuẩn mực hơn như sau: "Một chương trình máy tính CT được xem là học cách thực thi một lớp nhiệm vụ NV thông qua trải nghiệm KN, đối với thang đo năng lực NL nếu như dùng NL ta đo thấy năng lực thực thi của chương trình có tiến bộ sau khi trải qua KN" (máy đã học).

Học máy (Machine Learning) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Ví dụ như các máy có thể "học" cách phân loại thư điện tử xem có phải thư rác (spam) hay không và tự động xếp thư vào thư mục tương ứng. Học máy rất gần với suy diễn thống kê (statistical inference) tuy có khác nhau về thuật ngữ.

Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy

luận được xếp vào loại bài toán NP-khó, vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Học máy có hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (robot locomotion).

3.2. Tổng quan về học sâu

Học sâu (deep learning) là một chi của ngành máy học dựa trên một tập hợp các thuật toán để cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hoặc bằng cách khác bao gồm nhiều biến đổi phi tuyến.

Học sâu là một phần của một họ các phương pháp học máy rộng hơn dựa trên đại diện học của dữ liệu. Một quan sát (ví dụ như, một hình ảnh) có thể được biểu diễn bằng nhiều cách như một vector của các giá trị cường độ cho mỗi điểm ảnh, hoặc một cách trừu tượng hơn như là một tập hợp các cạnh, các khu vực hình dạng cụ thể,.. Một vài đại diện làm khiến việc học các nhiệm vụ dễ dàng hơn (ví dụ, nhận dạng khuôn mặt hoặc biểu hiện cảm xúc trên khuôn mặt) từ các ví dụ. Một trong những hứa hẹn của học sâu là thay thế các tính năng thủ công bằng các thuật toán hiệu quả đối với học không có giám sát hoặc nửa giám sát và tính năng phân cấp.

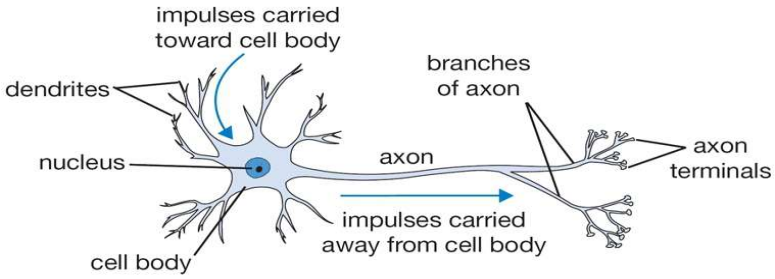
Các nghiên cứu trong lĩnh vực này cố gắng thực hiện các đại diện tốt hơn và tạo ra các mô hình để tìm hiểu các đại diện này từ dữ liệu không dán nhãn quy mô lớn. Một số đại diện được lấy cảm hứng bởi những tiến bộ trong khoa học thần kinh và được dựa trên các giải thích của mô hình xử lý và truyền thông thông tin trong một hệ thống thần kinh, chẳng hạn như mã hóa

thần kinh để cố gắng để xác định các mối quan hệ giữa các kích thích khác nhau và các phản ứng liên quan đến thần kinh trong não.

Nhiều kiến trúc học sâu khác nhau như mạng neuron sâu, mã mạng neuron tích chập sâu, mạng niềm tin sâu và mạng neuron tái phát đã được áp dụng cho các lĩnh vực như thị giác máy tính, tự động nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên, nhận dạng âm thanh ngôn ngữ và tin sinh học, chúng đã được chứng minh là tạo ra các kết quả rất tốt đối với nhiều nhiệm vụ khác nhau. [7]

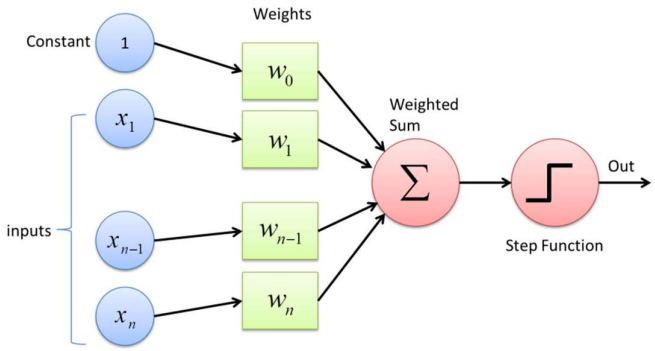
3.2.1. Mạng nơ-ron (Neural Network)

Một mạng nơ-ron được cấu thành bởi các nơ-ron đơn lẻ được gọi là các perceptron. Nơ-ron nhân tạo được lấy cảm hứng từ nơ-ron sinh học như hình mô tả bên dưới:



Hình 3.1: Mạng neuron sinh học

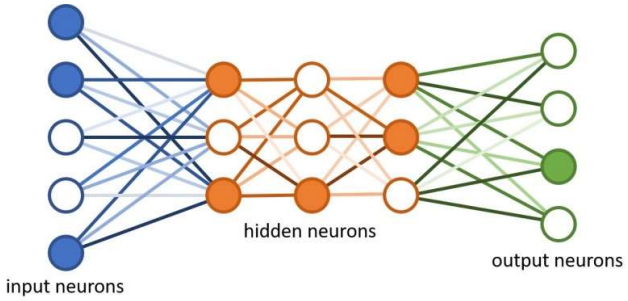
Ta có thể thấy một nơ-ron có thể nhận nhiều đầu vào và cho ra một kết quả duy nhất. Mô hình của perceptron cũng tương tự như vậy



Hình 3.2: Mô hình của perceptron

Một perceptron sẽ nhận một hoặc nhiều đầu x vào dạng nhị phân và cho ra một kết quả có dạng nhị phân duy nhất. Các đầu vào được điều phối tầm ảnh hưởng bởi các tham số trọng lượng tương ứng w của nó, còn kết quả đầu ra được quyết định dựa vào một ngưỡng quyết định b nào đó.

Mạng nơ-ron là sự kết hợp của của các lớp perceptron hay còn được gọi là perceptron đa lớp (multilayer perceptron) như hình vẽ bên dưới:

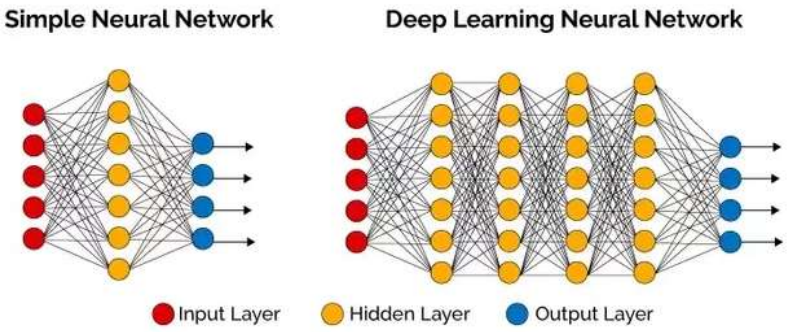


Hình 3.3. Mô hình mạng nơ-ron

Một mạng nơ-ron nhìn chung sẽ có 3 kiểu lớp:

- Lớp đầu vào (input layer): Là tầng bên trái cùng của mạng thể hiện cho các đầu vào của mạng.
- Lớp đầu ra (output layer): Là tầng bên phải cùng của mạng thể hiện cho các đầu ra của mạng.
- Lớp ẩn (hidden layer): Là tầng nằm giữa tầng vào và tầng ra thể hiện cho việc suy luận logic của mạng.

Một mạng nơ ron chỉ có 1 lớp vào và 1 lớp ra nhưng có thể có nhiều lớp ẩn. Mạng nơ-ron có nhiều lớp ẩn được xem là mạng nơ-ron học sâu.



Hình 3.4. Mạng nơron và mạng học sâu

3.2.2. Mạng nơ-ron tích chập (Convolution neural network)

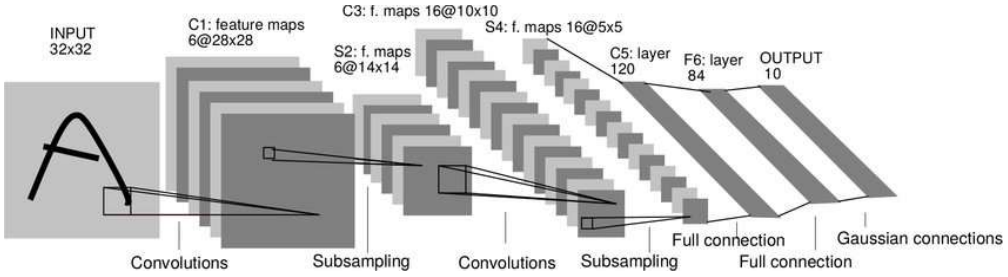
Sự ra đời của mạng CNN là dựa trên ý tưởng cải tiến cách thức các mạng nơ-ron nhân tạo truyền thống học thông tin trong ảnh. Do sử dụng các liên kết đầy đủ giữa các điểm ảnh vào node, các mạng nơ-ron nhân tạo truyền thẳng (Feedforward Neural Network) bị hạn chế rất nhiều bởi kích thước của ảnh, ảnh càng lớn thì số lượng liên kết càng tăng nhanh và kéo theo sự bùng nổ khối lượng tính toán. Ngoài ra sự liên kết đầy đủ này cũng là sự dư thừa khi với mỗi bức ảnh, các thông tin chủ yếu thể hiện qua sự phụ thuộc giữa các điểm ảnh với những điểm xung quanh nó mà không quan tâm nhiều đến các điểm ảnh ở cách xa nhau. Mạng CNN ra đời với kiến trúc thay đổi, có khả năng xây dựng liên kết chỉ sử dụng một phần cục bộ trong ảnh kết nối đến node trong lớp tiếp theo thay vì toàn bộ ảnh như trong mạng nơ-ron truyền thẳng.

CNN gồm một vài lớp tích chập kết hợp với các hàm kích hoạt phi tuyến (Nonlinear Activation Function) như ReLU hay tanh để tạo ra thông tin trừu tượng hơn (abstract/higher-level) cho các layer tiếp theo. Trong mô hình Feedforward Neural Network (mạng nơ-ron truyền thẳng), các layer kết nối trực tiếp với nhau thông qua một trọng số w (weighted vector). Các layer này

còn được gọi là có kết nối đầy đủ (fully connected layer) hay affine layer. Trong mô hình CNN thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution. Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Nghĩa là mỗi nơ-ron ở layer tiếp theo sinh ra từ filter áp đặt lên một vùng ảnh cục bộ của nơ-ron layer trước đó. Mỗi layer như vậy được áp đặt các filter khác nhau, thông thường có vài trăm đến vài nghìn filter như vậy tùy thuộc vào thiết kế của mạng. Một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu). Tuy nhiên, ta sẽ không định nghĩa các nhân tích chập của các lớp này. Trong suốt quá trình huấn luyện, CNN sẽ tự động học được các thông số cho các nhân tích chập. Ví dụ trong tác vụ phân lớp ảnh, CNN sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.

CNN có tính bất biến và tính kết hợp cục bộ (Location Invariance and Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các gốc độ khác nhau dịch chuyển, xoay, hay co giãn (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể. Lớp lấy mẫu thể hiện tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling). Tính kết hợp cục bộ biểu diễn phân cấp thông tin từ mức độ thấp đến mức độ cao và trừu tượng hơn thông qua nhân tích chập từ các bộ lọc. Đó là lý do tại sao CNN cho ra mô hình với độ chính xác rất cao. Cũng giống như cách con người nhận biết các vật thể trong tự nhiên. Ta phân biệt được một con chó với một con mèo nhờ vào các đặc trưng từ mức độ thấp (có 4 chân, có đuôi) đến mức độ cao (dáng đi, hình thể, màu lông).

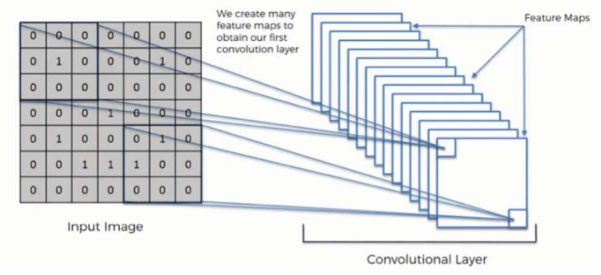
a. Kiến trúc mạng nơ-ron tích chập:



Hình 3.5: Mạng CNN LeCun 1989

Các lớp cơ bản trong một mạng CNN bao gồm: Lớp tích chập (Convolutional), Lớp kích hoạt phi tuyến (Nonlinear Activation), Lớp lấy mẫu (Pooling) và Lớp kết nối đầy đủ (Fully-connected), được thay đổi về số lượng và cách sắp xếp để tạo ra các mô hình huấn luyện phù hợp cho từng bài toán khác nhau. [8]

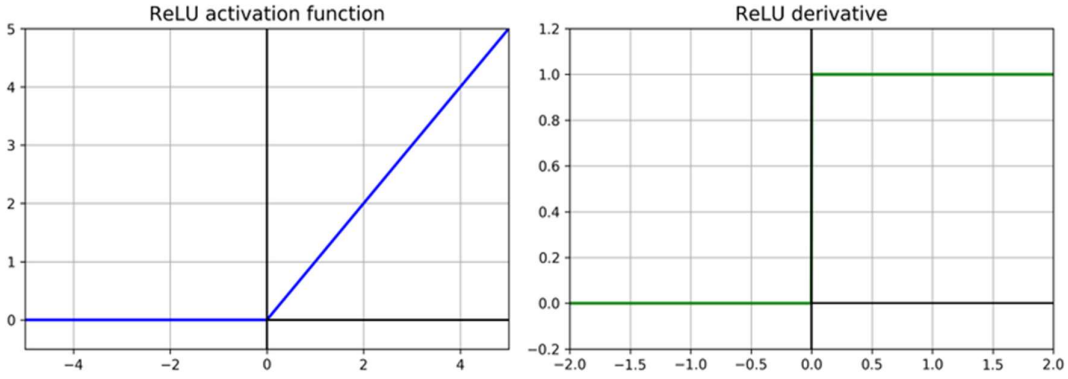
Lớp tích chập: là lớp chứa một số các lớp nhân tích chập, mỗi lớp như vậy được áp đặt các bộ lọc khác nhau. Một số layer khác như pooling/ subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu). Tuy nhiên, ta sẽ không đi sâu vào khái niệm của các layer này. Trong suốt quá trình huấn luyện, CNNs sẽ tự động học được các thông số cho các filter. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự ảnh gốc, cạnh, hình dạng, đặc trưng. Layer cuối cùng được dùng để phân lớp ảnh.



Hình 3.6: Lớp tích chập trong mạng CNN

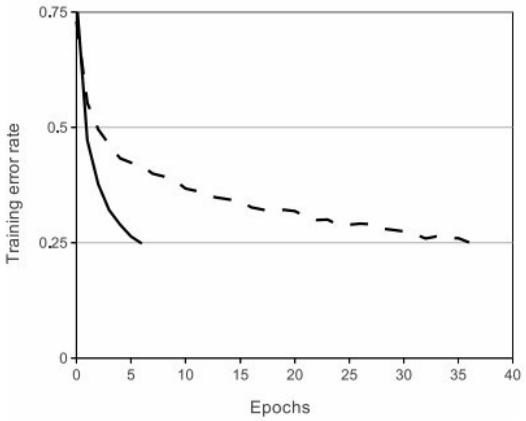
Như vậy, sau khi đưa một bức ảnh đầu vào cho lớp tích chập ta nhận được kết quả đầu ra là một loạt ảnh tương ứng với các bộ lọc đã được sử dụng để thực hiện phép tích chập. Các trọng số của các bộ lọc này được khởi tạo ngẫu nhiên trong lần đầu tiên và sẽ được cải thiện dần xuyên suốt quá trình huấn luyện.

Lớp kích hoạt phi tuyến: Lớp này có ý nghĩa là đảm bảo tính phi tuyến của mô hình huấn luyện sau khi đã thực hiện một loạt các tính toán tuyến ở lớp tích chập trước đó. Có một số hàm kích hoạt thưởng được sử dụng như là: ReLU, Sigmoid, Tanh … với chức năng là giới hạn vi phạm biên độ của giá trị đầu ra.



Hình 3.7: Hàm kích hoạt ReLU

Hàm kích hoạt ReLU (Rectified Linear Unit) được sử dụng rỗng rại gần đây vì tính đơn giản mà vẫn đảm bảo hiệu quả. Hàm kích hoạt ReLU được chứng minh là giúp cho việc training các mang học sâu nhanh hơn rất nhiều.



Hình 3.8: Đồ thị hàm mất mát sử dụng hàm ReLU

Hình 3.8 so sánh sự hội tụ của SGD (Stochastic Gradient Decenst) khi sử dụng hai hàm kích hoạt là tanh (nét đứt) và ReLU (nét liền). Sự tăng tốc này được cho ra là vì ReLU tính toàn gần như tức thời và gradient của nó cũng được tính toàn cực nhanh.

Hàm ReLU có một số biến thể như là Noisy ReLU, Leaky ReLU, PReLU …

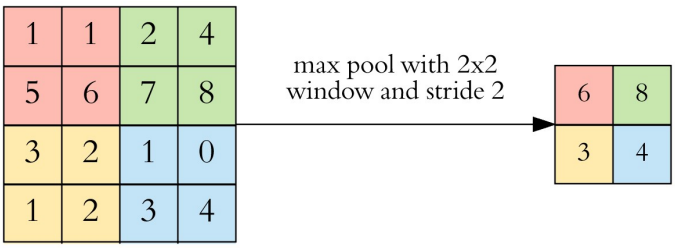
Hàm Softmax: là hàm nhận đầu vào là một vector và đầu ra là một vector có cùng số chiều, $f(x) : R_n => R_n$

$$a_i = f(x_i) = \frac{e^{x_i}}{\sum e^{x_j}}$$

Hàm Softmax thường được đặt ở lớp cuối cùng để tính xác suất nhân của dữ liệu đầu vào thuộc vào các lớp đầu ra.

Lớp lấy mẫu (pooling layer): là một trong những thành phần tính toán chính của cấu trúc CNN. Xét về mặt toán học, pooling thực chất là quá trình tính toán trên ma trận để giảm kích thước ma trận mà vẫn làm nối bật lên đặc trưng của ma trận đầu vào. Trong CNN toán tử pooling được thực hiện độc lập trên mỗi kênh màu của ma trận đầu vào.

Có nhiều toán tử pooling như Sum-Pooling, Max-Pooling, L2-Pooling, nhưng Max-Pooling thường được sử dụng phổ biến nhất. Về mặt ý nghĩa thì Max-Pooling xác định vị trí tín hiệu mạnh nhất khi áp dụng một loại lớp lọc.



Hình 3.9: Lớp lấy mẫu Max-Pooling

3.3. Mạng đối nghịch tạo sinh GANs (Generative Adversarial Network)

3.3.1. Giới thiệu

GAN được giới thiệu trong một bài báo của Ian Goodfellow và các nhà nghiên cứu khác tại Đại học Montreal vào năm 2014. Nhắc đến GANs, giám đốc nghiên cứu AI của Facebook, Yann LeCun nói rằng đây là ý tưởng thú vị nhất trong 10 năm qua trong lĩnh vực học máy

Tiềm năng của GANs rất lớn, bởi vì nó có thể học cách bắt chước bất kỳ phân phối dữ liệu nào. GANs có thể được dạy để tạo ra những thế giới tương tự như thế giới của chúng ta trong bất kỳ lĩnh vực nào: hình ảnh, âm nhạc, lời nói, văn xuôi. Có thể nói đó là những nghệ sĩ robot theo một nghĩa nào đó, và đầu ra của họ rất ấn tượng, thậm chí sâu sắc.

3.3.2. Kiến trúc mạng đối nghịch tạo sinh GANs (Generative Adversarial Network)

a. Khái niệm

GAN thuộc lớp bài toàn học không giám sát, mạng đối nghịch tạo sinh được hình thành trên cơ sở là sự cạnh tranh giữa hai mạng khác là mạng sinh

(Generative network) và mạng phân biệt (Discriminative network). Để hiểu được mạng GAN hoạt động ra sao ta phải tìm hiểu về hai mạng con bên trong nó:

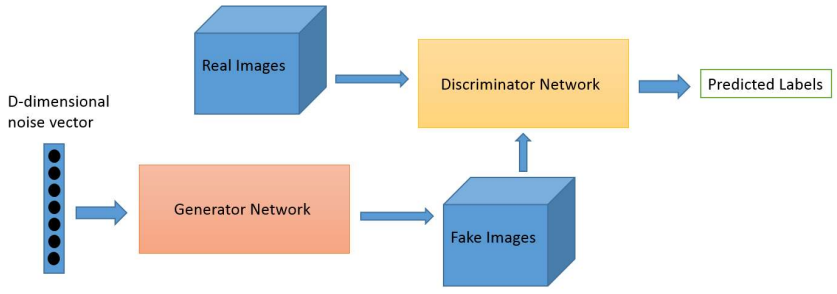
Mạng phân biệt: Là mạng thuộc lớp phân loại sử dụng mạng học sâu tích chấp, nhiệm vụ là phân biệt dữ liệu thật và dữ liệu giả hay tương đương với việc dự đoán nhãn của dữ liệu

Mạng sinh: Trái ngược với mạng phân biệt, mạng sinh cố gắng tạo ra bộ dữ liệu sao cho có phân phối giống nhất với dữ liệu thật.

Có thể hiểu hơn về mạng đối nghịch tạo sinh qua ví dụ sau:

Mạng sinh đóng vai trò như người tạo ra tranh giả cố gắng làm sao cho giống nhất có thể. Mạng phân biệt đóng vai trò như là người định giá tranh, tìm cach để phát hiện tranh giả dù có tinh vi đến đâu. Hai bên luôn luôn trong tình trạng đối nghịch nhau. Dẫn đến hệ quả là bức tranh giả được tạo ra ngày càng giống thật hơn, chi tiết hơn.

b. Cách thức hoạt động



Hình 3.10: Mô hình mạng đối nghịch tạo sinh GAN cơ bản

Mạng sinh lấy ngẫu nhiên một số đầu vào và cố gắng sinh ra mẫu dữ liệu giả. Như mô tả ở hình trên, mạng sinh $G(z)$ lấy đầu vào z từ $P_z(z)$, với z là sample thuộc phân phối xác xuất $p(z)$, được sinh ngẫu nhiên từ latent space, sau đó gán thêm nhiễu (noise). Mẫu được sinh ra từ $G(z)$ là đầu vào

của mạng phân biệt $D(x)$. Công việc của mạng phân biệt là từ tập train (real sample) và mẫu được sinh ra từ G (generated sample) và xác định xem mẫu nào mới là thật. Real sample x được lấy từ phân phối xác suất $P_{data(x)}$. $D(x)$ phân bằng cách sử dụng hàm sigmoid, trả về kết quả khoảng từ 0 đến 1, với xác suất đầu ra càng cao thì khả năng sample đó là thật (sample lấy từ tập data) càng lớn, và ngược lại. D được huấn luyện để tối đa xác suất gán đúng nhãn cho sample, đồng thời G lại được huấn luyện để tối thiểu khả năng phát hiện của D, tương đương tối thiểu $\log\left(1-D\left(G\left(z\right)\right)\right)$. Nói cách khác, việc huấn luyện D và G tương ứng với trò chơi lớn nhất nhỏ nhất giữa hai người cho hàm số: [9]

$$V\left(D,G\right)=E_{x\square Pdata(x)}\left[\log\mathrm{D}\left(x\right)\right]+E_{x\square Pdata(x)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right]$$

Trong hàm số $V\left(D,G\right)$:

$E_{x\square Pdata(x)}\left[\log\mathrm{D}\left(x\right)\right]$ là giá trị kỳ vọng khả năng sample từ phân phối training được D đánh giá dữ liệu thật. Giá trị này càng cao thì khả năng đánh giá dữ liệu training của D càng chính xác.

$E_{x\square Pdata(x)}\left[\log\left(1-D\left(G\left(z\right)\right)\right)\right]$ là giá trị kỳ vọng khả năng sample từ G (được G sinh ra từ phân phối pz) được D đánh giá là dữ liệu giả. Giá trị càng cao khả năng đánh giá dữ liệu sinh của D càng chính xác.

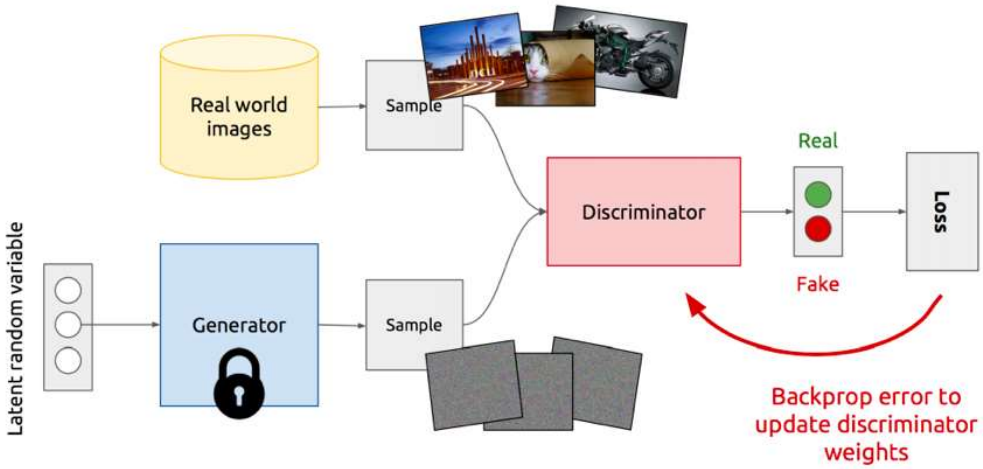
Tổng thể, D cố gắng maximize $V\left(D,G\right)$ trong khi G thì ngược lại. Quá trình huấn luyện sẽ hội tụ khi D không thể phân biệt được một sample là thật hay giả (xác suất đều là 0.5).

c. Quá trình huấn luyện

Quá trình huấn luyện của mạng GAN được chia làm hai giai đoạn lặp đi lặp lại liên tiếp nhau là quá trình huấn luyện mạng phân biệt và mạng đối

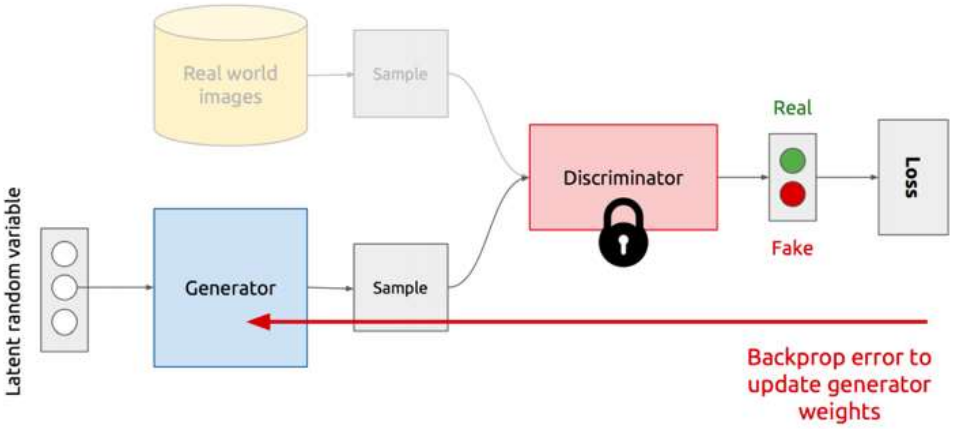
nghịch tạo sinh. Khi mạng đối nghịch tạo sinh sinh ảnh và mạng phân biệt phân biệt đúng thì tiếp tục training lại mạng đối nghịch tạo sinh và ngược lại. [9]

- Pha 1: Huấn luyện mạng phân biệt và cố định mạng đối nghịch tạo sinh (với mạng đối nghịch tạo sinh chỉ feed-forward và không back-propagation), D được update Stochastic gradient sau mỗi bước train bằng cách công thêm giá trị:



Hình 3.11: Quá trình huấn luyện mạng phân biệt

- Pha 2: Huấn luyện mạng đối nghịch tạo sinh và cố định mạng phân biệt, G được update Stochastic gradient bằng cách trừ đi giá trị:

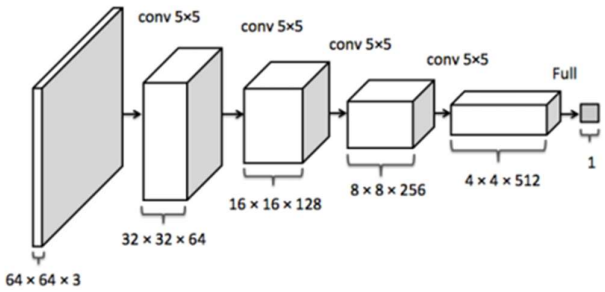


Hình 3.12: Quá trình huấn luyện mạng tạo sinh

d. Một số mạng GAN phổ biến

Trong những năm gần đây mạng GAN được coi như là một hiện tượng trong giới nghiên cứu học sâu do hướng tiếp cận đặc biệt cũng như những ứng dụng của nó trong thực tế.

- DCGAN [10] (Deep Convolutional GAN): Có thể nói DCGAN là một sự mở rộng của mạng GANs cơ bản, là tiền đề cho rất nhiều biến thể sau này nên không khó hiểu khi đây là mạng phổ biến nhất cũng là triển khai thành công nhất của GAN. Nó bao gồm các ConvNets thay cho các nhiều lớp perceptrons. ConvNets được triển khai mà không cần lấy mẫu tối đa (max pooling), trên thực tế được thay thế tích chập. Ngoài ra, các lớp không được kết nối đầy đủ (Fully-connected).



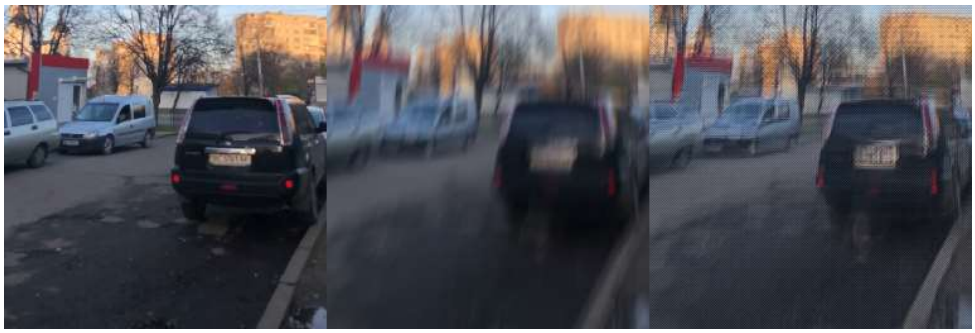
- ESRGAN [11] (Enhanced Super-Resolution Generative Adversarial Networks): là mô hình siêu phân giải ảnh sử dụng mạng GANs tuy nhiên thay vì sử dụng các khối dư và hoàn toàn không dùng Batch Normalize mà tác giả bài báo sử dụng RRDB (Residual-in-Residual Dense Block) nhằm giảm thiểu các tạo tác (artifacts) cũng như làm cho độ sáng của ảnh siêu phân giải trở nên chân thực hơn

Hình 3.13: Hình ảnh được sinh ra từ mạng CycleGAN

- EDSR [12] (Enhance Distortion Super Resolution): Tác giả bài báo cho rằng việc tăng chất lượng cảm quan của ảnh siêu phân giải và việc làm méo hình ảnh là một sự đánh đổi. Như vậy cần phải có một phương pháp để cần bằng (Trade-off) vấn đề này. Trong mạng EDSR sử dụng BNets để mô hình hóa phần ảnh bị méo.

- Conditional GAN (CGAN): Mạng GAN có điều kiện, DCGAN có thể được mô tả như một phương pháp học sâu trong đó một số tham số có điều kiện được đưa vào. Trong CGAN, một tham số bổ sung 'y' được thêm vào mạng tạo sinh để tạo dữ liệu tương ứng. Các nhãn cũng được đưa vào đầu vào cho mạng phân biệt để giúp phân biệt dữ liệu thực với dữ liệu được tạo giả.

- Debluring GAN: Khử mờ bằng mạng GAN, điểm đặc biệt của khử mờ bằng GAN là không cần dự đoán hạt nhân mờ, đầu vào để huấn luyện là cặp các ảnh gốc và ảnh mờ cùng một bố cục. Đặc biết hữu ích cho các trường hợp nhận dạng vật thể mờ di chuyển.



Hình 3.14: Hình ảnh được sinh ra từ mạng Debluring GAN

3.4. Mạng siêu phân giải ảnh SRGANs

3.4.1. Giới thiệu

Mục tiêu của đề tài khôi phục hình của độ phân giải cao HR nhiều chi tiết từ ảnh có độ phân giải thấp LR và phiên bản được tạo ra từ thuật toán gọi là ảnh siêu phân giải SR. Bài toán siêu phân giải nhận được rất nhiều sự quan tâm trong nghiên cứu thị giác máy thì tính ứng dụng của nó.

Nhìn chung có rất nhiều thuật toán siêu phân giải ảnh đơn khung, một số trong đó chạy rất nhanh và cho ra chất lượng tương đối tốt trên một khía cạnh nào đó. Nhưng bên cạnh đó còn một vấn đề rất lớn chưa được giải quyết đó là làm sao để khôi khục được chi tiết ảnh từ ảnh phân giải thấp khi đã mất đi các thông tin trong quá trình xử lí ảnh. Các phương pháp trước đây chủ yếu tập trung vào làm sao để đạt được tỉ số tín hiệu cực đại trên nhiễu là cao nhất nhưng điều này lại không phản ánh được sự mất các thành phần tần số cao và cảm quan bức ảnh không đạt được như độ phân giải yêu cầu và thường làm

ảnh bị mờ đi do bị mất chi tiết. Nên việc cần làm là phải chụp lại được sự khác biệt cảm quan giữa ảnh HR và ảnh SR.

Mạng SRGAN là mạng học sâu được thiết kế trên kiến trúc mạng tích chập CNN. Kiến trúc mạng càng sâu thì có thể việc huấn luyện càng phức tạp và khó khăn hơn nhưng mặt khác lại có tiềm năng nâng cao độ chính xác lên đáng kể vì nó cho phép mô hình hoá các ánh xạ có độ phức tạp rất cao. Thực nghiệm có thấy mạng càng sâu thì đem lại hiệu quả cao trong các bài toán siêu phân giải ảnh từ một ảnh đầu vào nên việc nghiên cứu để nâng cao hiệu xuất huấn luyện mạng học sâu rất được quan tâm. Ngoài ra tác giả sử dụng hàm mất mát cảm quan (Perceptual loss) sử dụng mạng tiền huấn luyện VGG-19 để nâng cao chất lượng ảnh siêu phân giải đầu ra.

Để nâng cao hiệu suất trong quá trình huyến luyện mạng sử dụng khối chuẩn hoá theo đợt (Batch Normalize) thường được sử dụng để chống lại sự thay đổi đồng biến nội bộ. Một lựa chon thiết kế mạnh mẽ khác là việc sử dụng các khối dư (Residual Block) để bỏ qua một số các kết nối, việc này giúp tín hiệu được bảo toàn khi đi qua mạng quá sâu.

3.4.2. Kiến trúc mạng

a. Hàm mất mát (loss function)

Mặc dù đã có bước tiến xa về độ chính xác và tốc độ về siêu phân giải đơn ảnh tuy nhiên còn một vấn đề rất lớn là việc khôi phục ảnh chất có độ phân giải thấp với tỉ lệ phóng to lớn. Các nghiên cứu gần đây chủ yếu tập trung vào tối ưu để giảm thiểu sai số toàn phương trung bình (MSE), kết quả là kết quả cho ra tỉ số tín hiệu trên lỗi là rất cao nhưng ảnh đầu ra thiếu các chi tiết tần số cao cần thiết và sự cảm nhận cảm tính không đạt được như kì vọng. Để khắc phục vấn đề trên trong luận văn sử dụng hàm mất mát là hàm mất mát cảm tính (Perceptual loss function) bao gồm hai hàm mất mát phân

biệt (Adversarial loss function) và hàm mất mát nội dung (Content loss function).

b. Mô tả thuật toán

Trong siêu phân giải đơn ảnh mục tiêu là ước lượng được ảnh độ phân giải cao ISR từ ảnh có độ phân giải thấp LR. Trong đó LR là ảnh được tạo ra bằng cách áp dụng bộ lọc Gaussian cho ảnh HR lấy mẫu xuống (down sampling) với tham số lấy mẫu r . Đối với ảnh màu C kênh màu thì ta định nghĩa kích thước ảnh là $W \text{ x } H \text{ x } C$ với LR và $rW \text{ x } rH \text{ x } C$ cho HR và SR

Mục tiêu cuối cùng là để huấn luyện một mạng sinh để có thể ước ảnh phân giải thấp đầu vào với ảnh phân giải cao tương ứng. Ta thực hiện huấn luyện một mạng sinh dựa trên mạng tích chập lan truyền xuôi G_{θ_G} với θ_G là $\{W, b\}$ là trọng số và phần bias của các lớp trong mạng. Tìm ra trọng số trong quá trình tối ưu hàm mất mát. [13]

$$\hat{\theta}_G = \arg\min \frac{1}{N} \sum_{n=1}^{N} l^{SR}\left(G_{\theta_G}\left(I_n^{LR}\right), I_n^{HR}\right)$$

c. Kiến trúc mạng phân biệt (Adversarial Network)

Bên cạnh mạng sinh, ta có xác định thêm mạng phân biệt D_{θ_D} sẽ được tối ưu cùng với mạng sinh G_{θ_G} để giải quyết vấn đề giảm thiểu đối nghịch.

$$\min_{\theta_G} \max_{\theta_D} E_{I^{HR} \square p_{train}\left(I^{HR}\right)}\left[\log D_{\theta_D}\left(I^{HR}\right)\right] + E_{I^{LR} \square p_G\left(I^{HR}\right)}\left[1 - \log D_{\theta_G}\left(I^{LR}\right)\right]$$

Ý nghĩa của công thức trên là huấn luyện mạng sinh G với mục tiêu là đánh lừa mạng phân biệt D (được sinh ra để phân biệt giữa ảnh phân giải cao và ảnh được sinh ra từ mạng G). Với cách tiếp cận này thì mạng sinh G sẽ học để tạo ra cách đánh lừa mạng phân biệt D, điều này tạo ra các phương án tối ưu về mặt cảm quan trong không gian con, đa tạp của hình ảnh tự nhiên. Đây cũng là phương pháp đối nghịch với giảm thiểu lỗi như MSE.

3.4.3. Hàm mất mát cảm quan (Perceptual loss)

Dưới đây là công thức hàm mất mát tổng quát bao gồm hai thành phần chính là hàm mất mát nội dung (content loss) và hàm mất mát đối nghịch (adversarial loss)

$$l^{SR} = \underbrace{\underbrace{l_{\mathrm{X}}^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

Ở vế bên trái ta có hàm mất mát nội dung, mạng tạo sinh sẽ tối ưu hàm mất mát này. Trong các phương pháp trước đây X thường được đo bằng MSE, ta có công thức bên dưới:

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I_{x,y}^{HR} - G_{\theta_G} \left(I^{LR} \right)_{x,y} \right)^2$$

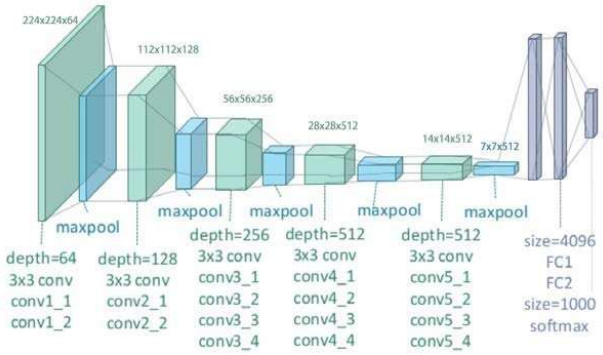
Hàm mất mát tương đồng: Hàm mất tương đồng tính tổng số lỗi tuyệt đối giữa mỗi pixel. Điều này có nghĩa là mỗi giá trị pixel về cơ bản được đo bằng các giá trị khác để tạo ra tổng đại diện cho mất pixel của hình ảnh.

Phương pháp thay thế, mất điểm nhận thức, tập trung nhiều hơn vào việc so sánh hình ảnh dựa trên các biểu diễn cấp cao. Các phương pháp trước đây hàm mất mát cố gắng mô tả sự tương đồng điểm ảnh (Pixel-wise similarity), là sự sai khác giữa các điểm ảnh và cố gắng giảm thiểu nó, nhưng thực tế cho thấy là ảnh siêu phân giải không thể hiện được các đặc trưng của nó. Việc giảm thiểu hàm mất mát này dẫn đến ảnh có độ tương đồng tốt tuy nhiên các thành phần tần số cao thì lại bị lược bỏ đi, hệ quả là ảnh sẽ bị hiện tượng mờ. Điều đó dẫn đến phải có một hàm mất cảm quan ra đời

Hàm mất mát cảm quan: Thể hiện sự sai khác giữa các lớp kích hoạt bộ lọc cụ thể trong mạng tích chập.

Trong đó hàm mất mát nội dung (content loss): được tính trên độ đo Euclid, tuy nhiên khi chỉ số tín hiệu trên lỗi trở nên rất tốt thì các phương pháp tối ưu MSE thường thiếu nội dung tần số cao và quá mịn

Nên thay vào đó ta sẽ định nghĩa hàm mất mát dựa trên hàm mất mát của mạng tiền huấn luyện VGG-19 được giới thiệu bởi Simonyan và Zisserman



Hình 3.15: Mô hình mạng VGG-19

Từ đó hàm mất mát cảm quan có dạng như sau:

$$l^{SR}_{VGG/i.j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}\left(I^{HR}\right)_{x,y} - \phi_{i.j}\left(G_{\theta_G}\left(I^{LR}\right)\right)_{x,y} \right)^2$$

Trong đó $\phi(\mathrm{i},\mathrm{j})$ chỉ ra vector đặc trưng thu được từ vector thứ j (sau lớp kích hoạt) trước lớp lấy mẫu trong mạng VGG-19 với độ đo Euclid. Công thức này tương đương với đo sự sai khác của từng vector đặc trưng tương ứng giữa ảnh HR và ảnh SR.

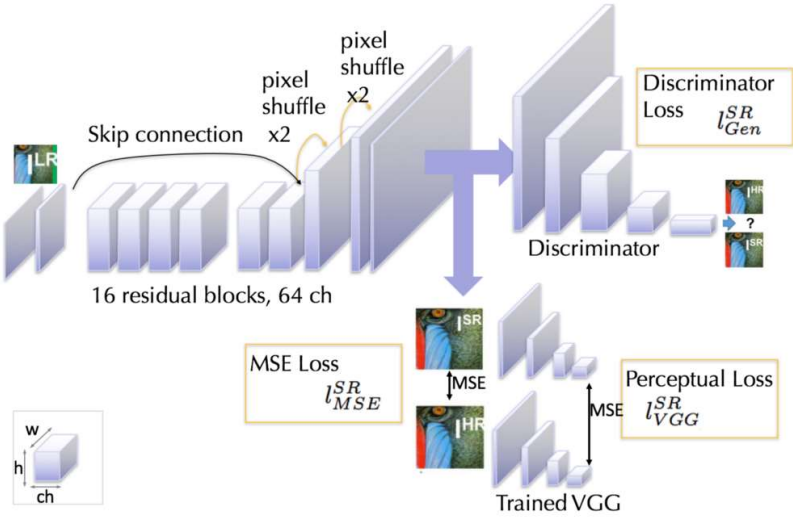
d. Hàm mất mát phân biệt (Adversarial loss)

Bên cạnh hàm mất mất mát nội dung, ta cần định nghĩa hàm mất mát phân biệt cho mạng phân biệt.

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_G}\left(G_{\theta_G}\left(I^{LR}\right)\right)$$

Với D_{θ_D}, $(G_{\theta_G}(I^{LR}))$ là xác suất mà ảnh sau khi lan truyền xuôi qua mạng sinh là ảnh tự nhiên phân giải cao. Công thức trên để giúp việc tối ưu thay vì $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$

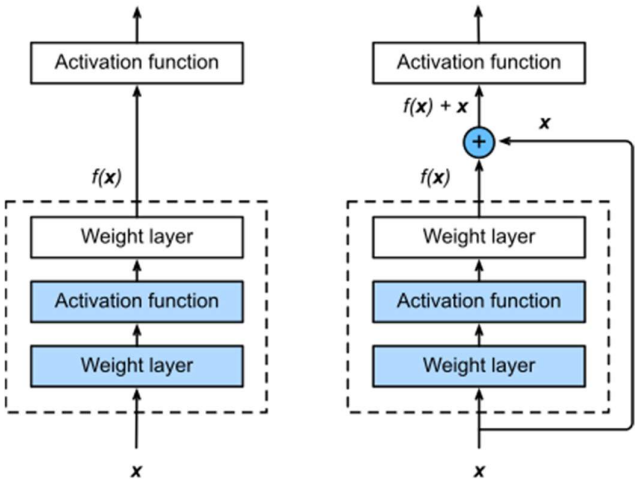
3.4.4. Thành phần trong kiến trúc mạng SRGANs



Hình 3.16: Mô hình hoạt động mạng SRGANs

Khối dư (Residual block):

Vấn đề của mạng học sâu: trong quá trình huấn luyện lặp lại tất cả các trọng số sẽ được cập nhật tương ứng với đạo hàm riêng của hàm lỗi với trọng số hiện tại. Tuy nhiên nếu gradient (đạo hàm của một hàm số) là rất nhỏ thì dẫn đến trọng số cập nhật không đáng kể và hoàn toàn kết thúc trong quá trình training. Hiện tượng trên được gọi là vanishing gradients, hay nói một cách khác là thông tin dữ liệu bị mất khi đi qua một mạng rất sâu



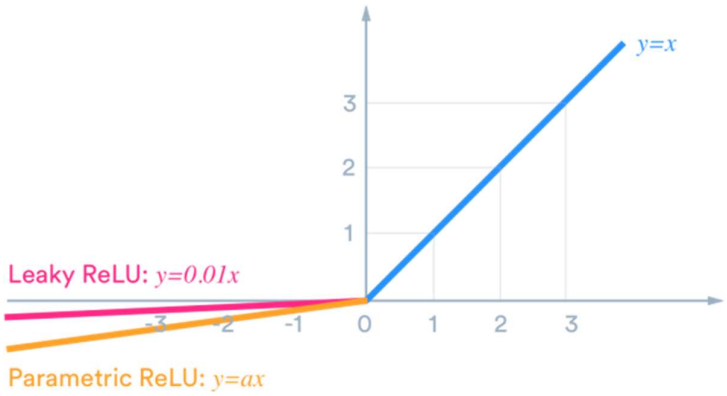
Hình 3.17. Mô hình khối dư skip connection

Residual network (ResNet): Những nhà nghiên cứu của Microsoft chỉ ra rằng nếu chia một mạng sâu ra thành 3 phần, và chuyển thẳng đầu vào của phần hiện tại vào đầu vào của phần tiếp theo cùng với đầu ra của phần hiện tại. Trừ đi cho nhau ta sẽ loại bỏ rất nhiều sự mất mát thông tin dữ liệu. Có 2 kiểu khối chính được sử dụng trong ResNet phục thuộc vào kích thước đầu vào là giống hay khác nhau.

Khối dư ra đời nhằm giải quyết vấn đề mất mát dữ liệu, ta hoàn toàn có thể huấn luyện những mạng CNN có độ phức tạp rất cao mà không lo mất mát gradient (vanishing gradient). Mấu chốt của khối dư là cứ sau hai lớp ta lại cộng đầu ra với đầu vào $f(x)+x$

Hàm kích hoạt PReLU (Parameterized Leaky ReLU):

Cũng giống như hàm kích hoat Leaky ReLU, nhưng PReLU tham số α có thể học trong quá trình huấn luyện (backpropagation) thay vì là một hyperparameter. Theo thực nghiệm cho thấy kết quả hội tụ của PReLU rất tốt trên bộ dữ liệu lớn nhưng có thể dẫn đến overfitting khi bộ dữ liệu nhỏ



Hình 3.18: Hàm kích hoạt PReLU

Khối chuẩn hóa theo tập (Batch Normalization):

Ý tưởng của BN là thêm một thao tác (operation) ngay trước hàm kích hoạt của mỗi lớp, tại đây mỗi operation có nhiệm vụ tính độ lệch chuẩn và phương sai của đầu vào trên các tập (mini batches), sau đó sẽ thực hiện chuẩn hóa và dịch chuyển tập dữ liệu về tâm 0 (zero-centering).

Lợi ích của batch normalization là:

- Giảm thiểu hiện tượng Vanishing/ Exploding gradients
- Giảm thiểu sự phụ thuộc vào quá trình khởi tạo trọng số ban đầu
- Có thể sử dụng tham số hệ số học cao hơn để tăng tốc huấn luyện

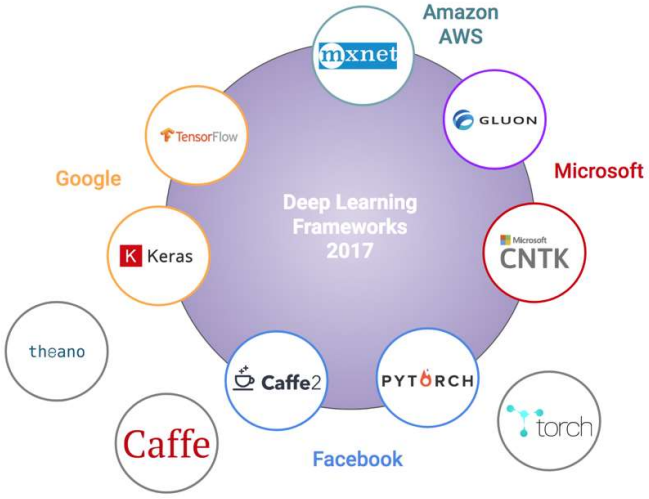
Chương 4. CÀI ĐẶT VÀ THỬ NGHIỆM

4.1. Tổng quan chương trình

4.1.1. Thư viện học sâu

Kể từ 2012 khi deep learning có bước đột phá lớn, hàng loạt các thư viện hỗ trợ deep learning ra đời. Cùng với đó, ngày càng nhiều kiến trúc deep learning ra đời, khiến cho số lượng ứng dụng và các bài báo liên quan tới deep learning tăng lên chóng mặt.

Các thư viện học sâu (deep learning) thường được phát triển từ những triển công nghệ lớn: Google (Keras, TensorFlow), Facebook (Caffe2, Pytorch), Microsoft (CNTK), Amazon (Mxnet), Microsoft và Amazon cũng đang bắt tay xây dựng Gluon (phiên bản tương tự như Keras). (Các hãng này đều có các dịch vụ cloud computing và muốn thu hút người dùng).

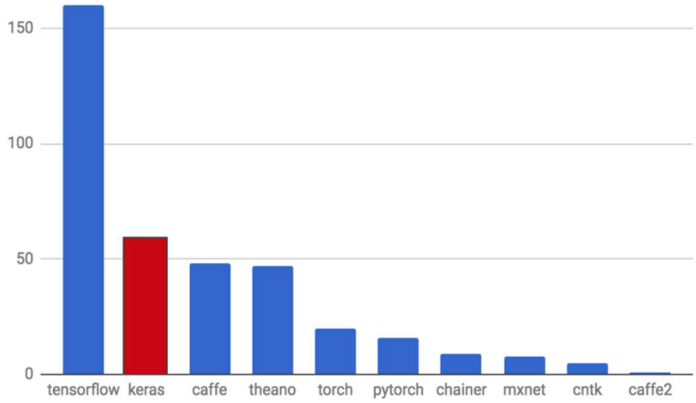


Hình 4.1: Các thư viện học sâu từ các hãng công nghệ lớn

Trong luận văn tác giả sử dụng thư viện Keras cho giao diện lập trình vì các lí do sau đây:

- Keras ưu tiên trải nghiệm của người lập trình;

- Keras đã được sử dụng rộng rãi trong doanh nghiệp và cộng đồng nghiên cứu;
- Keras hỗ trợ huấn luyện trên nhiều GPU phân tán;
- Keras hỗ trợ đa backend engines và không giới hạn bạn vào một hệ sinh thái.



Hình 4.2: Số lượng các bài báo trên arxiv có đề cập đến mỗi thư viện

4.1.2. Bộ dữ liệu

Siêu phân giải ảnh đã được nghiên cứu rất mạnh trong những năm gần đây vì nhu cầu cần thiết ảnh chất lượng cao nhiều chi tiết làm tiền đề cho các bước tiền xử lí nhằm nâng cao độ chính xác trong các bài toán nhận dạng, xác định vật thể, phần vùng, và các bài toán thính giác máy nói chung.

Trong luân văn tác giả chú trọng nghiên cứu siêu phân giải ảnh để làm đầu vào cho bài toán nhận dạng mặt người (face regconition), ngoài ra tác giả muốn áp dùng vào bài toán thực thế của người Việt Nam. Với mục đích như vậy tác giả sử dụng bộ dữ liệu mini-dataset VN-clebrity với 23.000 ảnh của 1000 người Việt Nam nổi tiếng (Hiện tại bộ dữ liệu đang được phát hành miễn phí cho mục đích học tập và nghiên cứu)

4.1.3. Mô tả quá trình huấn luyện

Bước 1: Chuẩn bị dữ liệu

Trong bộ dữ liệu có ảnh của 1000 người Việt Nam nổi tiếng tác giả thực hiện chọn mỗi người 3 ảnh có chất lượng hình ảnh cao nhất, vậy bộ dữ liệu là 3000 ảnh có kích thước 200x200 pixel. Thực hiện lấy mẫu giảm 4 lần kích thước ảnh gốc để thu được bộ ảnh huấn luyện bao gồm ảnh phân giải cao (HR) và ảnh phân giải thấp (LR)

Tác giả tiến hành chia bộ dữ liệu thành 2 phần là bộ huấn luyện và bộ kiểm tra với tỉ lệ 80:20 tương đương với 2400 ảnh cho bộ huấn luyện và 600 ảnh cho bộ kiểm tra

Bước 2: Huấn luyện

Quá trình huấn luyện diễn ra tuần tự như sau, cho ảnh LR vào mạng sinh để tạo ra ảnh siêu phân giải (SR), sau đó sử dụng mạng phân biệt để phân biệt giữa anh HR và ảnh SR. Sử dụng hàm mất mát GAN để lan truyền ngược và cập nhật các trọng số của mạng sinh nếu như mạng phân biệt phát hiện đúng và ngược lại.

Bước 3: So sánh và đánh giá

Việc so sánh đánh giá được chia làm 2 phần là so sánh định lượng và so sánh định tính.

So sánh định lượng: Trong luận văn tác giả sử dụng 2 tỉ số đo định lượng được sử dụng rộng dãi trong so sánh ảnh là tỉ số tín hiệu cực đại trên nhiễu (PSRN) và chỉ số tương đồng cấu trúc (SSIM). [3] [14]

PSNR (peak signal-to-noise ratio): tỉ số tín hiệu cực đại trên nhiễu là một thuật ngữ dùng để tính tỉ lệ giữa giá trị năng lượng tối đa của một tín hiệu và năng lượng nhiễu ảnh hướng đến độ chính xác của thông tin. Bởi vì có rất nhiều tín hiệu có phạm vi biến đổi rộng, nên PSNR thường được biểu diễn bởi đơn vị logarithm decibel (dB). Trong xử lý ảnh, PSNR được sử dụng để đo chất lượng tín hiệu khôi phục ảnh của các thuật toán (ví dụ: nén ảnh, khử nhiễu, khử mờ).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \hat{I}(i) \right)^2$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{L^2}{MSE} \right)$$

Trong đó L là giá trị pixel tối đa có thể (đối với ảnh RGB 8bits L=255)

Ta có thể thấy PSNR là tỉ lệ nghịch với logarit của sai số bình phương trung bình, dễ nhận thấy PSNR chỉ quan tâm đến sự khác biệt giữa các pixel chứ không quan tâm đến chất lượng cảm nhận

Khi so sánh các thuật toán khôi phục ảnh thường dựa vào sự cảm nhận gần chính xác của con người đối với dữ liệu được khôi phục, chính vì thế trong một số trường hợp dữ liệu được khôi phục của thuật toán này dường như có chất lượng tốt hơn những cái khác, mặc dù nó có giá trị PSNR thấp hơn (thông thường PSNR càng cao thì chất lượng dữ liệu được khôi phục càng tốt). Vì vậy khi so sánh kết quả của 2 thuật toán cần phải dựa trên codecs giống nhau và nội dung của dữ liệu cũng phải giống nhau.

SSIM (Structural Similarity Index): Chỉ số tương đồng cấu trúc. SSIM được sử dụng để đo độ tương tự giữa hai hình ảnh. SSIM được thiết kế để cải thiện các phương pháp truyền thống như tỷ lệ nhiễu tín hiệu cực đại (PSNR) và lỗi bình phương trung bình (MSE).

Sự khác biệt đối với các kỹ thuật khác được đề cập trước đây như MSE hoặc PSNR là các phương pháp này ước tính sai số tuyệt đối. Mặt khác, SSIM là mô hình dựa trên nhận thức, coi sự suy giảm hình ảnh là sự thay đổi nhận thức về thông tin cấu trúc, đồng thời kết hợp các hiện tượng nhận thức quan trọng, bao gồm cả thuật ngữ mặt nạ độ sáng và mặt nạ tương phản. Thông tin cấu trúc là ý tưởng rằng các pixel có sự phụ thuộc lẫn nhau mạnh mẽ đặc biệt là khi chúng ở gần nhau về mặt không gian. Những phụ thuộc này mang thông tin quan trọng về cấu trúc của các đối tượng trong cảnh thị giác. Mặt nạ

độ sáng là một hiện tượng trong đó các biến dạng hình ảnh có xu hướng ít nhìn thấy hơn ở các vùng sáng, trong khi mặt nạ tương phản là một hiện tượng mà các biến dạng trở nên ít nhìn thấy hơn khi có hoạt động hoặc "kết cấu" đáng kể trong ảnh.

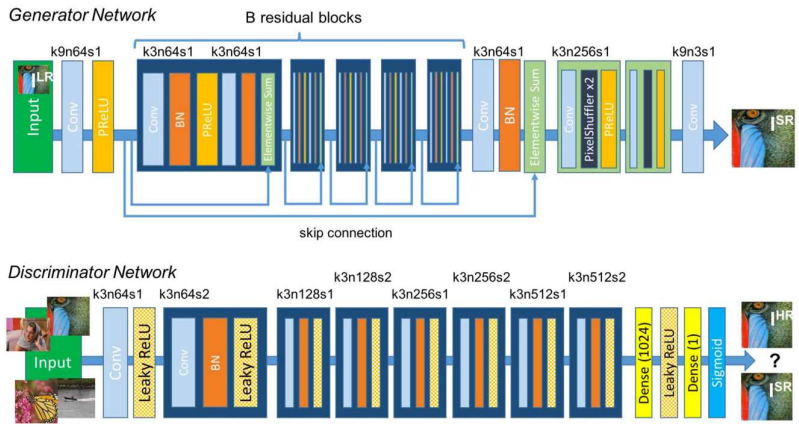
Tuy nhiên việc các thuật toán siêu phân giải gần đây quá chú trọng vào tối ưu MSE để tăng PSRN lên, kết quả là tỉ số PSRN rất tốt nhưng ảnh lại dường như bị mất chi tiết và không được như kì vọng. Để hạn chế điều này tác giả sẽ sử dụng thêm một độ đo nữa là chỉ số tương đồng cấu trúc (SSIM) dựa trên mô hình nhận thức, SSIM coi việc sự sai khác của ảnh là sự thay đổi nhận thực về thông tin cấu trúc (là sự phụ thuộc lẫn nhau của các pixel gần nhau về mặt không gian)

So sánh định tính: Tác giả sẽ so sánh thực nghiệm phương pháp siêu phân giải ảnh bằng SRGANs cùng vơi các phương pháp cổ điển để có thể thấy sự khác biệt rõ rệt bằng cảm quan thị giác.

4.2. Cài đặt

4.2.1. Mô hình chi tiết mạng SRGANs

Trong luận văn tác giả thực hiện lập trình theo mô tả gốc của mạng SRGANs bao gồm mạng sinh và mạng phân biệt



Hình 4.3: Mô hình chi tiết mạng SRGANs

Trong đó:

- B residual blocks: là lớp lớn bao gồm 16 khối dư tác dụng làm cho mạng sâu hơn đáng kể, cũng như làm cho việc huấn luyện hội tụ nhanh chóng hơn tăng hiệu suất
- PixelShuffler: là lớp lấy mẫu tăng, 2 lớp PixelShuffler làm tăng kích thước ảnh lên 4 lần
- PReLU (Parameterize ReLU): là lớp kích hoạt có chứa tham số
- k3n64s1: nghĩa là mặt nạ tích chập 3 x 3, bước nhảy 1 và 64 kênh
- BN (Batch Normalize): Đứng ngay trước lớp kích hoạt, có nhiệm vụ chuẩn hóa dữ liệu về lân cận 0

4.2.2. Thử nghiệm đánh giá và so sánh

4.2.2.1. Thử nghiệm

Tiến hành thử nghiệm chương trình sau khi đã tiến hành cài đặt mô hình mạng học sâu SRGANs như mô tả bên trên. Đầu vào bức ảnh có kích thước là 200 x 200 pixel, sau đó tác giả tiến hành lấy mẫu giảm ảnh gốc thành ảnh phân giải thấp 50 x 50 pixel. Dưới đây là một số bộ ảnh từ tập huấn luyện đã được xử lý nâng cao độ phân giải.

Hình 4.4: Một số ảnh siêu phân giải từ bộ dữ liệu huấn luyện

Hình 4.5: Một số ảnh siêu phân giải từ bộ dữ liệu kiểm tra

4.2.2.2. Đánh giá và so sánh

Để đánh giá được độ tốt của phương pháp siêu phân giải ảnh ảnh sử dụng mạng học sâu SRGANs, tác giả đưa ra 3 kịch bản để đánh giá như sau:

- Kịch bản 1: Tác giả lấy ngẫu nhiên một số ảnh ở tập huấn luyện (batch), tập dữ liệu kiểm tra để tiến hành chạy chương trình tạo ảnh siêu phân giải ảnh, sau đó tính trung bình cộng tỷ số PSNR, SSIM cũng như Sharpness của từng phương pháp siêu phân giải ảnh được trình bày trong luận văn để so sánh độ chênh lệch

- Kịch bản 2: Tác giả sẽ sử dụng ảnh thực tế để tiến hành siêu phân giải ảnh. Trong kịch bản này sẽ thể hiện được chi tiết ảnh sau khi phóng lên 4 lần

Kết quả kịch bản 1:

Bảng so sánh chỉ số tín hiệu cực đại trên nhiễu(PSNR) và chỉ số tương đồng cấu trúc(SSIM) batch ảnh (12 ảnh) lấy từ tập huấn luyện:

Phương pháp	PSNR	SSIM
Nearest Neighbor	24.83	0.79
Bilinear	26.5	0.85
Bicubic	28.33	0.89
SRGANs	27.92	0.88

Tập ảnh minh hoạ chất lượng ảnh đầu vào và đầu ra:

Hình 4.6: Batch ảnh phân giải cao HR đầu vào lấy từ tập huấn luyện

Hình 4.7: Batch ảnh siêu phân giải SR đầu ra

Bảng so sánh chỉ số tín hiệu cực đại trên nhiễu(PSNR) và chỉ số tương đồng cấu trúc(SSIM) batch ảnh (12 ảnh) lấy từ kiểm tra:

Phương pháp	PSNR	SSIM
Nearest Neighbor	25.14	0.86
Bilinear	27.28	0.93
Bicubic	30.15	0.95
SRGANs	25.71	0.92

Tập ảnh minh hoạ chất lượng ảnh đầu vào và đầu ra:

Hình 4.8: Batch ảnh phân giải cao HR đầu vào lầy từ tập kiểm tra

Hình 4.9: Batch ảnh siêu phân giải SR đầu ra

Nhận xét:

Tập huấn luyện: Chất lượng ảnh siêu phân giải đầu ra có chỉ số PSNR kém hơn một chút so với siêu phân giải nội suy Bicubic, tuy nhiên chỉ số tương đồng cấu trúc lại tương đương với Bicubic.

Tập kiểm tra: Chất lượng ảnh siêu phân giải đầu ra có chỉ số PSNR kém hơn rất nhiều chút so với Bicubic, nhưng thay vào chỉ số tương đồng cấu trúc vẫn tương đương với Bicubic.

Mạng SRGANs tiếp cận hàm mất mát có sự cái tiến hơn so với các phương pháp siêu phân giải trước đây là sự xuất hiện của hàm mất mát cảm quan (perceptual loss). Hàm này sử dụng một mạng pre-train VGG-19 có nhiệm vụ làm cho ảnh siêu phân giải tự nhiên hơn, chân thực đúng với độ phân giải cao nó đạt được.

PSNR và SSIM chỉ dựa vào sự khác biệt mức thấp giữa các pixel, và hoạt động với giả định tồn tại nhiễu Gauss, nhưng điều này có thể không phù hợp ảnh siêu phân giải.

Hình 4.10: Ảnh minh hoạ các phương pháp siêu phân giải trong chương trình

Kết quả kịch bản 2:

Tác giả sử dụng một ảnh khuôn mặt trên internet để tiến hành lấy mẫu lên nhưng chi tiết ảnh phải được đảm bảo

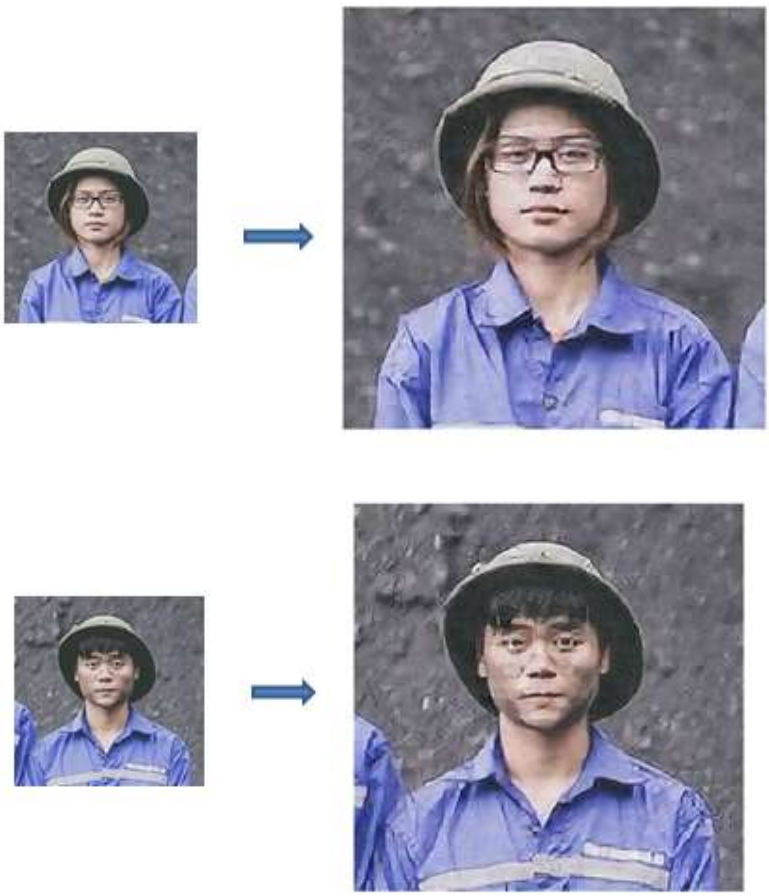
Hình 4.11: Ảnh thực tế có chứa nhiều khuôn mặt

Ta có thể thấy ảnh có chất lượng khá tốt, chi tiết khá sắc nét. Tuy nhiên khi phóng to ảnh vào một trong các khuôn mặt trong bức ảnh thì có hiện tượng bị aliasing, không thể nhìn rõ khuôn mặt sắc nét.

Hình 4.12: Hình ảnh được cắt từ ảnh 4.6 và up-sampling bằng NN

Hình 4.13: Hình ảnh được cắt từ ảnh 4.6 và up-sampling bằng Bicubic

Bằng mắt thường ta cũng có thể thấy rằng, nếu phóng to ảnh với phương pháp thông thường có thể làm cho ảnh bị hiện tượng răng cưa hoặc là bị mờ do trong quá trình lấy mẫu lên thông tin bị mất sẽ được thêm tùy thuộc vào thuật toán sử dụng tuy nhiên nói chung là sẽ dựa vào mức xám của các điểm ảnh lân cận. Điều này dẫn đến chất lượng khôi phục ảnh sẽ càng tệ hơn nếu chỉ số phóng đại lớn hơn



Hình 4.14: Hình ảnh được cắt từ ảnh 4.6 và up-sampling bằng SRGANs

Ảnh được khôi phục bằng SRGANs có độ sắc nét tốt hơn rõ rệt so với phương pháp là Nearest Neighbor và Bicubic. Các thành phần chi tiết tần số cao được thể hiện đầy đủ hơn rất nhiều

KẾT LUẬN VÀ KHUYẾN NGHỊ

1. Kết luận:

Luận văn đã thực hiện được các mục tiêu và nội dung đã đề ra đã đề ra. Nghiên cứu đề xuất đã có thấy sự phát triển tiến bộ theo chiều sâu và toàn cảnh về các vấn đề siêu phân giải ảnh đơn. Phương pháp siêu phân giải đơn ảnh sử dụng mạng học sâu SRGANs đã mang lại kết quả tốt hơn rõ rệt về mặt định tính cũng như chi tiết ảnh được thể hiện chân thực hơn rất nhiều so với các phương pháp cổ điển.

Những đóng góp thiết thực của luận văn có thể kể đến như sau:

- Lý thuyết cơ sở về siêu phân giải ảnh
- Định nghĩa hàm mất mát cảm quan dựa trên hàm mất mát của mạng tiền huấn luyện tỏ ra việc đánh giá hiệu quả hơn rất nhiều, tiềm năng còn rất lớn vì có thể sử dụng các pre-trained model khác nhau để so sánh đánh giá
- Đề xuất sử dụng các khối dư làm cho mang tạo sinh sâu hơn và phức tạp hơn, dẫn đến có thể khôi phục ảnh có bị giảm mẫu lớn hơn. Thêm vào đó các khối dư làm cho quá trình huấn luyện hội tụ nhanh hơn.

2. Khuyến nghị:

Đề tài luận văn có tính thực tiễn cao, là bước tiền xử lý quan trong cho các hệ thống thị giác máy khác. Tuy nhiên để đề tài có thể áp dụng vào thực tế cần phải sử dụng tập dữ liệu lớn nhằm nâng cao độ tin cậy

Phương pháp tác giả đưa ra đang được huấn luyện trên cho bộ dữ liệu khuôn mặt, áp vào hệ thống nhận dạng. Ngoài ra có thể mở rộng ra các lĩnh vực khác như siêu phân giải cho ảnh y tế, viễn thám …

TÀI LIỆU THAM KHẢO

Tiếng Anh:

[1] S. Chaudhuri, "Super-Resolution Imaging," Kluwer Academic Publishers, 2002.

[2] M. Elad and A. Feuer, Restoration of a single superresolution image from serveral blured, noisy and downsampled measured images, IEEE Trans. Image Processing, 1997.

[3] N. Nguyen, P. Milanfar and G. Golub, "A computationally efficient super resolution image reconstruction algorithm," IEEE Trans. Image Processing, 2001.

[4] X. Zhang and X. Wu, "Image interpolation by adaptive 2-d autoregressive," IEEE Trans. Image, 2008.

[5] R. Y. Tsai and T. S. Huang, "Multipleframe image reconstruction and Registration," Advances in Computer Vision and Image Processing, 1984.

[6] M. Ezhilarasan and P. Thambidurai, "Simplified Block Matching Algorithm for Fast Motion Estimation in Video Compression," Science Publications, 2008.

[7] Jürgen Schmidhuber, "Deep learning in neural networks," Elsevier Ltd, 2014.

[8] Y. Lecun and Y. Bengio, "Convolution Networks for Images, Speech and Time-Series," AT&T Bell Laboratories, 1998.

[9] I. J. Goodfellow, J. Pouget-Abadie and M. Mirza, "Generative Adversarial Nets," Universite de Montréal, 2014.

[10] Huang Bin, Chen Weihai and Wu Xingming, "High-Quality Face Image Super-Resolution Using Conditional Generative Adversarial Networks," Beihang University, 2017.

[11] Xintao Wang, Ke Yu, Shixiang Wu and Jinjin Gu, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," ECCV, 2018.

[12] Subeesh Vasu, Nimisha Thekke Madam and Rajagopalan A.N, "Analyzing Perception-Distortion Tradeoff using Enhanced Perceptual Super-resolution Network," Indian Institute of Technology, 2018.

[13] C. Ledig, L. Theis and F. Huszár, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial," 2017.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transaction on Image Processing, 2004.

[15] T. Acharya and P. S. Tsai, "Computational foundations of image interpolation," ACM Ubiquity, 2007.

[16] R. C. Gonzalez, R. E. Woods and S. Eddins, "Digital Image Processing Using MATLAB," Prentice-Hall, 2004.

LÝ LỊCH TRÍCH NGANG

Họ và tên: Vũ Anh Tú

Ngày tháng năm sinh: 20/06/1993 Nơi sinh: Hà Nội

Địa chỉ liên lạc: Số 16 ngõ 4 Hoàng Hoa Thám, Yên Phụ, Tây Hồ, Hà Nội

Quá trình đào tạo:

- Từ năm 2011 đến 2016: Học viện Kỹ Thuật Quân Sự – Hà Nội

- Từ năm 2017 đến nay: Học viên cao học chuyên ngành Hệ thống thông tin K29A tại Học viện Kỹ thuật Quân sự.

Quá trình công tác:

Từ 2016 – 2019: Công ty cổ phần phần mềm FPT Software

Từ 6/2019 đến nay: Cục Công nghệ thông tin và Dữ liệu Tài nguyên Môi trường – Bộ Tài nguyên và Môi trường

XÁC NHẬN QUYỂN LUẬN VĂN ĐỦ ĐIỆU KIỆN BẢO VỆ

Họ và tên tác giả luận văn: Vũ Anh Tú

Đề tài luận văn: Nghiên cứu mô hình mạng SRGAN trong nâng cao độ phân giải ảnh và ứng dụng

Chuyên ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Cán bộ hướng dẫn: TS. Nguyễn Văn Giang

Đã đủ điều kiện bảo vệ trước Hội đồng chấm luận văn.

CÁN BỘ HƯỚNG DẪN KHOA HỌC	**HỌC VIÊN**
(*Ký và ghi rõ họ tên*)	(*Ký và ghi rõ họ tên*)
TS. Nguyễn Văn Giang	**Vũ Anh Tú**
CHỦ NHIỆM KHOA (BỘ MÔN) QUẢN LÝ CHUYÊN NGÀNH (*Ký và ghi rõ họ tên*)	**CÁN BỘ KIỂM TRA** (*Ký và ghi rõ họ tên*)