

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á
KHOA CÔNG NGHỆ THÔNG TIN



BÀI TẬP LỚN

HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH

**Đề tài số 13: Xây dựng hệ thống nhận dạng và phân loại
động vật ăn thịt và ăn cỏ trong ảnh**

Giảng viên hướng dẫn: Lương Thị Hồng Lan

TT	Mã sinh viên	Sinh viên thực hiện	Lớp hành chính
1	20210611	Vũ Quang Huy	DCCNTT12.10.2
2	20210526	Lò Tiến Anh	DCCNTT12.10.2

Bắc Ninh, năm 2024

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ ĐÔNG Á
KHOA CÔNG NGHỆ THÔNG TIN

BÀI TẬP LỚN

HỌC PHẦN: XỬ LÝ ẢNH VÀ THỊ GIÁC MÁY TÍNH

Đề tài số 13: Xây dựng hệ thống nhận dạng và phân loại
động vật ăn thịt và ăn cỏ trong ảnh

Giảng viên hướng dẫn: Lương Thị Hồng Lan

TT	Mã sinh viên	Sinh viên thực hiện	Lớp hành chính
1	20210611	Vũ Quang Huy	DCCNTT12.10.2
2	20210526	Lò Tiến Anh	DCCNTT12.10.2

Bắc Ninh, năm 2024

PHIẾU CHẤM THI BÀI TẬP LỚN KẾT THÚC HỌC PHẦN

Mã đề thi: **13**

Tên học phần: **Xử lý ảnh và thị giác máy tính**

Lớp Tín chỉ: **XATGMT.03.K12.02.LH.C04.1_LT**

Cán bộ chấm thi 1

(Ký và ghi rõ họ tên)

Cán bộ chấm thi 2

(Ký và ghi rõ họ tên)

Lương Thị Hồng Lan

TT	TIÊU CHÍ	THANG ĐIỂM	Vũ Quang Huy	Lò Tiến Anh
			20210611	20210526
1	Nội dung báo cáo trên Word đầy đủ	3.5		
1.1	Có bố cục rõ ràng (mục lục, phần mở đầu, nội dung chính, kết luận).	0,5		
1.2	Nội dung phân tích rõ ràng, logic.	0,5		
1.3	Có dẫn chứng, số liệu minh họa đầy đủ.	0,5		
1.4	Ngôn ngữ và trình bày chuẩn, không lỗi chính tả.	0,5		

TT	TIÊU CHÍ	THANG ĐIỂM	Vũ Quang Huy	Lò Tiến Anh
			20210611	20210526
1.5	Có trích dẫn tài liệu tham khảo đúng quy cách.	0,5		
1.6	Được trình bày chuyên nghiệp (canh lề, font chữ, khoảng cách dòng hợp lý).	0,5		
1.7	Tài liệu đầy đủ, bám sát yêu cầu của đề bài.	0,5		
2	Nội dung thuyết trình đầy đủ	1.0		
2.1	Trình bày tự tin, phát âm rõ ràng, mạch lạc.	0,5		
2.2	Nội dung thuyết trình đúng trọng tâm, không lan man.	0,5		
3	Slides báo cáo đầy đủ nội dung + Hỏi đáp	3.0		
3.1	Slides có bố cục rõ ràng (mở đầu, nội dung, kết luận).	0,5		
3.2	Thiết kế slides đẹp, chuyên nghiệp (màu sắc, hình ảnh minh họa).	0,5		
3.3	Nội dung trên slides ngắn gọn, dễ hiểu, súc tích.	0,5		

TT	TIÊU CHÍ	THANG ĐIỂM	Vũ Quang Huy	Lò Tiến Anh
			20210611	20210526
3.4	Nội dung slides phù hợp với nội dung báo cáo.	0,5		
3.5	Trả lời câu hỏi đầy đủ, chính xác.	0,5		
3.6	Trả lời câu hỏi tự tin, thuyết phục.	0,5		
4	Code đầy đủ	2.5		
1.1	Code được trình bày rõ ràng, có chú thích đầy đủ.	0,5		
1.2	Code chạy đúng, không lỗi.	0,5		
1.3	Code tối ưu, không dư thừa.	0,5		
1.4	Đáp ứng đầy đủ các yêu cầu chức năng theo đề bài.	0,5		
1.5	Có tính sáng tạo hoặc cải thiện so với yêu cầu.	0,5		
TỔNG ĐIỂM BẢNG SỐ:		10		
TỔNG ĐIỂM BẢNG CHỮ:		<i>Mười tròn</i>		

LỜI MỞ ĐẦU

Trong những năm gần đây, sự phát triển vượt bậc của trí tuệ nhân tạo và học máy đã mở ra nhiều cơ hội mới trong việc nghiên cứu và phân tích dữ liệu hình ảnh. Đặc biệt trong lĩnh vực sinh học và nghiên cứu động vật, việc ứng dụng các công nghệ nhận dạng hình ảnh không chỉ giúp các nhà nghiên cứu thu thập và phân tích dữ liệu một cách hiệu quả mà còn góp phần quan trọng trong việc nghiên cứu, bảo tồn và phân loại các loài động vật trong tự nhiên.

Nhận thức được tầm quan trọng của công nghệ trí tuệ nhân tạo trong nghiên cứu sinh học, nhóm đã lựa chọn đề tài "Xây dựng hệ thống nhận dạng và phân loại động vật ăn thịt và ăn cỏ trong ảnh" làm nội dung nghiên cứu. Đề tài này không chỉ là cơ hội để nhóm áp dụng những kiến thức về học máy và xử lý ảnh vào thực tiễn mà còn giúp nhóm hiểu sâu sắc hơn về việc phát triển các hệ thống nhận dạng thông minh trong thực tế.

Bên cạnh việc phát triển hệ thống, nhóm cũng nhận thức được thách thức và khó khăn trong việc xây dựng một mô hình có độ chính xác cao, đặc biệt khi phải xử lý đa dạng các loại ảnh với điều kiện ánh sáng và góc chụp khác nhau. Do điều kiện thời gian có hạn và sự phức tạp của bài toán nhận dạng hình ảnh, đề tài của nhóm chúng em khó có thể tránh khỏi những hạn chế nhất định. Rất mong sự đóng góp của cô và các bạn để đề tài của nhóm có thể hoàn thiện thêm. Nhóm xin chân thành cảm ơn!

LỜI CẢM ƠN

Chúng em xin gửi lời tri ân sâu sắc đến Giảng viên Lương Thị Hồng Lan, người đã tận tình hướng dẫn và hỗ trợ chúng em trong quá trình thực hiện đề tài về "Xây dựng hệ thống nhận dạng và phân loại động vật ăn thịt và ăn cỏ trong ảnh".

Đặc biệt, Cô đã tạo điều kiện cho chúng em có cơ hội thực hành và áp dụng kiến thức lý thuyết vào thực tế, giúp chúng em phát triển kỹ năng về trí tuệ nhân tạo, xử lý ảnh và học sâu trong lĩnh vực nhận dạng hình ảnh. Những hướng dẫn chi tiết của Cô về các kỹ thuật học máy và xử lý dữ liệu đã giúp chúng em rất nhiều trong việc xây dựng mô hình phân loại hiệu quả.

Bên cạnh đó, chúng em cũng rất trân trọng sự kiên nhẫn và sự động viên của Cô trong quá trình vượt qua những thách thức và khó khăn trong nghiên cứu và thực hiện đề tài. Sự hỗ trợ của Cô không chỉ giúp chúng em hoàn thành tốt đề tài mà còn giúp chúng em tích lũy được nhiều kinh nghiệm quý báu trong lĩnh vực trí tuệ nhân tạo và thị giác máy tính.

Một lần nữa, chúng em xin bày tỏ lòng biết ơn chân thành đến Cô Lương Thị Hồng Lan, người đã luôn bên cạnh, hỗ trợ và khuyến khích chúng em trong suốt hành trình này. Chúng em sẽ luôn nhớ và trân trọng những kiến thức và kinh nghiệm mà Cô đã truyền đạt. Xin chân thành cảm ơn!

LỜI CAM ĐOAN

Những nội dung trong bài tập lớn này là thành quả từ sự nghiên cứu và được thực hiện dưới sự trực tiếp hướng dẫn của giảng viên Lương Thị Hồng Lan.

Bài tập lớn được thực hiện hoàn toàn mới, là thành quả của cả nhóm, không sao chép theo bất cứ bài tập lớn hay đồ án tương tự nào. Mọi sự tham khảo sử dụng trong bài tập lớn đều được trích dẫn các nguồn tài liệu trong báo cáo và danh mục tài liệu tham khảo.

Mọi sao chép không hợp lệ, vi phạm quy chế của nhà trường, chúng em xin hoàn toàn chịu trách nhiệm.

MỤC LỤC

LỜI MỞ ĐẦU	iv
LỜI CẢM ƠN.....	v
LỜI CAM ĐOAN	vi
MỤC LỤC	vii
DANH MỤC HÌNH	ix
DANH MỤC BẢNG	x
DANH MỤC CÁC TỪ VIẾT TẮT.....	xi
Chương 1: Giới thiệu tổng quan.....	1
1.1 Nhận dạng đối tượng trong thị giác máy tính.....	1
1.1.1 Nhận dạng đối tượng là gì?	1
1.1.2 Những vấn đề cần giải quyết trong bài toán.....	1
1.1.3 Ứng dụng trong các lĩnh vực	2
1.2 Các phương pháp áp dụng cho bài toán nhận dạng.....	2
1.2.1 Convolutional Neural Network (CNN)	2
1.2.2 Region-based CNN (R-CNN)	3
1.2.3 YOLO (You Only Look Once)	4
1.2.4 SSD (Single Shot Detector).....	4
1.2.5 ResNet50 (Residual Network).....	5
1.3 Ngôn ngữ lập trình và các thư viện	6
1.3.1 Python.....	6
1.3.2 PyTorch	7
1.3.3 Yolo	8
1.3.4 ResNet50	8
Chương 2: Xây dựng hệ thống nhận dạng và phân loại động vật ăn thịt và ăn cỏ trong ảnh	9
2.1 Yêu cầu của bài toán.....	9

2.2 Xây dựng hệ thống	11
Chương 3: Kết quả thực nghiệm	18
3.1 Dữ liệu	18
3.1.1 Nguồn dữ liệu	18
3.1.2 Tiền xử lý dữ liệu	18
3.1.3 Phân chia dữ liệu	19
3.2 Độ đo đánh giá.....	19
3.2.1 Độ chính xác (Accuracy).....	19
3.2.2 Độ chính xác theo lớp (Precision)	19
3.2.3 Thời gian xử lý	20
3.3 Kết quả thực nghiệm.....	20
3.3.1 Kết quả huấn luyện mô hình.....	20
3.3.2 Kết quả xây dựng mô hình	21
Kết luận.....	23
Tóm tắt nội dung trong đề tài	23
Kết quả nhận được	23
Hướng phát triển.....	23
Danh mục tài liệu tham khảo	25

DANH MỤC HÌNH

Hình 1.1. Phương pháp CNN	3
Hình 1.2. Mô tả phương pháp SSD	5
Hình 1.3. Mô tả phương pháp ResNet50.....	6
Hình 2.1. Hình ảnh đầu vào của hệ thống	9
Hình 2.2. Hình ảnh kết quả đầu ra của hệ thống	10
Hình 2.3. Mô hình hệ thống.....	11
Hình 2.4. Luồng xử lý	12
Hình 2.5. Module tiền xử lý ảnh đầu vào	13
Hình 2.6. Module phát hiện đối tượng với Yolo	15
Hình 2.7. Module phân loại với ResNet50.....	16
Hình 3.1. Hình ảnh dữ liệu thô được thu thập chưa qua xử lý	18
Hình 3.2. Phát triển ứng dụng trên nền tảng website	21
Hình 3.3. Kết quả đầu ra của hệ thống	22
Hình 3.4. Kết quả phân tích đầu ra của hệ thống	22

DANH MỤC BẢNG

Bảng 3.1 Kết quả huấn luyện mô hình	20
-------------------------------------------	----

DANH MỤC CÁC TỪ VIẾT TẮT

Chữ viết tắt	Thuật ngữ tiếng Anh/Giải thích	Thuật ngữ tiếng Việt/Giải thích
CNN	Convolutional Neural Network	Mạng thần kinh tích chập
R-CNN	Region-based Convolutional Neural Network	
YOLO	You Only Look Once	
SSD	Single Shot Detector	
AI	Artificial Intelligence	Trí tuệ nhân tạo
ML	Machine Learning	Học Máy

Chương 1: Giới thiệu tổng quan

1.1 Nhận dạng đối tượng trong thị giác máy tính

1.1.1 Nhận dạng đối tượng là gì?

Nhận dạng đối tượng là một kỹ thuật thị giác máy tính để xác định các đối tượng trong hình ảnh hoặc video. Nhận dạng đối tượng là đầu ra chính của thuật toán học sâu và học máy. Khi con người nhìn vào một bức ảnh hoặc xem video, chúng ta có thể dễ dàng phát hiện ra người, vật thể, cảnh và chi tiết trực quan. Mục tiêu là dạy máy tính làm những gì con người có thể làm một cách tự nhiên: đạt được mức độ hiểu biết về nội dung của hình ảnh. Nhận dạng đối tượng là công nghệ then chốt đằng sau xe không người lái, cho phép chúng nhận ra biển báo dừng hoặc phân biệt người đi bộ với cột đèn. Nó cũng hữu ích trong nhiều ứng dụng khác nhau như xác định bệnh trong hình ảnh sinh học, kiểm tra công nghiệp và thị giác robot.

1.1.2 Những vấn đề cần giải quyết trong bài toán

Thu thập và tiền xử lý dữ liệu

- Thu thập bộ dữ liệu ảnh phù hợp
- Chuẩn hóa kích thước ảnh
- Tăng cường dữ liệu (data augmentation)
- Loại bỏ nhiễu và cải thiện chất lượng ảnh

Trích xuất đặc trưng

- Xác định các đặc trưng quan trọng của đối tượng
- Sử dụng các phương pháp trích xuất đặc trưng thủ công hoặc tự động
- Chọn lọc đặc trưng phù hợp

Xây dựng và huấn luyện mô hình

- Lựa chọn kiến trúc mô hình phù hợp
- Thiết lập các tham số huấn luyện
- Thực hiện quá trình học

Đánh giá và tối ưu hóa

- Xác định các metrics đánh giá
- Kiểm tra hiệu suất mô hình

- Tinh chỉnh để cải thiện kết quả

1.1.3 Ứng dụng trong các lĩnh vực

Ứng dụng trong Y tế

- Phát hiện khối u từ ảnh X-quang, CT, MRI
- Phân tích tế bào trong ảnh kính hiển vi
- Nhận dạng bệnh lý qua hình ảnh da liễu
- Hỗ trợ chẩn đoán bệnh qua nội soi

Ứng dụng trong Quân sự và An ninh

- Giám sát biên giới tự động
- Nhận dạng mục tiêu quân sự qua ảnh vệ tinh
- Hệ thống nhận dạng khuôn mặt tại các điểm kiểm soát
- Phát hiện vũ khí qua máy quét an ninh

Ứng dụng trong Kinh tế và Thương mại

- Hệ thống thanh toán tự động trong siêu thị
- Kiểm tra chất lượng sản phẩm trên dây chuyền sản xuất
- Nhận dạng sản phẩm trong kho hàng tự động
- Phân tích hành vi mua sắm của khách hàng

Ứng dụng trong Nông nghiệp

- Phát hiện sâu bệnh trên cây trồng
- Phân loại nông sản tự động
- Giám sát sinh trưởng cây trồng

Ứng dụng trong Chăn nuôi

- Nhận dạng và đếm số lượng vật nuôi
- Theo dõi sự phát triển của đàn
- Định danh cá thể thông qua đặc điểm nhận dạng
- Phát hiện xâm nhập của động vật gây hại

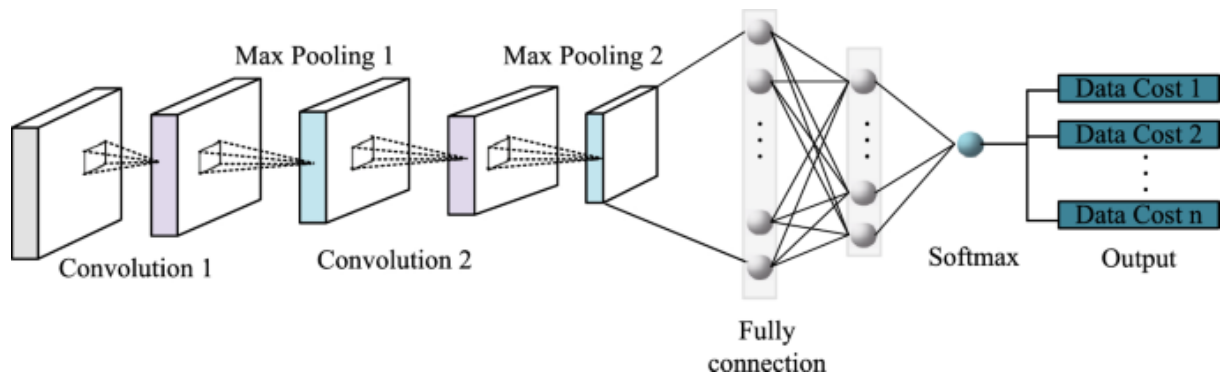
1.2 Các phương pháp áp dụng cho bài toán nhận dạng

1.2.1 Convolutional Neural Network (CNN)

a), Ý tưởng

Trích xuất đặc trưng tự động thông qua các lớp convolution. Giảm chiều dữ liệu qua các lớp pooling. Học các đặc trưng từ đơn giản đến phức tạp. Sử dụng fully connected layers để phân loại.

b), Mô tả phương pháp



Hình 1.1. Phương pháp CNN

c), Ưu điểm

- Khả năng tự động trích xuất đặc trưng
- Hiệu quả trong xử lý ảnh
- Tính bất biến với dịch chuyển và biến dạng
- Khả năng học các pattern phức tạp.

d), Nhược điểm

- Cần nhiều dữ liệu huấn luyện
- Dễ bị overfitting
- Chi phí tính toán cao
- Khó điều chỉnh hyperparameters

1.2.2 Region-based CNN (R-CNN)

a), Ý tưởng

Kết hợp vùng đề xuất (region proposal) với CNN. Trích xuất khoảng 2000 vùng từ ảnh đầu vào sử dụng selective search. Áp dụng CNN để phân loại từng vùng. Tính chỉnh bounding box bằng linear regression

b), Ưu điểm

- Độ chính xác cao trong việc định vị đối tượng
- Khả năng phát hiện nhiều đối tượng trong một ảnh
- Phù hợp với các bài toán detection phức tạp

c), Nhược điểm

- Tốc độ xử lý chậm do phải phân tích nhiều vùng

- Yêu cầu nhiều bước xử lý riêng biệt
- Tốn nhiều bộ nhớ
- Training phức tạp, nhiều giai đoạn

1.2.3 YOLO (*You Only Look Once*)

a), Ý tưởng

Chia ảnh thành lưới $S \times S$. Dự đoán bounding boxes và class probabilities cho mỗi cell. Thực hiện detection trong một lần đi qua mạng neural. Xử lý real-time với tốc độ cao

b), Ưu điểm

- Tốc độ xử lý nhanh, có thể chạy real-time
- Hiểu được context tổng thể của ảnh
- Dễ dàng triển khai và tối ưu hóa
- Architecture đơn giản, end-to-end training

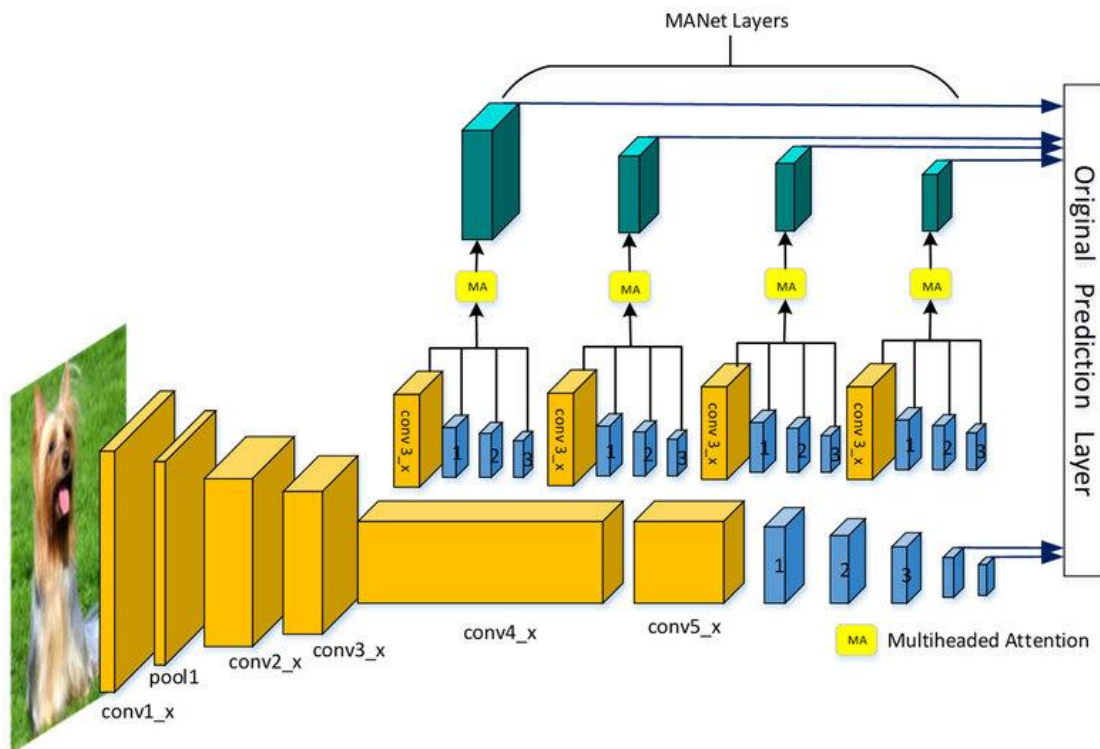
c), Nhược điểm

- Độ chính xác thấp hơn so với R-CNN
- Khó phát hiện các đối tượng nhỏ
- Hạn chế trong việc phát hiện các đối tượng gần nhau
- Nhạy cảm với grid size

1.2.4 SSD (*Single Shot Detector*)

a), Ý tưởng

Sử dụng nhiều feature maps ở các scale khác nhau. Tạo default boxes với tỷ lệ và kích thước khác nhau. Dự đoán class scores và box offsets cho mỗi default box. Kết hợp predictions từ nhiều layers



Hình 1.2. Mô tả phương pháp SSD

b), Ưu điểm

- Cân bằng giữa tốc độ và độ chính xác
- Khả năng phát hiện đối tượng ở nhiều scale
- Training đơn giản, end-to-end
- Hiệu quả với các đối tượng có kích thước khác nhau

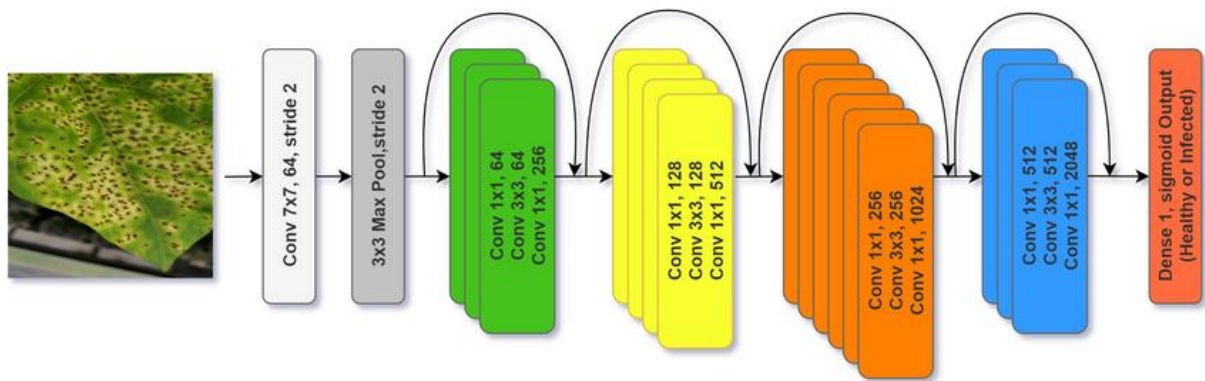
c), Nhược điểm

- Hiệu suất phụ thuộc vào chất lượng của base network
- Tốn nhiều bộ nhớ do nhiều feature maps
- Độ chính xác có thể thấp hơn với đối tượng rất nhỏ
- Cần điều chỉnh cẩn thận các default box parameters

1.2.5 ResNet50 (Residual Network)

a), Ý tưởng

Giải quyết vấn đề vanishing gradient trong các mạng neural sâu bằng cách sử dụng skip connections (residual connections). Cho phép xây dựng các mạng neural có độ sâu lên đến hàng trăm layer mà không gặp vấn đề về hiệu suất huấn luyện.



Hình 1.3. Mô tả phương pháp ResNet50

c), Ưu điểm

- Khắc phục được vấn đề vanishing gradient
- Khả năng xây dựng mạng neural rất sâu
- Hiệu suất cao trong các bài toán phân loại và object detection
- Dễ dàng huấn luyện và tối ưu hóa
- Có thể chuyển giao học (transfer learning) hiệu quả

d), Nhược điểm

- Số lượng tham số lớn
- Chi phí tính toán cao
- Yêu cầu nhiều dữ liệu huấn luyện
- Có thể bị redundancy do quá nhiều layer
- Khó giải thích các đặc trưng được học

1.3 Ngôn ngữ lập trình và các thư viện

1.3.1 Python

Python là một ngôn ngữ lập trình bậc cao, cú pháp đơn giản, dễ đọc và dễ học, mã nguồn mở và đa nền tảng. Python được sử dụng rộng rãi để phát triển các ứng dụng web, phát triển phần mềm, khoa học dữ liệu và máy học (ML).

Python trở thành ngôn ngữ ưu tiên trong lĩnh vực máy học và thị giác máy tính với nhiều ưu điểm:

Hệ sinh thái thư viện phong phú:

- Xử lý số liệu: NumPy, Pandas
- Máy học: Scikit-learn, TensorFlow, PyTorch
- Trực quan hóa: Matplotlib, Seaborn

- Xử lý ảnh: OpenCV, Pillow

Tích hợp linh hoạt:

- Dễ dàng kết nối các framework deep learning
- Hỗ trợ mạnh mẽ cho các thư viện khoa học
- Tích hợp tốt với các công cụ phân tích dữ liệu

Hiệu suất tính toán:

- Hỗ trợ tốt cho các thao tác ma trận
- Khả năng tối ưu hóa với Numba, Cython
- Hỗ trợ tính toán song song

1.3.2 PyTorch

PyTorch là một thư viện Máy học mã nguồn mở, được phát triển bởi nhóm nghiên cứu của Facebook. Bản phát hành đầu tiên của PyTorch là vào tháng 10 năm 2016.

Đặc biệt, trong các lĩnh vực nghiên cứu, nhiều tác giả hiện nay sử dụng PyTorch để triển khai bài toán của mình. PyTorch cho thấy lợi thế của nó trong lĩnh vực nghiên cứu bởi việc rất dễ dàng debug model.

PyTorch dễ sử dụng, được viết theo Python-based nên dễ tiếp cận đối với những người dùng có kiến thức lập trình cơ bản.

Các tính năng chính:

Tensor operations mạnh mẽ

- Hỗ trợ các phép toán ma trận phức tạp
- Tích hợp GPU một cách liền mạch
- Hỗ trợ các phép toán tự động trên GPU

Autograd system cho automatic differentiation

- Tự động tính toán gradient
- Hỗ trợ backpropagation động
- Linh hoạt trong việc xây dựng mô hình

Neural Networks module (torch.nn)

Data loading và preprocessing utilities

Hỗ trợ distributed training

1.3.3 Yolo

Yolo là một mô hình mạng CNN cho việc phát hiện, nhận dạng, phân loại đối tượng. Yolo được tạo ra từ việc kết hợp giữa các convolutional layers và connected layers. Trong đó các convolutional layers sẽ trích xuất ra các feature của ảnh, còn full-connected layers sẽ dự đoán ra xác suất đó và tọa độ của đối tượng.

Phương Pháp Phát Hiện Đối Tượng:

- Chia ảnh thành lưới (grid) $S \times S$.
- Mỗi cell dự đoán:
- Số lượng bounding boxes.
- Xác suất xuất hiện của từng class.
- Độ chính xác của mỗi box.

Quy Trình Xử Lý:

- Chia ảnh thành lưới đều.
- Dự đoán đồng thời vị trí và class của đối tượng.
- Lọc và tinh chỉnh các bounding boxes.
- Áp dụng non-maximum suppression.

1.3.4 ResNet50

ResNet (Residual Network) được giới thiệu đến công chúng vào năm 2015 và thậm chí đã giành được vị trí thứ 1 trong cuộc thi ILSVRC 2015 với tỉ lệ lỗi top 5 chỉ 3.57%. Không những thế nó còn đứng vị trí đầu tiên trong cuộc thi ILSVRC and COCO 2015 với ImageNet Detection, ImageNet localization, Coco detection và Coco segmentation. Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152,... Với tên là ResNet theo sau là một số chỉ kiến trúc ResNet với số lớp nhất định.

Cấu Trúc Bottleneck Block mỗi block ResNet50 bao gồm 3 lớp convolution:

- 1x1 convolution giảm chiều đặc trưng (dimensionality reduction).
- 3x3 convolution xử lý đặc trưng chính.
- 1x1 convolution khôi phục chiều ban đầu.

Residual Connection:

- Giải pháp cho bài toán vanishing gradient.
- Cho phép gradient truyền trực tiếp qua các layer.
- Kích hoạt học sâu hơn mà không làm suy giảm hiệu suất.

Chương 2: Xây dựng hệ thống nhận dạng và phân loại động vật ăn thịt và ăn cỏ trong ảnh

2.1 Yêu cầu của bài toán

Các hệ sinh thái tự nhiên được cấu thành từ nhiều loài động vật khác nhau với các đặc điểm và vai trò sinh thái riêng biệt. Trong đó, việc phân loại động vật theo tập tính ăn uống (động vật ăn thịt và động vật ăn cỏ) có ý nghĩa quan trọng trong nghiên cứu chuỗi thức ăn và mối quan hệ sinh thái. Tuy nhiên, việc phân loại thủ công từng loài động vật là một công việc tốn nhiều thời gian và công sức.

Từ thực trạng trên, việc xây dựng một hệ thống tự động có khả năng nhận dạng và phân loại động vật theo tập tính ăn uống từ ảnh là một nhu cầu thiết yếu. Hệ thống này không chỉ giúp các nhà nghiên cứu và bảo tồn động vật làm việc hiệu quả hơn, mà còn mang lại nhiều ứng dụng thực tiễn trong giáo dục và nghiên cứu sinh thái học.

Đầu vào của hệ thống sẽ là hình ảnh chứa đối tượng động vật, như minh họa trong hình 2.1 dưới đây:



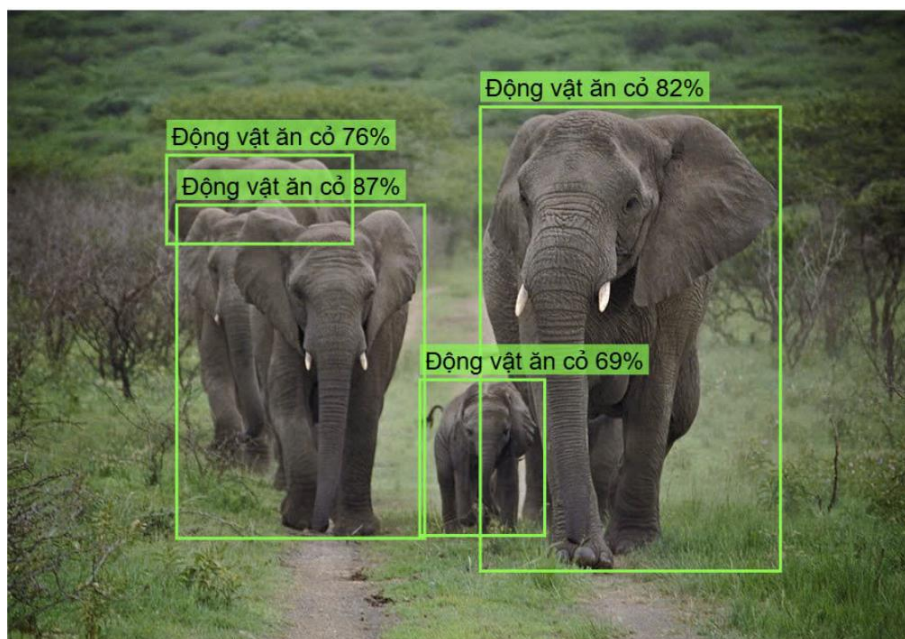
Hình 2.1. Hình ảnh đầu vào của hệ thống

Đầu vào của hệ thống sẽ là hình ảnh chứa đối tượng động vật. Hệ thống được thiết kế để xử lý hình ảnh chứa các đối tượng động vật như bò, dê, ngựa,... Các hình ảnh này có thể đến từ nhiều nguồn khác nhau, chẳng hạn như máy ảnh của người dùng, camera an ninh, hoặc các kho dữ liệu hình ảnh công khai.

Các hình ảnh đầu vào có thể có độ phân giải và chất lượng khác nhau, từ những hình ảnh sắc nét, đã được chụp ở góc nhìn thuận lợi, đến những hình ảnh bị nhiễu, góc chụp khó khăn hoặc kích thước nhỏ. Hệ thống cần phải có khả năng xử lý được các trường hợp này một cách linh hoạt.

Ngoài ra, hình ảnh có thể bao gồm một hoặc nhiều đối tượng động vật trong cùng một khung hình. Mỗi đối tượng có thể có kích thước, vị trí và đặc điểm riêng biệt.

Kết quả xử lý được minh họa trong hình 2.2.



Hình 2.2. Hình ảnh kết quả đầu ra của hệ thống

Đầu ra sẽ là thông tin chi tiết về việc nhận dạng và phân loại các đối. Dựa trên các hình ảnh đầu vào, hệ thống sẽ cung cấp các thông tin sau:

Phát hiện các đối tượng động vật:

Xác định chính xác vị trí của từng cá thể động vật trong ảnh.

Sử dụng các khung chữ nhật hay các mặt nạ phân vùng để đánh dấu và tách biệt các đối tượng.

Nhận dạng và phân loại các đối tượng: Xác định loài động vật.

Phân loại từng cá thể thành các nhóm khác nhau, chẳng hạn như động vật ăn cỏ hay ăn thịt.

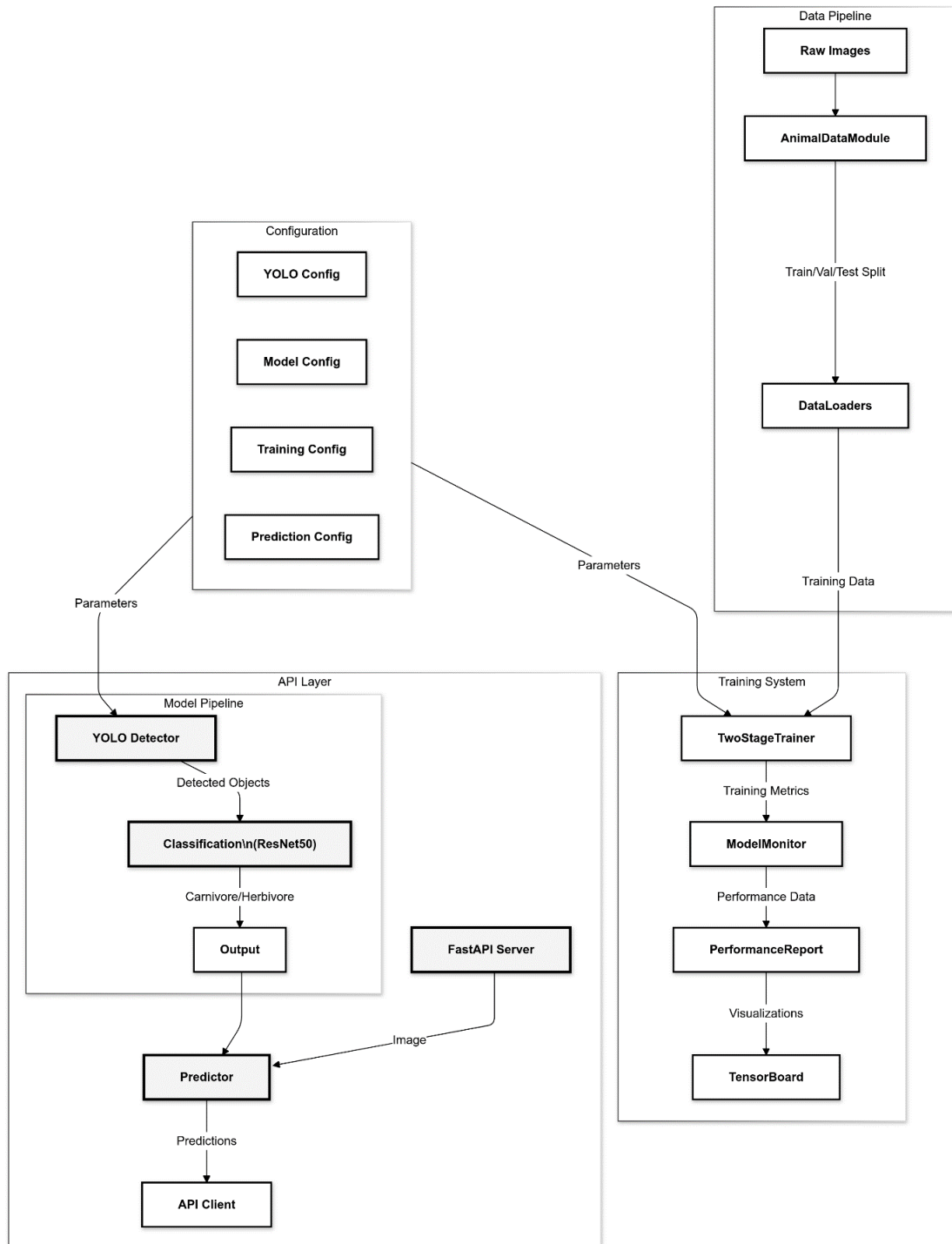
Thông tin bổ sung:

Điểm số tin cậy (confidence score) cho mỗi nhận dạng và phân loại, thể hiện mức độ chắc chắn của hệ thống.

2.2 Xây dựng hệ thống

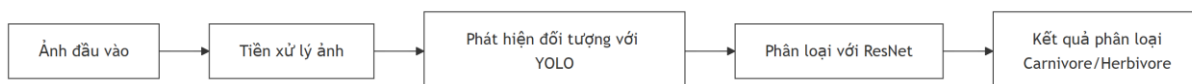
Dựa trên những tiến bộ gần đây trong lĩnh vực thị giác máy tính và học sâu, đặc biệt là các mô hình phát hiện đối tượng như YOLO (You Only Look Once) và các kiến trúc mạng phân loại tiên tiến như ResNet, nhóm đề xuất một hệ thống hai giai đoạn để giải quyết bài toán này.

Chi tiết mô hình hệ thống được thể hiện ở hình 2.3 dưới đây.



Hình 2.3. Mô hình hệ thống

Hệ thống được thiết kế theo luồng xử lý tuần tự như minh họa trong hình 2.4:



Hình 2.4. Luồng xử lý

Luồng xử lý của hệ thống bao gồm các bước chính sau:

a) Tiền xử lý ảnh:

- Chuẩn hóa kích thước ảnh.
- Chuẩn hóa giá trị pixel.
- Chuyển đổi không gian màu.

b) Phát hiện đối tượng với YOLO:

- Áp dụng mô hình YOLO để phát hiện vị trí đối tượng.
- Xác định bounding box chứa đối tượng.
- Tính toán độ tin cậy của phát hiện.

c) Phân loại với ResNet:

- Trích xuất vùng ảnh chứa đối tượng.
- Áp dụng mô hình ResNet để phân loại.
- Xác định loại động vật (ăn thịt/ăn cỏ).

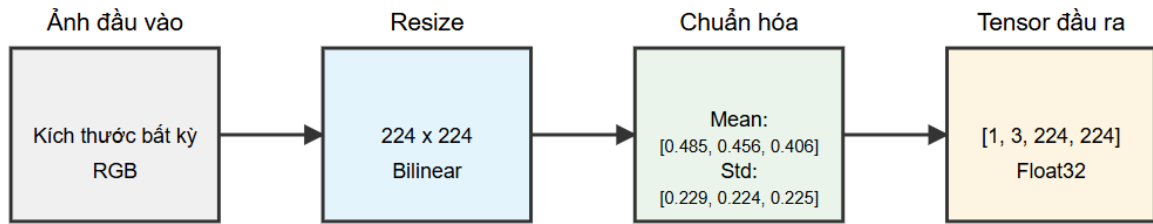
d) Kết quả cuối cùng:

- Tổng hợp kết quả phát hiện và phân loại.
- Tính toán độ tin cậy tổng thể.
- Trả về kết quả phân loại cuối cùng.

2.1.1 Chi tiết các thành phần

a) Module tiền xử lý đầu vào

Module tiền xử lý đầu vào đóng vai trò quan trọng trong việc chuẩn bị dữ liệu ảnh trước khi đưa vào mô hình học sâu. Module này thực hiện một chuỗi các phép biến đổi để đảm bảo ảnh đầu vào có định dạng phù hợp và được chuẩn hóa theo yêu cầu của mô hình. Quy trình xử lý được minh họa trong hình 2.5:



Hình 2.5. Module tiền xử lý ảnh đầu vào

Module thực hiện 3 phép biến đổi chính:

Thay đổi kích thước ảnh Ảnh đầu vào được thay đổi kích thước thành 224×224 pixels thông qua phép nội suy song tuyến. Với mỗi pixel (x,y) trong ảnh đích, giá trị được tính:

$$O(x, y) = \sum_{i,j \in R} I(i, j) \cdot k(x - i, y - j)$$

Trong đó:

- $O(x,y)$: Giá trị pixel tại vị trí (x,y) trong ảnh đích
- $I(i,j)$: Giá trị pixel tại vị trí (i,j) trong ảnh gốc
- $k()$: Hàm nội suy song tuyến
- R : Vùng 2×2 pixel lân cận trong ảnh gốc

Chuẩn hóa giá trị pixel Giá trị pixel được chuẩn hóa qua hai bước:

Chuẩn hóa về đoạn [0,1]:

$$X_{norm} = \frac{X}{255}$$

Chuẩn hóa theo kênh màu:

$$X_{final}^c = \frac{X_{norm}^c - \mu_c}{\sigma_c}$$

Với các tham số:

$c \in \{R,G,B\}$: Kênh màu

$\mu = [0.485, 0.456, 0.406]$: Giá trị trung bình

$\sigma = [0.229, 0.224, 0.225]$: Độ lệch chuẩn

Chuyển đổi định dạng Tensor đầu ra có dạng:

$$T \in \mathbb{R}^{C \times H \times W}$$

Trong đó:

- $C = 3$: Số kênh màu RGB
- $H = W = 224$: Chiều cao và rộng của ảnh

Kết quả là tensor 3 chiều với:

- Dữ liệu kiểu float32
- Kích thước $[3, 224, 224]$
- Giá trị đã được chuẩn hóa cho mỗi kênh màu

Các phép biến đổi này giúp:

Đồng nhất kích thước dữ liệu đầu vào

Giảm thiểu ảnh hưởng của độ sáng và độ tương phản

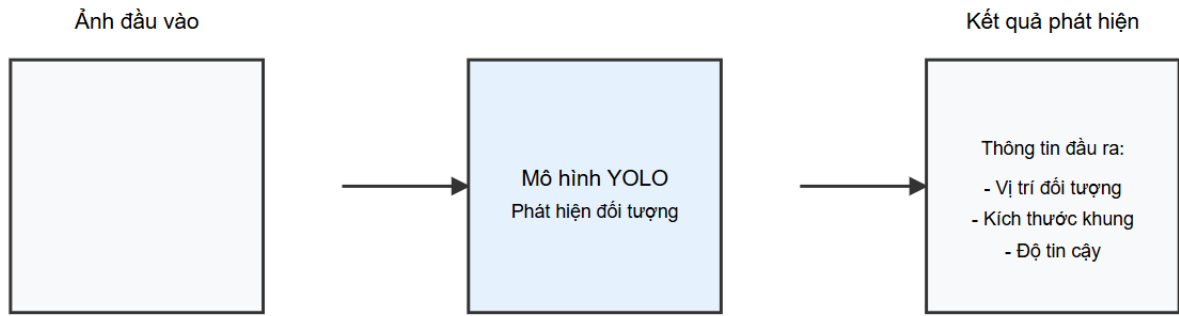
Tối ưu hóa cho quá trình học của mạng neural

Tương thích với pre-trained weights từ ImageNet

Quản lý hiệu quả bộ nhớ trong quá trình xử lý

b) Module phát hiện đối tượng với YOLO

Module phát hiện đối tượng sử dụng mô hình YOLO (You Only Look Once) để xác định vị trí của động vật trong ảnh. YOLO được lựa chọn vì khả năng phát hiện nhanh và chính xác các đối tượng trong ảnh. Module này chỉ tập trung vào nhiệm vụ xác định vùng chứa động vật (bounding box) và độ tin cậy của phát hiện, tạo tiền đề cho bước phân loại động vật ở giai đoạn sau. Quy trình phát hiện được minh họa trong hình 2.6:



Hình 2.6. Module phát hiện đối tượng với Yolo

Quy trình phát hiện được thực hiện qua các bước sau:

Xử lý ảnh đầu vào Ảnh sau khi được tiền xử lý sẽ được đưa qua mô hình YOLO. Mỗi ảnh đầu vào I có kích thước:

$$I \in \mathbb{R}^{H \times W \times 3}$$

Xác định vị trí đối tượng Mô hình YOLO dự đoán các bounding box B cho mỗi đối tượng, mỗi box được biểu diễn:

$$B = (x, y, w, h)$$

Trong đó:

(x, y) : Tọa độ tâm của box

(w, h) : Chiều rộng và chiều cao của box

Tính độ tin cậy Mỗi bounding box có một điểm tin cậy c được tính:

$$c = Pr(\text{Object}) \times \text{IOU}_{pred}^{truth}$$

Loại bỏ các box trùng lặp Áp dụng Non-Maximum Suppression (NMS) để loại bỏ các box chồng chéo bằng cách tính IOU:

$$\text{IOU}(B_i, B_j) = \frac{\text{area}(B_i \cap B_j)}{\text{area}(B_i \cup B_j)}$$

Box sẽ bị loại bỏ nếu:

$$IOU(B_i, B_j) > 0.45 \text{ và } c_i < c_j$$

Kết quả đầu ra Với mỗi ảnh, module trả về tập hợp các phát hiện:

$$D = \{(B_k, c_k) | k \in [1, N], c_k > 0.25\}$$

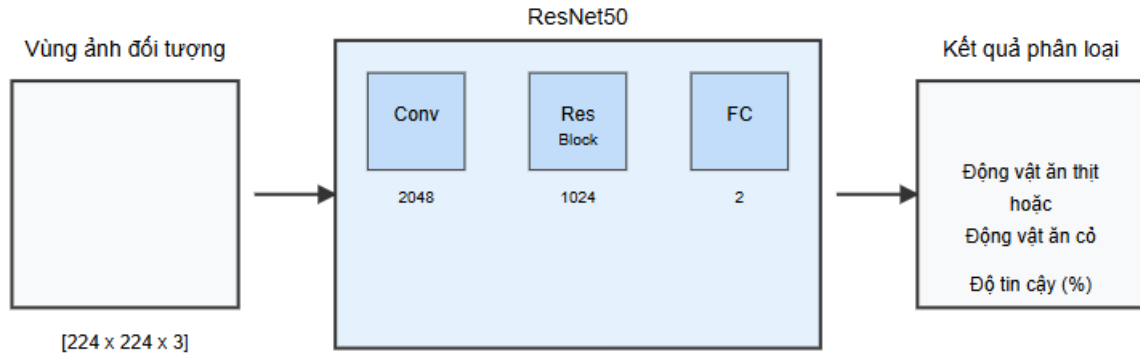
Trong đó:

N: Số đối tượng được phát hiện

0.25: Ngưỡng độ tin cậy tối thiểu

c) Module phân loại với ResNet50

Module phân loại sử dụng mạng neural tích chập ResNet50 để phân loại động vật thành hai nhóm: động vật ăn thịt và động vật ăn cỏ. ResNet50 được chọn vì khả năng trích xuất đặc trưng mạnh mẽ và hiệu quả đã được chứng minh trên nhiều bài toán phân loại ảnh. Quy trình phân loại được minh họa trong hình 2.7:



Hình 2.7. Module phân loại với ResNet50

Quy trình phân loại được thực hiện qua các bước sau:

Trích xuất đặc trưng ResNet50 chuyển đổi ảnh đầu vào thành vector đặc trưng F:

$$F = f_{ResNet}(I) \in \mathbb{R}^{2048}$$

Phân loại nhị phân Vector đặc trưng được đưa qua lớp fully connected cuối để dự đoán xác suất hai lớp:

$$p = \text{softmax}(Wf + b)$$

Trong đó:

$p = [p_1, p_2]$: Xác suất cho hai lớp (ăn thịt/ăn cỏ)

$W \in \mathbb{R}^{2 \times 2048}$: Ma trận trọng số

$b \in \mathbb{R}^2$: Vector bias

Kết quả đầu ra Cho mỗi ảnh, module trả về:

$$C = (\text{class}, \text{confidence})$$

Với:

class: Nhãn lớp được dự đoán (ăn thịt/ăn cỏ)

confidence: Độ tin cậy của dự đoán (0-1)

ResNet50 được huấn luyện transfer learning với:

Backbone đã được pre-train trên ImageNet

Fine-tuning lớp fully connected cuối

Cross-entropy loss cho bài toán phân loại nhị phân

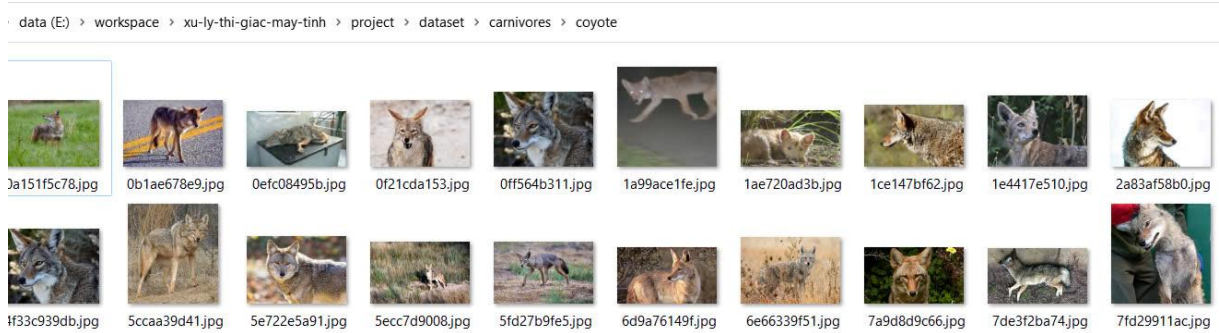
Chương 3: Kết quả thực nghiệm

3.1 Dữ liệu

3.1.1 Nguồn dữ liệu

Trong nghiên cứu này, nhóm sử dụng bộ dữ liệu hình ảnh động vật từ Kaggle (Animal Image Dataset - 90 Different Animals). Bộ dữ liệu này bao gồm:

- Tổng số 5,400 hình ảnh
- 90 loài động vật khác nhau
- Mỗi loài có 60 hình ảnh đại diện
- Đường dẫn nguồn: <https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals/data>



Hình 3.1. Hình ảnh dữ liệu thô được thu thập chưa qua xử lý

Bộ dữ liệu này được chọn vì:

- Tính đa dạng cao về loài và môi trường sống
- Chất lượng hình ảnh tốt và đồng đều
- Số lượng mẫu cân bằng giữa các loài
- Phù hợp với mục tiêu phân loại động vật ăn thịt và ăn cỏ

3.1.2 Tiền xử lý dữ liệu

Để chuẩn bị dữ liệu cho quá trình huấn luyện, nhóm đã thực hiện các bước tiền xử lý sau:

a), Chuẩn hóa kích thước ảnh:

- Resize tất cả ảnh về kích thước thống nhất 224x224 pixels
- Chuyển đổi không gian màu về RGB
- Chuẩn hóa giá trị pixel về khoảng [0,1]

b), Gán nhãn dữ liệu:

- Phân loại lại 90 loài động vật thành 2 nhóm: ăn thịt và ăn cỏ

- Tạo one-hot encoding cho nhãn phân loại
- Kiểm tra và đảm bảo tính cân bằng giữa các lớp

c), Augmentation dữ liệu:

- Xoay ảnh ngẫu nhiên trong khoảng $[-20^\circ, 20^\circ]$
- Lật ngang ảnh với xác suất 0.5
- Điều chỉnh độ sáng ngẫu nhiên $\pm 20\%$
- Thêm nhiễu Gaussian với độ lệch chuẩn 0.01
- Cắt ngẫu nhiên (random crop) với tỷ lệ giữ lại 0.8-1.0

3.1.3 Phân chia dữ liệu

Bộ dữ liệu được chia thành 3 tập như sau:

- Training set: 70% (3,780 ảnh)
- Validation set: 15% (810 ảnh)
- Test set: 15% (810 ảnh)

Quá trình phân chia được thực hiện với các đặc điểm:

- Sử dụng stratified sampling để đảm bảo tỷ lệ các lớp được giữ nguyên
- Phân chia ngẫu nhiên nhưng cố định random seed để đảm bảo tính tái lập
- Kiểm tra không có sự trùng lặp giữa các tập

3.2 Độ đo đánh giá

3.2.1 Độ chính xác (Accuracy)

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

- TP (True Positive): Số động vật ăn thịt được phân loại đúng
- TN (True Negative): Số động vật ăn cỏ được phân loại đúng
- FP (False Positive): Số động vật ăn cỏ bị phân loại nhầm thành ăn thịt
- FN (False Negative): Số động vật ăn thịt bị phân loại nhầm thành ăn cỏ

3.2.2 Độ chính xác theo lớp (Precision)

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

- Đánh giá khả năng phân loại chính xác động vật ăn thịt
- Giảm thiểu trường hợp phân loại nhầm động vật ăn cỏ thành ăn thịt

3.2.3 Thời gian xử lý

- Thời gian huấn luyện trung bình trên mỗi epoch
- Thời gian dự đoán trên một ảnh đầu vào
- Thời gian xử lý batch trong quá trình inference

3.3 Kết quả thực nghiệm

3.3.1 Kết quả huấn luyện mô hình

Việc huấn luyện diễn ra trong vòng 2 tiếng và cho ra kết quả như Bảng 2 bên dưới:

Bảng 3.1 Kết quả huấn luyện mô hình

Chu kỳ	Độ lỗi huấn luyện	Độ chính xác huấn luyện	Độ lỗi kiểm định	Độ chính xác kiểm định
1	0.701	66.52%	0.520	76.00%
3	0.447	81.80%	0.628	75.00%
6	0.289	88.25%	0.820	71.00%
8	0.119	96.07%	0.472	84.00%
10	0.080	97.28%	0.588	84.00%
13	0.039	98.30%	0.926	81.00%

Thông số training Model:

Đầu ra: Model được training cho ra kết quả ước tính với các trọng số như:

Loss: 0.0393

Accuracy: 98.30%

3.3.2 Kết quả xây dựng mô hình



**TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ ĐÔNG Á**
EAST ASIA UNIVERSITY OF TECHNOLOGY

Hệ Thống Nhận Dạng và Phân Loại Động Vật

Về Dự Án

Dự án này tập trung vào việc xây dựng một hệ thống trí tuệ nhân tạo có khả năng nhận dạng và phân loại động vật thành hai nhóm chính: động vật ăn thịt (carnivore) và động vật ăn cỏ (herbivore) từ hình ảnh. Hệ thống sử dụng các kỹ thuật deep learning tiên tiến để đạt độ chính xác cao trong việc phân loại.

Nhóm Thực Hiện

Giảng viên hướng dẫn:
Lương Thị Hồng Lan

Sinh viên thực hiện:

- Vũ Quang Huy - 20210611
- Lò Tiến Anh - 20210526

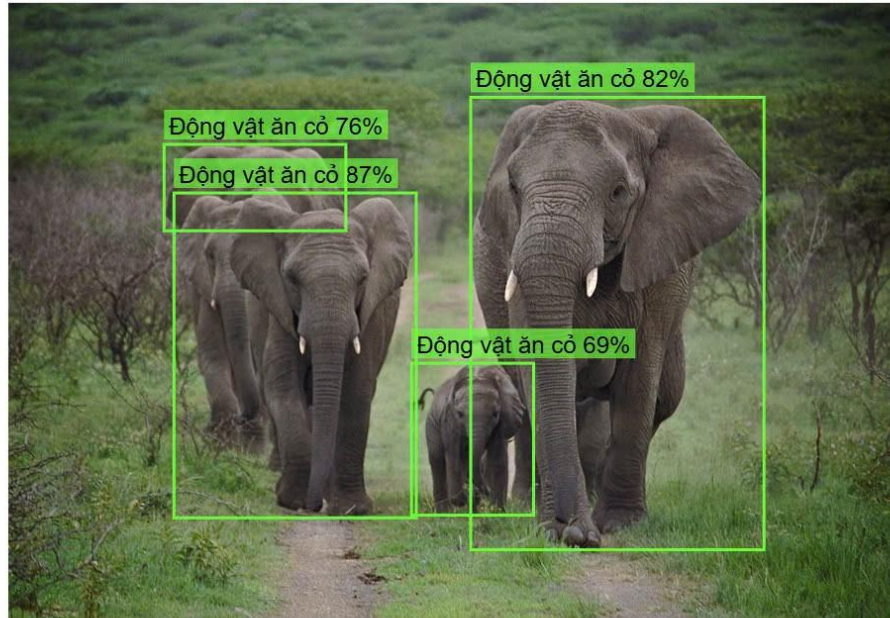
Tính năng chính:

- Nhận dạng động vật trong ảnh với độ chính xác cao
- Phân loại thành 2 nhóm: động vật ăn thịt và ăn cỏ
- Hiển thị vị trí của động vật trong ảnh (bounding box)
- Cung cấp độ tin cậy của kết quả phân loại

Hình 3.2. Phát triển ứng dụng trên nền tảng website

Demo Hệ Thống

Tải lên hình ảnh động vật và để AI phân loại giúp bạn



Hình 3.3. Kết quả đầu ra của hệ thống

Kết quả phân tích

Loại động vật Động Vật Ăn Cỏ	Độ tin cậy phát hiện 93.90%
Độ tin cậy phân loại 82.37%	Vị trí (bbox) [467, 95, 763, 552]
Loại động vật Động Vật Ăn Cỏ	Độ tin cậy phát hiện 90.11%
Độ tin cậy phân loại 68.99%	Vị trí (bbox) [407, 364, 530, 518]
Loại động vật Động Vật Ăn Cỏ	Độ tin cậy phát hiện 89.21%
Độ tin cậy phân loại 86.66%	Vị trí (bbox) [167, 192, 412, 520]
Loại động vật Động Vật Ăn Cỏ	Độ tin cậy phát hiện 77.31%
Độ tin cậy phân loại 76.33%	Vị trí (bbox) [157, 143, 341, 230]

Hình 3.4. Kết quả phân tích đầu ra của hệ thống

Kết luận

Tóm tắt nội dung trong đề tài

Trong quá trình thực hiện đề tài "Xây dựng hệ thống nhận dạng và phân loại động vật ăn thịt và ăn cỏ trong ảnh", nhóm đã nỗ lực và đạt được các mục tiêu chính sau:

Xây dựng thành công một hệ thống hoàn chỉnh có khả năng phát hiện và phân loại động vật trong ảnh thành hai nhóm: động vật ăn thịt và động vật ăn cỏ

Thực hiện tích hợp các kỹ thuật deep learning hiện đại như YOLO và CNN để nâng cao hiệu quả nhận dạng

Hoàn thiện pipeline xử lý từ tiền xử lý ảnh đầu vào đến post-processing kết quả đầu ra

Phát triển thành công API để triển khai hệ thống một cách linh hoạt

Kết quả nhận được

Qua quá trình thực nghiệm, chúng tôi đã thu được những kết quả đáng khích lệ:

Hệ thống đạt được độ chính xác cao trong việc phân loại, với 84% trên tập kiểm định tốt nhất

Khả năng xử lý đa dạng điều kiện ảnh đầu vào như góc chụp, ánh sáng, kích thước

Tốc độ xử lý được tối ưu hóa, phù hợp cho ứng dụng thực tế

Hệ thống hoạt động ổn định và có tính mở rộng cao

Hướng phát triển

Dựa trên kết quả nghiên cứu và phát triển, nhóm đề xuất một số hướng phát triển tiếp theo:

Mở rộng khả năng phân loại:

Phát triển thêm các nhóm phân loại chi tiết hơn

Mở rộng khả năng nhận dạng nhiều loài động vật hơn

Cải thiện khả năng phân biệt các trường hợp khó

Tối ưu hóa hiệu năng:

Tiếp tục cải thiện tốc độ xử lý

Nghiên cứu giảm thiểu yêu cầu tài nguyên

Tối ưu hóa để chạy được trên thiết bị di động

Nâng cao độ chính xác:

Thu thập và bổ sung thêm dữ liệu huấn luyện đa dạng

Thử nghiệm các kiến trúc mạng neural mới

Phát triển thêm các kỹ thuật augmentation dữ liệu

Phát triển ứng dụng:

Phát triển phiên bản cho thiết bị di động

Nghiên cứu tích hợp với các hệ thống giám sát động vật

Danh mục tài liệu tham khảo

- [1]. Banerjee, S. (2022). Animal Image Dataset (90 Different Animals). Retrieved from Kaggle.com website: <https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals/data>
- [2]. Sách Deep Learning cơ bản | Deep Learning cơ bản. (2021, February 16). Retrieved December 9, 2024, from Deep Learning cơ bản website: <https://nttuan8.com/sach-deep-learning-co-ban/>
- [3]. Tài liệu lưu hành nội bộ trường Đại học Công Nghệ Đông Á học phần Xử lý ảnh và Thị giác máy tính
- [4]. Python Software Foundation. (2019). 3.7.3 Documentation. Retrieved from Python.org website: <https://docs.python.org/3/>
- [5]. PyTorch. (2019). PyTorch documentation — PyTorch master documentation. Retrieved from Pytorch.org website: <https://pytorch.org/docs/stable/index.html>
- [6]. Jocher, G. (2020, May 18). YOLOv8 Documentation. Retrieved from docs.ultralytics.com website: <https://docs.ultralytics.com/>