

ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO ĐỒ ÁN LẬP TRÌNH TRÍ TUỆ NHÂN TẠO NÂNG CAO

Cao học khóa 2022

Đề tài:

Phân Loại Danh Lam Thắng Cảnh

Giảng viên:

PGS.TS. Trần Minh Triết
TS. Nguyễn Tiến Huy
TS. Lê Trung Nghĩa
TS. Lê Thanh Tùng

Thực hiện đề tài:

22C15018: Phạm Minh Thạch
22C15023: Nguyễn Thanh Tùng

TPHCM - 4/2023

MỤC LỤC

| | |
|---|----------|
| GIỚI THIỆU..... | 3 |
| 1. PHÂN LOẠI BẰNG SUPPORT VECTOR MACHINE..... | 3 |
| 1.1. Giới thiệu | 3 |
| 1.2. Tiền xử lý dữ liệu | 3 |
| 1.3 Huấn luyện mô hình | 3 |
| 1.4 Đánh giá mô hình..... | 4 |
| 2. PHÂN LOẠI BẰNG DEEP LEARNING..... | 4 |
| 2.1. Giới thiệu | 4 |
| 2.2 Xây dựng mô hình thứ nhất | 4 |
| 2.2.1 Ý tưởng và thư viện..... | 4 |
| 2.2.2 Tiền xử lý | 4 |
| 2.2.3 Huấn luyện mô hình..... | 4 |
| 2.2.4 Đánh giá mô hình | 5 |
| 2.3 Xây dựng mô hình thứ hai | 5 |
| 2.3.1 Ý tưởng và thư viện..... | 5 |
| 2.3.2 Tiền xử lý | 5 |
| 2.3.3 Huấn luyện mô hình..... | 5 |
| 2.3.4 Đánh giá mô hình | 5 |
| 3. SO SÁNH KẾT QUẢ CỦA SVM VÀ DEEP LEARNING..... | 5 |

GIỚI THIỆU

Xin chào quý thầy cô, ở đồ án này, nhóm sẽ thực hiện đề tài phân loại danh lam thắng cảnh bằng hai phương pháp Support Vector Machine (SVM) và Deep Learning.

Mã nguồn đồ án được lưu trữ tại github.com/thachmphan/landmark-classification và được tổ chức với cấu trúc như sau:

- *src*
 - *build_svm_model.ipynb* : build SVM model.
 - *build_deep_model_01.ipynb* : build deep learning model thứ nhất.
 - *build_deep_model_02.ipynb* : build deep learning model thứ hai.
- *report*
 - *report.pdf* : file báo cáo
 - *report.docx*
- *README.md* : hướng dẫn cài đặt môi trường và chạy chương trình.

Tập data để huấn luyện và kiểm tra được lấy từ: drive.google.com/file/d/1IASIW-wSzG4LtXai0vyUVEKTvgfqlhFL/view?usp=sharing.

Sau đây là bảng phân công công việc của các thành viên trong nhóm.

| | |
|---|------------------------------------|
| Phân loại bằng Support Vector Machine | Phạm Minh Thạch |
| Phân loại bằng mô hình Deep Learning 1 | Phạm Minh Thạch |
| Phân loại bằng mô hình Deep Learning 2 | Nguyễn Thanh Tùng |
| So sánh Support Vector Machine và Deep Learning | Nguyễn Thanh Tùng, Phạm Minh Thạch |

1. PHÂN LOẠI BẰNG SUPPORT VECTOR MACHINE

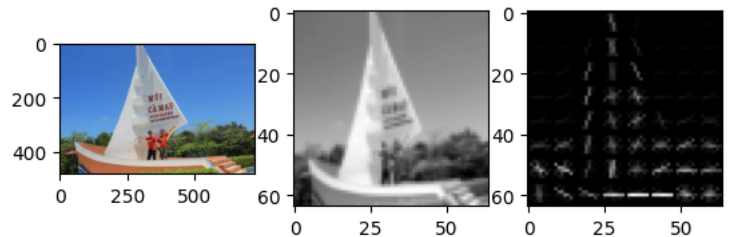
1.1. Giới thiệu

Support Vector Machine (SVM) là một thuật toán học máy có giám sát được sử dụng rất phổ biến ngày nay trong các bài toán phân lớp. Ý tưởng của SVM là tìm ra các đường kẻ để phân tách các điểm dữ liệu. Mỗi đường kẻ này sẽ chia không gian thành các miền khác nhau và mỗi miền sẽ chứa một loại dữ liệu.

1.2. Tiền xử lý dữ liệu

Ở bước tiền xử lý, ta kết hợp phương pháp Histogram of Object Gradient và Standard Scale. Trong đó, Histogram of Object Gradient là phương pháp giúp ta lấy ra các đường nét chính của bức ảnh (Hình 1). Standard Scale được sử dụng để chuẩn hoá dữ liệu, sau khi chuẩn hoá giá trị trung bình của dữ liệu là 0. Quá trình này giúp ta làm giảm được kích thước của dữ liệu đầu vào, đồng thời đưa dữ liệu về khoản thuận lợi cho việc tính toán của các thuật toán.

Input gồm 1500 ảnh, thời gian để load ảnh vào dataset là 2m33s.



Hình 1. Minh họa Histogram of Object Gradient

1.3 Huấn luyện mô hình

Trên tập huấn luyện gồm 1500 ảnh và 30 nhãn, thời gian huấn luyện là 5s.

1.4 Đánh giá mô hình

Trên tập kiểm tra gồm 1500 ảnh và 30 nhãn, thời gian để đánh giá là 4s, độ chính xác accuracy score là 59.47%.

2. PHÂN LOẠI BẰNG DEEP LEARNING

2.1. Giới thiệu

Các mạng Deep Neural Network bao gồm nhiều lớp neuron khác nhau, có khả năng thực hiện các tính toán có độ phức tạp rất cao. Deep Learning hiện đang phát triển rất nhanh và được xem là một trong những bước đột phá lớn nhất trong Machine Learning. Trong phần này, chúng ta sẽ xây dựng hai network khác nhau để thực hiện công việc phân loại danh lam thắng cảnh.

2.2 Xây dựng mô hình thứ nhất

2.2.1 Ý tưởng và thư viện

Mô hình đầu tiên được cài đặt với thư viện Pytorch.

2.2.2 Tiền xử lý

Ở bước tiền xử lý, ta sẽ chuẩn hoá dữ liệu theo giá trị trung bình và độ lệch chuẩn, để đưa dữ liệu về phân bố xung quanh điểm trung bình là 0. Việc này sẽ giúp ích cho việc tính toán của các thuật toán. Ảnh cũng được scale về kích thước 128x128. Thời gian thực hiện thao tác tiền xử lý là 46s.

2.2.3 Huấn luyện mô hình

Đầu vào của mô hình là một ảnh 3x128x128 (channel x width x height). Mô hình gồm 6 lớp Convolution và 2 lớp Linear. Các lớp Convolution được kết hợp với Max Pool, ReLu và Batch Norm để rút trích đặc trưng. Kết quả của lớp Linear được kết hợp với hàm Log Soft Max. Bảng sau cho ta tóm tắt về kiến trúc của model.

| Layer (type) | Output Shape |
|----------------|--------------------|
| Conv2d-1 | [-1, 16, 128, 128] |
| ReLU-2 | [-1, 16, 128, 128] |
| BatchNorm2d-3 | [-1, 16, 128, 128] |
| MaxPool2d-4 | [-1, 16, 64, 64] |
| Conv2d-5 | [-1, 32, 64, 64] |
| ReLU-6 | [-1, 32, 64, 64] |
| BatchNorm2d-7 | [-1, 32, 64, 64] |
| MaxPool2d-8 | [-1, 32, 32, 32] |
| Conv2d-9 | [-1, 64, 32, 32] |
| ReLU-10 | [-1, 64, 32, 32] |
| BatchNorm2d-11 | [-1, 64, 32, 32] |
| MaxPool2d-12 | [-1, 64, 16, 16] |
| Conv2d-13 | [-1, 128, 16, 16] |
| ReLU-14 | [-1, 128, 16, 16] |
| BatchNorm2d-15 | [-1, 128, 16, 16] |
| MaxPool2d-16 | [-1, 128, 8, 8] |
| Conv2d-17 | [-1, 256, 8, 8] |
| ReLU-18 | [-1, 256, 8, 8] |
| BatchNorm2d-19 | [-1, 256, 8, 8] |

| | |
|--------------|-----------------|
| MaxPool2d-20 | [-1, 256, 4, 4] |
| Flatten-21 | [-1, 4096] |
| Linear-22 | [-1, 1024] |

Mô hình được huấn luyện trên 20 epoch, với hàm loss được sử dụng là Negative Like và phương pháp được sử dụng để tối ưu bộ tham số là Adam Stochastic Optimization với learning rate là 0.001.

Thời gian thực hiện việc training là 18 phút 20 giây.

2.2.4 Đánh giá mô hình

Việc đánh giá mô hình được thực hiện trên tập gồm 1500 ảnh, tỉ lệ dự đoán chính xác là 44%, thời gian thực hiện là 41.5 giây.

2.3 Xây dựng mô hình thứ hai

2.3.1 Ý tưởng và thư viện

Sử dụng Transfer Learning, với pre-trained model là Xception.

Sử dụng thư viện tensorflow

2.3.2 Tiền xử lý

Chia 80% tập train dùng để huấn luyện, 20% cho validation

Dùng các kỹ thuật image augmentations để tăng số lượng dữ liệu cho tập huấn luyện. Ví dụ xoay ảnh ngang, xoay ảnh dọc, phóng to, ...

2.3.3 Huấn luyện mô hình

Hình nhận ảnh với 3x224x224 (channel x width x height). Thêm 4 lớp mới lần lượt là 1 Global Average Pooling, 1 lớp dense, 1 lớp dropout và lớp đầu ra.

Mô hình được huấn luyện với 25 epoch, với hàm loss là cross entropy, dùng Adam để tối ưu tham số.

Thời gian huấn luyện là 25 phút cho 25 epoch, mỗi epoch tầm 1 phút trên Macbook M1.

2.3.4 Đánh giá mô hình

Độ chính xác trên tập test 1500 ảnh là 87.4%, thời gian thực hiện 1 phút 25 giây.

Bởi vì mô hình có kích thước vượt quá giới hạn cho phép của github, nên ta lưu trữ tại link bên dưới:

drive.google.com/file/d/1MLeUKrWHRUchsqIM85KMw2bcIUg5w6T/view?usp=share_link

3. SO SÁNH KẾT QUẢ CỦA SVM VÀ DEEP LEARNING

Mô hình SVM có thời gian huấn luyện và thời gian đánh giá nhanh hơn mô hình Deep Learning. Tuy nhiên, mô hình Deep Learning thứ hai lại cho kết quả dự đoán có độ chính xác cao hơn so với mô hình SVM. Bảng sau đây thể hiện sự khác biệt giữa SVM và hai mô hình Deep Learning.

| | Mô hình SVM | Mô hình Deep Learning 1 | Mô hình Deep Learning 2 (Transfer Learning) |
|-------------------------|-------------|-------------------------|--|
| Thời gian huấn luyện | 5 giây | 18 phút 20 giây | 25 phút |
| Thời gian đánh giá | 4 giây | 41.5 giây | 1 phút 25 giây |
| Tỉ lệ dự đoán chính xác | 59.47% | 44% | 87.4% |

Với mô hình Deep Learning sử dụng Transfer Learning đạt được kết quả cao là do nó tận dụng được kiến thức được tìm ra bởi mô hình pre-trained (được huấn luyện trên một tập dữ liệu rất lớn, rất nhiều layer)