

INT3404E 20 - Image Processing

Planning and Proposal Development

Group 15

1 Thành viên nhóm

- **Ngô Tùng Lâm** MSV: 22028092 - Nhóm trưởng
- **Lương Thị Linh** MSV: 22028202 - Thành viên
- **Nguyễn Văn Sơn** MSV: 22028020 - Thành viên
- **Nguyễn Thị Thu Trang** MSV: 22028254 - Thành viên

2 Tổng quan bài toán

Bài toán cần giải quyết là phân loại kí tự Nôm, thuộc nhóm bài toán Image Classification (Phân loại ảnh). Image Classification là một nhiệm vụ phổ biến trong Computer Vision. Có tổng cộng 2138 kí tự Nôm tất cả. Và mục tiêu chính của bài toán này đó chính là phân loại một hình ảnh kí tự Nôm đầu vào (input) thành một nhãn (label) đầu ra (output).

3 Nghiên cứu các phương pháp

3.1 Phương pháp tiền xử lý dữ liệu (Data Preprocessing)

Các phương pháp thu thập dữ liệu thường không được kiểm soát chặt chẽ, dẫn đến các giá trị ngoại lệ và thiếu sót. **Tiền xử lý dữ liệu(Data preprocessing)** là quá trình làm sạch, chuẩn hóa và xử lý giá trị thiếu để cải thiện chất lượng dữ liệu trước khi sử dụng trong các bước xử lý chính.

Tuy nhiên, khi dữ liệu mất cân bằng, tiền xử lý truyền thống thường gặp khó khăn. Trong trường hợp này, **tăng cường dữ liệu(Data augmentation)** là một giải pháp phụ trợ hiệu quả. Phương pháp này tạo ra các bản sao biến đổi của dữ liệu hiện có bằng cách sử dụng các kỹ thuật như *color jitter*, *random crop*, *random grayscale*, *Gaussian blur*, *random rotation*, *horizontal and vertical flip*. Với bài toán này, chúng ta có thể sử dụng 3 kỹ thuật được đề cập ở trên là *color jitter*, *random grayscale* và *random rotation*.

3.2 CNN (Convolutional Neural Network)

CNNs gồm 1 lớp input, 1 lớp output, kèm theo các lớp ẩn bao gồm lớp convolutional, pooling, fully connected và normalization. CNNs xử lý hình ảnh gốc trực tiếp mà không cần tiền xử lý phức tạp. Các lớp của CNN có các neuron sắp xếp theo 3 chiều: chiều rộng, chiều cao và chiều sâu, giúp giảm hình ảnh thành một vector điểm số lớp.

Trong nhận dạng đối tượng hình ảnh, CNNs thường đạt hiệu suất tốt và bảo toàn mối quan hệ hàng xóm và tính cục bộ không gian, phát triển tốt với hình ảnh có kích thước thực tế và không phụ thuộc vào số chiều đầu vào.

3.3 Các pre-model cho bài toán phân loại ảnh (Image Classification)

Trong trường hợp bộ dữ liệu có kích thước quá nhỏ và khó có thể tìm kiếm và mở rộng thêm, việc sử dụng các pretrained-model với cùng tác vụ phân loại sẽ giúp các mô hình được huấn luyện dự báo tốt hơn. Dưới đây là các pretrained-model nổi tiếng trong nhiệm vụ phân loại ảnh:

- **VGG** : được phát triển bởi Karen Simonyan và Andrew Zisserman thuộc Đại học Oxford vào năm 2014, với kiến trúc đơn giản nhưng hiệu quả cao trong các nhiệm vụ nhận dạng hình ảnh, đặc biệt là phân loại hình ảnh. Tuy nhiên, kích thước của mô hình tương đối lớn, yêu cầu nhiều tài nguyên tính toán, đem lại hiệu suất không tốt trong một số nhiệm vụ nhất định
- **Inception** : Kiến trúc Inception của GoogLeNet được thiết kế để duy trì hiệu suất cao trong khi hoạt động dưới các ràng buộc nghiêm ngặt về bộ nhớ và ngân sách tính toán. Nó giảm đáng kể số lượng tham số so với các mô hình tiền nhiệm như AlexNet và VGGNet, làm cho nó trở nên khả thi hơn trong các kịch bản dữ liệu lớn hoặc trong các tình huống mà bộ nhớ hoặc khả năng tính toán có giới hạn.
- **ResNet** : được phát triển bởi nhóm nghiên cứu Kaiming He vào năm 2015, giúp giải quyết vấn đề "mất gradient" và đạt được độ chính xác cao trong các nhiệm vụ xử lý hình ảnh, đặc biệt là bài toán phân loại ảnh, tuy nhiên kiến trúc có độ phức tạp cao.
- **EfficientNet** : được giới thiệu vào năm 2019 bởi một nhóm các nhà nghiên cứu tại Google AI, được ứng dụng rộng rãi trong các lĩnh vực phân loại hình ảnh, nhận dạng vật thể, phân đoạn ảnh, xác định vị trí đối tượng và nhiều ứng dụng thị giác máy tính khác. Efficient Net sử dụng phương pháp tăng kích thước mạng mới gọi là Compound Scaling, mô hình này đạt được độ chính xác cao với số lượng tham số ít hơn so với các mô hình tương đương. Efficient Net cung cấp hiệu suất tính toán tốt và dễ dàng sử dụng. Tuy nhiên, cần phần cứng mạnh và tinh chỉnh cẩn thận cho các nhiệm vụ cụ thể.

4 Kế hoạch chung

- **Xử lý bộ dữ liệu** (Lương Thị Linh, Nguyễn Thị Thu Trang)
 - Tìm kiếm bộ dữ liệu bên ngoài
 - Phân tích bộ dữ liệu
 - Tăng cường dữ liệu
- **Xây dựng mô hình ban đầu** (Ngô Tùng Lâm, Nguyễn Văn Sơn)
 - Xây dựng mô hình CNN với các hàm mất mát khác nhau
 - Đánh giá độ chính xác của đầu ra
 - Phân tích, sửa lỗi
- **Thử nghiệm fine-tune với các pre-trained model, đánh giá độ hiệu quả** (Cả nhóm)
- **Viết báo cáo** (Lương Thị Linh, Nguyễn Thị Thu Trang)