

University of Engineering and Technology

VIETNAM NATIONAL UNIVERSITY



Midterm Project Image Processing

Class : INT3404E 20E

Group : 18

Members : Trịnh Công Trung

I) Giới thiệu

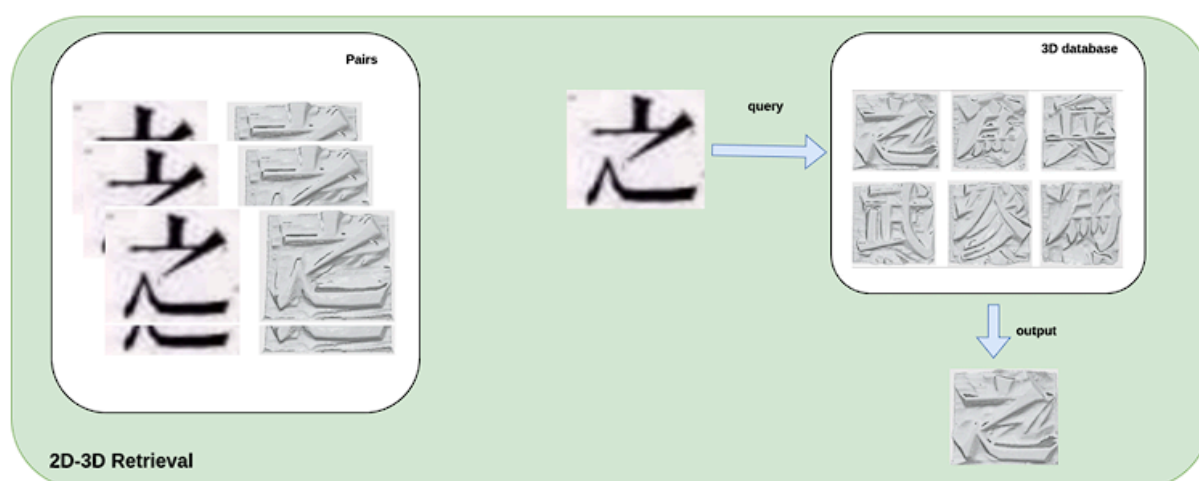
1.1. Chữ Nôm

Chữ Nôm (hay còn gọi là Hán-Nôm) là một hệ thống ký tự được người Việt tạo ra dựa trên chữ Hán. Sau khi Việt Nam độc lập khỏi Trung Quốc vào năm 939 CN, chữ Nôm được tạo ra trong khoảng thời gian này. Suốt một thiên niên kỷ sau đó, là khoảng thời gian từ thế kỷ 10 đến thế kỷ 20, chữ Nôm đã trở thành chữ viết chính thức của Việt Nam. Đã có rất nhiều tài liệu viết bằng chữ Hán-Nôm mô tả các khía cạnh về lịch sử, văn hóa, chính trị, văn học,... có giá trị.

Về sau, người ta tạo ra các mộc bản của chữ Nôm, bằng cách khắc ngược chữ Nôm vào bề mặt gỗ. Chúng được dùng để in lên sách tiếng Việt.

1.2. Bài toán

Báo cáo này tập trung vào việc phát triển một mô hình có thể xử lý, phân loại các hình ảnh chứa ký tự chữ Nôm. Mục tiêu của chúng ta là giúp mô hình có thể học cách ánh xạ các hình ảnh 2D in sang các hình ảnh khắc gỗ 3D tương ứng.



II) Dữ liệu

Bộ dữ liệu của bài tập này là 2D-3D Retrieval Dataset, bao gồm các hình ảnh ký tự Hán-Nôm được in (ảnh 2D, định dạng file .png) và các hình ảnh 3D được lưu dưới dạng file ảnh .stl

Trong mỗi hình ảnh chỉ thể hiện một ký tự Hán-Nôm riêng biệt. Như hình 1 dưới đây cho thấy 10 ảnh 2D kích thước 512x512 đầu tiên trong thư mục print:



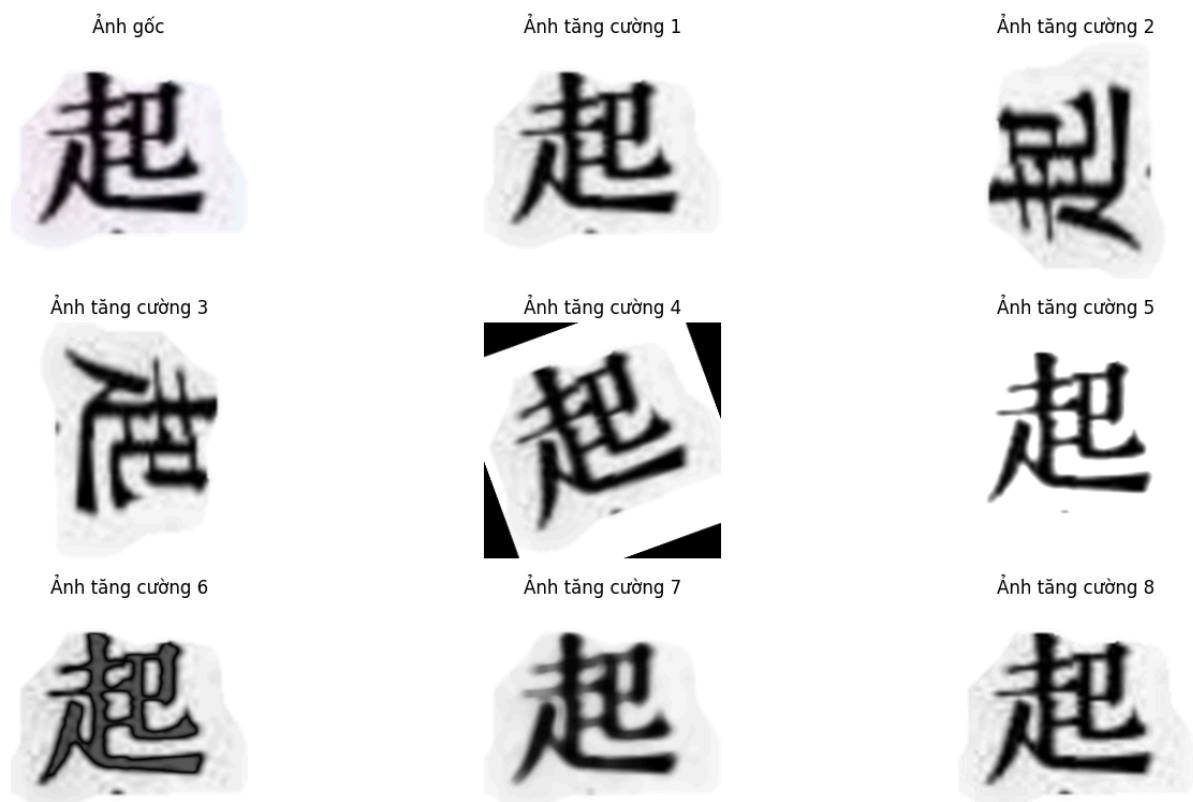
Hình 1

III) Phương pháp

3.1. Tiền xử lý dữ liệu

Tiền xử lý dữ liệu là quá trình chuẩn bị dữ liệu trước khi đưa vào mô hình. Ở đây, ta sử dụng các phép biến đổi để tăng cường dữ liệu ảnh 2D bao gồm: lật ảnh, xoay ảnh, tăng/giảm độ sáng, làm mờ/làm sắc nét ảnh.

Như hình dưới đây, ta đã xử lý ảnh 0.png theo các cách trên:



3.2. Chuẩn bị dữ liệu

Điều chỉnh kích thước ảnh thành 224x224, làm phẳng ảnh bằng OpenCv
Chia tập dữ liệu ra làm 2 phần: training set (80%) và validation set (20%)

```
#Chia dữ liệu thành tập huấn luyện và tập kiểm tra
X_train, X_test, y_train, y_test = train_test_split(features, labels, test_size=0.2, random_state=42)
```

3.3. Xây dựng mô hình

Mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) là một loại mạng nơ-ron sâu, được thiết kế đặc biệt để xử lý dữ liệu dạng lưới như hình ảnh. CNN sử dụng các phép toán tích chập để tự động và thích ứng học các đặc trưng từ dữ liệu đầu vào. Điều này giúp cho việc nhận dạng, phân tích hình ảnh trở nên hiệu quả hơn. CNN có khả năng học các đặc trưng ở nhiều cấp độ khác nhau, từ cục bộ đến toàn cục, qua nhiều lớp ẩn.

Mạng CNN có thể bao gồm nhiều lớp khác nhau, mỗi lớp có chức năng riêng:

- Lớp Tích Chập (Convolutional Layer): Đây là lớp thực hiện mọi tính toán và trích xuất các đặc trưng từ hình ảnh đầu vào. Lớp này sử dụng các bộ lọc (kernels) để duy trì mối quan hệ giữa các pixel và tìm hiểu các đặc trưng hình ảnh.
- Lớp Gộp (Pooling Layer): Lớp này giảm bớt số lượng tham số khi hình ảnh quá lớn, giúp giảm kích thước không gian mà vẫn giữ được các thông tin quan trọng.
- Lớp Kết Nối Đầy Đủ (Fully Connected Layer): Lớp này nhận đầu vào từ các lớp trước và thực hiện phân loại.

Kiến trúc mô hình

Layer (type)	Output Shape	Param #
conv2d_6 (Conv2D)	(None, 224, 224, 64)	320
max_pooling2d_6 (MaxPooling2D)	(None, 74, 74, 64)	0
conv2d_7 (Conv2D)	(None, 74, 74, 64)	16,448
max_pooling2d_7 (MaxPooling2D)	(None, 24, 24, 64)	0
conv2d_8 (Conv2D)	(None, 24, 24, 196)	50,372
max_pooling2d_8 (MaxPooling2D)	(None, 8, 8, 196)	0
dropout_2 (Dropout)	(None, 8, 8, 196)	0
flatten_2 (Flatten)	(None, 12544)	0
dense_2 (Dense)	(None, 228)	2,860,260

Total params: 2,927,400 (11.17 MB)
Trainable params: 2,927,400 (11.17 MB)
Non-trainable params: 0 (0.00 B)

Biên dịch mô hình: Mô hình được biên dịch với trình tối ưu hóa Adam, hàm mất mát (loss) 'sparse_categorical_crossentropy', được sử dụng cho các bài toán phân loại đa lớp, và độ chính xác (accuracy) được sử dụng làm số liệu đánh giá.

IV) Đánh giá

Ta thực hiện đào tạo mô hình với 15 epoch, vào epoch thứ 15, độ chính xác trên tập huấn luyện là accuracy=0.9; tập kiểm định là 0.644.

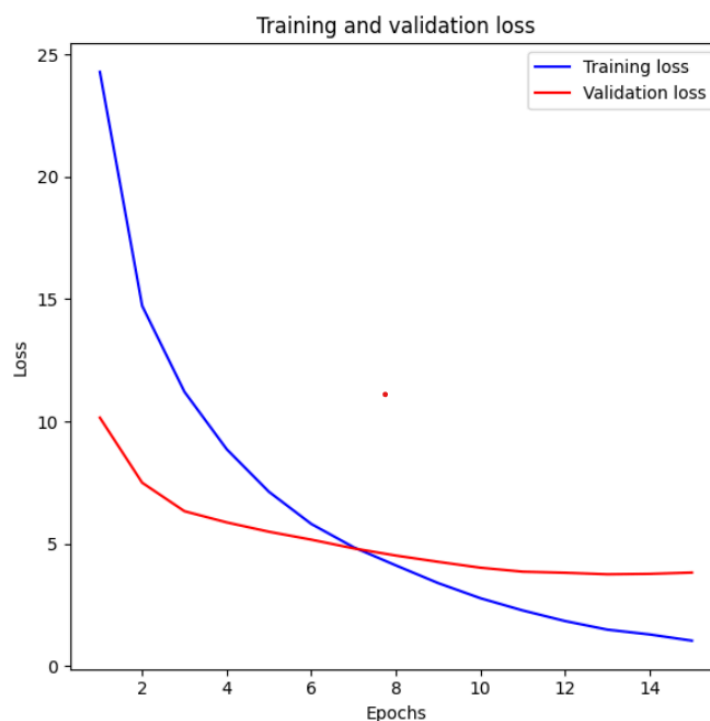
```
Epoch 12/15  
31/31 — 2s 49ms/step - accuracy: 0.7123 - loss: 1.3358 - val_accuracy: 0.6466 - val_loss: 3.8972  
Epoch 13/15  
31/31 — 2s 49ms/step - accuracy: 0.8007 - loss: 0.9852 - val_accuracy: 0.6466 - val_loss: 3.9649  
Epoch 14/15  
31/31 — 2s 49ms/step - accuracy: 0.8550 - loss: 0.7952 - val_accuracy: 0.6466 - val_loss: 4.0160  
Epoch 15/15  
31/31 — 2s 49ms/step - accuracy: 0.8955 - loss: 0.6534 - val_accuracy: 0.6438 - val_loss: 4.1764
```

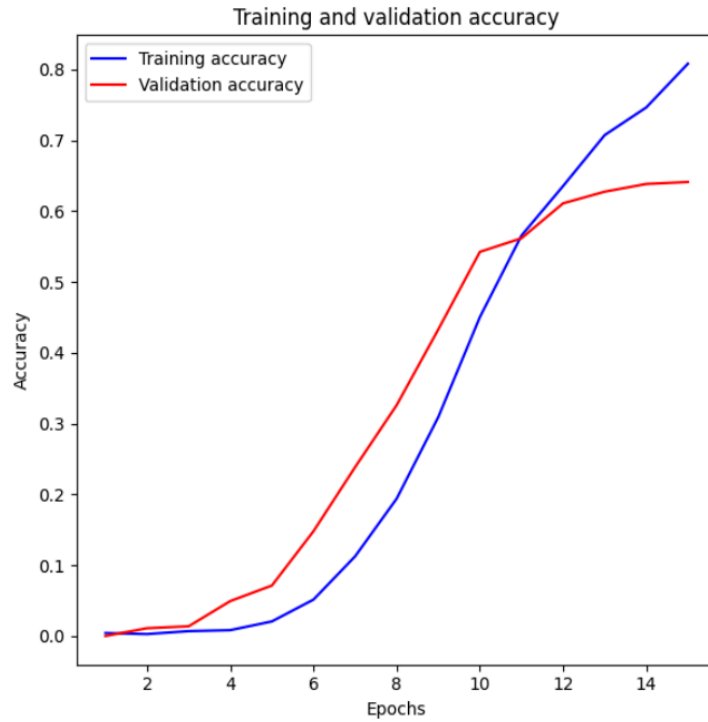
Đánh giá mô hình trên tập kiểm định:

```
test_loss, test_acc = model.evaluate(X_test, y_test)
```

- Accuracy = 0.64
- Loss = 4.16

Biểu đồ của accuracy và loss được thể hiện qua các tập dữ liệu huấn luyện và kiểm định như sau:





Nhận xét:

Mặc dù giải pháp này chỉ sử dụng mô hình CNN cơ bản với 3 lớp Conv2D, 3 lớp MaxPooling2D và 1 lớp Dropout nhưng đã cho ta thấy kết quả khá khả quan. Độ chính xác là khoảng 0.64 với hàm mất mát tương đối thấp là 4.16. Qua 2 biểu đồ trên thể hiện đường biểu diễn các giá trị accuracy, loss của tập đào tạo và tập kiểm định, mô hình không có dấu hiệu bị “overfitting” hay “underfitting”. Qua đó cho ta thấy CNN là mô hình đáng tin cậy đối với các vấn đề liên quan đến xử lý dữ liệu hình ảnh.