TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN HỌC PHẦN

PHẦN MỀM NHẬN DẠNG 5 ĐỐI TƯỢNG DỤNG CỤ HỌC TẬP

GIẢNG VIÊN HƯỚNG DẪN: ThS. Tôn Quang Toại

SINH VIÊN THỰC HIỆN:

Nguyễn Ngọc Anh Thiên – 21DH114521

Nguyễn Bảo Khang – 21DH110785

Lê Đăng Tùng – 21DH114280

TP Hồ Chí Minh, tháng 4 năm 2024

TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN HỌC PHẦN

PHẦN MỀM NHẬN DẠNG 5 ĐỐI TƯỢNG DỤNG CỤ HỌC TẬP

GIẢNG VIÊN HƯỚNG DẪN: ThS. Tôn Quang Toại

SINH VIÊN THỰC HIỆN:

Nguyễn Ngọc Anh Thiên – 21DH114521

Nguyễn Bảo Khang – 21DH110785

Lê Đăng Tùng – 21DH114280

TP Hồ Chí Minh, tháng 4 năm 2024

PHIẾU CHẤM ĐIỂM

Họ tên sinh viên 1 (SV1):_____ Mã SV: _____

Họ tên sinh viên 2 (SV2):_____ Mã SV: _____

Họ tên sinh viên 3 (SV3):_____ Mã SV: _____

CLO	Nội dung/Chuẩn đầu ra	ĐIỂM CỦA SV 1	ĐIỂM CỦA SV 2	ĐIỂM CỦA SV 3
1	Xây dựng và huấn luyện mạng neuron (Bài toán phân lớp ảnh, huấn luyện mô hình phân lớp…)			
2	Tinh chỉnh mô hình và thuật toán huấn luyện (Bài toán phân đoạn, tìm gốc quay, phát hiện vị trí, …)			
3	Vận dụng mạng neuron tích chập và mạng hồi quy (Triển khai mô hình trên web, desktop, mobile, …)			
4	Có khả năng giải quyết một số vấn đề thực tế (Thu thập dữ liệu, xử lý dữ liệu, …)			
5	Có năng lực trình bày giải pháp kỹ thuật (Thuyết trình, trình bày báo cáo, …)			
	Tổng			

Họ tên GV 1: _____ Ký tên: _____

Họ tên GV 2: _____ Ký tên: _____

TÓM TẮT ĐỒ ÁN

Bài toán đặt ra là việc nhận dạng và phân loại 5 đối tượng dụng cụ học tập, bao gồm tẩy, bút, chuốt bút chì, thước và đồ bấm gim. Bên cạnh đó là phải phân đoạn được các đối tượng có trong hình ảnh. Thách thức chính là xử lý ảnh đa dạng về cả màu sắc, hình dạng và góc chụp để nhận dạng chính xác.

Phương pháp được đề xuất dựa trên việc sử dụng mạng nơ-ron (deep neural network) để tiến hành huấn luyện, sau cùng triển khai ứng dụng nhận dạng các đối tượng trong ảnh. Quá trình này bao gồm các bước:

- Thu thập dữ liệu cần thiết cho quá trình hiện thực bài toán.
- Xử lí dữ liệu ảnh phù hợp cho quá trình huấn luyện.
- Xây dựng các mô hình học sâu và tiến hành huấn luyện với tập dữ liệu đã được chuẩn bị từ trước. Đối với bài toán phân loại ảnh các mô hình được triển khai ở đây là Multilayer perceptron, Convolutional Neural Network tự xây dựng, VGG16. Còn lại bài toán phân đoạn ảnh sử dụng mô hình U-Net.
- Triển khai các mô hình lên ứng dụng với thư viện Streamlit trên nền tảng Pycharm.

Kết quả đạt được sau quá trình trên đầu tiên là phần mềm phân loại ảnh vì chọn mô hình VGG16 có áp dụng kỹ thuật Transfer learning với độ chính xác khoảng 92% nên khi sử dụng thực tế cho ra kết quả rất tốt. Còn lại là phần mềm phân đoạn ảnh áp dụng U-Net cho ra kết quả không khả quan cho lắm khi độ sai lệch khá cao và có những trường hợp không nhận diện được. Nguyên nhân có thể do hình dạng cũng như màu sắc của các dụng cụ học tập vô cùng đa dạng và có những trường hợp tương tự nhau cũng như yếu tố màu sắc của vật thể với nền tương tự nhau cũng ảnh hưởng đến kết quả.

MỤC LỤC

Chương 1. **Giới thiệu bài toán**

1.1 Câu hỏi nghiên cứu

Bài toán nhằm giải quyết việc phân loại và phân đoạn các đối tượng cụ thể ở đây là 5 đối tượng dụng cụ học tập: tẩy, bút, chuốt bút chì, thước và bấm gim trong ảnh.

Đối với bài toán phân loại và phân đoạn ảnh input đầu vào là các ảnh chứa các dụng cụ học tập đã nêu ở trên. Các hình ảnh này có thể có độ phân giải và độ sâu màu sắc khác nhau, có thể chụp ở các góc độ và ánh sáng khác nhau.

Hình 1. Minh họa một ảnh trong data

Sau quá trình xử lí kết quả đầu ra đối với bài toán phân lớp là các nhãn liên quan như tẩy, bút, chuốt bút chì, thước và bấm gim.

Hình 2. Minh họa kết quả bài toán phân lớp

Đối với bài toán phân đoạn ảnh kết quả đầu ra là một ảnh chứa mask của ảnh đầu vào.

Hình 3. Minh họa input của bài toán phân đoạn

Hình 4. Minh họa output bài toán phân đoạn

Một số ví dụ minh họa:

- Đối với bài toán phân lớp:
    - Input: Hình ảnh chứa một cây bút. Output: Nhãn "pen".
    - Input: Hình ảnh chứa một thước đo. Output: Nhãn "ruler".
- Đối với bài toán phân đoạn:

  Input: Hình ảnh chứa cây bút, chuốt bút chì. Output: Hình ảnh chứa mask của cây bút và chuốt bút chì.

Dataset của cả 2 bài toán toàn bộ đều có định dạng ảnh là 224x224.

- Dataset của bài toán phân loại ảnh gồm 2710 ảnh được chia thành các tập train với 1841 ảnh, test với 373 ảnh và valid với 496 ảnh. Bên trong mỗi thư mục train, test, valid sẽ chứa 5 thư mục là các lớp tương đương như tẩy, bút, chuốt bút chì, thước và bấm gim bên trong các thư mục chứa ảnh tương ứng.
- Dataset của bài toán phân đoạn gồm 910 ảnh và 910 ảnh chứa nhãn của các tấm ảnh đó. Trong quá trình train ta sẽ chia ra 2 tập train và valid như sau.

```
(910, 224, 224, 1)
Training data shape is: (728, 224, 224, 6)
Validaing data shape is: (182, 224, 224, 6)
```

Hình 5. Minh họa cấu trúc ảnh bài toán phân đoạn

1.2 Giới hạn nghiên cứu

Số lượng đối tượng: Phần mềm chỉ có thể nhận dạng và phân loại 5 đối tượng cụ thể là bút, sách, bảng, thước, và viết. Các đối tượng khác ngoài phạm vi này sẽ có thể dẫn đến sai lệch.

Độ phức tạp của hình ảnh: Phần mềm có thể gặp khó khăn khi xử lý hình ảnh có độ phức tạp cao, chẳng hạn như hình ảnh có nền phức tạp, đối tượng bị che khuất, hoặc hình ảnh bị nhoè, nền và vật thể cùng màu cũng như các đối tượng có đa dạng kiểu .

Phân giải hình ảnh: Phần mềm có thể gặp khó khăn trong việc xử lý hình ảnh có độ phân giải thấp hoặc cao vượt quá giới hạn được xác định trước.

1.3 Bố cục đồ án

Đồ án được tổ chức thành ba chương chính.

Chương 1 giới thiệu bài toán nhằm nhận dạng và phân loại 5 đối tượng dụng cụ học tập từ các hình ảnh, bao gồm:

- 1.1 Câu hỏi nghiên cứu, phát biểu rõ ràng về bài toán cần giải quyết và giới thiệu dataset được sử dụng.
- 1.2 Giới hạn nghiên cứu, đưa ra các ràng buộc và giới hạn của bài toán
- 1.3 Bố cục đồ án, trình bày tổng quan về cấu trúc và nội dung của đồ án.

Chương 2 trình bày giải pháp đề xuất, bao gồm các mục sau:

- 2.1 Phân tích dữ liệu, mô tả quá trình phân tích và tiền xử lý dữ liệu;
- 2.2 Mô hình, trình bày cụ thể về kiến trúc của mạng nơ-ron sâu được sử dụng.

Chương 3 tập trung vào thực nghiệm, mô tả quá trình thử nghiệm và phân tích kết quả thu được.

Chương 2. Giải pháp đề xuất

2.1 Phân tích dữ liệu

Với 2 tập dữ liệu ảnh của 2 bài toán đều có kích thước là 224x224

- Tập dataset của bài toán phân loại cho thấy có tổng cộng 2710 mẫu. Dataset này bao gồm 5 lớp đối tượng dụng cụ học tập, bao gồm eraser (cục tẩy, 540 ảnh), pen (bút, 621 ảnh), ruler (thước, 672 ảnh), pencil_sharpener (đồ gọt bút chì, 521 ảnh), và stapler (cái ghim, 356 ảnh). Tập dữ liệu được chia thành ba phần: Train (1841 ảnh), Valid (496 ảnh), và Test (373 ảnh). Phân tích này cung cấp cái nhìn tổng quan về sự phân phối và kích thước của dữ liệu, giúp định hình được cách tiếp cận trong việc huấn luyện và kiểm tra mô hình.

  Một số ví dụ về hình ảnh có trong dataset:

Hình 6. Minh họa dữ liệu bài toán phân lớp

Cấu trúc dataset:

Bảng 1. Minh họa cấu trúc dataset bài toán phân lớp

Train	eraser	336
	pen	420
	pencil_sharpener	356
	ruler	453
	stapler	251
Test	eraser	101
	pen	116
	pencil_sharpener	96
	ruler	126
	stapler	62
Valid	eraser	76
	pen	88
	pencil_sharpener	72
	ruler	96
	stapler	46

- Đối với dữ liệu của bài toán phân đoạn ảnh gồm 910 ảnh được chia thành 2 tập train và valid sau bên trong ảnh chứa label thường sẽ có 6 lớp trong đó 5 lớp là dụng cụ học tập và một lớp nền:

```
(910, 224, 224, 1)
Training data shape is: (728, 224, 224, 6)
Validaing data shape is: (182, 224, 224, 6)
```

Hình 7. Minh họa cấu trúc data bài toán phân đoạn

Ví dụ về 1 ảnh kèm theo ảnh chứa nhãn của nó:

Hình 8. Minh họa input tập huấn luyện bài toán phân đoạn

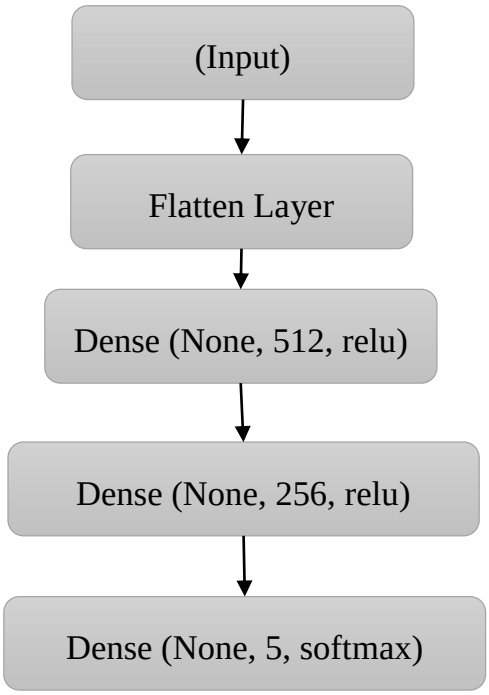
Hình 9. Minh họa nhãn input bài toán phân đoạn

Về tập dữ liệu do các đối tượng là dụng cụ học tập có nhiều màu sắc và kích thước khác nhau ví dụ như gôm có nhiều hình dạng khác cũng sẽ ảnh hưởng đến khả năng học của mô hình. Kèm theo nhiều vật có màu sắc gần tương tự với màu nền cũng là một khó khăn trong quá trình huấn luyện như ta có thể thấy là phần màu xanh ở cục tẩy trên ảnh và nền.

2.2 Mô hình

1. Mô hình MLP:

- Kiến trúc:



Hình 10. Cấu trúc mô hình MLP

- Mô tả chi tiết kiến trúc:

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 flatten_1 (Flatten)         (None, 150528)            0

 dense_3 (Dense)             (None, 512)               77070848

 dense_4 (Dense)             (None, 256)               131328

 dense_5 (Dense)             (None, 5)                 1285

=================================================================
Total params: 77203461 (294.51 MB)
Trainable params: 77203461 (294.51 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

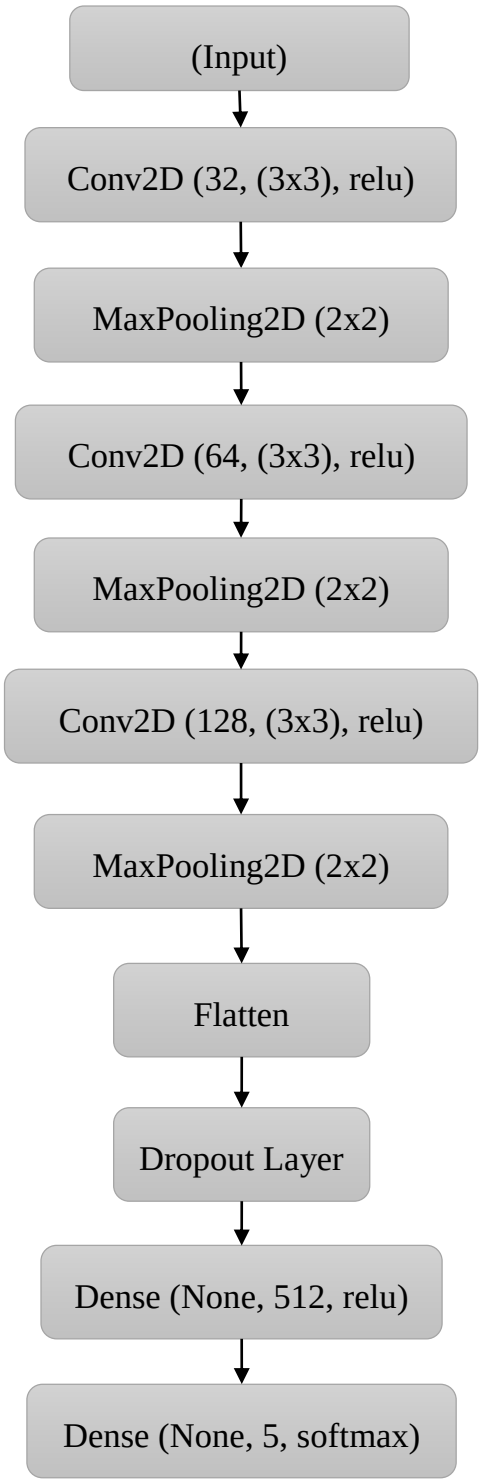
Hình 11. Chi tiết cấu trúc mô hình MLP

- Hàm activation của tầng cuối: softmax

- Hàm loss: categorical_crossentropy

2. Mô hình CNN:

- Kiến trúc:



Hình 12. Cấu trúc mô hình CNN

- Mô tả chi tiết kiến trúc:

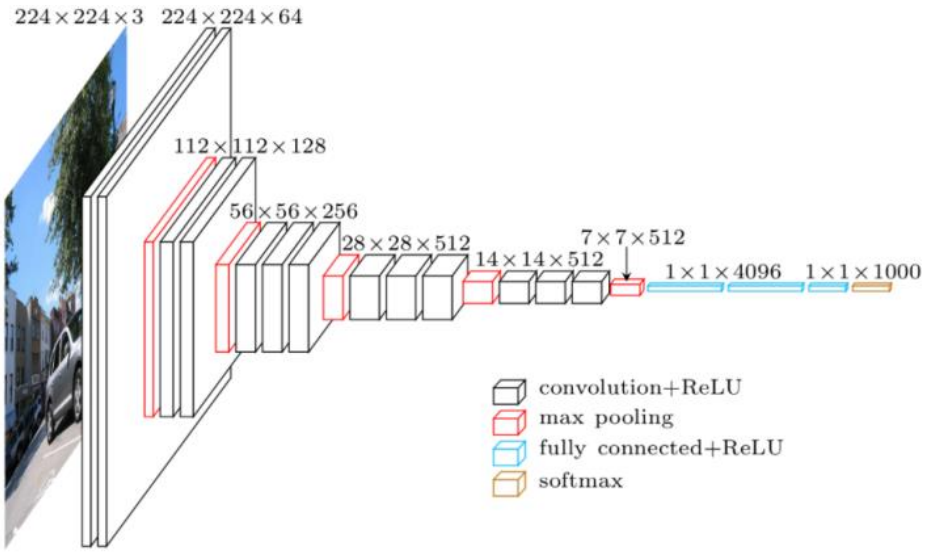
```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_19 (Conv2D)          (None, 222, 222, 32)      896

 max_pooling2d_11 (MaxPooli  (None, 111, 111, 32)      0
 ng2D)

 conv2d_20 (Conv2D)          (None, 109, 109, 64)      18496

 max_pooling2d_12 (MaxPooli  (None, 54, 54, 64)        0
 ng2D)

 conv2d_21 (Conv2D)          (None, 52, 52, 128)       73856

 max_pooling2d_13 (MaxPooli  (None, 26, 26, 128)       0
 ng2D)

 flatten_3 (Flatten)         (None, 86528)             0

 dropout_4 (Dropout)         (None, 86528)             0

 dense_7 (Dense)             (None, 512)               44302848

 dense_8 (Dense)             (None, 5)                 2565

=================================================================
Total params: 44398661 (169.37 MB)
Trainable params: 44398661 (169.37 MB)
Non-trainable params: 0 (0.00 Byte)
```

Hình 13. Chi tiết kiến trúc mô hình CNN

- Hàm activation của tầng cuối: softmax.
- Hàm loss: categorical_crossentropy.

3. Mô hình VGG16:

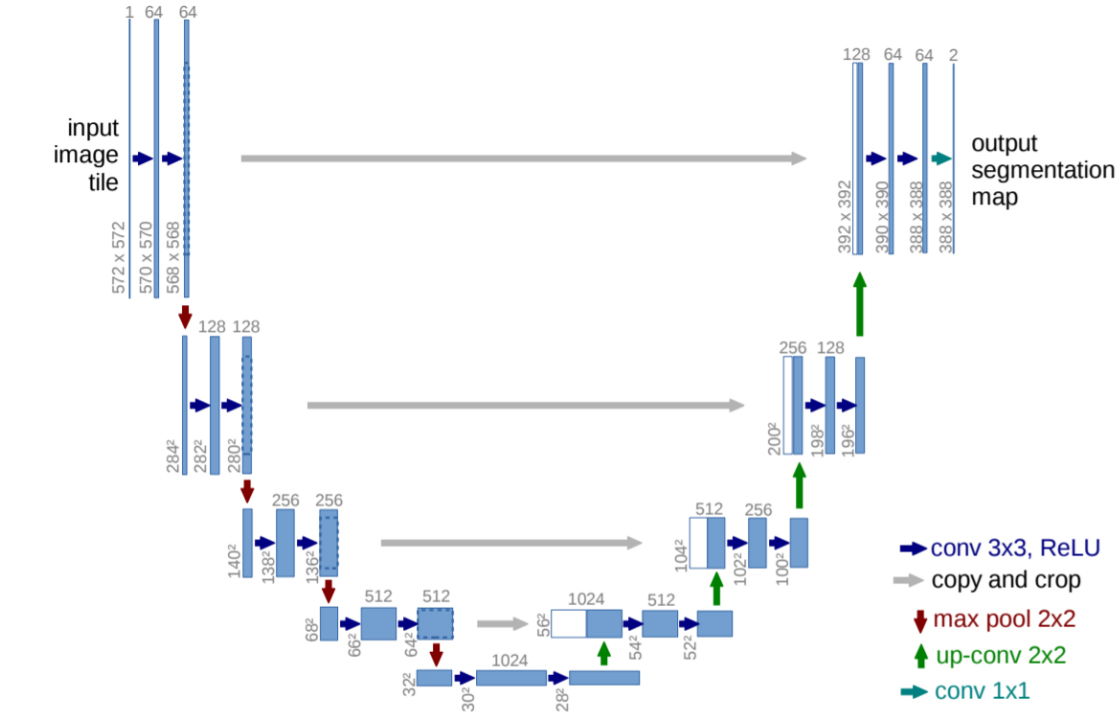
- Kiến trúc:



Hình 14. kiến trúc mô hình VGG16

- Mô hình này được biết đến với cấu trúc đơn giản nhưng hiệu quả, bao gồm 16 lớp có trọng số: 13 lớp tích chập (convolutional layers) và 3 lớp kết nối đầy đủ (fully connected layers). VGG-16 sử dụng các bộ lọc (kernel) kích thước nhỏ 3x3 trên tất cả các lớp tích chập, bước nhảy (stride) là 1 và sử dụng padding để giữ nguyên kích thước không gian của đầu vào qua mỗi lớp tích chập. Ngoài ra, mô hình cũng sử dụng hàm kích hoạt ReLU để thêm tính phi tuyến và Pooling layers để giảm kích thước không gian của đặc trưng sau mỗi vài lớp tích chập.
- Hàm activation của tầng cuối: softmax
- Hàm loss: categorical_crossentropy

4. Mô hình Unet:



Hình 15. Kiến trúc mô hình U-Net

- Kiến trúc mạng Unet bao gồm 2 phần là phần thu hẹp (contraction) ở bên trái và phần mở rộng (expansion) ở bên phải. Mỗi phần sẽ thực hiện một nhiệm vụ riêng như sau:
    - Phần thu hẹp: Làm nhiệm vụ trích lọc đặc trưng để tìm ra bối cảnh của hình ảnh. Vai trò của phần thu hẹp tương tự như một Encoder. Một mạng Deep CNN sẽ đóng vai trò trích lọc đặc trưng. Lý do nhánh được gọi là thu hẹp vì kích thước dài và rộng của các layers giảm dần. Từ input kích thước 224 chỉ còn 14x14. Đồng thời độ sâu cũng tăng dần từ 3 lên 512.

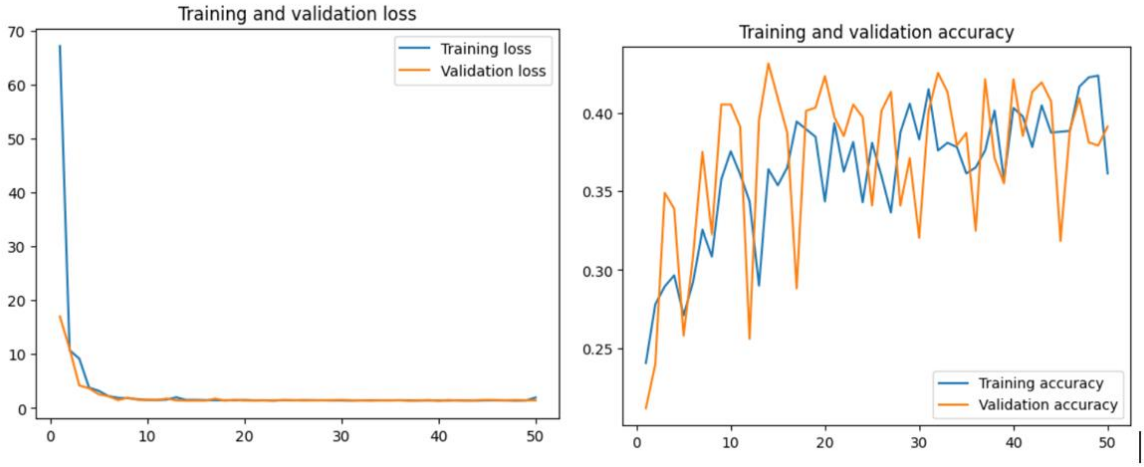
    - Phần mở rộng: Gồm các layer đối xứng tương ứng với các layer của nhánh thu hẹp. Quá trình Upsampling được áp dụng giúp cho kích thước layer tăng dần lên. Sau cùng ta thu được một ảnh mask đánh dấu nhãn dự báo của từng pixel.

      Đặc trưng riêng trong cấu trúc của Unet đó là áp dụng kết nối tắt đối xứng giữa layer bên trái với layer bên phải.
- Hàm activation của tầng cuối: softmax
- Hàm loss: categorical_crossentropy.

Chương 3. Thực nghiệm

1. Mô hình MLP:

- Các Siêu Tham Số:
  + Hàm loss: Categorical Cross-Entropy.
  + Thuật toán tối ưu hóa: Adam.
  + Số lượng epochs: 50.
- Biểu Đồ Loss và Accuracy:



Hình 16. Biểu đồ Loss và Accuracy MLP

- Kết Quả Thực Nghiệm:

  Độ chính xác (Accuracy): ≈ 39%

2. Mô hình CNN:

- Các Siêu Tham Số:
  + Hàm loss: Categorical Cross-Entropy.
  + Thuật toán tối ưu hóa: Adam.
  + Số lượng epochs: 80.
- Biểu Đồ Loss và Accuracy:

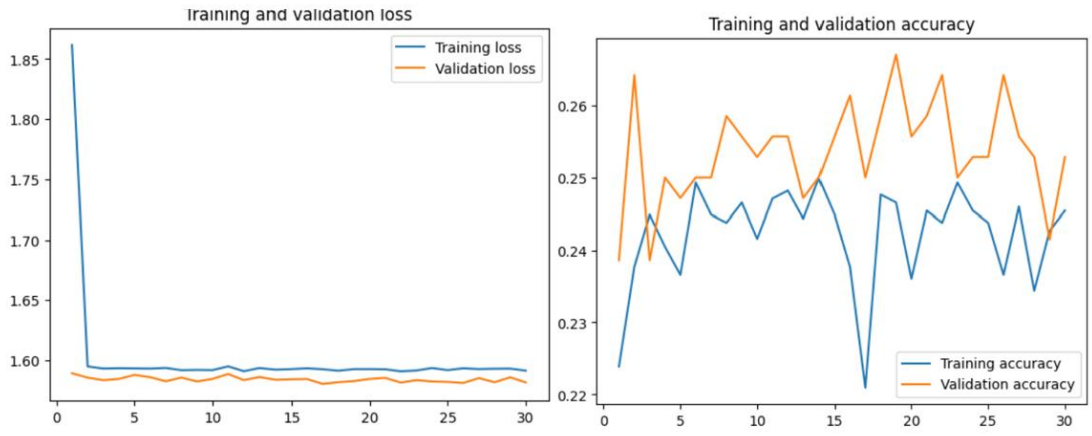
Hình 17. Biểu đồ Loss và Accuracy của CNN

- Kết Quả Thực Nghiệm:

  Độ chính xác (Accuracy) ≈ 70%

3. Mô hình VGG16:

- Các Siêu Tham Số:
  + Hàm loss: Categorical Cross-Entropy.
  + Thuật toán tối ưu hóa: Adam.
  + Số lượng epochs: 30.
- Biểu Đồ Loss và Accuracy:



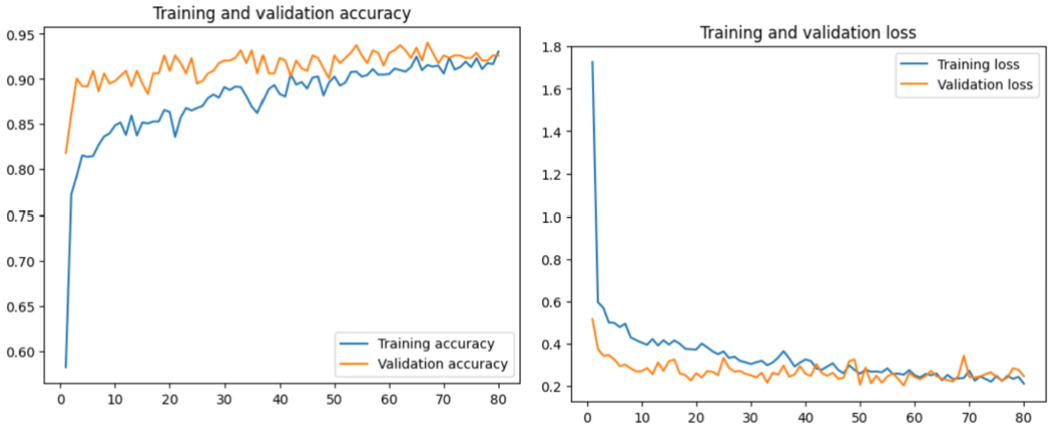
Hình 18. Biểu đồ Loss và Accuracy của VGG16

- Kết Quả Thực Nghiệm:

  Độ chính xác (Accuracy): ≈ 25%

4. Mô hình VGG16 áp dụng Transfer Learning:

- Các Siêu Tham Số:
  + Hàm loss: Categorical Cross-Entropy.
  + Thuật toán tối ưu hóa: Adam.
  + Số lượng epochs: 80.
- Biểu Đồ Loss và Accuracy:



Hình 19. Biểu đồ Loss và Accuracy của VGG16 áp dụng Transfer learning

- Kết Quả Thực Nghiệm:

  Độ chính xác (Accuracy) ≈ 92%

KẾT LUẬN

Những điều đã làm được

Trong đồ án này, nhóm đã tiến hành nghiên cứu về việc nhận dạng các đối tượng dụng cụ học tập từ hình ảnh với 2 bài toán được đặt ra là phân loại và phân đoạn ảnh, tập trung vào 5 loại đối tượng chính là tẩy, bút, chuốt bút chì, thước và bấm ghim. Nhóm đã đề xuất và triển khai một số phương pháp phân loại dựa trên các mô hình MLP, CNN, VGG16 và VGG16 Transfer learning và sử dụng một tập dữ liệu chứa 2710 hình ảnh đã được phân loại. Đối với bài toán phân đoạn nhóm em đã hiện thực lại mô hình U-Net để tiến hành phân đoạn với tập dữ liệu gồm 910 ảnh đã được gán nhãn. Sau đó, nhóm đã thực hiện huấn luyện và đánh giá các mô hình trên tập dữ liệu kiểm tra để đánh giá hiệu suất của các mô hình và chọn ra mô hình phù hợp nhất để triển khai ứng dụng.

Qua quá trình trên nhóm chúng em đã hiểu thêm về cấu trúc, cách thức hoạt động của các kiến trúc mạng như MLP, CNN, VGG16 và Unet cũng như độ hiệu quả của các kiến trúc trên đối với tập dữ liệu. Có thể phân loại với độ chính xác cao với mô hình VGG16 áp dụng kỹ thuật Transfer learning sau đó đã triển khai được bài toán lên ứng dụng. Về phân đoạn ảnh kết quả không được khả quan có thể là do ảnh hưởng của data. Và bài toán này cũng đã được triển khai lên ứng dụng.

Những điều chưa làm được

Đối với phần này có lẽ là độ chính xác của bài toán phân đoạn chưa được cao cho lắm cho ra khá nhiều lỗi và đôi lúc không nhận dạng được các vật thể.

Hướng phát triển

Cải thiện vấn đề độ chính xác của bài toán phân đoạn ảnh theo hướng tối ưu kỹ thuật xử lí ảnh, tăng cường dữ liệu cũng như tối ưu hóa mô hình phù hợp với điều kiện bài toán để có kết quả tốt hơn. Ngoài ra thì có một số hướng phát triển tiềm năng cho đồ án này. Như tiến hành một nghiên cứu chi tiết về hiệu suất của mô hình trên từng lớp

đối tượng để hiểu rõ hơn về khả năng nhận dạng của mô hình đối với từng loại đối tượng cụ thể. Điều này sẽ giúp xác định các điểm yếu và cải thiện mô hình một cách hiệu quả. Cũng như cần tiếp tục đánh giá sự ảnh hưởng của các yếu tố nhiễu như ánh sáng, góc chụp, và nhiễu khác đối với hiệu suất của mô hình để đảm bảo tính ổn định và độ tin cậy trong các điều kiện môi trường khác nhau. Bằng cách tăng cường dữ liệu huấn luyện và tối ưu hóa mô hình, chúng ta có thể cải thiện hiệu suất của mô hình và mở ra những cơ hội mới trong lĩnh vực trí tuệ nhân tạo và tự động hóa.

TÀI LIỆU THAM KHẢO

Tài liệu Internet

1. Bui Tien Dung, 15/1/2020, U-Net Kiến trúc mạnh mẽ cho Segmentation, https://viblo.asia/p/u-net-kien-truc-manh-me-cho-segmentation-1Je5Em905nL, 25/3/2024.
2. Great Learning team, 18/11/2022, Introduction to VGG16 | What is VGG16?, https://www.mygreatlearning.com/blog/introduction-to-vgg16/ , 25/3/2024.