Pixel-Level Dress Detection through Semantic Segmentation in Images

Swarag Reddy Pingili
Vishwaak Chandran

Abstract

This paper dives into using computer vision techniques to understand fashion images better. We focus on semantic segmentation, to analyze fashion images. We are particularly interested in pulling out important details from fashion images that can help the fashion industry.

In computer vision, there are three main tasks: figuring out what's in an image (classification), locating objects (detection), and precisely outlining them (segmentation). We're focusing on segmentation, which involves drawing perfect outlines around objects.

In this paper we use the U-Net and U-2Net Architecture to perform the image segmentation tasks. Both of these architectures are made up of convolutional layers which helps to create a mask on the image.

Introduction:

In computer vision there are three major sub-tasks: object detection, classification, and segmentation. All these tasks depend on inferring the different properties of an image.

In case of classification, an image needs to be identified what class it belongs to whereas in object detection the model draws a bounding box over the detected object, in case of segmentation the model needs to draw a perfect outline of an object in an image.

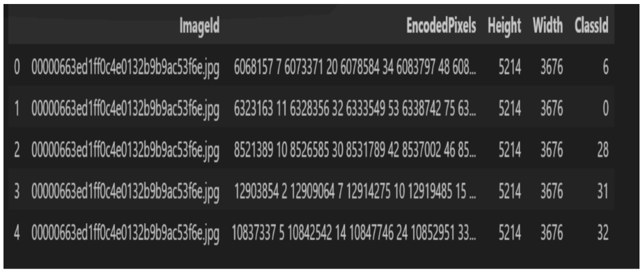
We our focusing on segmentation which deals with understanding the properties or characteristics of an object to draw an outline over it. There are two types of segmentation:

- **Semantic segmentation:** In semantic segmentation the objects are divided based on their semantic characteristics.Different instances of the same class are counted as one and there is no differentiation between them.
- **Instance Segmentation:** In instance segmentation objects of the same class are divided into their very own entities. Even though all the objects belong to the same class, an individual object can be highlighted.

1. Dataset description

We are using the iMaterialist (Fashion) 2019 at FGVC6 dataset from Kaggle in our project. The dataset has an extensive collection of image date and is 23.53 GB in size.

Files in the dataset: train.csv, label_description.csv, train folder, test folder.



	ImageId	EncodedPixels	Height	Width	ClassId
0	00000663ed1ff0c4e0132b9b9ac53f6e.jpg	6068157 7 6073371 20 6078584 34 6083797 48 608...	5214	3676	6
1	00000663ed1ff0c4e0132b9b9ac53f6e.jpg	6323163 11 6328356 32 6333549 53 6338742 75 63...	5214	3676	0
2	00000663ed1ff0c4e0132b9b9ac53f6e.jpg	8521389 10 8526585 30 8531789 42 8537002 46 85...	5214	3676	28
3	00000663ed1ff0c4e0132b9b9ac53f6e.jpg	12903854 2 12909064 7 12914275 10 12919485 15 ...	5214	3676	31
4	00000663ed1ff0c4e0132b9b9ac53f6e.jpg	10837337 5 10842542 14 10847746 24 10852951 33...	5214	3676	32

Figure 1: Train.csv

The different columns in the train.csv are:

- **ImageID:** This column provides a unique identification for each image.
- **EncodedPixels:** This column shows the area to mask in an image.
- **Height:** This column provides the height of an image.
- **Width:** This column provides the width of an image.
- **ClassId:** This column is an identifier for a class the image belongs, based on a list of categories.

There are 45 sub categories in total. Each sub category has an ID, name, level, and it belongs to some category. In total, there are 12 categories.

2. Project Description

2.1 **Description:** Our project focuses on the application of image segmentation to the fashion industry by applying semantic segmentation on dresses.

By segmenting dresses in images pixel by pixel, we plan to produce detailed insights into the design and patterns of the dresses. This can enhance the customer experience.

To build this project we plan to use advance CNN architectures, such as U-Net or U2Net, which are known

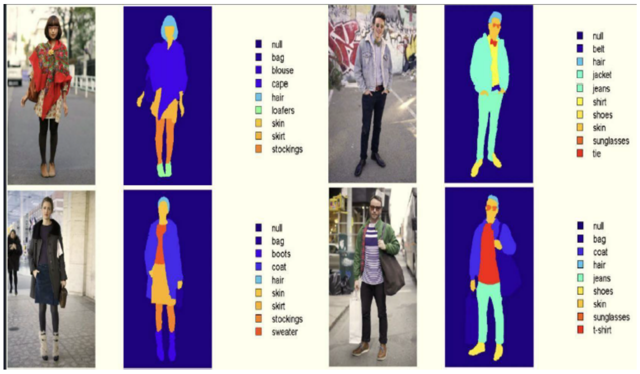


Figure 2: Sample Results

for their efficiency in image segmentation tasks. These models will be trained to recognize various patterns of dresses.
The main idea of our problem statement is to train these models to recognize and segment various patterns, styles, and types of dresses within various images.
Our project is about creating a tool that can help the fashion world. This tool can identify and analyze dresses in images. It improves the shopping experience for customers online by making it easier for them to find what they are looking for

2.2 References :

i Semantic Segmentation of Fashion Images Using Feature Pyramid Networks
This paper addresses the challenge of segmenting fashion images into various clothing categories, focusing on the importance of texture, shape and context. The paper utilizes a fully convolutional neural network model that uses Feature Pyramid networks for semantic segmentation. The model Is highlighted for its computational efficiency.

ii FashionSegNet: a model for high-precision semantic segmentation of clothing images
This paper utilizes ResNet50 backbone within an encoder-decoder structure, enhancing feature extraction and boundary identification. It achieves state-of-the art performance on the DeepFashion2 dataset. It is very efficient in handling complex segmentation tasks.

iii Segmentation task for fashion and apparel
This paper compared different state-of-the-art models on different datasets to understand the trade-off between them technique and which will yield the best result for dress segmentation problem.

iv U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection
It introduces a new network called U2Net which is a nested U-net for Salient Object Detection and introduces a new block called Residual U block. This architecture helps to get state-of-the art of performance without the use of any backbone architectures.

2.3 Difference in Approach :

- We implemented U-Net architecture from scratch in pytorch. The primary objective being segmenting fashion images.

- Experimented with different training parameters to understand the impact of them on the training of these networks. Some of the paramters we changes are optimizers (like RMSprop, Adam, SGD). This also helped us understand how they affect training speed, stability and overall model performance.

- We compared the U-Net architecture with U2Net architecture to see which one performs better for our use case.

- We trained both the models on completed different dataset than the one used in the references

2.4 Difference in Accuracy:

- **References Accuracies:**

i Semantic Segmentation of Fashion Images Using Feature Pyramid Networks: This paper has achieved an accuracy of 93.82 percent by using a feature pyramid network (FPN) with a ResNeXt backbone for the semantic segmentation of fashion images.

ii FashionSegNet: a model for high-precision semantic segmentation of clothing images: The paper mentions achieving a Mean Intersection over Union (mIoU) of 74.55 percent and Boundary Intersection over Union (Boundary IoU) of 57.51 percent on the simpli123 FashionSegNet dataset. Additionally, it mentions satisfactory performance on CCP, CFPD, and CIHP datasets.

iii Segmentation task for fashion and apparel: By using Adam optimizer, this paper has achieved an accuracy of 80.84 percent with the U-Net model at LR=0.01
By using the Atrous ResNet-50 they have achieved an accuracy of 93 percent. They have also achieved an accuracy of 89 percent by using the SegNet model.

iv U2-Net: Going Deeper with Nested U-Structure for Salient Object Detection: By using the U2-Net model, this paper has achieved a maximum accuracy of 92.8 percent.

- **Our Accuracy:** We were able to achieve a maximum accuracy of 96 percent by using the U2-Net architecture.

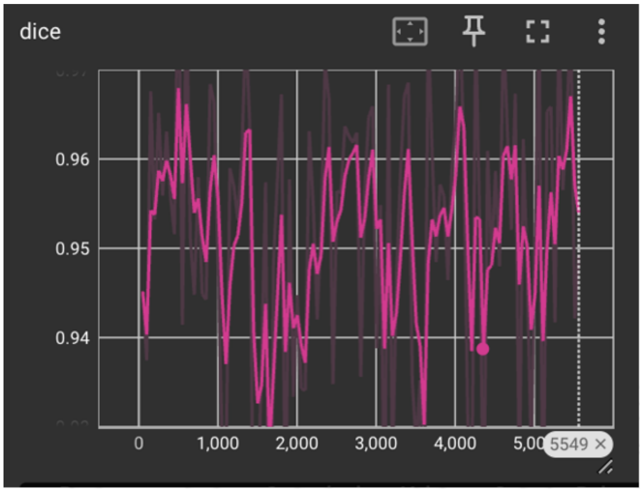


Figure 3: Dice score

Training approach for u-net/u2net: We used early stop to train the models, and also we also tried out different optimizes like Adam, SGD, RMS Prop with different learning rates 0.1,0.01,0.001.

Final list of parameters - unet:

- optimizer: SGD
- learning rate: 0.01
- momentum = 0.9
- weight_decay= 0.0005

Final list of parameters - u2net::

- optimizer: SGD
- learning rate: 0.01
- momentum = 0.9
- weight_decay= 0.0005

3. Analysis

Achievement

1. **Unet**: In case of Unet works wonderful on simple clothing but then when a clothing with complex patter or when some part of the image is occluded then the network is not performing well.
2. **U2Net**: The network can handle clothing with complex pattern, and it can manage well with the occlusion scenario.

Improvement

1. The dataset we used only use western wear dress. This is one of the main limitation of the models. The dataset can be improved with different cultural clothing like Indian traditional wear, Japanese dresses as well.
2. The dataset also mostly contained front facing images, it can also be expanded to different view of the models and also include all the age groups. This would create a universal dataset which could improve the overall usability of the model.

Future Work

1. We also trained the model on the smaller subset of the dataset around 35000 images. In the future, we would like to train the model on the entire dataset, without any limitation on compute.
2. Another possible direction to expand upon, would to experiment with U2net network with other networks to create a hybrid network with help of other networks like ResNet, AlexNet.

4. Conclusion:

In summary, our project has been a success. We have successfully built a model that has an accuracy of over 96 percent. Our model is able to properly segment fashion images and separate different clothing items.
This achievement is important because it means our model can be used for many things in the fashion world. Our project is capable of making online shopping easier for people by helping them find exact clothes they are looking for easily.
Looking ahead, there is still a lot more we could do to improve our model. We could train our model on different sets of images to further improve its performance in various situations.
We could also make our project work real time. This could possibly lead to virtual fitting rooms and personalized fashion recommendations.
To conclude, our project uses semantic segmentation to create a mask on fashion images which can be used to understand various aspects of fashion. We have also achieved a very high accuracy. Because of this our project is able to properly extract all valuable information from a fashion image.