Clothing Detection and Isolation

Hyungjun Kim
University of Michigan
500 S State St, Ann Arbor
hyungjun@umich.edu

Alex Ni
University of Michigan
500 S State St, Ann Arbor
axni@umich.edu

1. Introduction

1.1. Problem Statement

In today's digital age, people frequently encounter clothing items in images, be it on social media, fashion blogs, or online advertisements, that they wish to purchase but cannot easily identify or locate online. An automated solution to isolate and identify clothing from images would significantly enhance user experience by simplifying the search process. This project aims to develop a system that can automatically isolate clothing items from images and identify them by searching for matching or similar pieces on the internet.

To manage the project scope effectively, we focused on utilizing existing tools and datasets to detect and segment clothing items while testing their integration for image-based searches. The problem of clothing isolation and identification is both challenging and significant, as it requires a combination of advanced computer vision techniques and robust dataset handling. Through this project, we explored the real-world utility and limitations of such a system while proposing future improvements.

1.2. Related Work

Previous research has demonstrated the effectiveness of deep learning models in object detection and segmentation. Tools like Mask R-CNN have been widely used for instance segmentation, achieving high accuracy on various benchmarks. Fashion-specific datasets such as Fashionpedia provide annotated clothing data, making them ideal for developing and validating clothing recognition systems.

However, existing approaches often struggle with generalization due to the diversity of clothing styles, fabrics, and poses. Additionally, most systems require significant computational resources for fine-tuning and real-time processing. Our work builds upon these strengths by leveraging pre-trained models, while we acknowledge the limitation of computational resources as a significant challenge during our implementation.

Other algorithms, such as YOLO (You Only Look Once), have also been employed for clothing identification tasks. While YOLO excels at real-time object detection and achieves high-speed performance, it lacks the capability to produce pixel-wise segmentation masks. This limitation makes it unsuitable for applications requiring precise isolation of clothing items, as segmentation is critical for accurate feature extraction and subsequent image-based searches. [1]

In order to understand what we require out of our tools, a very simple trial program was created using the Fashion-MNIST dataset, which has low-res clothing items. However, this alone was insufficient, as the model was trained on low-resolution images and lacked the extensiveness required for general use. Also, it didn’t isolate and cut out the region of the clothes. This gave clear insight into what was required out of the potential neural net we would choose, as it would need to successfully create a bitmask of the articles of clothing, but do it on higher resolution images.

2. Method

2.1. High Level

The project utilized Detectron2’s pre-built Mask R-CNN model for clothing detection and segmentation. This model was chosen for its performance and pre-trained weights, which provided a strong baseline for experimentation. The Fashionpedia dataset was used for training and validation due to its comprehensive annotations tailored to the fashion domain.

Steps in the high level workflow involved:

1. Data Preparation: Images from Fashionpedia were preprocessed and formatted for easy use by the neural network.
2. Segmentation: Detectron2’s Mask R-CNN was applied to detect and isolate clothing items from the images.

3. Evaluation: Metrics such as Intersection over Union (IoU) and Mean Average Precision (mAP) were used to evaluate segmentation accuracy.

2.2. Mask RCNN

Mask R-CNN, or Mask Region-based Convolutional Neural Network, is an advanced deep learning architecture designed for both object detection and instance segmentation. It extends the Faster R-CNN framework by introducing an additional branch that generates pixel-level segmentation masks for each detected object, enabling fine-grained delineation of object boundaries.

The architecture comprises a backbone network (e.g., ResNet) for feature extraction, a Feature Pyramid Network (FPN) to enhance multi-scale representation, a Region Proposal Network (RPN) to identify candidate object regions, and ROI heads for refining bounding boxes and classifying objects. A key enhancement, ROIAlign, ensures precise feature alignment using bilinear interpolation, crucial for accurate segmentation. [2]

Mask R-CNN excels in applications requiring detailed scene understanding, such as medical imaging and fashion analysis. While powerful, it demands significant computational resources and large annotated datasets, making efficient deployment a challenge.

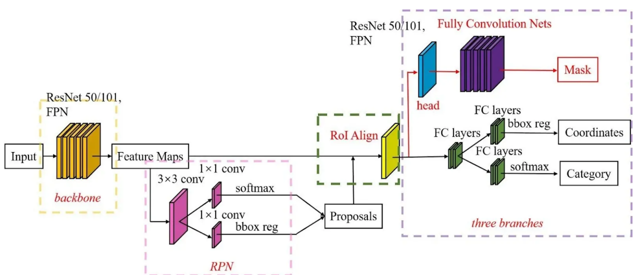

Figure 1: The Mask R-CNN framework for instance segmentation [2]

2.3. Model Details

The experiments utilized the Fashionpedia dataset, which provides comprehensive annotations of clothing items across 46 categories. The dataset was split into training and validation sets for training and evaluation. The annotations include pixel-level segmentation masks, making it ideal for instance segmentation tasks.

Hyperparameters: The model was configured with the following hyperparameters:

Model Architecture: Mask R-CNN with ResNet-50 as the backbone and Feature Pyramid Network (FPN).

Batch Size: 2 images per iteration.

Learning Rate: Set to 0.001 to ensure stable convergence during training.

Maximum Iterations: Set to 5000 allowing the model sufficient epochs to learn from the data.

Number of Classes: 46 classes corresponding to the categories in the Fashionpedia dataset.

ROI Head Batch Size per Image: Set to 512 to optimize the proposals used during training.

Mask Format: Input masks were formatted as bitmasks to align with Detectron2 requirements.

Evaluation: The model performance was evaluated on the validation set using metrics such as:

Intersection over Union (IoU): Measures the overlap between predicted and ground-truth masks.

Mean Average Precision (mAP): Evaluates detection and segmentation accuracy across all categories.

Baseline Comparison: The baseline used for comparison was the pre-trained Mask R-CNN model provided in the Detectron2 model zoo, fine-tuned on the COCO dataset. The custom training aimed to improve segmentation performance specifically for the Fashionpedia dataset categories.

2.4. Performance Results

The model demonstrated improvements in IoU for common clothing categories but struggled with overlapping and complex backgrounds. Resource constraints limited the maximum number of iterations and hyperparameter tuning, leaving room for improvement compared to fully optimized models.

Significant work remains to develop an accurate model, including refining the cropping stage. Since the bitmasks produced were of insufficient quality, the image extraction process has been postponed.

Images of the unsuccessful identifications have been attached in the appendix.

3. Conclusion

This project highlighted both the challenges and learning opportunities of utilizing advanced models like Mask R-CNN for clothing isolation and identification. While the segmentation of clothing items showed moderate success, the identification aspect was not fully realized due to computational limitations and the complexity of the task. These challenges emphasized the importance of fine-tuning hyperparameters, extending training iterations, and improving resource allocation. Despite these limitations, the process provided valuable insights into model behavior, dataset handling, and evaluation metrics, which can guide future improvements. With better computational resources and a deeper understanding of the model, future iterations can focus on refining segmentation quality, enhancing feature extraction, and integrating robust identification mechanisms. This project serves as a foundation for developing a more effective and optimized system for clothing analysis in subsequent efforts.

Appendix A

Figure 2 & 3: First Model's performance

Figure 3 & 4: Most recent model's performance

4. Bibliography

[1] P. Potrimba, "What is Mask R-CNN? The Ultimate Guide," Roboflow, 9 August 2023. [Online]. Available: https://blog.roboflow.com/mask-rcnn/. [Accessed 9 Dec 2024].

[2] J. a. F. A. Redmon, 2018. [Online]. Available: https://pjreddie.com/darknet/yolo/. [Accessed 10 December 2024].