Fashion Image Segmentation and Attribute Classification Using Mask R-CNN: Leveraging Kaggle and Fashionpedia Datasets

Jing Liu
CS444 Project Proposal
jingl17@illinois.edu

Abstract

The main goal of this project is to implement Mask R-CNN for fashion image segmentation and attribute classification using data from a Kaggle competition. Various methodologies will be explored to improve model performance, including ensembling techniques, hyperparameter tuning, model architecture tuning, and data augmentation. The Kaggle dataset will be used for training, while the Fashionpedia dataset will serve as the test set to evaluate generalization performance. These approaches will help address the unique challenges posed by fashion images, such as variations in clothing styles, patterns, and object sizes. The model will be evaluated using Intersection over Union (IoU) for segmentation and F1-score for attribute classification.

1. Introduction

Fashion image segmentation and attribute classification present unique challenges due to the wide variability in clothing styles, textures, and accessories. To tackle these challenges, this project aims to implement and fine-tune a Mask R-CNN model, which excels in instance segmentation tasks [2]. In addition to the core implementation, the project will explore several methods to improve performance, including ensembling, hyperparameter tuning, model architecture tuning, and data augmentation. Data from a Kaggle competition will be used as the primary training resource, while the Fashionpedia dataset [7] will be used for testing purposes to evaluate the model's ability to generalize across different datasets.

2. Resources

The Kaggle competition dataset will serve as the primary resource for training the model, and the Fashionpedia dataset will be used to test the model's performance in a real-world scenario. The Kaggle dataset offers a comprehensive set of fashion images with segmentation masks and attribute labels, making it ideal for training. The Fashionpedia dataset includes annotations for 46 clothing categories and 294 attributes, providing a robust test set for evaluating how well the model generalizes beyond the training data. The project will be implemented using PyTorch and Detectron2, with pre-trained weights from the COCO dataset leveraged for transfer learning [3]. Data augmentation will be applied using the Albumentations library to increase the model's robustness and generalization ability.

3. Work Overview

This section describes the key methodologies and steps involved in developing and optimizing the performance of the Mask R-CNN model for fashion image segmentation and attribute classification. Each subsection details specific aspects of the workflow, including data preprocessing, model implementation with transfer learning, architecture tuning, hyperparameter optimization, ensembling, and model evaluation.

3.1. Data Preprocessing and Augmentation

The Kaggle dataset will be preprocessed to resize images and prepare segmentation masks. Data augmentation techniques, including random flips, rotations, scaling, and color jittering, will be applied using the Albumentations library. These augmentations will simulate real-world variations in fashion images, helping the model generalize better across different viewing angles, lighting, and clothing types. The Fashionpedia dataset will be used for testing the model's generalization capability, helping to measure its performance on unseen data.

3.2. Model Implementation and Transfer Learning

The model will be based on Mask R-CNN, implemented using PyTorch and Detectron2. Transfer learning will be employed by using pre-trained weights from the COCO dataset [3], and fine-tuning the model on the Kaggle competition dataset. This will speed up training and improve the model's performance on fashion-related tasks, such as segmentation of garments and classification of attributes like fabric type

and color. After training, the model will be evaluated on the Fashionpedia dataset.

3.3. Model Architecture Tuning

Different backbone architectures, such as ResNet [1] and EfficientNet [6], will be explored to identify the most suitable architecture for fashion segmentation tasks. Feature Pyramid Networks (FPN) [4] will be tested to improve the model's ability to detect objects at multiple scales, a critical feature for fashion images that contain items of varying sizes. Attention mechanisms will also be experimented with to help the model focus on fine details, such as fabric textures and patterns.

3.4. Hyperparameter Tuning

The project will involve hyperparameter tuning of key factors such as learning rate, batch size, and optimizer choice. Both grid search and random search will be employed to find the optimal configuration. These tuning processes will help optimize model training and ensure it efficiently learns the complexities of fashion images.

3.5. Ensembling

To further enhance the model's performance, ensembling will be implemented by combining predictions from multiple models, such as Mask R-CNN, U-Net [5], and ResNet-based models. Model averaging or weighted voting will be used to combine the strengths of different architectures, resulting in more accurate segmentation and attribute classification predictions.

3.6. Evaluation and Error Analysis

The model will be evaluated using Intersection over Union (IoU) for instance segmentation and F1-score for attribute classification. The Kaggle dataset will be used for training, while the Fashionpedia dataset will be used for testing. A detailed error analysis will be conducted to identify common failure cases, such as misclassifications of small or intricate objects. Based on this analysis, targeted improvements will be made to the model and training pipeline.

4. Relationship to Background

This project is directly related to topics covered in CS444, including object detection, segmentation, and transfer learning. By implementing Mask R-CNN for fashion segmentation and attribute classification, this project provides an opportunity to apply these concepts to both the Kaggle competition dataset and the Fashionpedia test set. The exploration of techniques such as hyperparameter tuning, data augmentation, and ensembling will deepen my understanding of deep learning for computer vision.

5. Conclusion

This project will implement Mask R-CNN for fashion image segmentation and attribute classification using the Kaggle dataset for training and the Fashionpedia dataset for testing. By incorporating advanced techniques such as transfer learning, ensembling, hyperparameter tuning, model architecture tuning, and data augmentation, the project aims to optimize the model's performance to meet the challenges posed by fashion images.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2

[2] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2961–2969, 2017. 1

[3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 740–755, 2014. 1

[4] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2117–2125, 2017. 2

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015. 2

[6] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019. 2

[7] Li Zhu, Jinkun Tang, Licheng Ma, and Shuaiyi Wang. Fashionpedia: Ontology, segmentation, and an attribute detection dataset for fashion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3347–3357, 2020. 1