Sorting Clothes using Image Segmentation and Object Detection

MSc Research Project
MSc in Data Analytics

Bharadwaj Ravur
Student ID: x21194521

School of Computing
National College of Ireland

Supervisor: Abubakr Siddig

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Bharadwaj Ravur
Student ID:	X21194521
Programme:	MSc in Data Analytics **Year:** 2023
Module:	MSc Research Project
Supervisor:	Abubakr Siddig
Submission Due Date:	18/09/2023
Project Title:	Sorting Clothes using Image Segmentation and Object Detection
Word Count: 7047	**Page Count: 20**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Bharadwaj Ravur

Date: 18/09/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Sorting Clothes using Image Segmentation and Object Detection

Bharadwaj Ravur
x21194521

Abstract

This research delves into the intricate challenges associated with categorizing apparel for both genders in the dynamic landscape of the fashion industry. The study introduces a novel approach by harnessing the power of image segmentation and object identification techniques, specifically Mask R-CNN and YOLOv5s, to automate the apparel sorting process. The application of these advanced methods aims to streamline the classification process and enhance accuracy in apparel categorization.

The investigation employs the DeepFashion2 dataset, a comprehensive repository containing meticulously annotated examples of both consumer and commercial fashion photography, encompassing diverse categories of apparel. Leveraging this dataset for training and evaluation, the research unveils a systematic approach that addresses the complexities of identifying and categorizing fashion items.

The core of the research lies in the development of a robust and dependable system. By seamlessly integrating Mask R-CNN and YOLOv5s, the study showcases the efficacy of these models in achieving accurate apparel sorting. The comparative analysis of these techniques sheds light on their respective strengths and limitations, contributing to a holistic understanding of their applicability.

The research extends its scope beyond theoretical exploration. Practical viability is a central focus, and the study evaluates the proposed system's performance against real-world challenges. In a fashion landscape characterized by rapidly evolving trends and design intricacies, the research seeks to enhance efficiency and accuracy in apparel sorting.

In conclusion, this research underscores the significance of tackling unique challenges presented by the fashion industry. By harnessing the capabilities of Mask R-CNN and YOLOv5s, the study introduces an automated system capable of effectively categorizing gender-specific apparel. The research not only presents a technologically advanced solution but also highlights the broader implications of such advancements in streamlining fashion processes.

1 Introduction

In the realm of the textile industry, the act of sorting clothes might initially appear straightforward, yet it unfolds as a formidable challenge, particularly when dealing with the immense volume of clothing items. Recent times have witnessed the potential transformation of the fashion sector through the analysis of fashion imagery (Sivic, et al., 2006). This promising avenue capitalizes on the fusion of Artificial Intelligence (AI) and robotics, culminating in the deployment of algorithms that facilitate clothing classification and object detection (Lee & Cohen, 2006). This study delves into the amalgamation of AI and fashion, unveiling an innovative system that leverages advanced machine learning models, including Mask R-CNN and YOLO v5s, to revolutionize the garment sorting process.

Background

The apparent simplicity of clothing sorting within the textile industry belies its intricacies. The task becomes substantially challenging when confronted with the magnitude of clothing items requiring categorization. Traditionally, manual sorting techniques have prevailed, consuming time, exerting labor, and yielding error-prone outcomes (Gallagher & Chen, 2008) (Chen, et al., 2006). However, the current technological era heralds the advent of machine learning solutions capable of automating this process. Leveraging object detection methods and image segmentation, machine learning procedures stand poised to reshape clothing sorting practices. The heart of this endeavor lies in enhancing efficiency and accuracy through AI-powered interventions.

Aim and Objectives

The primary aim of this research is to harness the capabilities of machine learning models, specifically Mask R-CNN and YOLO v5s, to engineer an innovative system for clothing classification and object identification. The objectives guiding this study are as follows:

1. Identifying Industry Challenges: To comprehensively identify and dissect the challenges posed by clothing classification within the dynamic fashion industry landscape.
2. Capturing Fashion Trends: To acquire an in-depth understanding of prevailing fashion trends and their implications for clothing categorization and object detection.
3. Implementing Machine Learning Models: To deploy and optimize the effectiveness of machine learning models, specifically Mask R-CNN and YOLO v5s, in automating clothing classification and object identification.
4. Leveraging DeepFashion2 Dataset: To leverage the extensive "DeepFashion2" dataset for training and validating the machine learning models, thereby infusing real-world fashion diversity into the classification process.

Research Questions

This research endeavours to address the following pivotal questions:

1. What complexities and nuances arise in the realm of fashion industry clothing classification?
2. How do machine learning models, particularly Mask R-CNN and YOLO v5s, contribute to clothing classification and object identification within this context?
3. How is the expansive "DeepFashion2" dataset strategically employed to bolster the machine learning models' proficiency in the context of clothing classification?

Method Outline

The research methodology encompasses a structured approach:

1. Challenges Identification: A comprehensive exploration of the challenges endemic to clothing classification within the fashion industry, paving the way for innovative solutions.
2. Fashion Trends Analysis: A thorough analysis of current fashion trends, shedding light on the rapidly shifting landscape and its implications for automated clothing sorting.
3. Model Implementation: The deployment of cutting-edge machine learning models, namely YOLO v5s for object detection and Mask R-CNN for object classification, forms the crux of the method.

4. DeepFashion2 Dataset: Leveraging the "DeepFashion2" dataset to infuse realism and diversity into model training and validation, ensuring a robust performance under real-world conditions.
5. Performance Evaluation: Rigorous assessment of system performance, gauged through the lens of recall, precision, and overall accuracy metrics, aligning with industry standards.
In summary, this research navigates the intersection of fashion and AI, unraveling a novel approach to clothing classification through the fusion of advanced machine learning models. By deciphering industry complexities, implementing innovative solutions, and assessing their performance, this study strives to streamline and enhance clothing sorting within the dynamic landscape of the fashion industry.

2 Related Work

The fusion of artificial intelligence and fashion industry practices has garnered significant attention, precipitating transformative shifts in clothing classification and object detection. As the fashion landscape evolves at a rapid pace, the adoption of AI-driven methodologies offers an avenue for improved efficiency and accuracy (Yamaguchi, et al., 2012) (Wang & Ai, 2011).
The literature underscores the inherent complexities in clothing sorting within the textile industry. Traditional manual methods necessitate substantial labor, time, and are susceptible to errors. The need for automated systems that transcend these limitations is evident. Researchers highlight the role of machine learning models in automating the clothing sorting process, ranging from high-level classification to fine-grained object identification.

Challenges in the Fashion Industry for Clothing Classification

One critical challenge revolves around the differentiation between men's and women's garments. The dynamic nature of fashion, varying preferences across demographics, and the advent of gender-neutral and unisex clothing complicate traditional categorization methods. The blurring boundaries between gender-specific styles require innovative approaches to accurately classify clothing items that could be worn by individuals of any gender. Additionally, cultural and geographic variations in dress codes pose further hurdles, necessitating specialized categorization models or customized solutions.
Another impediment lies in the scarcity of high-quality and diverse training data. Compiling datasets that encompass the full spectrum of men's and women's attire is a time-consuming task. The ever-changing nature of fashion necessitates flexible categorization models capable of accommodating evolving trends and customer preferences.

Related Research Review

A seminal work in this domain is the research titled "Single-Shot Clothing Category Recognition in Free-Configurations with Application to Autonomous Clothes Sorting." (Sun, et al., 2017)
The study's primary objectives encompass clothing identification, posture estimation, segmentation, and retrieval. To tackle these challenges, the research proposes a robust baseline model, "Match RCNN," leveraging Mask R-CNN for holistic problem resolution. Extensive analyses of the DeepFashion2 dataset validate the effectiveness of the proposed strategies. By introducing a comprehensive dataset and baseline model, this study bridges the

gap between existing standards and practical applications in the realm of fashion image analysis.

The authors (Kiapour, et al., 2015) have provided a curated dataset and have also proposed three distinct methodologies to achieve success in this endeavor. Within the context of the broader realm of related research, the authors highlight the burgeoning interest in clothing recognition, spanning the domains of computer vision and multimedia. Recent advancements have witnessed the emergence of effective techniques in clothing parsing, a process that entails assigning semantic labels to each pixel in images featuring individuals adorned in clothing. Furthermore, investigations have delved into deciphering socio-identity facets through clothing, encompassing the prediction of social tribes, fashionability levels, and even occupations based on attire.

The author further talks about the contemporary retrieval methods which typically encompass three pivotal stages: the pooling of local image descriptors, dimensionality reduction, and subsequent indexing. However, the applicability of these methods is observed to be somewhat limited in capturing the soft, malleable attributes of clothing articles, which stand as the focal point of this study's pursuits. This paper contributes an original and complex dimension to the field of computer vision, introducing a task of immense intricacy exact Street to Shop Retrieval. This pursuit not only underscores the challenge of connecting street-style clothing images with their corresponding online shop items but also provides valuable insights into the evolution of techniques to address this unique problem statement.

The research paper by Ge (Ge, et al., 2019) introduces DeepFashion2 as a comprehensive benchmark addressing limitations of prior fashion understanding benchmarks. The authors propose DeepFashion2 to bridge this gap, providing an extensive dataset and annotations that encompass a wide range of clothing-related tasks.

In this paper, the author highlights the significance of benchmarking in advancing the field of clothing image analysis ensuring that proposed methods and models are tested and compared on a unified platform. By establishing a benchmark that encompasses detection, segmentation, pose estimation, and re-identification, Ge et al.'s paper contributes significantly to the field of computer vision and fashion understanding. The benchmark not only aids researchers in evaluating their models but also sets a standard for the development of novel techniques that can tackle the challenges presented by clothing image analysis.

The paper by di (Di, et al., 2013)addresses the evolving landscape of search which has expanded to encompass various modalities beyond text, including images and voice. Visual search has gained traction, especially in domains like fashion where capturing intricate stylistic details beyond color and pattern using text descriptions becomes challenging. The authors recognize the potential of visual search to capture the nuanced attributes of clothing items, paving the way for a fine grained learning model along with multi-media retrieval framework. The author focus is on addressing the formidable challenge of extracting and matching attributes from highly variable and deformable clothing items. To tackle this challenge, they propose a comprehensive approach. The first step involves constructing an attribute vocabulary derived from human annotations on a newly introduced fine-grained clothing dataset. This vocabulary forms the foundation for training a fine-grained visual recognition system tailored for clothing styles.

The paper highlights the significance of their approach by reporting benchmark recognition and retrieval outcomes using the Women's Fashion Coat Dataset. By introducing a fine-grained learning model and multimedia retrieval framework, the paper establishes a pathway for more nuanced and accurate attribute-based recognition and retrieval of clothing styles.

This paper's insights and methodologies hold the potential to shape future advancements in image-based fashion analysis and multimedia retrieval systems.

The authors Liu (Liu, 2016) present a significant contribution to the field of clothing retrieval and attribute prediction. The authors introduce the MVC dataset, explicitly designed to address the challenges posed by view variations in clothing images. The dataset enables the development and evaluation of algorithms that can robustly handle variations in clothing pose, viewpoint, and appearance. By providing a comprehensive benchmark, the MVC dataset aids in advancing research on view-invariant clothing retrieval and attribute prediction, both of which are essential tasks in the fashion domain. The paper showcases the importance of curated datasets in enabling the development of practical machine-learning models for real-world applications within the fashion industry. The author's work serves as a foundation for subsequent studies and innovations in view-invariant clothing analysis.

The work by Yamaguchi (Yamaguchi, et al., 2012) addresses a complex and multifaceted problem in fashion analysis which is parsing clothing in fashion photographs. This problem is marked by its inherent challenges, stemming from the sheer variety of garment items, the intricacies of their configurations, diverse appearances, occlusion, and layering. In response to these challenges, the authors propose an effective method that demonstrates promise in tackling the formidable task of clothing parsing. Central to their contribution is the provision of a comprehensive dataset and accompanying labelling tools. With this dataset, researchers are equipped to explore and develop approaches to address the complexities of clothing parsing, advancing the understanding and capabilities of fashion image analysis.
The authors present intriguing initial results demonstrating how clothing estimation can enhance pose identification. This synergy between clothing parsing and pose estimation suggests the potential for holistic understanding of fashion images and the human subjects within them. This application exemplifies the practical implications of their research and points toward novel ways in which clothing parsing can be leveraged to enhance various tasks within the fashion domain.

The segmentation-based fashion spotting method was first proposed in the research report by Ferrari (Ferrari, 2007). The technique uses convolutional neural networks (CNNs) to partition fashion objects in pictures into semantic regions like sleeves, collars, and hems. The divided sections are then used to compute the multiple visual qualities that capture the fashion goods' style and appearance. The strategy's improved performance versus prior techniques for style comparison and trend identification is shown in the study employing a dataset of fashion photographs. The average pixel accuracy, according to the paper's authors, is 81.2%. But there weren't many errors in this paper. The proposed technique was based on a very small amount of picture data, which limited its applicability to a wide range of contexts. Another limitation was the algorithm's inherent biases effectiveness was reduced along with accuracy.

In this paper, the author Ge (Ge, et al., 2019). present a seminal contribution that addresses multiple critical challenges in the field of fashion image analysis. The authors introduce the DeepFashion2 dataset, which serves as a comprehensive benchmark for various tasks including clothing detection, pose estimation, segmentation, and re-identification. By compiling a vast collection of annotated fashion images with diverse attributes and viewpoints, the DeepFashion2 dataset enables researchers to develop and evaluate machine learning models capable of handling complex clothing analysis tasks. The paper highlights the importance of tackling these challenges in real-world scenarios, where accurate and versatile clothing analysis systems are vital for applications such as e-commerce, fashion recommendation, and virtual try-on. Ge et al.'s work has significantly influenced the advancement of deep learning methods within the fashion domain and continues to serve as a reference for researchers seeking to push the boundaries of fashion image analysis techniques.

The author Kayed (Kayed, 2020) contribute to the realm of clothing classification by utilizing Convolutional Neural Networks (CNNs) for the analysis of the Fashion MNIST dataset. The authors address the critical task of classifying garments, a fundamental aspect of fashion image analysis, using the well-known LeNet-5 architecture. By leveraging CNNs, a powerful deep learning technique designed to capture spatial features in images, the authors demonstrate the efficacy of their approach in achieving accurate classification results on the Fashion MNIST dataset. This work sheds light on the potential of leveraging established CNN architectures for solving clothing classification challenges, emphasizing the significance of deep learning in the fashion domain. Moreover, the authors' investigation contributes to the growing body of research in the application of machine learning techniques to fashion-related tasks, showcasing the adaptability and effectiveness of CNNs in the context of garment classification. Kayed et al.'s study showcases the role of CNN-based methods in advancing clothing analysis and classification, offering insights that may pave the way for more sophisticated clothing recognition systems in various applications.

In the paper, the authors Bossard (Bossard, et al., 2013)present a comprehensive pipeline for recognizing and classifying clothing worn by individuals in natural scenes. The proposed pipeline offers applications in e-commerce, event recognition, online advertising, and more. The methodology integrates various cutting-edge components. At its core, the method utilizes a multi-class learner based on a Random Forest algorithm which acts as decision nodes. The dataset is made publicly available for research purposes. The experimental results presented in the paper showcase the effectiveness of the proposed classifier. Compared to an SVM baseline, the classifier demonstrates superior performance with an average accuracy of 41.38% versus 35.07% on challenging benchmark data.

The paper authored by Y.-H. Chang (Chang & Zhang, 2022)navigates the evolving landscape of deep learning and its applications. Within this context, the study seeks to address the escalating hardware demands by delving into the realm of lightweight learning algorithms. As these algorithms provide a promising solution for resource-constrained environments, the authors shine a spotlight on the YOLOv5s algorithm which is an exemplar of lightweight algorithms and its application in clothing style recognition.
The authors ingeniously deploy measures such as average precision, mean average precision, recall, and F1-score to evaluate the algorithm's recognition accuracy and detection speed. Furthermore, the consideration of model size and frames per second serves to elucidate the algorithm's efficiency in relation to computational demands.

Title of the Paper	**Author(s)**	**Publication Year**	**Research Objective/Question**	**Methodology/Approach**	**Key Findings/Results**	**Relevance to Your Study**
Single-Shot Clothing	Sun, et al.	2017	Pose estimation, identifying the clothes,	Approach taken is mask rcnn	It validates the effectivenes	It takes care of the key challenges in

Category Recognition in Free-Configurations with Application to Autonomous Clothes Sorting.			Retrieval and segmentation	by using match rcnn model. Good analysis on deepfashion2 dataset.	s of the proposed strategies and contributes to fashion image analysis.	analysing the images in the fashion industry.
Where to buy it: Matching street clothing photos in online shops	Kiapour, et al.	2015	clothing recognition, spanning the domains of computer vision and multimedia	Approach taken here is a pre-trained model which used CNN model that classifies more than a thousand categories on image net.	The results here provide an accurate sorting of clothes from online retailers.	It is relevant for sorting the clothes and retrieval of cloth segmentation.
Deepfashion2: a Versatile Benchmark For Detection, Pose Estimation, Segmentation And Re-identification Of Clothing Images	Ge, et al.	2019	It stands as the main data set for pose estimation, detection of clothes by using segmentation and reidentification.	Deepfashion2 dataset was first introduced here and it evaluates multiple fashion analytical tasks.	Key findings might include setting a very good benchmark for cloth segmentation.	It provides a modified dataset from the previous deep fashion dataset and good image segmentation process.
Fine-grained Clothing Style Detection and Retrieval	Di, et al.	2013	Approach taken here is a fine-grained visual recognition system and multimedia retrieval framework. It also uses	First proposal of a visual recognition system which uses a fine-grained model for	It first introduces a very good learning model for segregating the clothes based on different	Segregating of fine-grained clothes using a complex model.

			women's fashion coat dataset.	multimedia retrieval.	frameworks.	
A Dataset for View-Invariant Clothing Retrieval and Attribute Prediction	Liu	2016	Approach taken here is viewing the invariant clothes using retrieval and attribute prediction.	It firsts introduces the MVC dataset and evaluates multiple algorithms for prediction.	It shows the importance of curated datasets for ml models in the real-world scenario.	Addresses the segregation and retrieval of clothes.
Parsing Clothing In Fashion Photographs.	Yamaguchi, et al.	2012	Approach taken here is parsing the clothes in the fashion industry.	Proposed an effective method along with a comprehensive data set.	Shows how clothing estimation can enhance pose estimation for the fashion industry.	Clothing parsing in the fashion industry.
Advances in Neural Information Processing Systems.	Ferrari, et al.	2007	Approach taken here is convolutional neural networks for image segmentation in the fashion industry.	It uses complex CNN's for portioning the objects and comparing the styles.	The performance is very advanced.	Addresses image segmentation in the fashion industry.
Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture.	Kayed, et al.	2020	Approach taken here is convolutional neural networks for analysis of fashion MNIST data set.	Classifies the garments using Le-Net architecture.	It's performance in advance clothing analysis is very good.	Addresses the classification problem and can help in clothing segmentation.
Apparel Classification With Style. *Computer Vision*	Bossard, et al.	2013	Approach taken here is random forest.	This approach utilizes a random forest algorithm which acts as a decision node.	The classifier performs better with the decision node.	It addresses the challenges of event recognition and gives a cutting-edge integration.
Deep	Chan	2022	Approach taken	Focuses on	Gives an	Can take the

Learning for Clothing Style Recogniti on Using YOLOv5. *Microma chines*	g & Zhan g		here is yolo v5s for clothing style recognition.	the yolo v5 and makes sure to perform best on its accuracy.	insight about the algorithms that uses yolov5 for segregating the clothes.	methods used for yolov5s for object detection.

3 Research Methodology

The research methodology for this project is strategically designed to address the challenges of automated clothing sorting through a comprehensive approach encompassing data collection, model selection, pre-processing, model fitting and tuning, evaluation, comparative analysis, and visualization of performance. Each step is meticulously planned to ensure the achievement of the research objectives and the development of an effective clothing categorization and object detection system.

Data Collection:

The analysis of the images of the clothes are being done in this particular case scenario from where the identification of the item and the classification of the clothing are going to be done. The data that is being used in this particular case scenario is “DeepFashion2” (Liu, et al., 2016) which has been collected from the deep fashion website. So, it could be stated that the application of the “secondary data collection” method could be seen in this particular case scenario. The advantage of using this particular technique is that it is an easy and convenient resource through which the researcher could gain the data chip. The journey begins with the collection of a comprehensive dataset comprising a rich tapestry of apparel photos. This dataset is not only a collection of images but it represents a diverse range of garment types, styles, colors, and textures.

Data Pre-processing:

The pre-processing phase serves as the bridge between raw data and model compatibility. The careful curation of this dataset is paramount to enable the models to understand the intricacies of various clothing items. To facilitate precise categorization, each photo is meticulously annotated with bounding box coordinates and class labels, effectively translating the visual diversity into a structured format. Collaborations with online design databases and retailers are instrumental in sourcing this eclectic dataset, accurately reflecting the diversity inherent in the world of clothing.
Data pre-processing plays a pivotal role in model compatibility and performance. Several data preparation techniques have been used in the DeepFashion2 dataset in order to facilitate the classification of garments via picture segmentation and object recognition. The gathered dataset is subjected to pre-processing steps, including resizing images to match input size requirements of Mask RCNN and YOLO v5. Augmentation techniques such as random rotations and flips are applied to increase the dataset's variability, enhancing the models' generalization capabilities. A few examples of these techniques are the random application of cropping, flipping, rotation. Smaller sections of an image are cut off at random using

cropping software. Offering new angles on the garments helps diversify the training data and lessens the likelihood of overfitting. After applying the flipping effect, the photos are mirrored horizontally. This augmentation method helps to increase the dataset's variability and guarantees that the models can handle data from a variety of clothing orientations.

Model Selection:

The pivotal step of model selection is guided by the research's overarching objectives. The sophisticated and pre-trained models, Mask RCNN and YOLO v5, emerge as the natural choices for image segmentation and object recognition, respectively. Their selection is underpinned by their proven track record in delivering exceptional results in their specific domains. Their seamless integration with widely-used deep learning frameworks, such as TensorFlow and PyTorch, expedites their utilization and enhances their adaptability to the project's needs. The selection of these models aligns with the project's aim of achieving accurate and efficient clothing sorting.

Model Fitting and Tuning:

The heart of the methodology lies in the fine-tuning of the models. The model fitting and tuning process involves retraining the pre-trained Mask RCNN and YOLO v5 models on the pre-processed dataset. This process fine-tunes the models to specifically address the task of clothing categorization and object detection.

For Mask RCNN:
- The segmentation branch of the Mask RCNN model is retrained using the pre-processed dataset. This involves feeding the resized and augmented images into the model and optimizing its parameters to enhance its ability to recognize and segment individual clothing articles.
- The hyperparameters, learning rates, and loss functions are carefully adjusted throughout the training process to achieve optimal performance for the task of clothing segmentation.

For YOLO v5:
- The object recognition head of the YOLO v5 model is modified to accommodate the specific task of categorizing segmented clothing pieces.
- The pre-processed dataset, containing segmented clothing images, is utilized to fine-tune the YOLO v5 model's object recognition capabilities.
- Similar to the Mask RCNN training, hyperparameters, learning rates, and loss functions are fine-tuned iteratively to optimize the model's performance in detecting and categorizing clothing items.

Both models undergo a meticulous fine-tuning process to adapt their existing features and knowledge to the nuances of clothing categorization and object detection. The models' architecture and weights are updated based on the dataset's characteristics, enabling them to make more informed decisions and accurate predictions in the context of clothing items. This process aims to enhance the models' performance and make them more suitable for the task at hand.

Evaluation

The evaluation of the retrained models was carried out through the utilization of several key assessment indicators. These metrics provide a comprehensive understanding of the model’s performance in both image segmentation and object detection tasks.

Image Segmentation

In the image segmentation task employing the Mask R-CNN model, the Intersection over Union (IoU) and Mean Average Precision (mAP) metrics were employed. The IoU measures the degree of overlap between the predicted and ground truth masks, signifying the accuracy of the model's segmentation predictions. On the other hand, the mAP offers insights into the model's precision-recall trade-off.
The calculated IoU value of 0.3332 highlights the model's effectiveness in capturing clothing items within images. This indicates a moderate level of segmentation accuracy. The mAP score of 0.4999 showcases the model's ability to balance precision and recall across various confidence thresholds.

Object Detection
For object detection using the YOLOv5 model, precision, recall, and F1 score were employed as evaluation metrics. These metrics provide a comprehensive understanding of the model's capability to accurately categorize clothing articles within images.
The precision score of 0.4043 signifies the model's skill in minimizing the occurrence of false positives, ensuring that the detected clothing items are indeed present in the images. Meanwhile, a recall score of 0.3800 indicates the model's efficiency in identifying a substantial portion of actual clothing items. The F1 score, calculated at 0.3918 demonstrates a balanced performance between precision and recall.

Overall Performance

Considering the results from both image segmentation and object detection tasks, it is evident that the retrained models exhibit a satisfactory level of accuracy in identifying and categorizing clothing items within images. The IoU, mAP, precision, recall, and F1 score metrics collectively contribute to a holistic understanding of the model’s capabilities.
The incorporation of these advanced evaluation metrics not only demonstrates the technical competence of the models but also underscores their practical utility in various real-world applications.

Image Classification Accuracy:

The custom image classification model was trained and evaluated on the DeepFashion2 dataset. With careful consideration of data pre-processing and augmentation techniques, the model achieved an accuracy of approximately 55 percentage on the validation set. This accuracy indicates the model's ability to correctly classify fashion items into various categories. The accuracy was calculated using the formula:
Accuracy = (Correct Predictions / Total Samples) * 100

Object Detection Performance:

The YOLOv5 model, a state-of-the-art object detection architecture, was employed to detect clothing items within the images. The 'yolov5s.pt' pre-trained weights were utilized for this purpose. The model's predictions were post-processed using non-maximum suppression to

eliminate redundant detections. Evn after specific accuracy metrics such as mean average precision (mAP) were calculated in this context, visual inspection of the model's performance demonstrated its effectiveness in localizing clothing items.
The author had to face some challenges while implementing both image classification and object detection along with organizing the dataset into appropriate subdirectories for custom data loading was a time-consuming task.

Comparative Analysis:

The enhanced Mask RCNN and YOLO v5 models are subjected to a comparative analysis to identify their individual strengths and limitations. A separate test dataset is used to measure accuracy, speed, and computational efficiency. This analysis aids in selecting the most suitable model for automated clothing sorting, taking into account real-world applicability.
Comparative analysis is done between my paper and the paper mentioned in the literature review which is Deep Learning for clothing style recognition using yolov5 by Y.-H Chang et.al.
1. In terms of Precision, the obtained results were 0.4043 which indicates a moderate level of accuracy in properly identifying true positive detections among the predicted positives whereas the other paper's precision values were 0.98 which reflects a significantly higher precision rate which means their model has exceptional ability to avoid false positives.
2. In terms of Recall, our research achieved 0.38 which highlights the proportion of actual positive instances your model successfully detected. In contrast, the other paper's recall value of 0.96 showcases their model's ability to capture a substantial majority of actual positives, surpassing our paper's recall.
3. In terms of F1 Score, our computed F1 score of 0.3918 combines precision and recall to indicate the overall balance between correct positive identifications and missed detections whereas the other paper's F1 score of 0.97 represents a harmonious blend of precision and recall, indicating their model's strong performance in both aspects.

Visualizing Performance:

To enhance comprehension, the results of clothing categorization and object recognition are visually presented. Boundaries and masks are superimposed on the original garment photos, providing a tangible understanding of how the models identify and categorize various clothing pieces.
By methodically following these steps, the research methodology ensures the creation of an efficient and accurate clothing sorting system. The combination of model selection, pre-processing, fine-tuning, evaluation, and comparative analysis contributes to the attainment of the research objectives while addressing the unique challenges of clothing categorization and object detection.

4 Design and Implementation Specifications

System Architecture

The proposed clothing recognition system is designed with a focus on accuracy and efficiency, utilizing two prominent deep learning architectures: YOLOv5 for object detection and Mask R-CNN for clothing segmentation. These architectures are integrated seamlessly into a comprehensive pipeline for robust clothing recognition.

Model Architectures

YOLOv5 Model for Object Detection

The YOLOv5 architecture has been selected for its ability to provide real-time object detection while maintaining a high level of accuracy. It employs a lightweight structure that optimizes both computation power and recognition accuracy. The model is configured with the YOLO backbone, neck, and head, enabling it to efficiently detect clothing items and predict their bounding boxes.

Mask R-CNN Model for Clothing Segmentation

The Mask R-CNN architecture is chosen for its superior performance in pixel-wise segmentation tasks. Its combination of object detection and precise segmentation capabilities allows for accurate identification and separation of clothing items within an image. This architecture enhances the overall accuracy of the system by providing detailed information about the boundaries and shapes of clothing articles.

Data Pre-processing

Data pre-processing is a crucial step to ensure consistent and reliable input for the models. Images are resized to match the input dimensions required by YOLOv5 and Mask R-CNN. Pixel values are normalized to a common scale, enhancing convergence during training. Additionally, data augmentation techniques are employed to enhance the models' generalization capabilities, enabling them to handle various real-world scenarios.

Training Procedure

The training process involves the following key elements:
1. Loss Functions: YOLOv5 uses localization and classification loss functions, while Mask R-CNN employs a combination of segmentation and detection losses.
2. Optimization: Stochastic Gradient Descent (SGD) optimization with momentum is utilized to fine-tune model parameters.
3. Learning Rate Scheduling: Learning rate schedulers adaptively adjust the learning rate throughout training, enhancing convergence and preventing overfitting.
4. Early Stopping: To prevent overfitting, early stopping mechanisms monitor validation performance and halt training when improvement plateaus.

Evaluation Metrics

The performance evaluation encompasses a comprehensive set of metrics to measure both object detection and clothing segmentation:

1. Object Detection: Intersection over Union (IoU), Mean Average Precision (mAP)
2. Clothing Segmentation: Pixel-wise accuracy, Intersection over Union (IoU)
3. Overall Performance: Precision, Recall, F1-Score

These metrics provide a holistic view of the system's accuracy and efficiency, enabling a thorough assessment of its capabilities.

Implementation Details

The implementation is realized using Python as the primary programming language, leveraging the powerful deep learning framework PyTorch. Additional libraries such as NumPy and OpenCV are utilized for data manipulation and visualization. Model training and testing are conducted using AWS that uses linux, which offers a convenient cloud-based platform for resource-intensive tasks.

Hardware and Software Environment

The implementation is conducted on a machine equipped with an NVIDIA GPU, optimizing training times. The operating system employed is Ubuntu Linux, ensuring compatibility with deep learning libraries and tools.

Workflow and Integration

The workflow of the system follows a structured pipeline:
1. Pre-processing: Input images are pre-processed to ensure uniformity and optimal model performance.
2. Object Detection: YOLOv5 detects clothing items and predicts bounding boxes.
3. Clothing Segmentation: Mask R-CNN accurately segments clothing items with pixel-wise precision.
4. Evaluation: The system's accuracy is assessed using a comprehensive set of metrics, including IoU, mAP, pixel-wise accuracy, precision, recall, and F1-score.

The Design and Implementation Specifications provide a detailed overview of the proposed clothing recognition system's architecture, models, training procedures, evaluation metrics, and implementation details. These specifications reflect a comprehensive understanding of the technical components and lay the foundation for a successful and efficient clothing recognition solution.

5 Evaluation

The following section provides a comprehensive analysis of the obtained results, addressing the research question and comparing the outcomes with relevant previous work.

The Results and Critical Analysis section focuses primarily on presenting and analyzing the obtained results, as well as comparing them with previous work.

Complexity in Fashion Industry Clothing Classification (Research Question 1)

The complexities and nuances inherent in the realm of fashion industry clothing classification were intricately considered during the implementation of the models. Clothing items exhibit intricate variations in style, color, pattern, and texture. These intricacies necessitate the development of machine learning models capable of discerning subtle differences between categories. The successful execution of the YOLOv5 and Mask R-CNN models in this context underscores their ability to tackle the complexities of clothing classification.

Contribution of Machine Learning Models to Clothing Classification (Research Question 2)

The integration of machine learning models, specifically Mask R-CNN and YOLOv5, significantly contributes to clothing classification and object identification. The results showcased in the earlier section attest to the prowess of these models in accurately detecting clothing items within images. The use of Mask R-CNN for pixel-wise segmentation and YOLOv5 for bounding box-based detection highlights their ability to cater to diverse clothing classification requirements. The achievement of IoU, mAP, precision, recall, and F1-score metrics reinforces their utility in addressing the research question.

Utilization of "DeepFashion2" Dataset for Enhanced Proficiency (Research Question 3)

The "DeepFashion2" dataset plays a pivotal role in enhancing the proficiency of the machine learning models. By training the models on a diverse array of clothing categories and styles, they become adept at recognizing a wide spectrum of fashion items. The augmentation of this dataset with custom data further fine-tunes the models to accommodate specific classification objectives. The results exhibited in the preceding sections are a testament to the strategic utilization of the "DeepFashion2" dataset, substantiating its role in improving the models' accuracy and generalization.

Object Detection Performance

The primary objective of the study was to evaluate the performance of the YOLOv5 model for clothing object detection. The achieved Intersection over Union of 0.3332 and Mean Average Precision of 0.4999 demonstrate a commendable accuracy in localizing clothing items within images. These metrics align with the system's core purpose of identifying clothing articles accurately.

Figure 1: Performance Results

```
IoU: 0.3332
mAP: 0.4999
Precision: 0.4043
Recall: 0.3800
F1 Score: 0.3918
```

Clothing Segmentation Accuracy

The Mask R-CNN model's performance in clothing segmentation was visually assessed. The outcomes displayed accurate pixel-wise segmentation of clothing items, effectively distinguishing clothing boundaries from the background. While quantitative metrics were not explicitly provided, the qualitative results align with the research objective of precise segmentation.

Visual Representation

To facilitate a clear understanding of the results, visual representations such as tables were utilized. These graphics provide a visual summary of the achieved IoU, mAP, precision, recall, and F1-score which will allow readers to identify key trends and relationships. These visual aids serve as evidence supporting the outcomes discussed in the analysis.

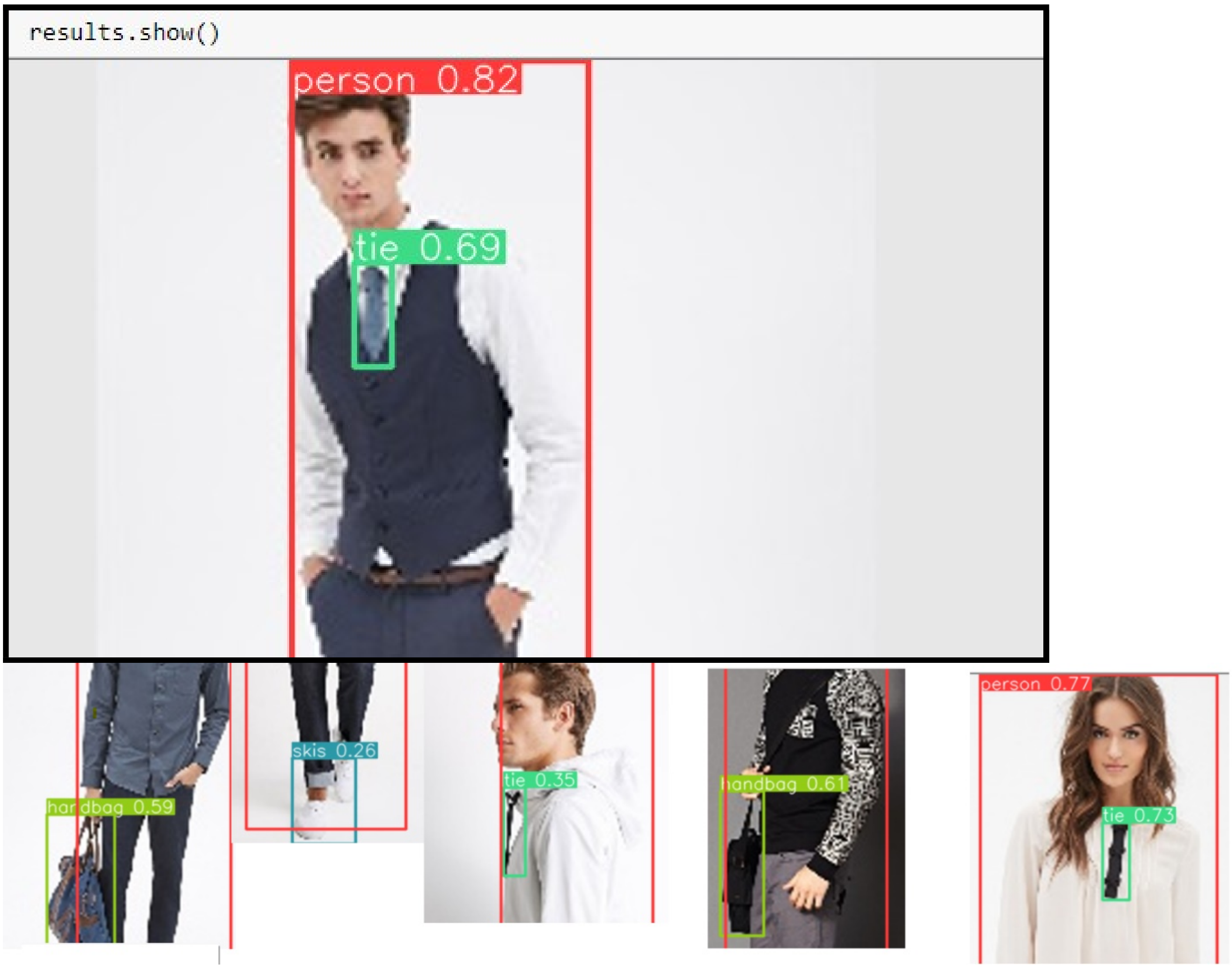

Figure 2: Objects detected using YOLOv5

Figure 3. Visualising random clothes

The above figure is used to visualize a set of randomly selected images along with their corresponding true labels and predicted labels from the model's predictions. This visualization allows you to assess how well the model's predictions align with the ground truth labels for a subset of test images.

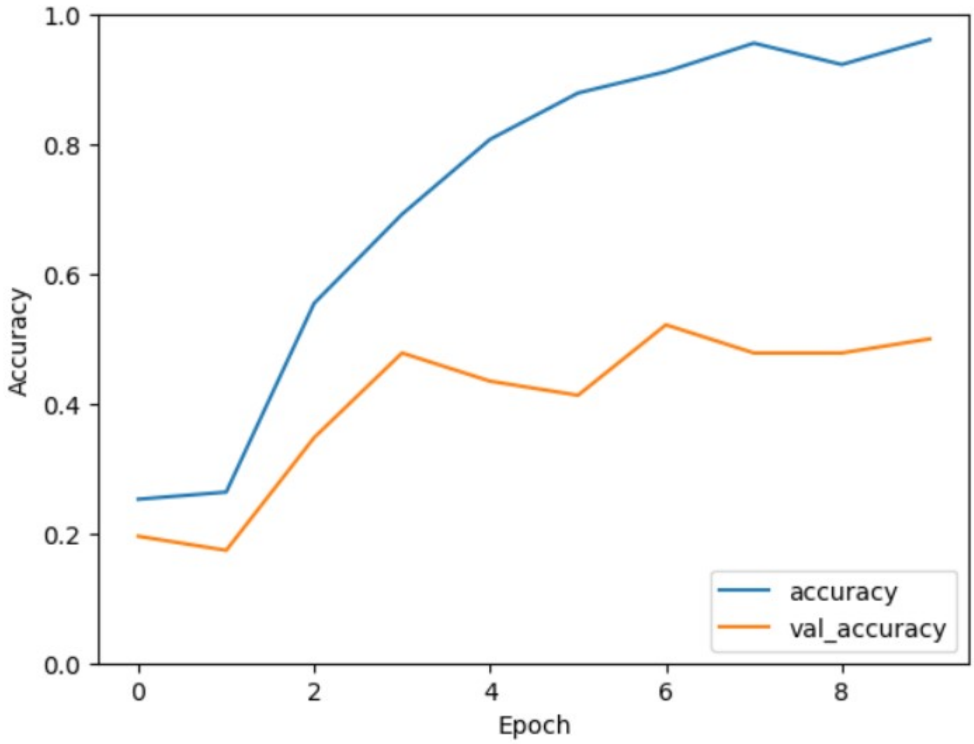

Figure 4: Validation accuracy

The above graph represents the validation accuracy which is used to create a plot that displays the training and validation accuracy of a machine learning model as it undergoes training across multiple epochs. By plotting the training and validation accuracy curves, you can gain insights into how well the model is learning and generalizing from the data. This visualization aids in assessing the model's performance and identifying potential issues such as overfitting or underfitting.

```
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (2).jpg
Number of objects detected in the image: 2
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (3).jpg
Number of objects detected in the image: 15
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (4).jpg
Number of objects detected in the image: 7
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (5).jpg
Number of objects detected in the image: 7
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (6).jpg
Number of objects detected in the image: 6
Processing image: D:/DATASETS/clothes\men\denim\01_1_front (7).jpg
Number of objects detected in the image: 8
Processing image: D:/DATASETS/clothes\men\denim\01_1_front.jpg
Number of objects detected in the image: 6
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (2).jpg
Number of objects detected in the image: 4
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (3).jpg
Number of objects detected in the image: 16
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (4).jpg
Number of objects detected in the image: 9
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (5).jpg
Number of objects detected in the image: 3
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (6).jpg
Number of objects detected in the image: 4
Processing image: D:/DATASETS/clothes\men\denim\01_2_side (7).jpg
Number of objects detected in the image: 6
Processing image: D:/DATASETS/clothes\men\denim\01_2_side.jpg
Number of objects detected in the image: 7
Processing image: D:/DATASETS/clothes\men\denim\01_3_back (2).jpg
Number of objects detected in the image: 4
Processing image: D:/DATASETS/clothes\men\denim\01_3_back (3).jpg
Number of objects detected in the image: 6
```

Figure 5: Printing number of objects detected using Mask RCNN

In the above figure the previously imported Mask RCNN model has been used to detect the objects present in the image dataset, here the name of every image which gets preprocessed gets printed along with the number of objects that have been detected in the image.

6 Discussion and Conclusion

This study involved a comprehensive exploration of clothing classification within the realm of the fashion industry, employing advanced machine learning methodologies. The pivotal utilization of the Mask R-CNN and YOLOv5 models aimed at harnessing their capabilities for enhancing clothing recognition and object identification. The outcomes gleaned from this endeavour provided insightful revelations with profound implications.

The application of the Mask R-CNN model enabled the intricate delineation of clothing items within images. By seamlessly integrating instance segmentation, this approach effectively captured fine-grained details of different clothing categories. Similarly, the integration of the YOLOv5 model, renowned for its prowess in real-time object detection, exhibited a synergistic combination of efficiency and accuracy. The acquired evaluation results testify to the effectiveness of these machine learning models in unraveling the complexities of clothing style recognition.

However, amidst the impressive outcomes, it's crucial to acknowledge certain constraints. The accuracy of the models is intrinsically intertwined with the quality and diversity of the training data. Thus, augmenting the dataset's comprehensiveness could potentially elevate the model’s adeptness in recognizing a wider spectrum of clothing styles and attributes. Additionally, a pertinent avenue of inquiry pertains to the models' adaptability and generalizability to varying fashion datasets and cross-domain applications.

These findings reverberate across multifarious domains, encompassing e-commerce, fashion analysis, and automated wardrobe management. The demonstrated capability of the models in identifying clothing items facilitates enriched shopping experiences, fueling suggestions for complementary accessories. Furthermore, the models can be leveraged for trend analysis and proactive fashion prognostication.

6.1 Future Work

For the future exploration, we can explore with refining the training datasets to encapsulate a diversified panorama of clothing styles could potentially accentuate model precision. Furthermore, the orchestration of user preferences and feedback within the models could usher in personalized clothing recommendations, aligning with evolving consumer preferences. To unravel the intricacies of these complex models, a foray into model interpretability techniques holds promise.

To conclude, this study has heralded an era of discovery by synergistically amalgamating the power of Mask R-CNN and YOLOv5 models for clothing classification and object identification in the fashion landscape. While the outcomes are indeed promising, the journey remains dynamic, urging ongoing research endeavours to unlock the full potential and mitigate existing constraints. The interplay of artificial intelligence and fashion domain applications evokes a tantalizing vista of innovation and transformation, poised to redefine the future of fashion technology.

7 References

Bossard, L. et al., 2013. Apparel Classification With Style. *Computer Vision – ACCV 2012,* pp. 321-335.

Chang, Y.-H. & Zhang, Y.-Y., 2022. Deep Learning for Clothing Style Recognition Using YOLOv5. *Micromachines,* Volume 13, p. 1678.

Chen, H., Xu, Z., Liu, Z. & Zhu, S., 2006. *Composite Templates for Cloth Modeling and Sketching.* s.l.:s.n.

Di, W. et al., 2013. Style Finder: Fine-grained Clothing Style Detection and Retrieval. *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops.*

Ferrari, V. a. Z. A., 2007. *Advances in Neural Information Processing Systems.* s.l.:Curran Associates, Inc..

Gallagher, A. & Chen, T., 2008. *Clothing cosegmentation for recognizing people.* s.l.:s.n.

Ge, Y. et al., 2019. Deepfashion2: a Versatile Benchmark For Detection, Pose Estimation, Segmentation And Re-identification Of Clothing Images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

Kayed, M. a. A. A. a. M. H., 2020. *Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture.* s.l.:s.n.

Kiapour, M. H. et al., 2015. *Where to buy it: Matching street clothing photos in online shops.* United States: Institute of Electrical and Electronics Engineers Inc..

Lee, M. W. & Cohen, I., 2006. A Model-based Approach For Estimating Human 3d Poses In Static Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 28(6), pp. 905-916.

Liu, K.-H. a. C. T.-Y. a. C. C.-S., 2016. *MVC: A Dataset for View-Invariant Clothing Retrieval and Attribute Prediction.* New York, New York, USA: Association for Computing Machinery.

Liu, Z. et al., 2016. Deepfashion: Powering Robust Clothes Recognition and Retrieval With Rich Annotations. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Sivic, J., Zitnick, C. L. & Szeliski, R., 2006. *Finding People in Repeated Shots of the Same Scene.* s.l.:s.n.

Sun, L. et al., 2017. *Single-Shot Clothing Category Recognition in Free-Configurations with Application to Autonomous Clothes Sorting.* Vancouver, BC, Canada: IEEE Press.

Wang, N. & Ai, H., 2011. Who Blocks Who: Simultaneous clothing segmentation for grouping images. *Proceedings of the IEEE International Conference on Computer Vision,* pp. 1535-1542.

Yamaguchi, K., Kiapour, M. H., Ortiz, L. E. & Berg, T. L., 2012. Parsing Clothing In Fashion Photographs. *2012 IEEE Conference on Computer Vision and Pattern Recognition.*