iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.42022

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Fine-Grained Identification of Clothing Apparels

Kushal Kothari[1], Ajay Arjunwadkar[2], Hitesh Bhalerao[3], Savita Lade[4]
[1, 2, 3]*Student,* [4]*Guide, Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai,*

Abstract: ***The rapid use of smartphones and tablet computers, search is now not just limited to text,but moved to other modalities such as voice and image. Extracting and matching this attributes still remains a daunting task due to high deformability and variability of clothing items.***
Visual analysis of clothings is a topic that has received attention due to tremendous growth of e-commerce fashion stores. Analyzing fashion attributes is also crucial in the fashion design process.
This paper addresses the solution of recognition of clothes and fashion related attributes related to it using better image segmentation RCNN based algorithms.
Keywords: Computer Vision, Fine grained identification, Clothing apparel detection, Convolutional Neural Network, Mask RCNN, Detectron2

I. INTRODUCTION

Visual analysis of clothings is a topic that has received increasing attention in computer vision communities in recent years. There's already an oversized body of research on clothing modeling, recognition, parsing, retrieval, and proposals.
The project is completed with the motivation to form recognition of garments and fashion related attributes associated with it. It might be utilized by retail malls or shops to classify clothes and will be employed by companies who analyze and predict fashion related trends .
Analyzing fashion attributes is important within the fashion design process. Current fashion forecasting firms, like WGSN, utilize information from all round the world (from fashion shows, visual merchandising, blogs, etc).
They gather information by experience, by observation, by media scan, by interviews, and by exposure to new things. Such an information analyzing process is named abstracting, which recognizes similarities or differences across all the clothes and collections.
In fact, such abstraction ability is beneficial in many fashion careers with different purposes. Fashion forecasters abstract across design collections and across time to spot fashion change and directions; designers, product developers and buyers abstract across group of clothes and collections to develop a cohesive and visually appeal lines; sales and marketing executives abstract across product each season to acknowledge selling points; fashion journalist and bloggers abstract across runway photos to acknowledge symbolic core concepts that may be translated into editorial features.
Thus our research aims to assist apparel industry professionals to ease their task of analyzing clothing attributes.

II. PROPOSED SYSTEM

Here, the fundamental task is instance segmentation with attribute localization, which unifies instance segmentation (detect and segment each object instance) and fine-grained visual attribute categorization (recognize one or multiple attributes). The proposed project requires both localizing and describing properties for particular clothing instances and identifying its fashion apparels(attributes) In order to solve this challenging task, a novel Attribute-Mask R-CNN model is proposed to jointly perform instance segmentation and localized attribute recognition, and provide a novel evaluation metric for this task.
The core algorithm used is Mask RCNN. It is a deep neural network aimed to solve instance segmentation problems in machine learning or computer vision. In other words, it can separate different objects in an image or a video.It gives you the object bounding boxes, classes and masks when you give it an image.
Thus our project returns exact location and properties of fine grained attributes of clothes, identifying fashion terminology for the style based on a well defined fashion ontology.

III.ALGORITHM USED

Algorithm used for the project is Mask RCNN, which is a deep neural network aimed to resolve instance segmentation problems in machine learning or computer vision. In other words, it can separate different objects in a picture or a video. You provide it a picture, it gives you the item bounding boxes, classes and masks.

©IJRASET: All Rights are Reserved

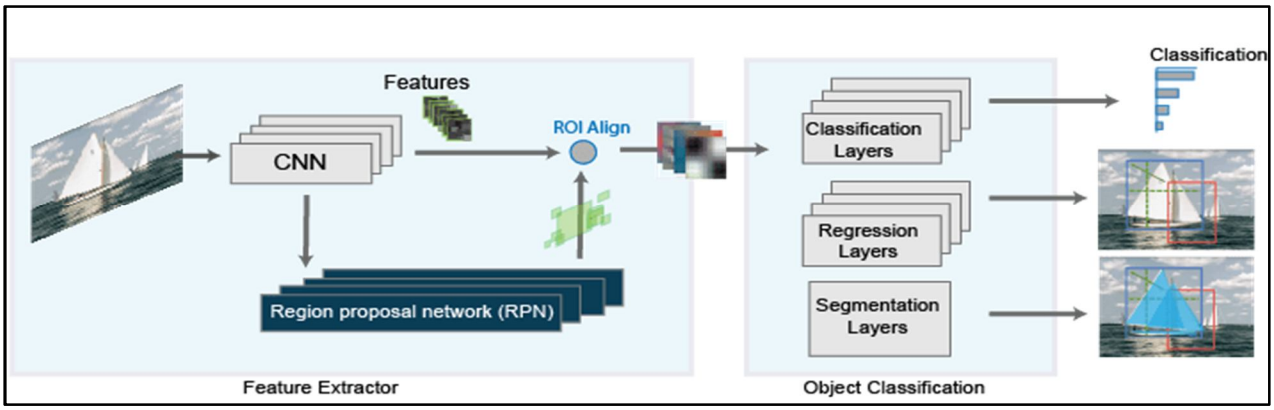


Fig. 1 Mask RCNN

There are two stages of Mask RCNN. First, it generates proposals about the regions where there may be an object supporting the input image. Second, it predicts the category of the item, refines the bounding box and generates a mask in pixel level of the object based on the primary stage proposal. Both stages are connected to the backbone structure.

The first stage is described as follows:
A light weight neural network called RPN scans all FPN top-bottom pathways(hereinafter spoken feature map) and proposes regions which may contain objects. That's all it's. While scanning feature maps is an efficient way, we want a way to bind features to its raw image location. Here come the anchors. Anchors are a collection of boxes with predefined locations and scales relative to photographs. Ground-truth classes(only object or background binary classified at this stage) and bounding boxes are assigned to individual anchors consistent with some IoU value. As anchors with different scales bind to different levels of feature map, RPN uses these anchors to work out where the feature map 'should' get an object and what size of its bounding box is. Here we may agree that convolving, downsampling and upsampling would keep features staying the identical relative locations because the objects within the original image, and wouldn't mess them around.

The second stage is described as follows:
At the second stage, another neural network takes proposed regions by the primary stage and assigns them to many specific areas of a feature map level, scans these areas, and generates objects classes(multi-categorical classified), bounding boxes and masks. The procedure looks the same as RPN. Differences are that without the assistance of anchors, stage-two used a trick called ROIAlign to locate the relevant areas of the feature map, and there's a branch generating masks for every object in pixel level. Work completed.

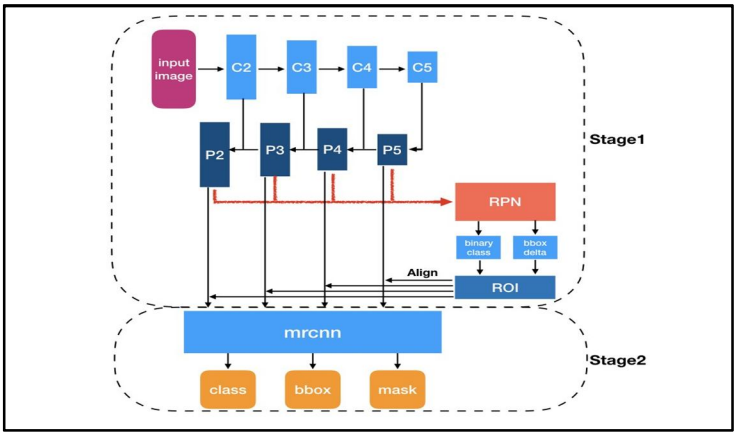


Fig. 2 Illustration of Mask RCNN structure

The most inspiring thing I found about Mask RCNN is that we could actually force different layers within the neural network to find features with different scales, similar to the anchors and ROIAlign, rather than treating layers as black boxes.

IV. METHODOLOGY

A. Mask RCNN Algorithm Steps

The steps involved in Mask RCNN Algorithm are as follows:

1) Initial layers are basically used for feature extraction in the RCNN model.
2) These extracted features are fed into Region Proposal Network(RPN) which gives potential bounding boxes for finding objects faster.
3) The proposed regions and the extracted features are then passed on to the ROIAlign to locate the relevant areas of the feature map, and there is a branch generating masks for each object in pixel level.
4) Finally based on the object detection task the Classification, Instance Segmentation layers are added at the end.

B. Implementation Details

We compared all the variants of backbone for the model available with the help of detectron2 computer vision library in kaggle jupyter notebook environment.

For Faster/Mask R-CNN, we tried out 3 different backbone combinations and each has a resnet of 50 layers. The model is trained for 1000 iterations in each backbone combination. The backbone combinations considered are explained as follows:

1) *FPN:* Use a ResNet+FPN backbone with standard conv and FC heads for mask and box prediction, respectively. It obtains the best speed/accuracy tradeoff, but the other two are still useful for research.
2) *C4:* Use a ResNet conv4 backbone with conv5 head. The original baseline in the Faster R-CNN paper.
3) *DC5 (Dilated-C5):* Use a ResNet conv5 backbone with dilations in conv5, and standard conv and FC heads for mask and box prediction, respectively. This is used by the Deformable ConvNet paper.

TABLE I
Backbone combinations for mask rcnn on fashionpedia dataset

Backbone combination	Box AP	Mask AP	Box AP50	Mask AP50	Box AP75	Mask AP75
FPN	3.427	3.114	6.133	5.139	3.707	3.394
C4	4.669	4.428	8.421	7.041	4.419	5.204
DC5	4.205	4.174	8.188	6.607	3.982	4.735

Based on the analysis we conducted for mask_rcnn_R_50 which uses resnet with depth 50 and a variety of networks as backbone. The selection of backbone would depend on requirements, whether accuracy or faster computation is crucial. The FPN is comparatively faster but for a small number of iterations provides less accuracy, on the other hand DC5 and C4 have slow computation, but better accuracy. In industrial scenarios, where training duration is longer and a balanced dataset is used the accuracy for FPN would theoretically increase and be similar to other backbones. So, FPN backbone combination is better for industrial use cases according to our understanding.

The model implemented in the jupyter notebook environment is downloaded and used to make inferences on input images in web apps developed for end users. The app is developed using streamlit, a python module for data based web app development. The app takes an image of a person wearing a dress as input. App then provides an output image with a bounding box and recognized class. The app also provides details regarding the model used and its accuracy based parameters. Thus it serves as a good interface for naive users to use the model.

V. CONCLUSION AND FUTURE WORK

Recent advances in deep learning have triggered a range of business applications based on computer vision. There are many industry segments where deep learning tools and techniques are applied in object recognition so as to create the business process much faster. The fashion apparel industry is one among them. By presenting the image of any apparel, the trained deep learning model can predict the name of that apparel and this process can be repeated at a much faster speed in so as to tag thousands of apparels in very less time with high accuracy. The proposed project requires both localizing and describing properties for particular clothing instances and identifying its fashion apparels(attributes). In order to solve this challenging task, a Mask R-CNN model is proposed to jointly perform instance segmentation and localized attribute recognition, and provide a novel evaluation metric for this task.

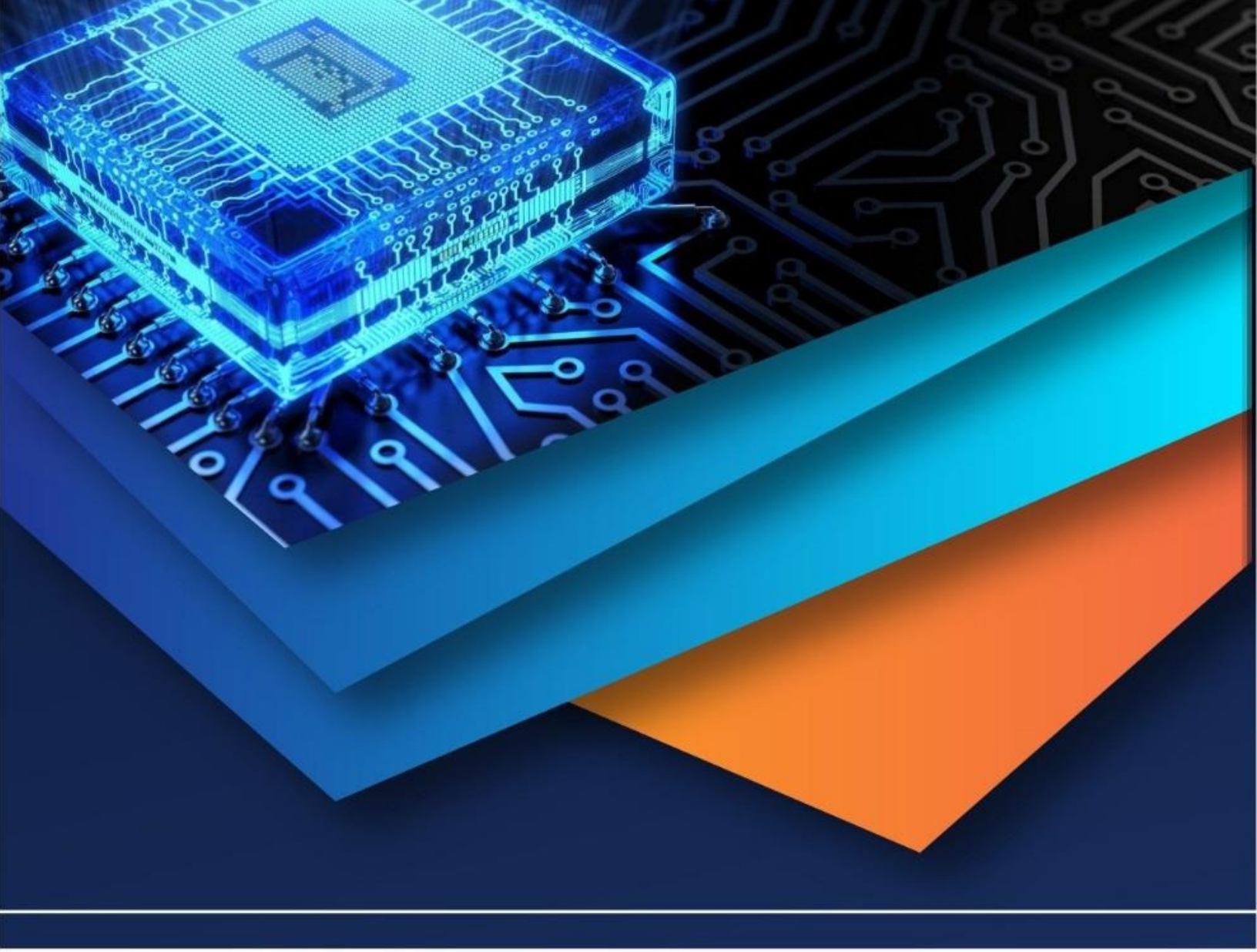
©IJRASET: All Rights are Reserved

Thus our project returns exact location and properties of fine grained attributes of clothes, identifying fashion terminology for the style based on a well defined fashion ontology. The project involves use of the Mask RCNN model, further in future the architecture of Generalized RCNN could be modified to improve accuracy. More attributes of fashion could be included, along with this quality of images can also be improved. The components like head and body could be modified so that lesser training data is used for training. Further an application could be developed which would take in a dataset of fashion trends, detect the attributes and provide detailed statistics for analyzing trends.

REFERENCES

[1] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Bharath Hariharan, Claire Cardie, Serge Belongie, "Fashionpedia TechReport"
[2] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui,Claire Cardie , Bharath Hariharan, Hartwig Adam, Serge Belongie, "Fashionpedia: Ontology, Segmentation, and an Attribute Localization Dataset"
[3] Menglin Jia, Yichen Zhou, Mengyun Shi, Bharath Hariharan, "A Deep-Learning-Based Fashion Attributes Detection Model"
[4] Fashionpedia. [Online]. Available: https://fashionpedia.github.io/home/
[5] Simple Understanding of Mask RCNN. [Online].Available: https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95
[6] Detectron2 Model Zoo and Baselines. [Online]. Available: https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)