applied sciences

MDPI

Article

A Two-Phase Fashion Apparel Detection Method Based on YOLOv4

Chu-Hui Lee * and Chen-Wei Lin

Department of Information Management, Chaoyang University of Technology, 168, Jifeng E. Rd., Wufeng District, Taichung 413310, Taiwan; s10814604@gm.cyut.edu.tw

* Correspondence: chlee@cyut.edu.tw

Abstract: Object detection is one of the important technologies in the field of computer vision. In the area of fashion apparel, object detection technology has various applications, such as apparel recognition, apparel detection, fashion recommendation, and online search. The recognition task is difficult for a computer because fashion apparel images have different characteristics of clothing appearance and material. Currently, fast and accurate object detection is the most important goal in this field. In this study, we proposed a two-phase fashion apparel detection method named YOLOv4-TPD (YOLOv4 Two-Phase Detection), based on the YOLOv4 algorithm, to address this challenge. The target categories for model detection were divided into the jacket, top, pants, skirt, and bag. According to the definition of inductive transfer learning, the purpose was to transfer the knowledge from the source domain to the target domain that could improve the effect of tasks in the target domain. Therefore, we used the two-phase training method to implement the transfer learning. Finally, the experimental results showed that the mAP of our model was better than the original YOLOv4 model through the two-phase transfer learning. The proposed model has multiple potential applications, such as an automatic labeling system, style retrieval, and similarity detection.

Keywords: object detection; YOLOv4; fashion apparel; deep learning; transfer learning

Citation: Lee, C.-H.; Lin, C.-W. A Two-Phase Fashion Apparel Detection Method Based on YOLOv4. *Appl. Sci.* **2021**, *11*, 3782. https://doi.org/10.3390/app11093782

Received: 13 March 2021
Accepted: 20 April 2021
Published: 22 April 2021

1. Introduction

An image is worth a thousand words. The fashion apparel industry is one of the fields that have a large image usage. These clothing images appear in newspapers, magazines, e-commerce platforms, social media, and even advertising boards. There are various object detection applications that extract information from images and apply it to apparel recognition, apparel detection, fashion recommendation, and online search for the fashion apparel industry, which can be used to flourish human life.

In recent years, deep learning technology in computer vision has many applications and researches, for example, image classification [1], object detection [2], semantic segmentation [3], and instance segmentation [4]. Among them, object detection technology is a rapidly developing research topic. Object detection methods and other computer vision algorithms are important to the fashion apparel industry. The clothing information marking system based on the object detection method can automatically mark images in social media or online stores and create information tags [5]. The online store can use the created tags for clothing quick search [5] or similar clothing recommendations [6] in the future. Additionally, occupation recognition [7], fashion style recognition [8], and fashion style recommendation [9] are all extended applications in the field of computer vision.

Clothing images have a massive distinction with different characteristics of clothing appearance, style, and posture [10]. Different types of clothing might be similar in types of material and color. Thus, the computer finds it difficult to recognize different types of clothing. Due to this difficulty, the task of detecting clothing by using computer vision technology becomes a difficult challenge. With the improvement of GPU's computational capabilities, the field of machine learning and deep learning has had a huge breakthrough.

Therefore, deep learning scientists can construct deeper neural networks to solve the unstructured data problem that machine learning algorithms cannot handle.

The mainstream object detection algorithms are based on convolution neural networks (CNN) that are divided into one-stage detection and two-stage detection, by using different feature extraction methods. Object detection algorithms that use a two-stage detection method include R-CNN [11], Fast R-CNN [12], and Faster R-CNN [13], which divide the detection task into (1) region proposal and (2) classification. The two-stage detection method has a high detection accuracy, but the problem of high time complexity leads to a longer detection time. Therefore, the two-stage detection method is hard to use in real-time detecting applications. The one-stage detection method integrates region proposal and classification into one step, which reduces the detection time. The mainstream methods of one-stage detection are SSD [14], YOLO [15], YOLOv2 [16], YOLOv3 [17], and YOLOv4 [18]. These methods improve detection speed without losing too much accuracy and are increasingly becoming popular object detection algorithms.

Transfer learning is the process of applying knowledge and skills learned in previous tasks to target tasks. According to research by Pan et al. [19], transfer learning is divided into three categories by different domains and tasks (1) Inductive Transfer Learning, (2) Transductive Transfer Learning, and (3) Unsupervised Transfer Learning.

In this research, we proposed a fashion apparel detection model based on the You Only Look Once (YOLOv4) algorithm named YOLOv4-TPD (YOLOv4 Two-Phase Detection model for fashion apparel) by using the characteristics of inductive transfer learning. The contributions of this paper are as follows. First, we proposed an object detection model by using two-phase transfer learning for detecting fashion apparel images with complex background. Second, the experimental results proved that implementing two-phase transfer learning and the CLAHE image enhancement method could improve the accuracy of model detection. Third, the detection accuracy of our model was better than the states of the art methods in the field of fashion apparel.

The rest of this paper is organized as follows. Section 2 reviews the related works on apparel image recognition. Section 3 introduces the overview and steps of the YOLOv4-TPD model. Section 4 describes the dataset, experimental, and results. Finally, Section 5 includes the study conclusion.

2. Related Work

In this section, the related works were introduced and divided into three parts--fashion apparel detection, YOLO, and transfer learning. The subsection of fashion apparel detection is the first subsection, where previous research is presented. The second and third subsections explain the technology used in this paper.

2.1. Fashion Apparel Detection

This subsection lists some existing and past research on the fashion apparel field. First of all, the scholars used traditional machine learning methods to classify fashion apparel or detect them in real-time tasks. Chen et al. [20] proposed a seven categories clothing classifier by using scale-invariant feature transform and the Support Vector Machine (SVM). Image features were extracted by the image processing method and the categories were classified by the machine learning algorithm. The model accuracy was between 44% and 81%. Yang and Yu [21] proposed a real-time clothing detection system in 2011. The classification task was constructed by using Linear SVM with the Histogram of Oriented Gradients (HOG) and dividing the categories of clothing into eight categories. As per the results of the detection system, the detection speed achieved was 16–20 fps in the 480p resolution video and the recall of each category was between 29.1% and 94.2%. Surakarin and Chongstitvatana [22] proposed a clothing classifier in 2015, which was constructed by Support Vector Machine (SVM). The classification accuracy of this model was between 57% and 73%. Synthesizing the above research, the classification and detection tasks of fashion apparel were successfully achieved. However, the accuracy of the model was not

satisfactory. The accuracy of traditional machine learning algorithms still has limitations in the field of complex clothing images.

With the advancement of deep learning technology, some problems of traditional machine learning methods were addressed. Two-stage detection through deep learning, R-CNN, Fast R-CNN, and Faster R-CNN were provided at first. Those two-stage detection methods increased the accuracy. However, the shortcomings of the two-stage detection method were also proved in later research. In research by Lao and Jagadeesh [23], they used the R-CNN model to detect apparel images. The accuracy of the model achieved after fine-tuning was 93.4% but the small object "belt" was easy to mix with the complex background causing error detection result. The R-CNN model needs more time for detection than the previous models. Z. Liu [24] compared the Faster R-CNN of the two-stage detection method and the SSD algorithm of the one-stage detection method. The experimental results showed that the SSD algorithm had a comparable detection accuracy and recall to Faster R-CNN, and the detection speed was faster than Faster R-CNN. The one-stage detection algorithm had more advantages in the fashion apparel detection task.

In recent research, the one-stage detection method was used to implement the fashion apparel detection task and achieved satisfactory results. Liu et al. [25] proposed a clothing brand predictive model based on YOLOv3, which detected twenty-five categories of fashion clothing. Experimental results proved that the detection accuracy of YOLOv3 was better than the two-stage detection method R-CNN, Fast R-CNN, and Faster R-CNN. Feng et al. [26] used the YOLOv2 algorithm to detect five categories of clothing on the CCP dataset. The apparel categories were divided into five categories, which were trousers, skirts, coats, T-shirts, and bags. Their model had a good performance in clothing detection, where the average precision and recall achieved were 83.9% and 73%. The detection speed achieved 56 milliseconds per image which was better than the Faster R-CNN which needs 268 milliseconds per image. However, the model had false recognition problems in the images with dark backgrounds. The model might need to enhance the images during the data pre-processing phase to improve this problem. Liu et al. [27] proposed a fashion apparel detection model using YOLOv3. The mAP was 92.98% after modifying and optimizing the model structure. However, the detection object only had skirts. Liu's model could not detect other types of fashion apparel.

In related research on the hierarchical structure, Kumar et al. [28] and Seo et al. [1] define the hierarchical classification to classify the clothing into coarse and fine categories. The experimental results showed that the model training with hierarchical categories could improve the accuracy of object classification.

2.2. YOLO

To solve the problem of slow detection speed of the two-stage detection algorithm, scientists proposed a one-stage detection algorithm. You Only Look Once was a one-stage object detection algorithm based on the CNN architecture. Redmon et al. [15] proposed the first version of the YOLO algorithm in 2016, which treats object detection tasks as a single regression problem. YOLO only needs to perform one convolutional calculation on the image to determine the target category and location. Compared to the two-stage detection method that performs multiple steps, YOLO has the advantage of faster detection speed. Redmon et al. continued to improve the YOLO network in the next two years. They proposed the YOLOv2 [16] in 2017 and YOLOv3 [17] in 2018. In 2020, Bochkovskiy et al. [18] proposed the YOLOv4 that improved on YOLOv3. The YOLOv4 algorithm added several optimization methods to increase the accuracy of target detection, such as data enhancement, normalization methods, data imbalance processing, replacing activation functions, and optimizing bounding box regression problems. The optimization of the network improved the detection accuracy of the network and reduced the requirements for hardware. According to the experimental results of YOLOv4 [18], YOLOv4 obtained 43.5% average precision in the test of the Microsoft COCO dataset. Compared to YOLOv3, YOLOv4 was increased by 10% on average precision and 12% on

detection speed. YOLOv4 integrates many novel optimization technologies. Compared to existing object detection algorithms, the high detection accuracy and detection speed of the YOLOv4 algorithm are among the best object detection algorithms. Therefore, we use the YOLOv4 as the base to implement our Two-Phase Fashion Apparel Detection method.

2.3. Transfer Learning

In research by Pan et al. [19], the definition of transfer learning could be divided into the following three categories—(1) Inductive Transfer Learning, (2) Transductive Transfer Learning, and (3) Unsupervised Transfer Learning. The purpose of transfer learning was to transfer knowledge from the source domain to the target domain that could improve the effect of tasks in the target domain.

According to [19], Table 1 shows the types of transfer learning based on different situations between the source and target, domains and tasks. In inductive transfer learning, both the source domain and the target domain are the same, and the source task is different from the target task. In transductive transfer learning, although the source task and target task are the same, the source domain and the target domain are different. In unsupervised transfer learning, the source domain and the target domain are different but related, and so is the source task and the target task. The definition of transfer learning is as follows—given a source domain Ds and task Ts, a target domain Dt and task Tt. Transfer learning refers to improving the target prediction function ft(·) of the target domain Dt, using the knowledge in the source domain Ds and task Ts, where Ds $\neq$ Dt, or Ts $\neq$ Tt.

Table 1. The definition of transfer learning.

Transfer Learning	Source and Target Domains	Source and Target Tasks
Inductive Transfer Learning	the same	different but related
Transductive Transfer Learning	different but related	the same
Unsupervised Transfer Learning	different but related	different but related

3. Proposed Method

In related work [25], it was found that the YOLOv3 has a better detection efficiency than R-CNN, Fast R-CNN, and Faster R-CNN in fashion clothing detection. Another study [26] showed the YOLOv2 has a faster detection speed than Faster R-CNN, in the field of fashion apparel detection. Synthesizing the above arguments and previous researches, the YOLO algorithm was found to be more suitable than the two-stage detection algorithms for fashion apparel detection. Additionally, the detection model might improve the effect of apparel detection tasks through the characteristics of inductive transfer learning. Therefore, we proposed a two-phase fashion apparel detection model named YOLOv4-TPD, based on the YOLOv4 algorithm and transfer learning, to detect the apparel in complex background. The proposed detection model needs to detect five fashion apparel categories (jacket, top, pants, skirt, and bag) and determine the location of the target in the image, showed in Figure 1. The detection task was to detect apparel features using its contour and appearance. Therefore, the detection effect of the model was not affected by the color of the apparel. Figure 2 shows the architecture of the two-phase detection model. The training process of the proposed model was divided into the data preparation phase and the model training phase. In the data preparation phase, data labeling and data preprocessing were performed to prepare the training data. In the training phase, the prepared training data were used to train the model. First, the three categories classifier was trained. Second, the five categories classifier was trained by using transfer learning.

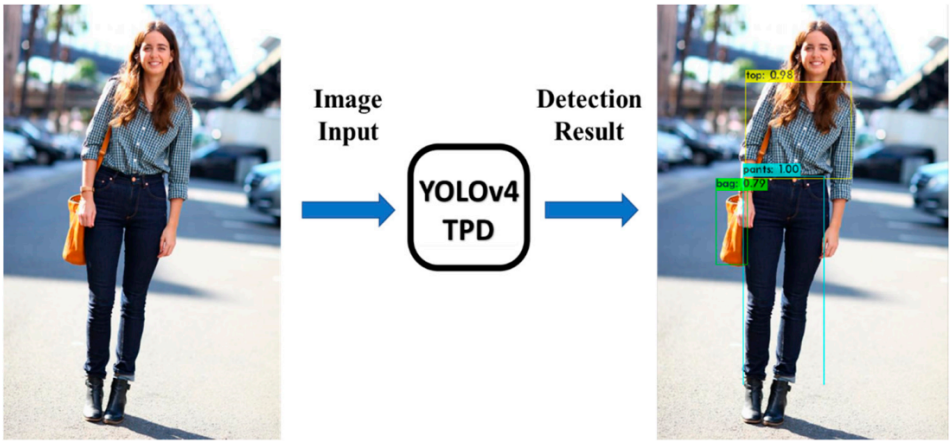

Figure 1. Fashion apparel detection with YOLOv4-TPD.

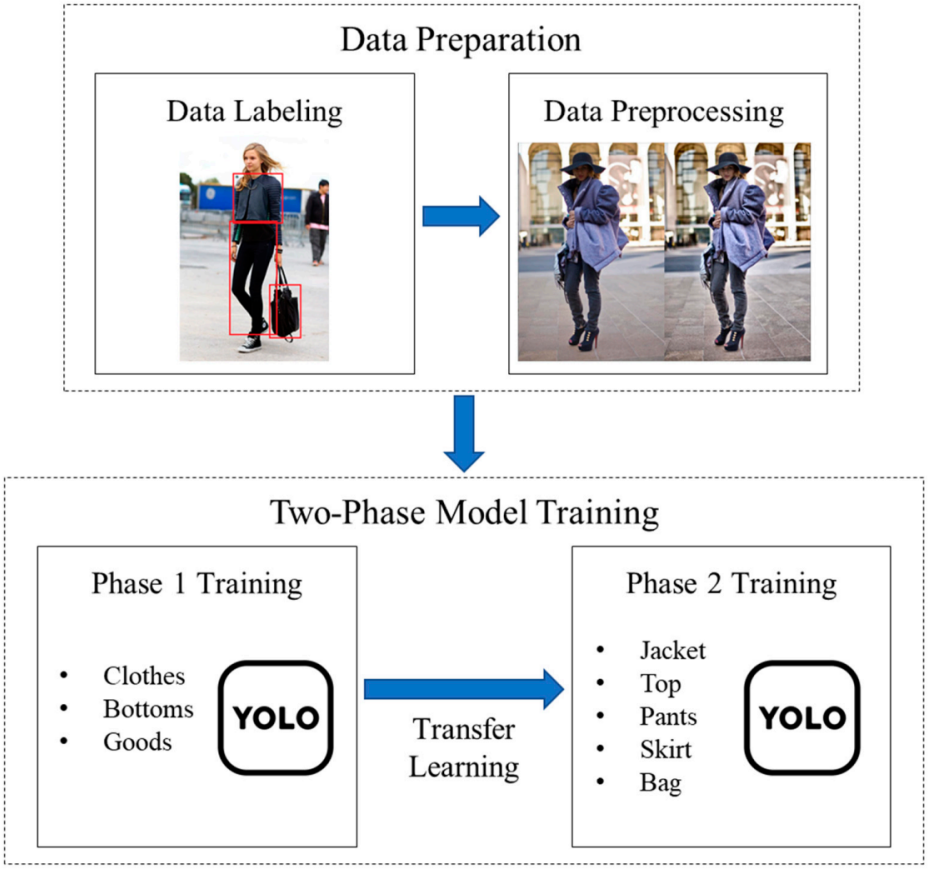

Figure 2. Training process of the YOLOv4-TP.

3.1. Data Preparation

In this phase, the data preparation was divided into two subsections. In the first subsection, labeling the images was needed for the training set and the testing set. The object detection algorithm was constructed by the CNN network, which belongs to supervised learning. The correct answer was provided to the model during the training phase. The YOLO network required that the label information of the image must be stored in a text format file. If there were multiple targets in one image, all information could be stored in the same file, and each image corresponded to only one label file. The label file contained the object number and object coordinates on this image. During the training process, the network could determine the target in the image by using the object information stored

in the label file. The two-phase detection model proposed in this research was trained by using inductive transfer learning. Therefore, it was necessary to label the data for both phases to provide the training data needed. Three categories of labeled data were used in phase 1 model training, and five categories of labeled data were used in phase 2.

In the second subsection, the image was implemented through preprocessing methods to enhance the detail of the image. The apparel images used in this research include various kinds of apparel and accessories, such as hats, glasses, parasols, and other items that might intercept the light. The image might have some excessive shadow or overexposure due to the environment of the shoot. In an area with a lot of glass or metal materials, the image might have overexposure due to reflection of light. To avoid the above potential problems from affecting the training results of the model, this research enhanced the image quality through the image preprocessing method of Contrast-Limited Adaptive Histogram Equalization. Contrast-Limited Adaptive Histogram Equalization (CLAHE) [29] is an image preprocessing technique that is used to improve contrast in images. The CLAHE method computes several histograms that are distinct sections of the image and redistributes the luminance values of the image. The CLAHE is suitable for improving the local contrast and enhancing the definitions of edges.

3.2. Two-Phase Model Training

This research proposes a fashion apparel detection model based on the YOLOv4 algorithm. Referring to the definition of fashion apparel in [26], the classification of the apparel in the research was similar to the actual situation in reality. Therefore, this research aimed to detect five different categories of fashion apparel as follows—jacket, top, pants, skirt, and bag. According to the definition of inductive transfer learning, a two-phase training method was proposed to improve the accuracy of the detection model.

The categories of the fashion apparel were divided into two phases, as shown in Figure 3. In the first phase, referring to the previous research [1] of the hierarchical classification on fashion apparel, we simplified the five categories of the final goal into three coarse categories of clothes, bottoms, and goods. Each first phase coarse category could be extended to the second-phase fine categories, where "clothes" could be subdivided into "jacket" and "top", "bottoms" could be subdivided into "pants" and "skirt", and "bag" belonged to the coarse categories of "goods".

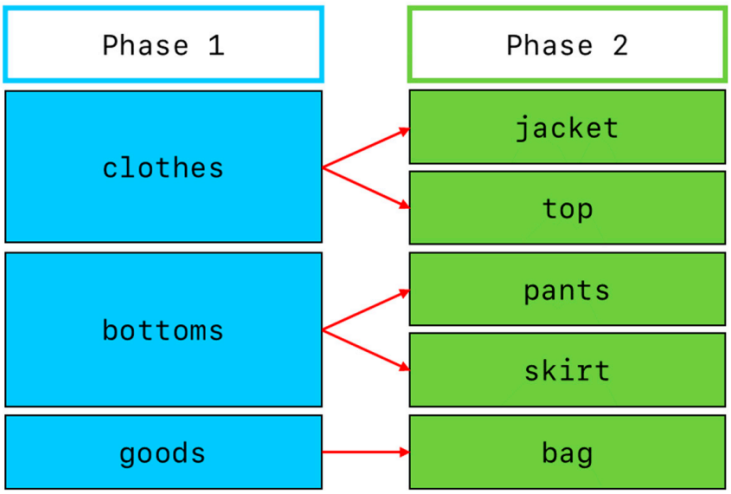

Figure 3. Category structure of two-phase fashion apparel model.

In the training process of the phase 1 model, the labeled data of three categories and pre-trained model of YOLOv4.conv.137 were used for training. The YOLOv4.conv.137 was the pre-trained model provided by YOLOv4 authors training through the Microsoft COCO dataset. After model training of phase 1 was over, the model weight was saved

and used as the new pre-trained model for the training of the phase 2 model. During the phase 2 model, labeled data of five categories and the trained model of phase 1 were used for transfer learning. The method proposed in this research used the characteristics of inductive transfer learning to improve the detection effect by relearning, based on the weight of the phase 1 model.

4. Experimental Results

In this section, the corresponding experiment materials are presented. Section 4.1 illustrates the experimental environment. Section 4.2 explains the dataset used in the research. Section 4.3 is about the hyperparameter setting in YOLOv4. Section 4.4 gives the evaluation of our proposed method. Section 4.5 concludes the experiment results.

4.1. Experimental Environment

The experimental environment of this research was implemented on a personal computer with a Windows 10 operating system. The system equipment was NVIDIA RTX 2070 super GPU using cuda 10.2 and cudnn 7.6.0, Intel i5-9600K CPU, and 32G DDR4 memory.

4.2. Dataset

The Clothing Co-Parsing (CCP) dataset was an open-source dataset that was constructed by Liang et al. [30]. The CCP dataset contained 2098 high-resolution street snaps of fashion apparel with a complex background. Each image was a full-color image and had a uniform size (550 × 830). The images in this dataset included various kinds of apparels with a complex background, which could not only suit our proposed model but also conformed to the real situation of the street.

In the data preparation, the LabelImg software was used to label the data. The target information could be exported to an XML format file including image name, path, size, target quantity, type, and coordinates. Next, using the Python language, the XML format was transformed to the text format, which was accepted by YOLO.

In the data preprocessing, the CLAHE method was used to enhance the image quality and increase the contrast. Figure 4 shows the effect of the CLAHE method. The contrast of the image on the right side was higher than the original image. The edges and contours of the apparel were also obvious.

Figure 4. The effect of image pre-processing by CLAHE method.

4.3. Hyperparameter Setting

This subsection lists the hyperparameters during the two-stage model training. Table 2 shows the different settings in the two-phased model. The other settings were as follows. (1) The image size was set to 416 × 416 recommended by YOLO; (2) the batch size was set to 2 due to equipment limitations; (3) the initial learning rate was set to 0.001, and (4) the momentum and decay were referred to the original setting by the YOLOv4 model. In phase 1, the classes parameter was set to 3, to conform to our goal. The iterations parameter was set to 6000 and the steps parameter was set to 4800, 5400. The learning rate would be decreased to 0.0001 after 4800 steps and to 0.00001 after 5400 steps. Explanations for the phase 2 model settings are the same as the phase 1 model.

Table 2. The hyperparameter setting of the model.

Parameters	Phase 1 Model	Phase 2 Model
classes	3	5
iterations	6000	10,000
steps	4800, 5400	1000, 8000, 9000

4.4. Evaluation Criterion

Precision and recall are the most common evaluation indicators for evaluating object detection models. According to [31], the definition of precision and recall are in Equations (1) and (2).

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The confusion matrix divides the model detection results into the following four categories—true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The precision measured the percentage of correct positive predictions among all predictions made. The recall measured the percentage of correct positive predictions among all positive cases in reality.

Intersection over Union (IoU) was the ratio between the intersection and the union of the predicted boxes and the ground truth boxes. Referring to previous research [32], the IoU formula is described by Equation (3).

$$IoU = \frac{area\left(B_P \cap B_{gt}\right)}{area\left(B_P \cup B_{gt}\right)} \tag{3}$$

Equation (3) shows the calculation of IoU, where B_{gt} is a ground truth bounding box and B_P is a predicted bounding box. By calculating the IoU, we could tell that the detection result was valid (TP) or not (FP). The most commonly used threshold was 0.5. If the IoU was >0.5, it was considered a TP, else it was considered an FP.

The mAP (mean average precision) was an indicator for evaluating object detection models. The AP (average precision) was defined as the mean of the precision values and recall values. The mAP hence was the mean of all average precision values across all classes.

4.5. Results

During the model training, the CCP dataset was divided into 90% for the training set and 10% for the validation set. Table 3 shows the experimental results of this research. The proposed YOLOv4-TPD model was through the two-phase transfer learning and the CLAHE image enhancement method. The YOLOv4-TL model only performed two-phase transfer learning. The YOLOv4-CLAHE model only implemented the CLAHE image enhancement method. The YOLOv4 model was the original model without any optimization method.

Compared to the original YOLOv4 model, the mAP was increased to 0.9%, by implementing the CLAHE image enhancement method and 1.67% by performing two-phase transfer learning. The proposed YOLOv4-TPD model increased the mAP from 92.57% to 96.01% by using both methods. The recall and the precision were also increased, getting 0.94 and 0.85. In IoU, the effect of adding only two-phase transfer learning or CLAHE alone had little difference. Adding both methods at the same time could increase the value of the IoU.

Table 3. The detection results of fashion apparel.

	YOLOv4-TPD	YOLOv4-TL	YOLOv4-CLAHE	YOLOv4
Two-phase	o	o	x	x
CLAHE	o	x	o	x
mAP	96.01%	94.24%	93.47%	92.57%
Recall	0.94	0.90	0.93	0.90
Precision	0.85	0.81	0.78	0.81
IoU	72.36%	66.91%	67.50%	68.07%

The detection results of the five categories of apparel are shown in Figure 5. The bars from left to right in each apparel category are YOLOv4-TPD, YOLOv4-TL, YOLOv4-CLAHE, and YOLOv4. As can be seen, the detection results of the YOLOv4-TPD model had the best average precision. The experimental results proved that our proposed model was effective in fashion apparel detection.

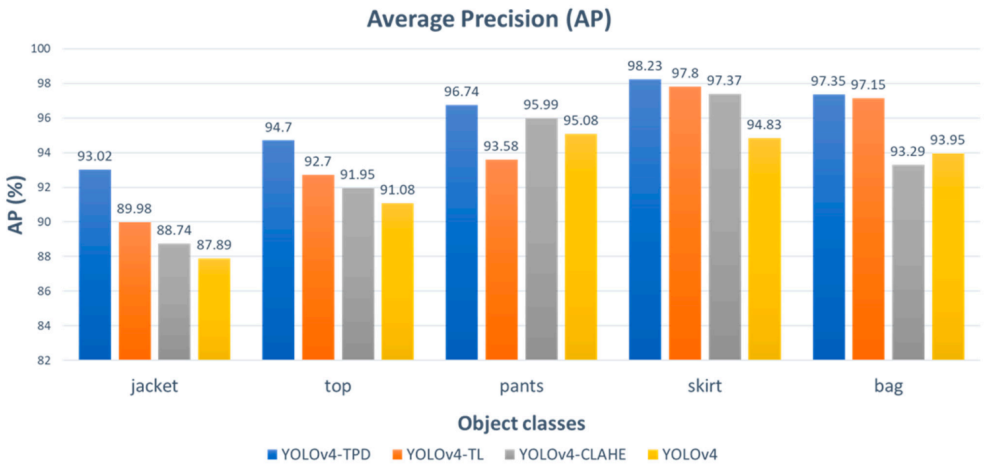

Figure 5. The detection results of the five categories of apparel.

The total detection time of the YOLOv4-TPD for fifty images was 781.654 milliseconds. The average time for one image was 15.633 milliseconds, which was equal to 64 frames per second (FPS). The detection efficiency of the model was fully suitable for real-time detection.

Table 4 shows the comparison of experimental results with previous methods. The previous work [26] used YOLOv2 in fashion apparel detection with five categories. The precision and recall achieved were 83.9% and 73%. The detection speed was 56 milliseconds per image. In research [27], the YOLOv3 was only used to detect skirts. The results showed that the model obtained 92.98% in mAP. In this study, we proposed the YOLOv4-TPD model to improve detection efficiency and accuracy. The experimental results showed that our proposed model was better than [26,27] in fashion apparel detection.

Table 4. The comparison of experimental results with previous methods.

	YOLOv4-TPD	Previous Method [26]	Previous Method [27]
Architecture	YOLOv4	YOLOv2	YOLOv3
Categories	5	5	1
mAP	96.01%	-	92.98%
Precision	0.85	0.839	-
Recall	0.94	0.73	-
Detection time	15.633 ms	56 ms	-

Figure 6 shows the detection results of our model. As can be seen in Figure 6a,b, our model had a good detection effect. Our model detected the apparel regions accurately and had satisfactory confidence scores. However in Figure 6c, the person was sideways, which prevented the model from correctly detecting the upper clothing. In Figure 6d, the model could not detect the apparel because the scarf covered a large apparel area. Therefore, our model was relatively poor at detecting the side of the apparel or apparel undercover. We supposed that the training data were not comprehensive enough, which caused this problem.

(**a**) (**b**) (**c**) (**d**)

Figure 6. The result images of object detection—(**a**,**b**) correct detection, (**c**) sideways apparel, and (**d**) object occlusion.

Figure 7 shows the detection results of our proposed YOLOv4-TPD in different hues. The image hue was randomly adjusted to change the color of the apparel. The experimental results showed that the model had correct detection results in different hues. The confidence of each apparel was also satisfactory. Therefore, the experimental results verified the previous statement that the proposed model performed the detection task through the contour and appearance feature of the apparel, the detection effect of the model was not affected by the color of the apparel.

Figure 7. The detection results in different hues.

5. Conclusions

In this research, we proposed a two-phase fashion apparel detection model based on the YOLOv4 algorithm. The experimental results showed that the mAP of YOLOv4-TPD was 3.03% higher than the original YOLOv4 model. The values of other validated indicators such as recall, precision, and IoU also increased. Compared to other existing research of clothing detection, our model had the advantage of high detection accuracy in fashion clothing detection with complex backgrounds.

In today's society, the types of fashion apparel are gradually increasing. It is not enough that the apparel detection model could only detect five categories. Therefore, in future work, we aim to increase the category of the apparel and consider detecting the inner clothing, such as a shirt under a jacket.

Author Contributions: Conceptualization, C.-H.L.; methodology, C.-H.L. and C.-W.L.; validation, C.-W.L.; resources, C.-H.L.; data curation, C.-W.L.; writing—original draft preparation, C.-W.L.; writing—review and editing, C.-H.L. and C.-W.L.; supervision, C.-H.L. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in «Lin, C. W. A Study of Fashion Apparel Detection based on YOLO. Master Thesis, Information Management, Chaoyang University of Technology, Taiwan, 2021».

Acknowledgments: This research was supported by the Ministry of Science and Technology of Taiwan under Grant MOST-109-2221-E-324-023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seo, Y.; Shin, K. Hierarchical Convolutional Neural Networks for Fashion Image Classification. *Expert Syst. Appl.* **2019**, *116*, 328–339. [CrossRef]
2. Benjdira, B.; Khursheed, T.; Koubaa, A.; Ammar, A.; Ouni, K. Car Detection Using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3. In Proceedings of the 2019 1st International Conference on Unmanned Vehicle Systems-Oman, UVS, Muscat, Oman, 5–7 February 2019; pp. 1–6.
3. Lyu, H.; Fu, H.; Hu, X.; Liu, L. Esnet: Edge-Based Segmentation Network for Real-Time Semantic Segmentation in Traffic Scenes. In Proceedings of the 2019 IEEE International Conference on Image Processing, ICIP, Taipei, Taiwan, 22–25 September 2019; pp. 1855–1859.
4. Yi, J.; Wu, P.; Hoeppner, D.J.; Metaxas, D. Pixel-Wise Neural Cell Instance Segmentation. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018; pp. 373–377.
5. Wang, X.; Zhang, T. Clothes Search in Consumer Photos via Color Matching and Attribute Learning. In Proceedings of the MM'11—2011 ACM Multimedia Conference and Co-Located Workshops, Scottsdale, AZ, USA, 28 November–1 December 2011; ACM Press: New York, NY, USA, 2011; pp. 1353–1356.
6. Liu, S.; Song, Z.; Liu, G.; Xu, C.; Lu, H.; Yan, S. Street-to-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3330–3337.
7. Song, Z.; Wang, M.; Hua, X.S.; Yan, S. Predicting Occupation via Human Clothing and Contexts. In Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1084–1091.
8. Yamamoto, T.; Nakazawa, A. Fashion Style Recognition Using Component-Dependent Convolutional Neural Networks. In Proceedings of the 2019 International Conference on Image Processing, ICIP, Taipei, Taiwan, 22–25 September 2019; pp. 3397–3401.
9. Shin, Y.G.; Yeo, Y.J.; Sagong, M.C.; Ji, S.W.; Ko, S.J. Deep Fashion Recommendation System with Style Feature Decomposition. In Proceedings of the 2019 IEEE 9th International Conference on Consumer Electronics—Berlin, ICCE-Berlin, Berlin, Germany, 8–11 September 2019; pp. 301–305.
10. Eshwar, S.G.; Gautham Ganesh Prabhu, J.; Rishikesh, A.V.; Charan, N.A.; Umadevi, V. Apparel Classification Using Convolutional Neural Networks. In Proceedings of the 2016 International Conference on ICT in Business, Industry, and Government, ICTBIG, Indore, India, 18–19 November 2016.
11. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
12. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016*; Springer: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
15. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
16. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
17. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
19. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
20. Chen, H.; Gallagher, A.; Girod, B. Describing Clothing by Semantic Attributes. In *Proceedings of the Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7574, pp. 609–623.
21. Yang, M.; Yu, K. Real-Time Clothing Recognition in Surveillance Videos. In Proceedings of the 2011 18th International Conference on Image Processing, ICIP, Brussels, Belgium, 11–14 September 2011; pp. 2937–2940.
22. Surakarin, W.; Chongstitvatana, P. Classification of Clothing with Weighted SURF and Local Binary Patterns. In Proceedings of the ICSEC 2015—19th International Computer Science and Engineering Conference: Hybrid Cloud Computing: A New Approach for Big Data Era, Chiang Mai, Thailand, 23–26 November 2015.
23. Lao, B.; Jagadeesh, K. Convolutional Neural Networks for Fashion Classification and Object Detection. Available online: http://cs231n.stanford.edu/reports/2015/pdfs/BLAO_KJAG_CS231N_FinalPaperFashionClassification.pdf (accessed on 11 March 2021).
24. Liu, Z. A Deep Learning Method for Suit Detection in Images. In Proceedings of the 2018 14th International Conference on Signal Processing Proceedings, ICSP, Beijing, China, 12–16 August 2018; pp. 439–444.

25. Liu, K.H.; Liu, T.J.; Wang, F. Cbl: A Clothing Brand Logo Dataset and a New Method for Clothing Brand Recognition. In Proceedings of the 2020 28th European Signal Processing Conference, Amsterdam, The Netherlands, 24–28 August 2021; pp. 655–659.
26. Feng, Z.; Luo, X.; Yang, T.; Kita, K. An Object Detection System Based on YOLOv2 in Fashion Apparel. In Proceedings of the 2018 IEEE 4th International Conference on Computer and Communications, ICCC, Chengdu, China, 7–10 December 2018; pp. 1532–1536.
27. Liu, R.; Yan, Z.; Wang, Z.; Ding, S. An Improved YOLOV3 for Pedestrian Clothing Detection. In Proceedings of the 2019 6th International Conference on Systems and Informatics, ICSAI, Shanghai, China, 2–4 November 2019; pp. 139–143.
28. Kumar, S.; Zheng, R. Hierarchical Category Detector for Clothing Recognition from Visual Data. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops, ICCVW, Venice, Italy, 22–29 October 2017; pp. 2306–2312.
29. Zuiderveld, K. Viii.5.-Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems*; Heckbert, P.S., Ed.; Academic Press: Cambridge, MA, USA, 1994; pp. 474–485. [CrossRef]
30. Yang, W.; Luo, P.; Lin, L. Clothing Co-Parsing by Joint Image Segmentation and Labeling. In Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3182–3189.
31. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *Proceedings of the 27th European Conference on Advances in Information Retrieval Research (ECIR), Santiago de Compostela, Spain, 21–23 March 2005*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.
32. Everingham, M.; van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]