An Optimization-Based Technology Applied for Face Skin Symptom Detection

Yuan-Hsun Liao, Po-Chun Chang, Chun-Cheng Wang and Hsiao-Hui Li

1 Department of Computer Science, Tunghai University, Taichung 407224, Taiwan
2 Bachelor’s Degree Program in Chain Store Management, Tainan University of Technology, Tainan 710302, Taiwan
* Correspondence: xiasohui@gmail.com

Abstract: Face recognition segmentation is very important for symptom detection, especially in the case of complex image backgrounds or noise. The complexity of the photo background, the clarity of the facial expressions, or the interference of other people’s faces can increase the difficulty of detection. Therefore, in this paper, we have proposed a method to combine mask region-based convolutional neural networks (Mask R-CNN) with you only look once version 4 (YOLOv4) to identify facial symptoms by this new method. We use the face image dataset from the public image databases DermNet and Freepic as the training source for the model. Face segmentation was first applied with Mask R-CNN. Then the images were imported into ResNet-101, and the facial features were fused with region of interest (RoI) in the feature pyramid networks (FPN) structures. After removing the non-face features and noise, the face region has been accurately obtained. Next, the recognized face area and RoI data were used to identify facial symptoms (acne, freckle, and wrinkles) with YOLOv4. Finally, we use Mask R-CNN, and you only look once version 3 (YOLOv3) and YOLOv4 are matched to perform the performance analysis. Although, the facial images with symptoms are relatively few. We still use a limited amount of data to train the model. The experimental results show that our proposed method still achieves 57.73%, 60.38%, and 59.75% of mean average precision (mAP) for different amounts of data. Compared with other methods, the mAP was more than about 3%. Consequently, using the method proposed in this paper, facial symptoms can be effectively and accurately identified.

Keywords: facial symptom detection; skin condition; YOLOv3; YOLOv4; Mask R-CNN

1. Introduction

It is human nature to love beauty. With the rapid development of technology and the economy, consumers are paying ever greater attention to skin care products, especially facial care. Skin care products have transformed from luxury items to indispensable necessities in daily life. According to a report by Grand View Research, Inc. published in March 2022, the global skin care market was worth US$130.5 billion in 2021 and is expected to grow at a compound annual growth rate (CAGR) of 4.6% from 2022 to 2030 [1].

With the prevalence of coronavirus disease 2019 (COVID-19) in the last few years, it has changed the operating model of many companies and consumer buying behavior [2]. Many consumers switched to online purchases instead of physical channels. According to a report by Euromonitor International, e-commerce will expand by another $1.4 trillion by 2025, accounting for half of global retail growth [3]. As consumption patterns change, numerous brands have begun to use artificial intelligence (AI), augmented reality, virtual reality, and other technologies to serve their customers.

In the past, consumers in the physical channel often relied on the advice of salespeople in making product purchases, but when online shopping is conducted, consumers can only make product selections according to their own preferences. Since everyone’s
skin condition is different, some consumers with sensitive skin may experience allergic reactions after using products unsuited to them [4]. According to a survey report, 50.6% of 425 participants had experienced at least one adverse reaction to product use in the past two years, experiencing conditions including skin redness (19%), pimples (15%), and itching (13%), and 25% of these participants had problems caused by the use of unsuited skin care products [5]. Thus, the use of unsuitable skin care products can not only seriously harm consumers’ skin but also have a terrible impact on a manufacturer’s reputation [6].

On the other hand, as the COVID-19 epidemic has swept the world in recent years, people will wear masks whenever they go out [7,8]. Because of that, people wear masks for a long time every day, and the problems with facial skin are increasing daily [9,10]. In particular, the proportion of medical staff who have facial skin problems has greatly increased, among which contact dermatitis, acne and pimples, and rosacea are the most common [11–13]. In order to help people take care of their own facial health while cooperating with the epidemic prevention policy, we hope this research can offer cogent advice on skin care issues.

In the past, it was not easy to create an intelligent skin care recommendation platform due to the limitations imposed by image processing techniques [14,15]. With the vigorous development of deep learning in recent years, image-processing techniques have become more mature. There is a thus glimmer of light on this issue. There are three common facial skin problems, which are acne, spots, and wrinkles [16–18]. We chose these three as the feature categories for this study.

The first type, acne, is caused by abnormal keratinization of pores and strong secretion of sebaceous glands, resulting in excessive oil that cannot be discharged and blocked hair follicles. The main reasons are insufficient facial cleansing, endocrine disorders, and improper use of cosmetics and skin care products [19,20].

The second type, freckles, is mainly caused by excessive sun exposure. When the skin’s melanocytes are overstimulated by ultraviolet light, it causes the cells to produce more melanin, which in turn causes freckles. Other causes include endocrine disorders and bodily aging [21–23].

For the third type, wrinkles, the most common cause is dryness. Dry lines often appear on the cheeks and around the corners of the mouth. Due to the lack of water in the skin, the outermost sebum film cannot play a protective role, the moisturizing ingredients (ceramides) under the stratum corneum are reduced, and the skin cannot retain water and begins to shrink and sag. Moisturizing as soon as you notice fine lines will most likely eliminate them [24–26].

Face recognition segmentation is very important for symptom detection, especially in the case of complex image backgrounds or noise. The complexity of the photo background, the clarity of the facial expressions, or the interference of other people’s faces can increase the difficulty of detection. In the past, Adjabi et al. [27] pointed out that the two-dimensional face recognition methods have holistic methods, local (geometrical) methods, local texture descriptors-based methods, and deep learning-based methods. Among them, the deep learning method is the current development direction, so in our face recognition method, we selected the deep learning algorithm Mask R-CNN [28–31].

Mask R-CNN is an instance segmentation algorithm. It can identify each object instance for every known object within an image. Because of that, we use it to detect where a face is and turn the region that Mask R-CNN was not predicted to black. After detecting the region of the face, YOLOv3 [32–41] and YOLOv4 [11,42,43] are deployed to detect facial symptoms. The reason for choosing this method in order to solve the problem is that pictures of facial lesions are not easy to obtain. The study results show our proposed method is effective in improving recognition rates. Although the facial images with symptoms are relatively few, we still use a limited amount of data to train the model. The experimental results show that our proposed method still achieves 57.73%, 60.38%, and 59.75% of mean average precision (mAP) for different amounts of data. Compared
with other methods, the mAP was more than about 3%. Consequently, using the method proposed in this paper, facial symptoms can be effectively and accurately identified.

The organization of this paper is as follows. In Section 2, this paper describes the related work of Mask R-CNN, YOLOv3, and YOLOv4. In Section 3, the materials and our method are described. Next, we discuss the experimental results. Finally, we provide our results, discussion, and future work. The detailed abbreviations and definitions used in the paper are listed in Table 1.

Table 1. List of abbreviations and acronyms used in the paper.

| Abbreviation | Definition |
|--------------|------------|
| AP           | average precision |
| BB           | bounding box |
| BEC          | bottleneck and expansion convolution |
| Faster R-CNN | faster region-based convolutional neural networks |
| FCN          | fully convolutional networks |
| FPN          | feature pyramid networks |
| GPU          | graphics processing unit |
| IoU          | intersection over union |
| Mask R-CNN   | mask region-based convolutional neural networks |
| PAN          | path aggregation network |
| RoI          | region of interest |
| RPN          | region proposal network |
| RUS          | random under-sampling |
| SOTA         | state-of-the-art |
| SPP          | spatial pyramid pooling |
| Tiny-SPP     | tiny spatial pyramid pooling |
| WCL          | wing convolutional layer |
| YOLO         | you only look once |
| YOLOv3       | you only look once version 3 |
| YOLOv4       | you only look once version 4 |

2. Related Work

Object Detection is an important aspect of image recognition. Many results have been reported in object recognition, vehicle recognition, person recognition, and face recognition. The Object detector model applied to object detection consists of four parts: Input, Backbone, Neck, and Head, as shown in Figure 1. The Backbone part is usually composed of a trained neural network that aims to capture basic features to improve the performance of target detection. The neck part is used to extract different feature maps at different stages of the backbone. The last part of the Head can be divided into Dense Prediction (one-stage) and Sparse Prediction (two-stage).

Figure 1. Object detector architecture.
There are several common two-dimensional face recognition methods: holistic methods, local (geometric) methods, methods based on local texture de-labeling, and methods based on deep learning [27]. Deep learning methods are the current trend. To improve face skin symptom detection, we conduct a deep learning review to introduce Mask R-CNN, YOLOv3, and YOLOv4.

2.1. Mask R-CNN

The two-stage model of Mask R-CNN combines the two-stage model of Faster Region-based Convolutional Neural Networks (Faster R-CNN) [44], and the Feature Pyramid Networks (FPN) [45] method uses feature maps with high feature levels in different dimensions for prediction as Figure 2. It also improves the shortcomings of Region of Interest (RoI) pooling in Faster R-CNN so that the longitude of the bounding box and object positioning can truly reach the pixel level, increasing the accuracy rate by 10~50%.

Mask R-CNN consists of:
- Backbone: ResNet-101 [46];
- Neck: FPN [45];
- Head: Dense Prediction(one-stage): RPN [44]
  Sparse Prediction(two-stage): Mask R-CNN [28].

There have been many previous studies using Mask R-CNN. Zhang et al. [47] created a publicly available large-scale benchmark underwater video dataset to retrain the Mask R-CNN deep model, which in turn was applied to the detection and classification of underwater creatures via random under-sampling (RUS), achieving a mean Average Precision (mAP) of 62.8%. Tanoglidis et al. [48] use Mask R-CNN to solve the problem of finding and masking ghosts and scattered-light artifacts in DECam astronomical images.

2.2. YOLOv3

The YOLOv3 [49] detector was developed to ensure symptoms detection would be more objective. The backbone of YOLOv3 is Darknet-53 which is more powerful than Darknet-19. The neck part includes FPN [45], which aggregates the deep feature maps of DarkNet-53.

YOLOv3 consists of:
- Backbone: Darknet-53 [49];
- Neck: FPN [45];
- Head: YOLO [50].

In the field of YOLOv3, Khan et al. [51] used this method and Microsoft Azure face API to perform face detection and face recognition, respectively, with a real-time automatic attendance system for face recognition, and this system enjoys a high accuracy rate in most cases. Menon et al. [52] implemented face recognition using both R-CNN and YOLOv3 algorithms. Compared with other algorithms, it has a higher processing speed.
2.3. YOLOv4

YOLOv4 improves various parts of YOLOv3, including the Backbone, Neck, and Head. Not only does it build an efficient and powerful object detection model that can be trained using a 1080Ti or 2080Ti GPU, but it also verifies the influence of the Bag-of-Freebies and Bag-of-Specials target detection methods of State of the Art (SOTA) and improves some tricks and SOTA methods, making it more efficient, and able to train on a single GPU.

YOLOv4 consists of the following:

- **Backbone**: CSPDarknet-53 [53];
- **Neck**: SPP [54], PAN [55];
- **Head**: YOLOv3 [49].

Prasetyo, Suciati, and Fatichah [56] discussed the application of YOLOv4-tiny to the identification of fish body parts. Since the author of this article found that the accuracy of identifying specific parts of fish is relatively low, the author modified Yolov4-tiny using wing convolutional layer (WCL), tiny spatial pyramid pooling (Tiny-SPP), bottleneck and expansion convolution (BEC), and additional third-scale detectors. Kumar et al. [57] used tiny YOLOv4-SPP to achieve better performance in mask detection than the original tiny YOLOv4, tiny YOLOv3, etc., and the mAP reached 64.31%. Zhang et al. [58] found that compared with YOLOv4, the proposed weight Improved YOLOv4 has a 3.45% increase in mAP, while the weight size is only 15.53% of the baseline model, and the number of parameters is only 15.84% of the baseline model.

3. Materials and Methods

In this paper, our experimental procedure approaches showed in Figure 3, including data collection, pre-processing of the dataset, feature extraction, and training and testing of the target detection algorithm. The detailed procedure is as follows:

![Figure 3. The overall schematic of the proposed model.](image-url)

3.1. Data Collection

Data collection is divided into simple faces and pictures of faces with symptoms. A large amount of data can be obtained from the parts of the human face, but disease data is not easy to obtain. At present, we have collected about 1500 pieces of symptom data, all of which are taken from the public databases DermNet [59] and Freepic [60]. Because of that, this paper will focus on how to use a small amount of data to drive a more efficient model.
3.1. Data Collection

Data collection is divided into simple faces and pictures of faces with symptoms. A large amount of data can be obtained from the parts of the human face, but disease data is not easy to obtain. At present, we have collected about 1500 pieces of symptom data, all of which are taken from the public databases DermNet [59] and Freepic [60]. Because of the difficulty in obtaining large amounts of data, we selected 600 pictures with 150 types of acne for the training set. We selected acne, freckles, and wrinkles as the primary characteristics of this study, as shown in Table 2.

| Acne Type | Size       | Color       | Pus | Inflammatory | Comments                                                                 | Image |
|-----------|------------|-------------|-----|--------------|---------------------------------------------------------------------------|-------|
| White-head| Tiny       | Whitish     | No  | No           | A chronic whitehead is called milia                                       |       |
| Papule    | <5 mm      | Pink        | No  | Yes          | Very common                                                               |       |
| Pustule   | <5 mm      | Red at the base with a yellowish or a whitish center | Yes | Yes          | Very common                                                               |       |
| Nodule    | 5–10 mm    | Pink and red| No  | Yes          | A nodule is similar to a papule but is less common                         |       |

The technical terms for freckles are Ephelides and Lentigines, and we define both as freckles for analysis, as defined in Table 3.

| Freckle Type | Ephelides | Lentigines |
|--------------|-----------|------------|
| Appearance   | First visible at 2–3 years of age after sun exposure, partially disappears with age | Accumulate with age, common after 50, stable |
| Effects of sun| Fade during winter | Stable |
| Size         | 1–2 mm and above | mm-cm in diameter |
| Borders      | Irregular, well defined | Well defined |
| Color        | Red to light brown | Light yellow to dark brown |

Table 2. Types of common acne in training sets [18].

Table 3. Types of Face Wrinkle in training sets [20].
Table 3. Types of Face Wrinkle in training sets [20].

| Freckle Type | Ephelides | Lentigines |
|--------------|-----------|------------|
| Skin type    | Caucasians, Asians, skin type I-II. | Caucasians, Asians, skin type I-III |
| Etiology     | Genetic   | Environmental |

The last type, wrinkles, has a total of 12 kinds. We selected six horizontal forehead lines, Glabellar frown lines, Periodic lines, Nasolabial folds, Cheek lines, and Marionette lines as the characteristics of wrinkles in this paper, Table 4.

Table 4. Characteristics of ephelides and lentigines [23].

| Wrinkle Type         | Position       | Image |
|----------------------|----------------|-------|
| Horizontal forehead lines | Forehead       |       |
| Glabellar frown lines          | Between eyebrows |   |
| Periorbital lines          | Canthus        |       |
| Nasolabial folds          | Nose to mouth  |       |
| Cheek lines               | Cheek          |       |
| Marionette lines          | Corner of mouth |     |

We used three different numbers of datasets to train the diagnosis of symptoms with splitting the training data into acne: 50%, freckles: 25%, and wrinkles: 25%, as shown in Table 5.
### Table 5. Distribution of symptom training datasets.

| Training datasets | 500  | 1000  | 1500  |
|-------------------|------|-------|-------|
| Acne              | 250  | 500   | 750   |
| Freckles          | 125  | 250   | 375   |
| Wrinkles          | 125  | 250   | 375   |

### 3.4. Detect Symptoms

The methods proposed in this research are divided into those with face recognition and those without face recognition. For the face recognition algorithm, we use Mask R-CNN to maintain the primary color of the identified Region of Interest (RoI) and turn the uninteresting areas into black. Then we use YOLOv4 to identify the symptoms and further compare the accuracy of YOLOv4 in the identification of symptoms. The detection structure is shown in Figure 4.

![Image](image.png)

**Figure 4.** The structure of detection symptoms.

#### 3.4.1. Face Detection Process

The face recognition uses Mask R-CNN to keep the primary color of the RoI being recognized and transform the uninteresting regions into black. The detailed procedures are shown in Algorithm 1. The results of the face detection process are shown in Figure 5. Figure 5a shows the original image. After the Mask, R-CNN recognizes and marks the position of the face, as in Figure 5b. At last, the part outside the face area is transformed to black by using a color splash, as in Figure 5c.

#### Algorithm 1: Face detection procedures

**Input:** Original image  
**Output:** The image with only the face  
1. Set the environment variables to match the features of the face.  
2. Adjust the image to fit the requirements.  
3. Pass the processed image into ResNet-101 and obtain the corresponding feature map.  
4. FPN corrects the size of the RoIs in the feature map.  
5. RPN classifies these RoIs and filters out the background, and BB regression corrects the BB of the RoI.  
6. Use RoI alignment to split the remaining RoIs into facial RoIs and non-facial RoIs.  
7. Use BB Regression to fix the BB of RoI again and generate the mask with FCN after classification.  
8. Keep the facial part after the non-facial part becomes black.  
9. End.
YOLOv4 further. The detailed procedures are shown in Algorithm 2. The results of the
previous section of face recognition to identify skin symptoms, as in Figure 6a. YOLOv4
is a popular metric used to measure the performance of models. AP values are calculated
over recall values from 0 to 1.

3.5. Mean Average Precision for Evaluation

Mean average precision (mAP) or simply just referred to as Average Precision (AP), is
a popular metric used to measure the performance of models. AP values are calculated
over recall values from 0 to 1.
Traditional Intersection over union (IoU) = 0.5, IoU means the ratio of the area of overlap and area of union of the predicted bounding box (BB) and test data label bounding box.

\[
\text{IoU} = \frac{\text{area of overlap}}{\text{area of union}}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} = \frac{TP}{\text{total positive results}}
\]

TP: True positive (IoU > 0.5 with the correct classification).
FP: False positive (IoU < 0.5 with the correct classification or duplicated BB).

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

FN: False negative (No detection at all or the predicted bounding box has an IoU > 0.5 but was the wrong classification).

The general definition for the Average Precision (AP) is finding the area under the precision-recall curve above.

\[
\text{AP} = \frac{1}{N} \sum_{i} \text{Recall}_i \times \text{Precision}(\text{Recall}_i)
\]

N: the number of queries.

4. Results and Discussion

First, we used images collected from public databases on the internet for training, using YOLOv3 and YOLOv4 to train symptom recognition, respectively, and used Mask R-CNN to train face labels.

We compared the model trained by YOLOv3 and the model trained by YOLOv4, determined which feature identification can achieve better results, and also tested whether the number of images in the training set influences the training model.

For the training set of YOLOv3, we used the training set of 500, 1000, and 1500 to train and generate the results of Table 6. For the model of 1500, we obtained the average value of mAP and the model with the highest value in the YOLOv3 training set.

Table 6. YOLOv3 Analysis of the training process (mAP = 0.5, Testing sets = 166).

| Training | 500  | 1000 | 1500 |
|----------|------|------|------|
| Iteration of model (×10k) | 39   | 43   | 52   |
| Best     | 55.50| 52.78| 57.52|
| Worst    | 45.09| 44.35| 50.47|
| Average  | 51.30| 48.56| 53.57|

The training results of YOLOv4 are also in line with the conclusions we have drawn from YOLOv3. The training of 1500 images obtains a better model in the symptom labels, as in Table 7.

Table 7. YOLOv4 Analysis of the training process (mAP = 0.5, Testing sets = 166).

| Training | 500  | 1000 | 1500 |
|----------|------|------|------|
| Iteration of model (×10k) | 88   | 115  | 100  |
| Best     | 59.90| 60.29| 62.87|
| Worst    | 48.96| 49.00| 47.90|
| Average  | 55.84| 54.96| 57.83|

Among the training results of Mask R-CNN, the training set with the largest number of sheets is also the best in this study, which is also in line with the conclusions we have
drawn from YOLOv4 and YOLOv3. The more training sets, the better the results. The training results are presented in Table 8.

Table 8. Mask R-CNN Face Detection Accuracy (mAP = 0.5, Testing sets = 160).

| Training Epoch (×100) | 100 | 250 | 500 |
|-----------------------|-----|-----|-----|
| Accuracy              | 83.38 | 85.74 | 85.84 |

According to the above three statistical charts, we use the best model in YOLOv3 and YOLOv4 to identify the pictures of symptoms with complex backgrounds in the picture set. Then analyze whether YOLO's symptom identification will be as we expected after the Mask R-CNN removes the parts other than the face in these image sets. It is better than the original YOLO to directly identify images with complex backgrounds. The accuracy of statistics is presented in Table 9.

Table 9. Method Comparison (mAP = 0.5, Testing sets = 61).

| Method\Training | 500  | 1000 | 1500  |
|-----------------|------|------|-------|
| YOLOv3          | 54.52 | 50.01 | 55.68 |
| YOLOv4          | 58.74 | 56.98 | 56.29 |
| Mask R-CNN + YOLOv3 (100 images for training) | 50.03 | 47.43 | 56.53 |
| Mask R-CNN + YOLOv3 (250 images for training) | 50.38 | 46.54 | 53.70 |
| Mask R-CNN + YOLOv3 (500 images for training) | 55.02 | 52.39 | 58.13 |
| Mask R-CNN + YOLOv4 (100 images for training) | 55.97 | 57.64 | 53.68 |
| Mask R-CNN + YOLOv4 (250 images for training) | 55.10 | 56.48 | 53.19 |
| Mask R-CNN + YOLOv4 (500 images for training) | 57.73 | 60.38 | 59.75 |

From the experimental results, we can find that when the number of training sets is inconsistent, there will be different results. When the number of training sets is too small, it is more likely the trained model will have unstable recognition. In Table 9, we can find that the Mask R-CNN training data set is unstable at 56.53% of 100 images and 53.70% of 250 images but stable and improved at 58.13% of 500 images. Therefore, the experimental results in Table 9 can demonstrate our designed model adequately. We first use Mask R-CNN to remove the parts other than the face in these images. Then, we use YOLO to identify the symptoms of the face more effectively. Compared with YOLOv3 alone, the results were only 54.52%, 50.01%, and 55.68%. Using our method (Mask R-CNN + YOLOv3), the results are 55.02%, 52.39%, and 58.13%, which are at least a 1% mAP improvement.

With YOLOv4 alone, only 58.74%, 56.98%, and 56.29% were achieved. Using our method (Mask R-CNN + YOLOv4), the results are 57.73%, 60.38%, and 59.75%, which are at least a 3% mAP improvement. Therefore, our proposed method can effectively improve the results of facial symptom recognition for symptom pictures with complex backgrounds.

Therefore, the results of our method process as shown in Figure 7. First, we input the images that may have noise, as in Figure 7a. Then, we use Mask R-CNN to remove the parts other than the face, as in Figure 7b. Finally, symptom recognition can be predicted with YOLOv4, as in Figure 7c.
Author Contributions: Conceptualization, Y.-H.L. and H.-H.L.; methodology, Y.-H.L. and H.-H.L.; software, P.-C.C. and C.-C.W.; validation, Y.-H.L., P.-C.C., C.-C.W., and H.-H.L.; formal analysis, Y.-H.L., P.-C.C., C.-C.W., and H.-H.L.; investigation, P.-C.C. and C.-C.W.; resources, Y.-H.L. and H.-H.L.; data curation, P.-C.C. and C.-C.W.; writing—original draft preparation, Y.-H.L., P.-C.C., C.-C.W., and H.-H.L.; writing—review and editing, Y.-H.L. and H.-H.L.; visualization, P.-C.C. and C.-C.W.; supervision, Y.-H.L. and H.-H.L.; project administration, Y.-H.L.; funding acquisition, Y.-H.L. All 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: Not applicable.

In general, AI research requires large training data sets. However, in the field of “Face Skin Symptom Detection” research, there are problems from insufficient image data and uneasy acquisition. The proposed approach can be used to train a model with fewer data sets and time and has good identification results.

Under the influence of COVID-19, consumers’ shopping habits have changed. With so many skin care products available on the internet, choosing the right skin care product is an important issue. Through our research, consumers can understand their own skin symptoms to facilitate the proper selection of skin care products and avoid purchasing unsuitable products that may cause skin damage.

In the future, we will combine the results of our research on face skin symptom detection into a product recommendation system. The detection results will be used in a real product recommendation system. We expect to design an App for facial skincare product recommendations in the future.

Figure 7. The results of detect symptoms process.

5. Conclusions

In this study, we compared YOLOv3, YOLOv4, Mask R-CNN + YOLOv3, and Mask R-CNN + YOLOv4 with the same number of diseases datasets and found that the accuracy of our method has improved significantly. At the same time, we experimented with Mask R-CNN before YOLO identified symptoms. The results indicate that our proposed method still achieves 57.73%, 60.38%, and 59.75% of mean average precision (mAP) for different amounts of data. Compared with only using YOLOv4 to symptom detect the image has noise, the mAP was more than about 3%.

Instead of relying on one image recognition algorithm for training, we combine multiple algorithms. We choose different algorithms in each stage according to the different features of the images. First, we segment the complex images to remove redundant images and noise. Then, we enhance the required image features for the detection of skin symptoms. This study reduces the difficulty of model training and model training time and increases the success rate of detailed feature identification.

In this study, we compared YOLOv3, YOLOv4, Mask R-CNN + YOLOv3, and Mask R-CNN + YOLOv4 with the same number of diseases datasets and found that the accuracy of our method has improved significantly. At the same time, we experimented with Mask R-CNN before YOLO identified symptoms. The results indicate that our proposed method still achieves 57.73%, 60.38%, and 59.75% of mean average precision (mAP) for different amounts of data. Compared with only using YOLOv4 to symptom detect the image has noise, the mAP was more than about 3%.

Instead of relying on one image recognition algorithm for training, we combine multiple algorithms. We choose different algorithms in each stage according to the different features of the images. First, we segment the complex images to remove redundant images and noise. Then, we enhance the required image features for the detection of skin symptoms. This study reduces the difficulty of model training and model training time and increases the success rate of detailed feature identification.

In general, AI research requires large training data sets. However, in the field of “Face Skin Symptom Detection” research, there are problems from insufficient image data and uneasy acquisition. The proposed approach can be used to train a model with fewer data sets and time and has good identification results.

5. Conclusions

In this study, we compared YOLOv3, YOLOv4, Mask R-CNN + YOLOv3, and Mask R-CNN + YOLOv4 with the same number of diseases datasets and found that the accuracy of our method has improved significantly. At the same time, we experimented with Mask R-CNN before YOLO identified symptoms. The results indicate that our proposed method still achieves 57.73%, 60.38%, and 59.75% of mean average precision (mAP) for different amounts of data. Compared with only using YOLOv4 to symptom detect the image has noise, the mAP was more than about 3%.

Instead of relying on one image recognition algorithm for training, we combine multiple algorithms. We choose different algorithms in each stage according to the different features of the images. First, we segment the complex images to remove redundant images and noise. Then, we enhance the required image features for the detection of skin symptoms. This study reduces the difficulty of model training and model training time and increases the success rate of detailed feature identification.

In general, AI research requires large training data sets. However, in the field of “Face Skin Symptom Detection” research, there are problems from insufficient image data and uneasy acquisition. The proposed approach can be used to train a model with fewer data sets and time and has good identification results.

Under the influence of COVID-19, consumers’ shopping habits have changed. With so many skin care products available on the internet, choosing the right skin care product is an important issue. Through our research, consumers can understand their own skin symptoms to facilitate the proper selection of skin care products and avoid purchasing unsuitable products that may cause skin damage.

In the future, we will combine the results of our research on face skin symptom detection into a product recommendation system. The detection results will be used in a real product recommendation system. We expect to design an App for facial skincare product recommendations in the future.

| Algorithm Combination | Training Data | mAP (0.5 IoU) | mAP (0.75 IoU) | mAP (0.5:0.95 IoU) |
|------------------------|---------------|--------------|----------------|-------------------|
| YOLOv3                 | 500 images    | 55.97        | 57.64          | 53.68             |
| YOLOv4                 | 500 images    | 57.73        | 60.38          | 59.75             |
| Mask R-CNN + YOLOv3    | 250 images    | 55.10        | 56.48          | 53.19             |
| Mask R-CNN + YOLOv4    | 100 images    | 50.38        | 46.54          | 53.70             |
| Mask R-CNN + YOLOv3    | 500 images    | 55.02        | 52.39          | 58.13             |
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grand View Research. Available online: https://www.grandviewresearch.com/industry-analysis/skin-care-products-market (accessed on 22 September 2022).
2. Dey, B.L.; Al-Karaghouli, W.; Muhammad, S.S. Adoption, adaptation, use and impact of information systems during pandemic time and beyond: Research and managerial implications. Inf. Syst. Manag. 2020, 37, 298–302. [CrossRef]
3. Euromonitor International. Available online: https://www.euromonitor.com/article/e-commerce-to-account-for-half-the-growth-in-global-retail-by-2025 (accessed on 22 September 2022).
4. Qin, J.; Qiao, L.; Hu, J.; Xu, J.; Du, L.; Wang, Q.; Ye, R. New method for large-scale facial skin sebum quantification and skin type classification. J. Cosmet. Dermatol. 2021, 20, 677–683. [CrossRef] [PubMed]
5. Lucca, J.M.; Joseph, R.; Al Kubaish, Z.H.; Al-Maskeen, S.M.; Alokaili, Z.A. An observational study on adverse reactions of cosmetics: The need of practice the Cosmetovigilance system. Saudi Pharm. J. 2020, 28, 746–753. [CrossRef] [PubMed]
6. Li, H.-H.; Liao, Y.-H.; Huang, Y.-N.; Cheng, P.-J. Based on machine learning for personalized skin care products recommendation engine. In Proceedings of the 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 13–16 November 2020.
7. Song, Z.; Nguyen, K.; Nguyen, T.; Cho, C.; Gao, J. Spartan Face Mask Detection and Facial Recognition System. Healthcare 2022, 10, 87. [CrossRef] [PubMed]
8. Sertic, P.; Alahmar, A.; Akilan, T.; Javorac, M.; Gupta, Y. Intelligent Real-Time Face-Mask Detection System with Hardware Acceleration for COVID-19 Mitigation. Healthcare 2020, 10, 873. [CrossRef]
9. Park, S.R.; Han, J.; Yeon, Y.M.; Kang, N.Y.; Kim, E. Effect of face mask on skin characteristics changes during the COVID-19 pandemic. Ski. Res. Technol. 2021, 27, 554–559. [CrossRef]
10. Gefen, A.; Oussey, K. Update to device-related pressure ulcers: SECURE prevention. COVID-19, face masks and skin damage. J. Wound Care 2020, 29, 245–259. [CrossRef]
11. Jose, S.; Cyriac, M.C.; Dhandapani, M. Health problems and skin damages caused by personal protective equipment: Experience of frontline nurses caring for critical COVID-19 patients in intensive care units. Indian J. Crit. Care Med. Peer-Rev. Off. Publ. Indian Soc. Crit. Care Med. 2021, 25, 134–139.
12. Elston, D.M. Occupational skin disease among health care workers during the coronavirus (COVID-19) epidemic. J. Am. Acad. Dermatol. 2020, 82, 1085–1086. [CrossRef] [PubMed]
13. Etgu, F.; Onder, S. Skin problems related to personal protective equipment among healthcare workers during the COVID-19 pandemic (online research). Cutan. Ocul. Toxicol. 2021, 40, 207–213. [CrossRef]
14. Lin, T.Y.; Chan, H.T.; Hsia, C.H.; Lai, C.F. Facial Skincare Products’ Recommendation with Computer Vision Technologies. Electronics 2022, 11, 143. [CrossRef]
15. Hong, W.; Wu, Y.; Li, S.; Wu, Y.; Zhou, Z.; Huang, Y. Text mining-based analysis of online comments for skincare e-commerce. J. Phys. Conf. Ser. 2021, 10, 012008. [CrossRef]
16. Beauregard, S.; Gilchrest, B.A. A survey of skin problems and skin care regimens in the elderly. Arch. Dermatol. 1987, 123, 1638–1643. [CrossRef]
17. Berg, M.; Lidén, S.; Axelson, O. Facial skin complaints and work at visual display units: An epidemiologic study of office employees. J. Am. Acad. Dermatol. 1990, 22, 621–625. [CrossRef] [PubMed]
18. Park, J.H.; Choi, Y.D.; Kim, S.W.; Kim, Y.C.; Park, S.W. Effectiveness of modified phenol peel (Exoderm) on facial wrinkles, acne scars and other skin problems of Asian patients. J. Dermatol. 2007, 34, 17–24. [CrossRef]
19. Ramli, R.; Malik, A.S.; Hani, A.F.M.; Jamil, A. Acne analysis, grading and computational assessment methods: An overview. Ski. Res. Technol. 2012, 18, 1–14. [CrossRef]
20. Bhat, K.; Williams, H.C. What’s new in acne? An analysis of systematic reviews published in 2011–2012. Clin. Exp. Dermatol. 2014, 39, 273–278. [CrossRef]
21. Praetorius, C.; Sturm, R.A.; Steingrimsson, E. Sun-induced freckling: Ephelides and solar lentigines. Pigment. Cell Melanoma Res. 2014, 27, 339–350. [CrossRef]
22. Copley, S.M.; Giamei, A.F.; Johnson, S.M.; Hornbecker, M.F. The origin of freckles in unidirectionally solidified castings. Metall. Trans. 1970, 1, 2193–2204. [CrossRef]
23. Fowler, A.C. The formation of freckles in binary alloys. IMA J. Appl. Math. 1985, 35, 159–174. [CrossRef]
24. Lemperle, G. A classification of facial wrinkles. Plast. Reconstr. Surg. 2001, 108, 1735–1750. [CrossRef] [PubMed]
25. Langton, A.K.; Sherratt, M.J.; Griffiths, C.E.M.; Watson, R.E.B. A new wrinkle on old skin: The role of elastic fibres in skin ageing. Int. J. Cosmet. Sci. 2010, 32, 330–339. [CrossRef] [PubMed]
26. Hatzis, J. The wrinkle and its measurement: A skin surface Profilometric method. Micron 2004, 35, 201–219. [CrossRef] [PubMed]
27. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. Electronics 2020, 9, 1188. [CrossRef]
28. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
29. Lin, K.; Zhao, H.; Lv, J.; Li, C.; Liu, X.; Chen, R.; Zhao, R. Face detection and segmentation based on improved mask R-CNN. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 9242917. [CrossRef]

30. Anantharaman, R.; Velazquez, M.; Lee, Y. Utilizing mask R-CNN for detection and segmentation of oral diseases. In *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine*, Madrid, Spain, 3–6 December 2018.

31. Chiao, J.Y.; Chen, K.Y.; Liao, K.Y.K.; Hsieh, P.H.; Zhang, G.; Huang, T.C. Detection and classification the breast tumors using mask R-CNN on sonograms. *Medicine* **2019**, *98*, e132000. [CrossRef]

32. Zhao, L.; Li, S. Object detection algorithm based on improved YOLOv3. *Electronics* **2020**, *9*, 537. [CrossRef]

33. Singh, S.; Ahuja, U.; Kumar, M.; Kumar, K.; Sachdeva, M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimed. Tools Appl.* **2021**, *80*, 19753–19768. [CrossRef]

34. Kumar, A.; Puniitha, R. Yolov3 and yolov4: Multiple object detection for surveillance applications. In *Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology*, Tirunelveli, India, 20–22 August 2020.

35. Ju, M.; Luo, H.; Wang, Z.; Hui, B.; Chang, Z. The application of improved YOLO V3 in multi-scale target detection. *Appl. Sci.* **2019**, *9*, 3775. [CrossRef]

36. Won, J.-H.; Lee, D.-H.; Lee, K.-M.; Lin, C.-H. An improved YOLOv3-based neural network for de-identification technology. In *Proceedings of the 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jeju, Republic of Korea, 12 August 2019.

37. Gong, H.; Li, H.; Xu, K.; Zhang, Y. Object detection based on improved YOLOv3-tiny. In *Proceedings of the 2019 Chinese Automation Congress (CAC)*, Hangzhou, China, 22–24 November 2019.

38. Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-time face mask detection method based on YOLOv3. *Electronics* **2021**, *10*, 837. [CrossRef]

39. Yang, Y.; Deng, H. GC-YOLOv3: You only look once with global context block. *Electronics* **2020**, *9*, 1235. [CrossRef]

40. Mao, Q.C.; Sun, H.M.; Liu, Y.B.; Jia, R.S. Mini-YOLOv3: Real-time object detector for embedded applications. *IEEE Access* **2019**, *7*, 133529–133538. [CrossRef]

41. Zhao, H.; Zhou, Y.; Zhang, L.; Peng, Y.; Hu, X.; Peng, H.; Cai, X. Mixed YOLOv3-LITE: A lightweight real-time object detection method. *Sensors* **2020**, *20*, 1861. [CrossRef] [PubMed]

42. Albahli, S.; Nida, N.; Itaza, A.; Yousaf, M.H.; Mahmoud, M.T. Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour. *IEEE Access* **2020**, *8*, 198403–198414. [CrossRef]

43. Li, P.; Yu, H.; Li, S.; Xu, P. Comparative Study of Human Skin Detection Using Object Detection Based on Transfer Learning. *Appl. Artif. Intell.* **2022**, *35*, 2370–2388. [CrossRef]

44. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]

45. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

46. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016.

47. Zhang, D.; O’Connor, N.E.; Simpson, A.J.; Cao, C.; Little, S.; Wu, B. Coastal fisheries resource monitoring through A deep learning-based underwater video analysis. *Es-Tuarine Coast. Shelf Sci.* **2022**, *269*, 107815. [CrossRef]

48. Tanoglidis, D.; Ćiprijanović, A.; Drlica-Wagner, A.; Nord, B.; Wang, M.H.; Amsellem, A.J.; Downey, K.; Jenkins, S.; Kafkes, D.; Zhang, Z. DeepGhostBusters: Using Mask R-CNN to detect and mask ghosting and scattered-light artifacts from optical survey images. *Astron. Comput.* **2022**, *39*, 100580. [CrossRef]

49. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv 2018*, arXiv:1804.02767.

50. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 12 December 2016.

51. Khan, S.; Akram, A.; Usman, N. Real time automatic attendance system for face recognition using face API and OpenCV. *Wirel. Pers. Commun.* **2020**, *113*, 469–480. [CrossRef]

52. Menon, S.; Gerohe, A.; Aswathy, N.; James, J. Custom face recognition using YOLO V3. In *Proceedings of the 2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, Coimbatore, India, 13–14 May 2021.

53. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, J.-H. CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, 28 July 2020.

54. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]

55. Liu, S.; Qi, L.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 16 December 2018.

56. Prasetyo, E.; Sucitani, N.; Fatichah, C. Yolov4-tiny with wing convolution layer for detecting fish body part. *Comput. Electron. Agric.* **2022**, *198*, 107023. [CrossRef]

57. Kumar, A.; Kalia, A.; Sharma, A.; Kaushal, M. A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system. *J. Ambient. Intell. Humaniz. Comput.* **2021**, 1–14. [CrossRef]
58. Zhang, C.; Kang, F.; Wang, Y. An Improved Apple Object Detection Method Based on Lightweight YOLOv4 in Complex Backgrounds. *Remote Sens.* **2022**, *14*, 4150. [CrossRef]

59. DermNet. Available online: https://dermnetnz.org (accessed on 22 September 2022).

60. Freepik. Available online: https://www.freepik.com (accessed on 22 September 2022).