Analysis of Clothing Image Classification Models: A Comparison Study between Traditional Machine Learning and Deep Learning Models

Jun Xu[1], Yumeng Wei[1], Aichun Wang[1], Heng Zhao[2]*, Damien Lefloch[1]*

[1] Tiangong University, Tianjin 300387, China
[2] Shenzhen Technology University, Shenzhen 518118, China
* Corresponding authors. E-mails: 15186958102@163.com, prof.d.le.floch@gmail.com

Abstract

Clothing image in the e-commerce industry plays an important role in providing customers with information. This paper divides clothing images into two groups: pure clothing images and dressed clothing images. Targeting small and medium-sized clothing companies or merchants, it compares traditional machine learning and deep learning models to determine suitable models for each group. For pure clothing images, the HOG+SVM algorithm with the Gaussian kernel function obtains the highest classification accuracy of 91.32% as compared to the Small VGG network. For dressed clothing images, the CNN model obtains a higher accuracy than the HOG+SVM algorithm, with the highest accuracy rate of 69.78% for the Small VGG network. Therefore, for end-users with only ordinary computing processors, it is recommended to apply the traditional machine learning algorithm HOG+SVM to classify pure clothing images. The classification of dressed clothing images is performed using a more efficient and less computationally intensive lightweight model, such as the Small VGG network.

Keywords

E-commerce, clothing image classification, traditional machine learning, CNN, HOG+SVM, small VGG network.

1. Introduction

In the era of information technology, image information has become essential for transmitting and obtaining information in clothing e-commerce sales [1]. As online clothing transactions have become increasingly popular these days, the number of clothing images is also increasing, requiring them to be classified. On the one hand, the accurate classification of clothing images facilitates the storage of large amounts of data [2], but on the other it has potential in automatic clothing recognition, clothing retrieval [3,4] and clothing recommendation [5]. Therefore, many scholars have been trying to devise an efficient and accurate classification model for the massive clothing dataset in e-commerce.

Clothing images can be divided into pure clothing and dressed clothing images [6]. Pure clothing images are simply displayed clothing without a human model, with a large solid background, such as clothing flat display images. Classifying pure clothing images is relatively easy but still has a low discriminability across different textures, colours, and fabric features [7]. In contrast, dressed clothing images often have a large portion of the complex background with a human model, for example, "seller's show" and "buyer's show." In these images, the cluttered background, deformation and obscuration by the human body cause difficulty in classification [8]. As Internet shopping develops, user needs tend to be diversified. As consumers are more likely to use dressed clothing images as a resource when making purchases, dressed clothing images have gradually become crucial when selling clothing online. [9]. Manually labeling attributes for such a large number of complex clothing images is labour-intensive and time-consuming, which is also easily influenced by subjective judgment [10]. The application of image classification technology to clothing images allows us to identify the deeper features of clothing images. However, the author still faces two questions: 1) whether traditional machine learning algorithms or deep learning models are more suitable for the classification of pure clothing images, and 2) which deep learning model is more suitable for the classification of dressed clothing images.

This paper aims to determine proper classification models for each pure clothing and dressed clothing image, thereby enabling the classification of clothing images in the e-commerce industry. First, the author summarises and analyses the existing classification models for pure clothing and dressed clothing images. Then, several models are selected, shown in *Figure 1(a)*, which are described in detail. In order to determine proper models, the author conducted two experiments (*Figure 1(b)*) for pure clothing and dressed clothing images, respectively. A different dataset was used for each experiment, shown in *Figure 1(c)*. In the first experiment, a typical traditional machine learning algorithm, HOG+SVM, was compared with two deep learning models. For the HOG+SVM algorithm, the author tested four cases. The rbf in parentheses means that the SVM uses rbf (Gaussian kernel) as the kernel function, and Linear in parentheses means that a linear SVM is used. 4x4 and 8x8 in parentheses indicate the size of each cell for HOG features. In the second experiment, the HOG+SVM algorithm was compared with the small VGG network, large VGG network, and GostNet network. And for retail end-users, the author selected different numbers of data sets to experiment on and test the effect. Furthermore, the
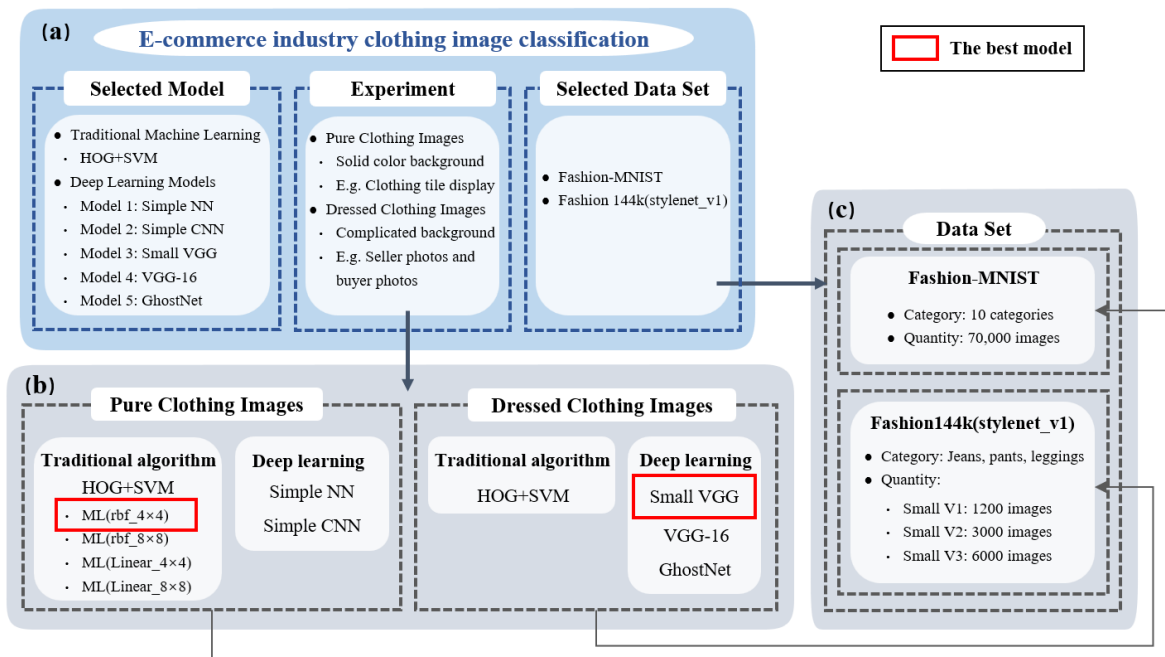
Fig. 1. Research object and methodology

author summarises the factors that affect the accuracy of the models.

The remainder of this paper is organised as follows. Section 2 provides an overview of related work, and Section 3 describes the theoretical model used in the experiments. The experimental design is given in Section 4. The results and discussion are proveded in Section 5. Finally, Section 6 summarises the conclusions of this paper and outlines potential future research directions.

2. Literature Review

Due to its increasing significance in the e-commerce industry, considerable research has been carried out on clothing image classification. The existing methods for clothing image classification can be divided into two categories: a) traditional machine learning algorithms and b) deep learning models. Nevertheless, existing clothing classification methods still face three fundamental challenges in their practical application. First, the increased variation of clothes in style, texture, and details leads to a complex image information base. Second, clothes are often distorted

and creased when worn on the human body. Third, clothing images appear quite different in different scenes. This section discusses studies to overcome these challenges from the two aspects of pure clothing image classification and dressed clothing image classification.

2.1. Pure clothing image classification

At the early stage of e-commerce development, pure clothing images occupied a relatively large proportion. Traditional machine learning algorithms were commonly used in the early studies for clothing image classification, which manually extracted image features for classification [11]. However, the traditional machine learning algorithms showed their shortcomings in handling a large volume of image data. To overcome the limitation, deep learning models have been actively studied to classify pure clothing images [12,13].

The traditional machine learning algorithms extract artificially pre-defined features (edges, corner points, colours, etc.). Typical feature extractors include the Scale-Invariant Feature

Transform (SIFT) [14], Speeded-Up Robust Features (SURF) [15], and the Histogram of Oriented Gradient (HOG) [16]. Thewsuwan et al. [17] established a preprocessing technique based on the Local Binary Pattern (LBP) and Gabor filter for a clothing classification system. Sha et al [18] adopted Uniform Local Binary Pattern (ULBP) features for pattern attributes and the Pyramid Histogram of Oriented Gradients (PHOG), Fourier, and GIST features for the collar and sleeve attributes. Surakarin et al [19] proposed the Bag of Features (BoF) model based on SURF and the Local Directional Pattern (LDP) to identify clothing types. Li et al. [20] proposed a Dragonfly Algorithm (DA)-optimised Online Sequential Extreme Learning Machine (OSELM)-based clothing image classification technique for pure clothing image classification.

The NN model mimics the characteristics of biological NNs. By simulating the architecture and function of a human NN, a complex NN consists of a large number of neurons, which simulates and inhibits neurons to complete complex operations. Recently, deep neural networks have been advanced and surpassed the traditional multilayer perceptron neural networks.

Deep neural networks automatically learn discriminative features from data without requiring a hand-crafted feature construction process. Especially, CNN has been used in clothing image classification, with its powerful feature extraction capabilities. Typical, CNN includes LeNet [21], AlexNet [22], GoogleNet [23], VGGNet [24], and GhostNet [25]. Lao et al. [26] proposed a clothing classification model based on the AlexNet model and a clothing target detection model based on R-CNN. Dong et al. [27] adopted VGG-Net and Spatial Pyramid Pooling (SPP) for multi-scale clothing image classification. Xiang et al. [28] constructed a dataset containing approximately 100,000 shirt images and proposed an RCNN-based classification framework. Di et al. [29] used four Neural Network (NN) models, such as a fully connected neural network, the Convolutional Neural Network (CNN), MobileNetV1, and MobileNetV2, to classify clothing images based on the Fashion-MNIST dataset.

According to literature, a traditional machine learning algorithm, such as HOG, requires only a small number of samples to obtain high accuracy and certain generalisation ability. However, there is still a strong dataset dependency or scene-specific dependency, and it can only classify pure clothing images with solid colour backgrounds. Although the NN model, with its automatic learning feature, achieves a better performance in classifying pure clothing images, it requires a large amount of data to train the model, which is not suitable for small- scale pure clothing images.

2.2. Dressed clothing image classification

Dressed clothing images have been taken more portion in clothing image data as the e-commerce industry develops and flat models emerge. Accordingly, many studies have been conducted on dressed clothing image classification.

Due to the complexity of dressed clothing images, only a few studies have been based on traditional machine

learning algorithms to achieve human silhouette and clothing recognition in clothing images. In the study by Bossard et al. [30] , HOG and LBP features were fused and fed into the SVM classifier to distinguish the background and human body of the dressed clothing images in natural scenes. Huo et al. [31] proposed a method based on the Deformable Part based Model (DPM) and key point regression method for locating the head and shoulders and the human torso in dressed clothing images. Liu et al. [32] proposed Colour-Fashion, which combines the human pose estimation module and colour information based on SIFT and HOG.

Most existing methods are mainly based on deep learning models to effectively handle the complex background, noise, and relatively small image area of dressed clothing images. Liu et al. [33] constructed a large-scale and semantically labeled comprehensive clothing dataset - DeepFashion. Also, they proposed a FashionNet based on global convolutional features and local keypoint features, incorporating four types of supervised information: broad categories, attributes, clothing IDs, and key-points for clothing feature learning. Wang et al. [34] combined CNN and Recurrent Neural Networks (RNN) to build a CNN-RNN framework for multi-label image classification. Nawaz et al. [35] used an Inception module-based deep network to automatically classify traditional Bangladeshi ethnic costume images. Cychnerski et al. [36] proposed a joint system for the accurate detection and classification of clothing images, providing results for fine-grained five attributes that outperformed SqueezeNet [37] and ResNet on the DeepFashion dataset???. Zhang et al. [38] proposed an effective deep learning network to classify a dataset containing 9,339 dressed clothing images. Rohrmanstorfer et al. [39] constructed a dressed clothing image dataset containing 2,567 images and employed CNN as a classification model. Hodecker et al. [40] evaluated the accuracy of CNN for clothing classification as compared to non-convolutional models.

As discussed, most existing dressed clothing image classification methods are based on the deep learning model. The performance of the traditional machine learning algorithm is limited, whose capability is only to distinguish the background and the body of dressed clothing images according to the body outline. CNNs can automatically extract image features and recognise highly distorted dressed clothing images with translation, scaling, and distortion invariance. However, a deep learning model still requires a large number of data for training.

In summary, for pure clothing image classification, there are more studies using traditional machine learning algorithms rather than deep learning models. For dressed clothing image classification, only a few studies use traditional machine learning algorithms to segment human silhouettes and clothing in dressed clothing images, and most of the studies are based on deep learning models. Both traditional machine learning and deep learning methods have achieved good performance for clothing image classification. However, the general trend of clothing image classification is toward more complex and a greater amount of dressed clothing image classification, and there are still many difficulties and challenges to be solved. For example, despite good performance in pure clothing image classification, the accuracy rates differ for different sizes of data sets. In dressed clothing image classification, deep learning models provide better performance than traditional machine learning algorithms. However, the performance of different deep learning models varies depending on the dataset and target tasks. Therefore, in this paper, two types of classification methods were explored in subsequent experiments to test the classification of pure clothing images and dressed clothing images, respectively.

3. Theoretical model

The essence of clothing image classification is to use image features to determine the category of clothing through a classification model. The flow
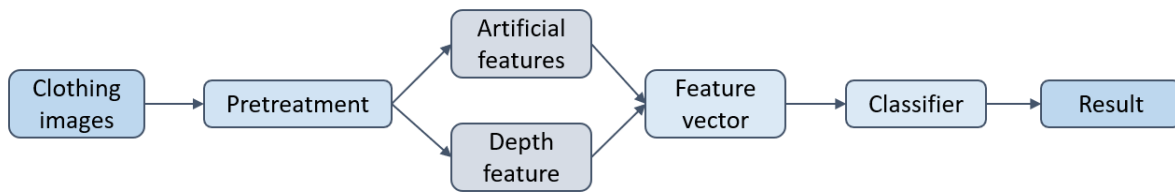
Fig. 2. Flow chart of clothing image classification

chart of clothing image classification is shown in **Figure 2.** First, the clothing image is preprocessed, and then the image features are extracted using traditional machine learning algorithms or deep learning models. The extracted features are fed into the classifier, obtaining the output category. This paper analyses the representative traditional machine learning algorithm, HOG+SVM, and the following effective deep learning models: NN, CNN, Small VGG, VGG-16, and GhostNet. In the subsequent sections, they are described in detail.

3.1. Clothing image classification based on traditional machine learning

Since traditional machine learning algorithms are based on artificially pre-defined features (edges, corner points, colours, etc.), appropriate features should be selected for different target tasks. SIFT, SURF, and HOG are commonly used as feature descriptors, and SVM, the Extreme Learning Machine (ELM), and Random Forest (RF) are commonly used as a classifier. Especially, HOG is one of the most successful feature descriptors in object detection, which constructs features based on the histogram of gradient directions in local regions of an image. SVM is a generalised linear classifier for supervised data classification. The combination of HOG and SVM has been widely used in many types of classification tasks, mainly due to its robustness and accuracy. In this framework, the input image is first divided into many small connected areas called cell units. Then, the gradient direction of each pixel in the cell unit is accumulated to construct a histogram of directional gradients. The histogram is normalised by the density of each

histogram for illumination robustness. The normalised block descriptor is defined as the HOG descriptor. The HOG descriptors of all blocks in the detection window are combined to form the final feature vector, and then the SVM classifier is used for image detection and classification. A schematic diagram of this process is depicted in **Figure 3** [16].

## 3.2. Clothing image classification based on deep learning models

A basic NN model consists of neurons with weights and biases, simulating biological neurons for information processing. During the training process, the weights and biases of neurons of the model are adjusted to process the input information into output in accordance with training data. With the rapid development of deep learning in recent years, CNN has achieved great success in image processing and computer vision tasks, with a deeper and feed-forward architecture suitable for image data [41] .

3.2.1. Model 1 — simple NN

The architecture of a simple NN is shown in **Figure 4**. As shown in **Figure 4**, the model consists of a flatten layer, a fully connected (FC) layer with 128 neurons, followed by Relu activation, and an FC layer with 10 neurons, followed by softmax activation. The flatten layer converts the multidimensional input into one-dimensional input.

3.2.2. Model 2 — CNN model

As shown in **Figure 5**, the CNN model consists of an input layer, a convolutional
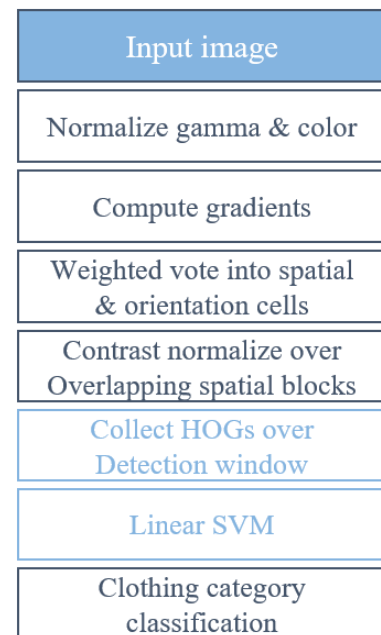
Above should be '...overlapping..'...detection....'

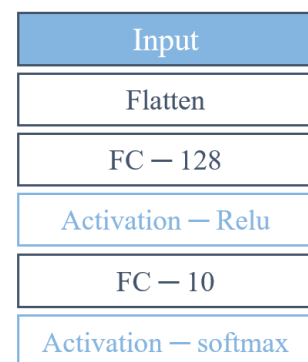*Fig. 3. Schematic diagram of HOG+SVM algorithm*

Fig. 4. Architecture of a simple NN

layer with 32 3x3 kernels, a flatten layer, and an FC layer. The convolutional layer is followed by Relu activation, a BatchNormalisation layer, max-pooling layer, and dropout layer. The FC layer is followed by the softmax activation

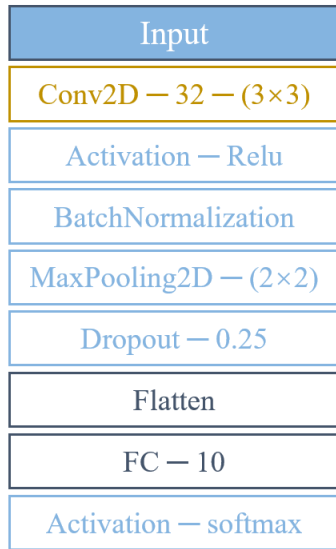| Input |
| :---: |
| Conv2D — 32 — (3×3) |
| Activation — Relu |
| BatchNormalization |
| MaxPooling2D — (2×2) |
| Dropout — 0.25 |
| Flatten |
| FC — 10 |
| Activation — softmax |

Fig. 5. Architecture of the CNN

function. BatchNormalisation serves to normalise the data, which can speed up the convergence and prevent the overfitting problem. The dropout layer can mitigate overfitting by randomly deactivating a certain number of neurons per iteration.

3.2.3. Model 3 — Small VGG network model

The small VGG network model was made by Adrian Rosebrock [42]. As shown in **Figure 6**, the small VGG network is a lightweight version of the full VGG-Net, which is faster in the compromise of accuracy. The small VGG network consists of repeated ConvBlocks and MiddleBlocks, where the ConvBlock consists of a convolutional layer, followed by the Relu activation

function and BatchNormalization layer. The MiddleBlock consists of two ConvBlocks, a MaxPooling layer, and dropout layer. The class_len in the penultimate layer indicates the number of labels in the dataset, which is 3 in this case.

3.2.4. Model 4 — VGG-16 network model

The VGG network model was proposed by the Visual Geometry Group at Oxford [43] . The VGG network has two structure variants: VGG-16 and VGG-19, which have a similar structure but different network depths. In this paper, the VGG-16 network is used. The VGG-16 network consists of 13 convolutional layers and 3 FC layers.
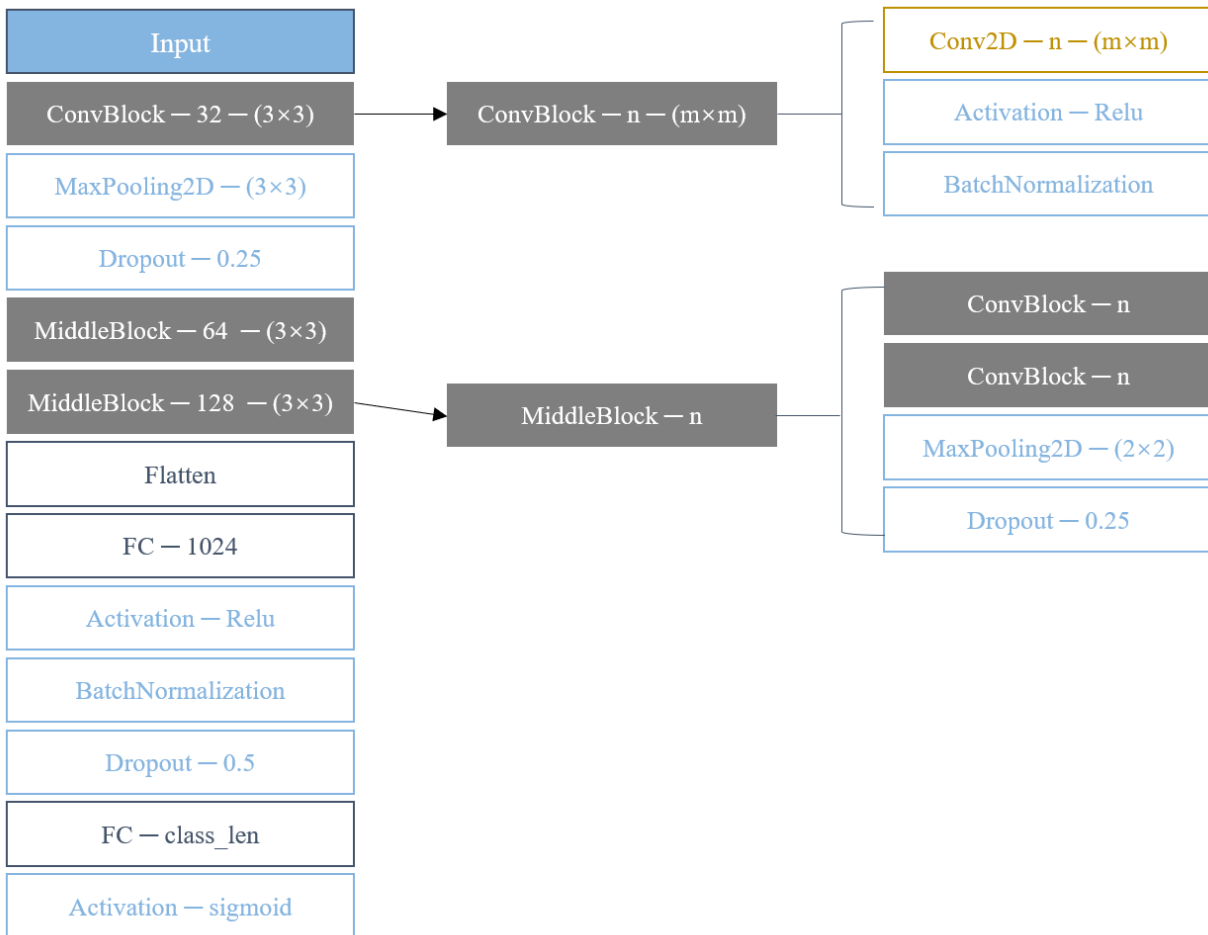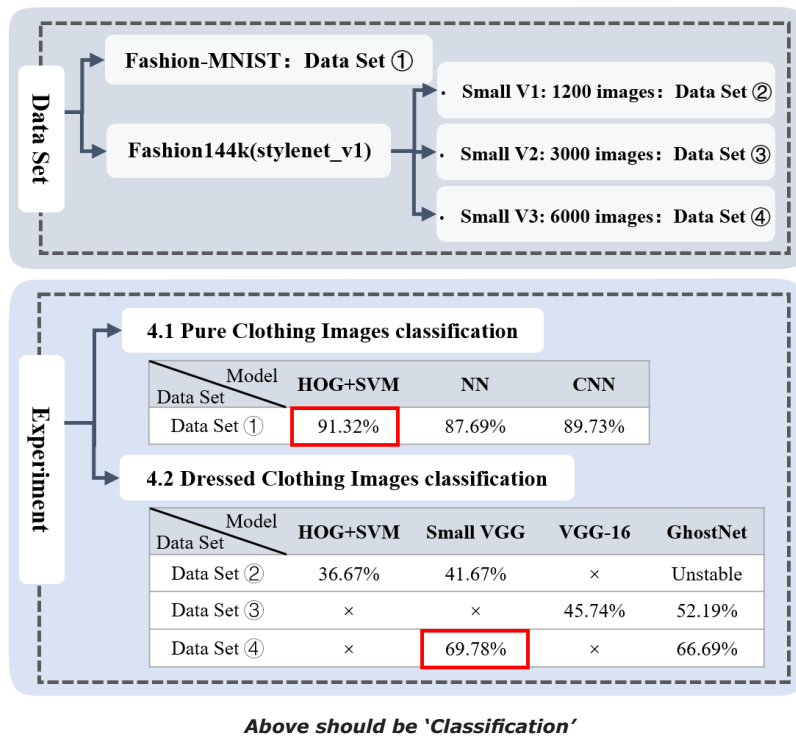
Fig. 6. Architecture of small VGG

Model / Data Set	HOG+SVM	NN	CNN
Data Set ①	91.32%	87.69%	89.73%

4.2 Dressed Clothing Images classification

Model / Data Set	HOG+SVM	Small VGG	VGG-16	GhostNet
Data Set ②	36.67%	41.67%	×	Unstable
Data Set ③	×	×	45.74%	52.19%
Data Set ④	×	69.78%	×	66.69%

Above should be 'Classification'

Fig. 7. Experimental flow chart

3.2.5. Model 5 — GhostNet model

The GhostNet is a lightweight but effective model built by Huawei Noah's Ark Laboratory. The GhostNet has the advantage of fast training and fewer hardware requirements, but to the detriment of accuracy. The author chose GhostNet to represent the lightweight network in this paper.

In summary, the author selected a representative traditional machine learning algorithm: HOG+SVM, and representative deep learning models: a simple NN, a simple CNN, a small VGG, a large VGG-16, and a light GhostNet. The above algorithms and models perform differently in different image domains and sample sizes. The purpose of our experiments is to find suitable models for clothing image classification tasks.

4. Experimental Design

After the literature review and theoretical model above, we designed two-part experiments for pure clothing images

and dressed clothing images. For the classification of pure clothing images, experiments were mainly conducted using traditional machine learning algorithms with different parameters, and compared with simple deep learning models. For the classification of dressed clothing images, different kinds of deep learning models were mainly used for experiments and compared with traditional machine learning algorithms. In particular, the experiment of the second part was carried out in two steps. First, the traditional HOG-SVM algorithm is compared with the small VGG network. Then, in the second step, the Small VGG network was compared with the large VGG-16 network and light GhostNet. **Figure 7** depicts a flow chart of the experiments.

4.1. Pure clothing image classification

4.1.1. Dataset

The Fashion-MNIST dataset is used to evaluate the models for pure clothing images. The Fashion-MNIST dataset was constructed by the research division of Zalando [44] , which consists of 70,000

28x28 pixel grayscale images. The categories and some example images are shown in **Figure 8**. These 70,000 images were labeled in ten categories: t-shirt/top, trousers, pullover, dress, coat, sandal, shirt, sneaker, bag, ankle boot, which are labeled as 0~9 (**Table 1**). In our experiments, the author used 60,000 images as the training set and 10,000 images as the validation set.

4.1.2. Experimental result

The traditional machine learning algorithm HOG+SVM was compared with the simple NN and simple CNN. In order to analyse the accuracy of the three methods for different numbers of training samples, the data size was gradually increased in the experiment: from 500 to 30,100. The training and testing accuracies are depicted in Figure 9, where ML denotes the HOG+SVM algorithm, NN denotes the simple NN model, and CNN denotes the CNN model. For the HOG+SVM algorithm, the author tested four cases with the necessary additions in parentheses. The rbf in parentheses means that the SVM uses rbf (Gaussian kernel) as the kernel function, Linear in parentheses that a linear SVM is used, and 4x4 and 8x8 in parentheses indicate the size of each cell for HOG features. Among them, since the HOG+SVM algorithm with the Gaussian kernel converges, the author increased the interval of the number of test samples when the ML (rbf_4x4) reached the saturation period (7,700), selected 13300, 18900, 24500, and 30100 samples for testing and continued to plot the line segments.

The results of the comparison show that ML (rbf 4x4) provides the best accuracy, while ML (Linear 8x8) provides the worst accuracy, and the NN and CNN are moderately accurate, but have some fluctuations due to dataset effects. The rbf kernel's results outperformed those of the linear kernel, indicating that the Gaussian kernel brings higher accuracy than a linear SVM. Also, the small size of the cell (4x4) provided higher accuracy than the large size of the cell (8x8), indicating that 4x4 is suitable for this dataset.
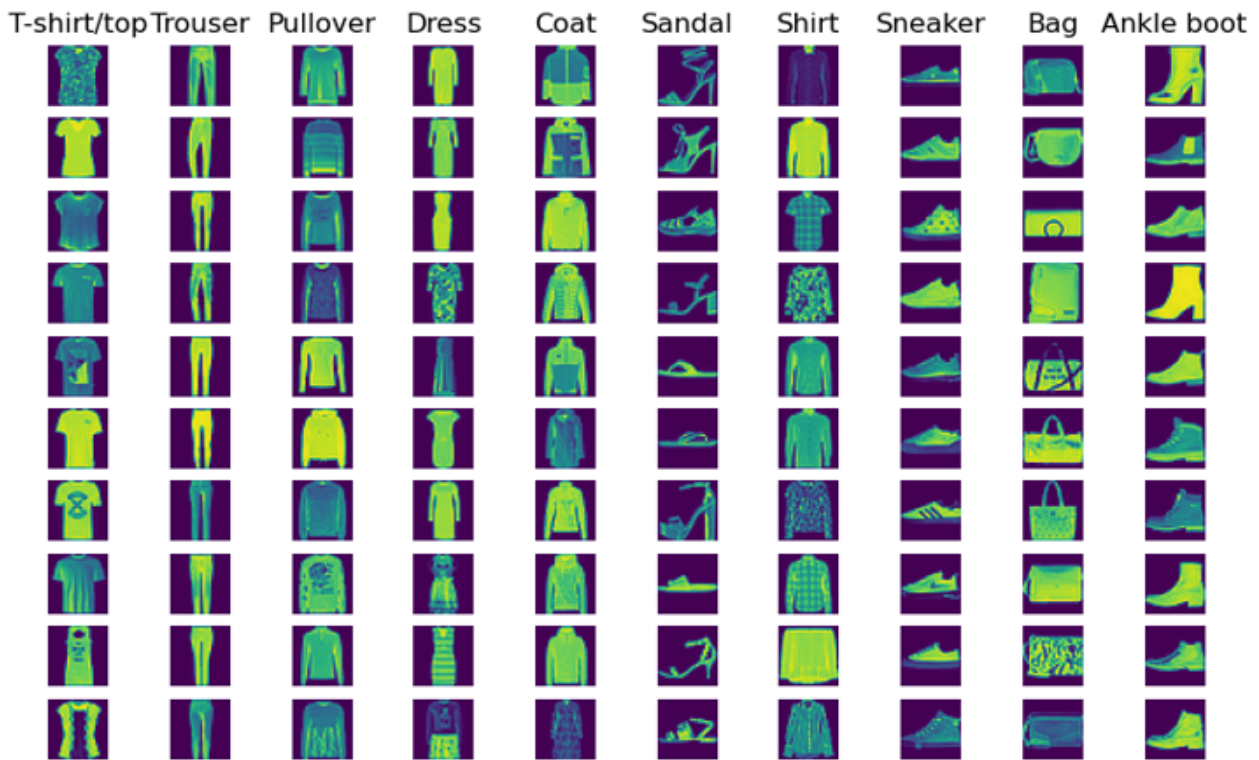
Fig. 8. Examples and categories of the Fashion-MNIST dataset

Table 1. Fashion-MNIST dataset label description

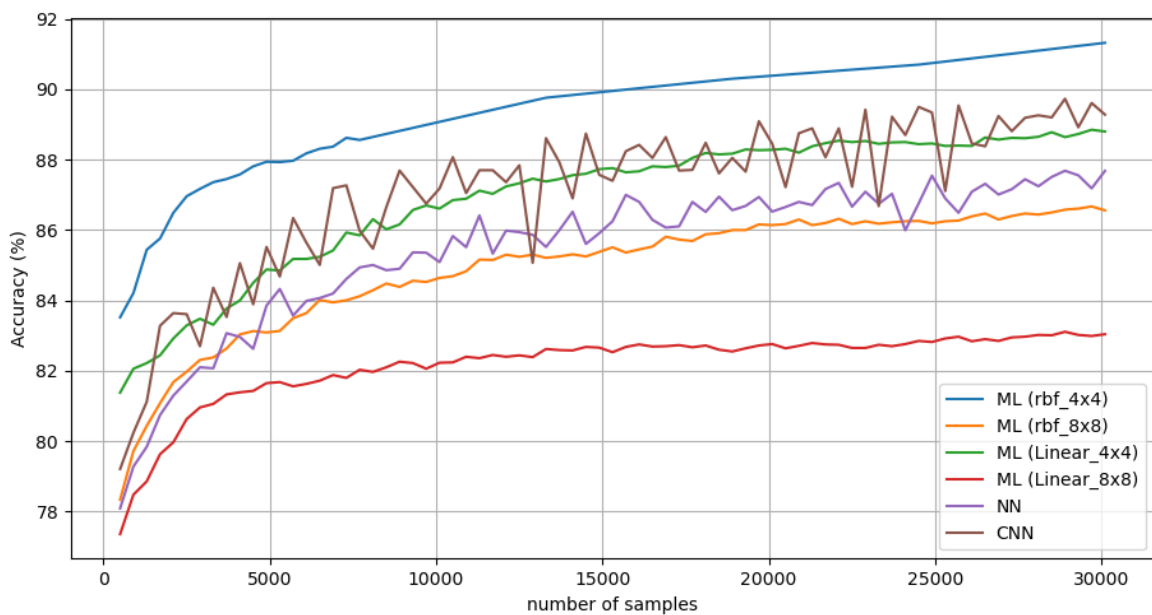| Description	T-shirt/top	Trouser	Pullover	Dress	Notes
Label	0	1	2	3	4
Description	Sandal	Shirt	Sneaker	Bag	Ankle boot
Label	5	6	7	8	9

Fig. 9. Accuracy comparison of HOG+SVM, NN, and CNN.

Table 2 compares the highest accuracy of different models, showing that the ML (Linear_4x4) outperforms the other methods, including NN and CNN, when the number of samples is very small. It shows that better results can be obtained by adjusting the parameters of traditional machine learning algorithms. But it should also be noted that using the rbf kernel and increasing the cell size of HOG features will increase the computational effort.

Table 2. Highest accuracy comparison of different models

Name	Highest Accuracy
ML (rbf_4x4)	91.3%
CNN	89.7%
ML (Linear_4x4)	88.9%
NN	87.7%
ML (rbf_8x8)	86.7%
ML (Linear_8x8)	83.1%

4.2. Dressed clothing image classification

4.2.1. Dataset

In this experiment, the Fashion144k (stylenet_v1) dataset [45] was used, which is a sub-dataset of Fashion144k. It contains more than 89,000 colour clothing images, 123 clothing feature labels, and 3,179 colour feature labels. The dataset images are 256x384 in size, which contains complex backgrounds. Unlike the Fashion-MNIST dataset, this dataset images contain the full body clothing of the person and more tags. **Figure 10** shows example images of the dataset. By observing the images in the dataset, the authors found that most of the images have more than one top, which will greatly interfere with the recognition of the model, therefore we filtered out all bottom images for the experiment. In order to train our model step by step, the filtered images were preprocessed and then divided into several sub-datasets. Each sub-dataset is split into a training set and validation set for subsequent experiments.

Dataset#1—SmallV1. Redundant labels were eliminated from the Fashion144k (stylenet_v1) dataset, thereby retaining only jeans, pants, and leggings. MRCNN [46] was then used to extract the clothing locations, and thus 400 images from each category were collected. The mislocated and misclassified images were manually eliminated, and the remaining 1,065 images were used as the dataset. The dataset, so-called SmallV1, was split into the training set and validation set at a ratio of 8:2. Example images of the SmallV1 dataset are presented in **Figure 11**.

Dataset#2 — SmallV2. The SmallV2 sub-dataset was further extended using the MRCNN, increasing each category of jeans, pants, and leggings to approximately 1,000 images. The 3,000 images of the SmallV2 dataset were also split into training and validation sets at a ratio of 8:2.

image 1 image 2 image 3 image 4 image 5 image 6
image 7 image 8 image 9 image 10 image 11 image 12
image 13 image 14 image 15 image 16 image 17 image 18
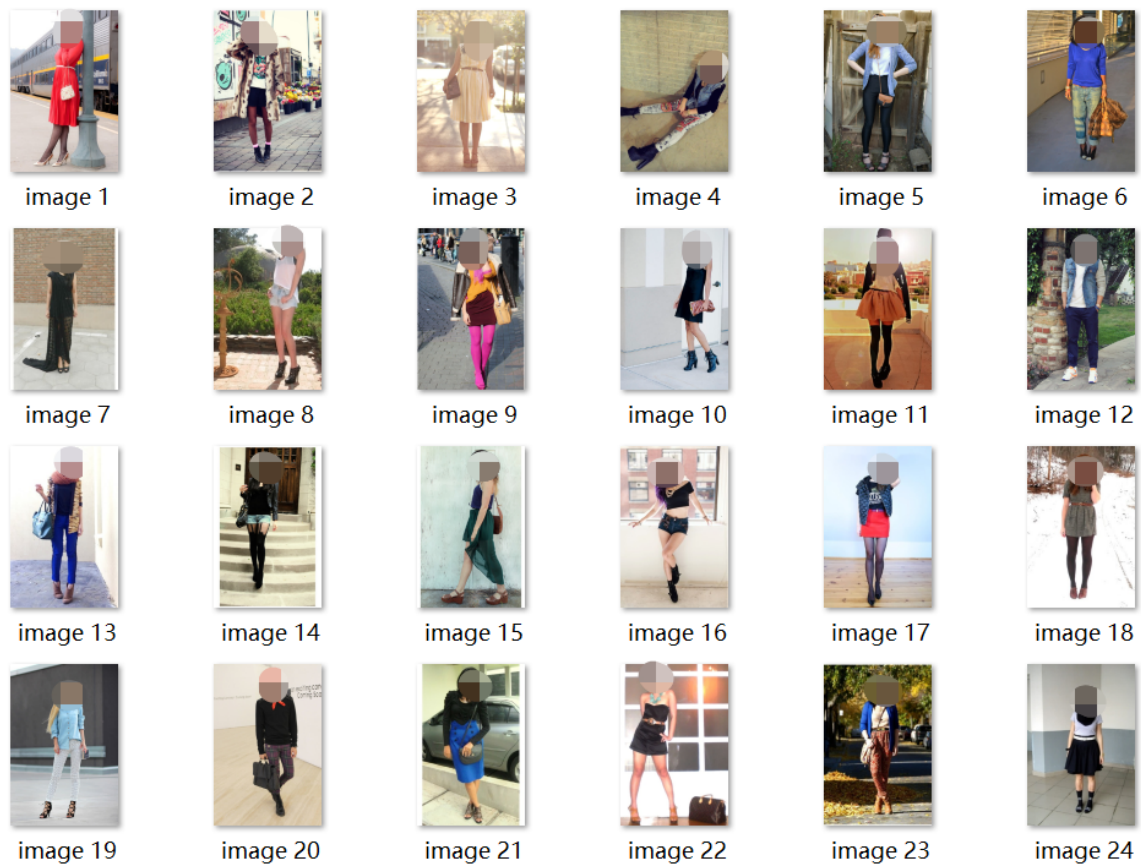image 19 image 20 image 21 image 22 image 23 image 24

Fig. 10. Example images of the Fashion144k dataset

Fig. 11. Example images of the SmallV1 dataset

Dataset#3 — SmallV3. The lower 2/3 height of the images within the Fashion144k (stylenet_v1) dataset was collected as the training set. Thus, 2000 images per category were obtained (2000, 2000, 1932 images, respectively). The 6,000 images of the SmallV3 dataset were also split into training and validation sets at a ratio of 8:2.

4.2.2. Experimental result

In the first experiment, the HOG+SVM algorithm was compared with the Small VGG network model for the SmallV1 dataset. *Figure 12* depicts a recognition accuracy comparison of the two methods, revealing that neither have high recognition accuracy for the validation set. As shown in *Figure 12*, the HOG+SVM algorithm performs poorly for the validation set despite good accuracy for the training set. In contrast, the recognition accuracy of the Small VGG network model is low for both the training set and validation set. Even

though the Small VGG network obtains a bit higher accuracy than the HOG+SVM algorithm for the validation dataset, the recognition accuracy is very limited, only 41.67%. One possible reason could be the small size of the SmallV1 dataset, only 1,065 images. The results of the first experiment show that the advantages of CNN only appear when the amount of data is larger. Therefore, the SmallV2 dataset with a medium number of images was not used in the next experiments, while the SmallV3 dataset, which has the largest number of images, was used directly .

As shown in *Figure 13*, the recognition accuracy of the Small VGG network model was improved to 69.78% for the SmallV3 dataset. The result indicates that the recognition accuracy of the CNN can be improved when the number of data increases.

4.2.3. Further Experiment result

Since the Small VGG network did not provide high enough accuracy in the previous section, further experiments were conducted to find more suitable models. Two additional deep learning models were adopted: VGG-16 and GhostNet, which are a large network and light network, respectively.

VGG-16 (large network). Since VGG-16 is a large network model, it was trained and tested for the SmallV2 dataset with a medium sample size. First, the author used Adam as the optimiser for training, and the VGG-16 model was trained for the SmallV2 dataset from scratch. It appears that the model does not learn properly. Then, the author retrained the VGG-16 model using the SGD optimiser. The model achieved an accuracy of 33.12% after 10 epochs and 45.74% after 40 epochs. Despite its large size, VGG-16 does not provide high accuracy in this experiment, from which it can
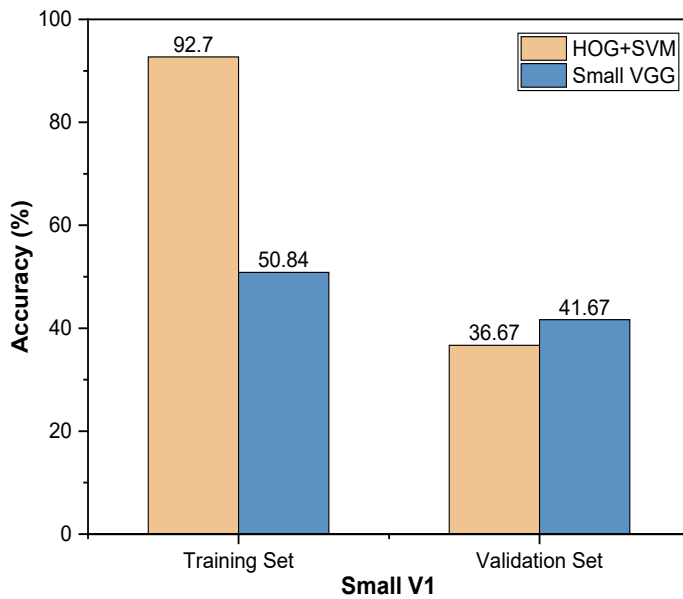
Fig. 12. Recognition accuracy of HOG+SVM and Small VGG for the SmallV1 dataset
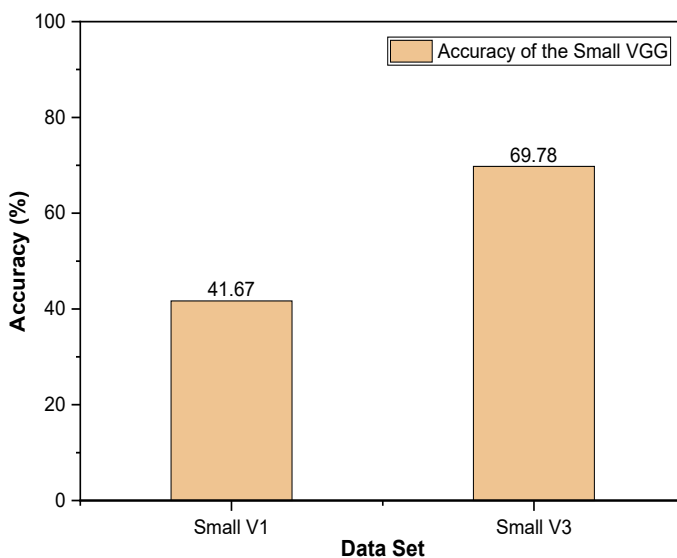

Fig. 13. Recognition accuracy of Small VGG network models in different datasets

be concluded that it is not suitable for clothing classification in e-commerce.

GhostNet (lightweight network). Since the GhostNet model is a lightweight network, it was trained and tested for the SmallV1 dataset. The GhostNet obtained a recognition accuracy of 92.90% for the training set but only 38.86% for the validation set. A further experiment was conducted adjusting the network parameters as follows: the Dropout to 0.35; a fully connected layer was added before the output layer, and the relu activation function and l2 regularisation were used. As a result, the model obtained a recognition accuracy of 58.46% for the training set and 45.02% for the validation set. Another experiment was conducted with another adjustment as follows: the Dropout to 0.45, and the increased epochs provided a recognition accuracy of 64.02% for the training set and 37.44% for the validation set. From those three experiments, it is found that the classification accuracy of the GhostNet network model for the SmallV1 data set is not stable enough.

In contrast, the GhostNet trained for the SmallV2 dataset obtained a recognition accuracy of 86.49% for the training set and 52.19% for the validation set, as shown in *Figure 14*. The results show that such a lightweight network as GhostNet can also be used for dressed clothing image classification, but the stability is slightly less compared with the Small VGG network models.

5. Discussion

5.1. Pure clothing image classification

The comparison analysis of the HOG+SVM algorithm, simple NN, and simple CNN for the Fashion-MNIST dataset shows that the traditional machine learning algorithms can achieve better accuracy than deep learning models for the small size dataset. However, when the number of samples increases, the accuracy of the CNN can also reach 89.73% accuracy and gradually improve with an increase in the number of samples. Therefore, the author concludes that the traditional machine learning HOG+SVM algorithm is more suitable for pure clothing image classification, which can better capture local shape information and has good invariance to geometric and optical changes. However, in the era of big data, as a large number of clothing datasets appear, using the CNN model is also one of the good choices.

5.2. Dressed clothing image classification

The first experimental comparison shows that the Small VGG network model obtains higher recognition accuracy than the HOG+SVM algorithm. Also, the recognition accuracy increased for the SmallV3 dataset with the largest number of images. In further experiments, two other types of deep learning models were analysed: a large network - VGG-16 and a lightweight network - GhostNet. The results show that the VGG-16 network is not suitable for dressed clothing image classification, while the GhostNet obtains
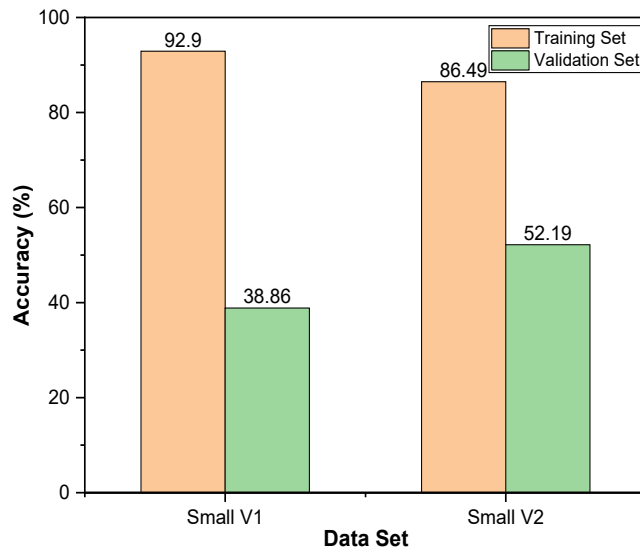
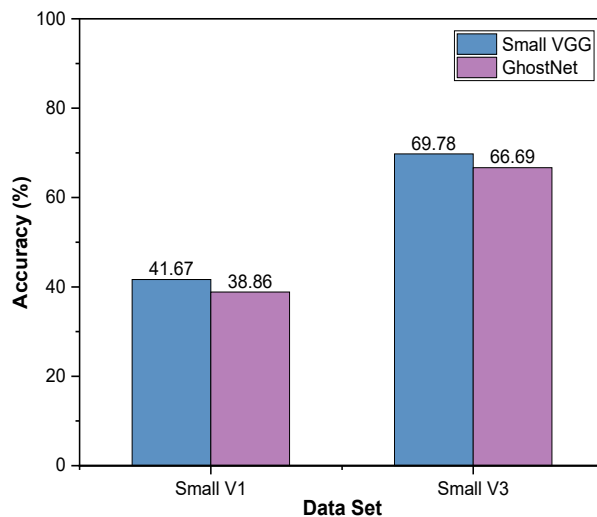Fig. 14. Recognition accuracy of the GhostNet model for different datasets

Fig. 15. Accuracy of Small VGG and GhostNet for different numbers of datasets

a comparable recognition accuracy but is unstable compared to the Small VGG network. Therefore, the Small VGG has the best results in dressed clothing image classification experiments and is expected to further improve its accuracy as the number of samples increases.

5.3. Further analysis of other factors affecting accuracy

Based on the experimental results, the author believes that the accuracy of the Small VGG network model can be further improved. There are many factors that affect the recognition accuracy of the

model, including clothing style, image orientation, image colour, number of datasets, and so on. However, since the image orientation is consistent in this paper, and the purpose of the experiment is to classify clothing styles, this section mainly summarises the influence of image colour and the number of data sets on the recognition accuracy of deep learning models.

First, the influence of colour on accuracy was analysed. The results obtained with colour and grayscale images of the SmallV2 dataset are compared. Note that the original colour images were converted into grayscale in the comparison, and the

network was trained with 20 epochs. The results show no major difference, indicating no significant contribution of colour information to clothing image classification.

Second, the influence of the sample size of the training dataset on accuracy was analyzed. To this end, the Small VGG network and GhostNet network were compared for the SmallV1 and SmallV3 datasets. As shown in **Figure 15**, the accuracy of clothing image classification increases as the number of training images increases. Accordingly, the quality and quantity of the training dataset significantly impact the accuracy of the network model.

6. Conclusion

In the age of data, accurate automated classification of images has been considered crucial to improve the operational efficiency of the e-commerce apparel industry. This paper, targeting small and medium-sized clothing companies or merchants, compares traditional machine learning algorithms and deep learning models to determine suitable models for each group. The experimental results demonstrate that the HOG+SVM algorithm with the Gaussian kernel function obtains the highest accuracy of 91.32% in the classification of pure clothing images. In contrast, the CNN model obtains a higher accuracy than the HOG+SVM algorithm in the dressed clothing image classification. Accordingly, for end-users with only ordinary computing processors, it is recommended to apply the traditional machine learning algorithm HOG+SVM to classify a limited number of pure clothing images. The classification of dressed clothing images with complex image information can be performed using a more efficient, shorter training time and a less computationally intensive lightweight model, such as the Small VGG network. In addition, for small and medium-sized clothing companies or merchants, they need to manage incoming and outgoing goods on a daily basis, add tags to clothing images and upload them to the cloud to save data.

But at the same time, a classification model with low cost and fast computing is needed. In this case, the model in the paper can be adapted to clothing retail in Amazon and similar categories.

Our future works will include constructing a large clothing image dataset with more style labels and more images to test different models, which will provide a reference for large mobile clients. And the experiments will also include tests of the

model's operational efficiency, stability and other performance. In addition, the author will further conduct a study of clothing video classification due to the current prevalence of live streaming industry videos.

References

1. Jing Y, Liu D, Kislyuk D, Zhai A, Xu J, Donahue J, Tavel S. Visual search at Pinterest, In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*. Association for Computing Machinery, New York, NY, USA, 2015; pp. 1889–1898.

2. Gao Z, Han L. Clothing image classification based on random erasing and residual network. *Journal of Physics: Conference Series* 2020; Vol. 1634 No. 1, pp. 012136.

3. Feng Y, Du C, Hua A, Jiang M, Wei X, Peng T, Hu X. EnCaps: Clothing Image Classification Based on Enhanced Capsule Network. *Applied Sciences* 2021; Vol. 11 No. 22, pp.11024.

4. Kiapour MH, Han X, Lazebnik S, Berg AC, Berg TL. Where to buy it: Matching street clothing photos in online shops, in: *IEEE International Conference on Computer Vision (ICCV)*, IEEE, Santiago, Chile 2015; pp. 3343-3351.

5. Liu KH, Chen TY, Chen CS. MVC: A Dataset for View-Invariant Clothing Retrieval and Attribute Prediction, in: *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval (ICMR '16)*, Association for Computing Machinery, New York, NY, USA 2016; pp. 313–316.

6. Deng LL. Pre-detection Technology of Clothing Image Segmentation Based on GrabCut Algorithm. *Wireless Pers Commun* 2018; Vol. 102 No. 2, pp. 599–610.

7. Yue X, Zhang C, Fujita H, Lv Y. Clothing fashion style recognition with design issue graph. *Appl Intell* 2020 ; Vol 51 No. 10, pp. 1-13.

8. Li Y, He Z, Wang S, Huang W. Multideep Feature Fusion Algorithm for Clothing Style Recognition. *Wirel Commun Mob Com* 2021 ;Vol. 2021 No. 4, pp. 1-14.

9. Kashilani D, Damahe LB, Thakur NV. An Overview of Image Recognition and Retrieval of Clothing items, in: *IEEE International Conference on Research in Intelligent and Computing in Engineering (RICE)*, IEEE, San Salvador, El Salvador 2018; pp. 1-6.

10. Donati L, Iotti E, Mordonini G, Prati A. Fashion Product Classification through Deep Learning and Computer Vision. *Applied Sciences*, 2019; Vol. 9 No. 7, pp. 1385.

11. Stricker M, Orengo M. Storage and Retrieval for Image and Video Databases III. *SPIE International Society for Optics and Photonics* 1995; Vol. 1995, pp. 381-392.

12. Hu MK. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on* 1962; Vol. 8 No. 2, pp.179-187.

13. Li Y, Xu S, Luo X, Lin S. A new algorithm for product image search based on salient edge characterization. *Journal of the Association for Information Science & Technology* 2014 ; Vol. 65 No. 12, pp. 2534–2551.

14. Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 2004; Vol. 60 No. 2, pp. 91-110.

15. Li KL, Liu ZD. Clothing Information System Design Based on Images Retrieval, in: *International Conference on Advanced Educational Technology and Information Engineering (AETIE 2015)*, Computer Information Centre, Beijing Institute of Fashion Technology 2015; Vol. 2015, pp. 677-683.

16. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection, in: *IEEE Computer Society Conference on Computer Vision & Pattern Recognition (CVPR'05)*, IEEE, San Diego, CA, USA2005; pp. 886-893.

17. Thewsuwan S, Horio K. Preprocessing Techniques Based on LBP and Gabor Filters for Clothing Classification, in: *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, IEEE, Phuket, Thailand, Vol. 2016, pp. 1-6.

18. Sha D, Wang D, Zhou X, F Shi, Ge Y. An Approach for Clothing Recommendation Based on Multiple Image Attributes, in: Cui, B., Zhang, N., Xu, J., Lian, X., & Liu, D. (Eds.), *International Conference on Web-Age Information Management*, Springer, Cham 2016; pp. 272-285.

19. Surakarin W, Chongstitvatana P. Predicting Types of Clothing Using SURF and LDP based on Bag of Features, in: *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, IEEE, Hua Hin, Thailand 2015; pp. 1-5.

20. Li JQ, Shi WM, Yang DH. Clothing Image Classification with a Dragonfly Algorithm Optimised Online Sequential Extreme Learning Machine. *Fibres Text East Eur* 2021 ; Vol. 29 No. 3, pp. 90-95.

21. Tabik S, Peralta D, Herrera-Poyatos A, Herrera F. A snapshot of image pre-processing for convolutional neural networks: case study of MNIST. *International Journal of Computational Intelligence Systems* 2017 ; Vol. 10 No. 1, pp. 555.

22. Bui HM, Lech M, Cheng E, Neville K, Burnett IS. Object Recognition Using Deep Convolutional Features Transformed by a Recursive Network Structure. *IEEE Access* 2016 ; Vol. 4, pp. 10059–10066.

23. Tang P, Wang H, Kwong S. G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. Neurocomputing. *Neurocomputing* 2017 ; Vol. 225 FEB.15, pp. 188-197.

24. Ghazi MM, Yanikoglu B, Aptoula E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing* 2017 ; Vol. 235 APR.26, pp. 228-235.

25. Han K, Wang Y, Tian Q, Guo J, Xu C. GhostNet: More Features From Cheap Operations, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Seattle, WA, USA 2020 ; pp. 1577-1586.

26. Lao B, Jagadeesh K. Convolutional Neural Networks for Fashion Classification and Object Detection.2015.

27. Dong C, Shi Y, Tao R. Convolutional Neural Networks for Clothing Image Style Recognition. in: *Proceedings of 2018 International Conference on Computational, Modeling, Simulation and Mathematical Statistics (CMSMS)*, Xi'an, China 2018, pp. 608-613.

28. Xiang J, Dong T, Pan R,, Gao W. Clothing Attribute Recognition Based on RCNN Framework Using L-Softmax Loss. *IEEE Access* 2020 ; Vol. 8, pp. 48299–48313.

29. Di W. A comparative research on clothing images classification based on neural network models, in: *IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, IEEE, Weihai, China 2020; pp. 495-499.

30. Bossard L, Dantone M, Leistner C, Wengert C, Quack T, Gool LV. Apparel Classification with Style, in: Lee, K.M., Matsushita, Y., Rehg, J. M., Hu, Z.Y. (Ed.s), *Asian conference on computer vision (ACCV)*, Springer-Verlag, Berlin, Heidelberg 2012; pp. 321–335.

31. Huo P, Wang Y, Liu Q. A part-based and feature fusion method for clothing classification, in: *Pacific Rim Conference on Multimedia*, Springer, Xi'an 2016; pp. 231–241.

32. Liu S, Feng J, Domokos C, Xu H, Huang J, Hu Z, Yan S. Fashion Parsing with Weak Colour-category Labels. *IEEE Transactions on Multimedia* 2014; Vol. 16 No. 1, pp. 253-265.

33. Liu Z, Luo P, Qiu S, Wang X, Tang X. DeepFashion: Powering ro-bust clothes recognition and retrieval with rich annotations, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA 2016; pp. 1096–1104.

34. Wang J, Yang Y, Mao J, Huang Z, Huang C, Xu W. CNN-RNN: A unified framework for multi-label image classification, in: *IEEE conference on computer vision and pattern recognition (CVPR)*, IEEE, Las Vegas, NV, USA 2016; pp. 2285-2294.

35. Nawaz M, Hasan R, Hasan MA, Hassan M, Rahman RM. Automatic Categorization of Traditional Clothing Using Convolutional Neural Network, in: *IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, IEEE, Singapore 2018; pp. 98-103.

36. Cychnerski J, Brzeski A, Boguszewski A, Marmolowski M, Trojanowicz M. Clothes detection and classification using convolutional neural networks, in: *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, Limassol, Cyprus 2017; pp. 1-8.

37. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *Under Review as a Conference Paper at ICLR*. 2016.

38. Zhang X, Song C, Yang Y, Zhang Z, Zou Q. Deep learning based human body segmentation for clothing fashion classification", in: *Chinese Automation Congress (CAC)*, Shanghai, China2020; Vol. 2020 Nov. 6-8, pp. 7544-7549.

39. Rohrmanstorfer S, Komarov M, Modritscher F. Image Classification for the Automatic Feature Extraction in Human Worn Fashion Data. *Mathematics* 2021 ; Vol. 9 No. 6, pp. 624.

40. Hodecker A, Fernandes A, Steffens A, Crocker P, Leithardt V. Clothing Classification Using Convolutional Neural Networks, in: *15th IEEE Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE, Seville, Spain 2020; pp. 1-6.

41. Basavanhally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, Bhanot G, Madabhushi, A. Computerized image-based detection and grading of lymphocytic infiltration in HER2-breast cancer histopathology. *IEEE Transactions on Biomedical Engineering* 2010; Vol. 57 No. 3, pp. 642-653.

42. Adrian R. Multi-label classification with Keras, available at: https://www.pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/ (accessed 7 May 2018)

43. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science* 2014; pp. 1-14.

44. Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms".2017.

45. Simo-Serra E, Ishikawa H. Fashion Style in 128 Floats: Joint Ranking and Classification Using Weak Data for Feature Extraction, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* IEEE, Las Vegas, NV, USA 2016; pp. 298-307.

46. Wang L, Guo S, Huang W, Xiong Y, Qiao Y. Knowledge Guided Disambiguation for Large-Scale Scene Classification with Multi-Resolution CNNs, in: *IEEE Transactions on Image Processing*, IEEE 2017; Vol. 26 No. 4, pp. 2055-2068.