Multiclass Apparel Identification using HOG Feature Extractor and SVM

Yash Kerkar, Vijayalaxmi Kadroli, Amit Haral and Kapil Jain

Department of Information Technology, Terna Engineering College, Nerul Navi Mumbai, India

Abstract. Image classification and recognition plays an important role in many applications, like online shopping, driverless cars, automation, similar item retrieval queries, etc. In this project we have presented the identification of fashion items in an image. Given an image our model can identify whether it contains any fashion item or not. It can identify items like shirt, shoes, t-shirt, trousers, handbag and 6 other items. Our model consists of two things which are a feature extractor and a classifier. Based on research and experimental work we have selected HOG (Histogram of Oriented Gradients) as feature extraction method and Support Vector Machine(SVM) as classifier. It is a very tough task to select appropriate model for classification. It requires training and testing various models and techniques. However, we are able to achieve excellent results using our model.

Keywords: Multiclass Apparel Identification · Machine Learning · Object Identification · Image Classification · SVM Classifier · Softmax Classifier

1 Introduction

The most popular applications in computer vision are object classification and object recognition. Object classification means to classify an image into its right class. Classification comes handy in lot of automation tasks like inventory management, items retrieval for queries and also driverless cars, online shopping etc. There are many models for classification like logistic regression, Random forest , decision tree, SVM, Softmax etc. There are also various pre-processing algorithms like LBP, HOG, SIFT, smooth etc. Though the performance of HOG feature descriptor is excellent compared to others. In this project we have classified the images in Fashion-MNIST dataset using HOG feature extractor and SVM classifier.

2 Literature Survey

Fashion-MNIST(F-MNIST) [3] is a dataset consisting of 70000 fashion images. This dataset is developed by Zalando Research Company. The data is been divided into two parts which are – 1) 60,000 images for training 2) 10,000 images for testing. The dataset consists of 10 different classes, some of which are –T-shirts, trousers, shoes, coat, handbag etc. All images in the dataset are greyscaled. The images in the dataset are very small of the size 28 by 28 pixel. They are greyscaled which means every pixel in the image is represented by only one value ranging between 0-255 which represent shades of grey. Though the images are small the dataset is large enough for efficient training. HOG was developed by

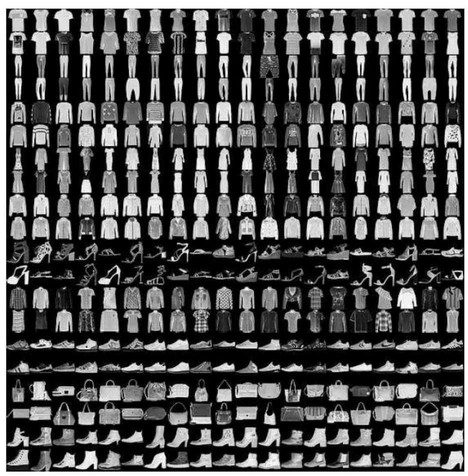

Fig. 1. Figure 1: Random Images from Fashion MNIST dataset

Dalal and Triggs (2005) [4] for the human detection and it is one of the most popular and successful feature extractors in pattern recognition and computer vision. Refering to research paper presented by Ebrahimzadeh and Jampour[5] very high accuracy was achieved on Hand written digits dataset using SVM + HOG. In one of the papers presented by Khan, H.A (2017)[7] he has introduced a new method called Multiple Cell Size (MCS) for improving feature vectors of HOG. By using MCS approach along with HOG excellent results have been achieved on classifications problems.Also the combination of HOG and SVM works very well for classification problems. Improvements based on Chain Code Histogram (CCH) for recognition of handwritten digits was proposed by Qian, Y. and Xichang (2013)[6] improves the speed of training and recognition and this reduces the feature dimension.

3 Proposed Methodology

3.1 Pre-processig - Histogram Of Oriented Gradients (HOG)

There are various methods for pre-processing like smooth, dilate, max etc. We have chosen HOG as pre-processor because experimental results of HOG show excellent performance. Before advancing to training the SVM classifier and evaluating the results a preprocessing task is introduced to decrease noise artifacts produced while collecting samples of images. The concept of HOG is that it captures the shape of the object in the image using the edges and corners of the object. HOG is a simple yet effective feature descriptor for images. HOG performs very well compared to SIFT and LBP which are also feature extractors. It is also proved that HOG feature vectors are very useful for object detection. Using HOG the shape and appearance of the main object in the image can be captured.

It divides the image into small cells like n-by-n and computes the edge directions. Normalizing the histograms can also help in increasing accuracy. HOG basically finds the distribution of directions of gradients in the feature vector. In the first step we find the Gradients(x and y derivative) of the image, for which the magnitude is large around the edges and corners (regions of intensity change) because edges pack in more information about the shape of an object rather than than flat regions.

In figure 2 the features extracted from a single image based on different cell sizes is given. Looking at the images we can easily conclude that cell size 2-by-2 contains more information than cell size 8-by-8. But 2-by-2 cell increases the size of the feature vector. A good choice is 4-by-4 cells. By using 4-by-4 cell size we are also able to capture good amount of information along with a small sized feature vector. After applying HOG on images using 4-by-4 cell size we get feature vector of size 1296 for each image. We now have a new dataset in which these features represent the image.

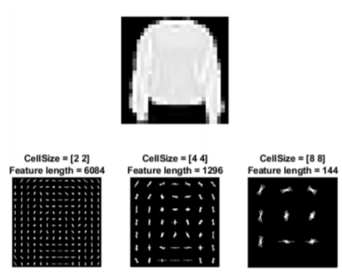

Fig. 2. Extracted Features of an Image

3.2 Support Vector Machines (SVM)

SVM is one of the most popular and powerful classifiers in machine learning. It shows good performance in various applications. SVM is famous because of the kernel trick. Kernels are used to solve a non-linear problem with a linear classifier. Kernels work very fast and efficiently with SVM using the kernel trick. Though here we are not using a kernel as the dataset is not very complex. Various applications of SVM include pattern recognition, text recognition, face recognition etc. In this part we utilize SVM for building the model. The main goal of any classifier is to separate classes with a decision boundary so that that a point which lies on either side of the boundary is classified into its proper class. But SVM not only finds this boundary but also tries to keep a good gap between the classes so that we get a boundary which generalizes well for the dataset.

3.3 Softmax Classifier

Softmax predicts the probability of all the classes given an image and these probabilities sum to one. The class with the highest probability is the predicted class. It is a multiclass version of Logistic Regression. In machine learning one of the most widely used supervised classification algorithms is Softmax which can be used for classification and regression task. Softmax has been applied in various fields like the field of pattern recognitions, face recognition, text recognition and so on. The experimental results of Softmax show similar performance to SVM. Soo in this part we utilize Softmax algorithm for classification. We have tested Softmax with two variations of data which are original pixel vector and HOG feature vector to correctly test its performance. The HOG feature vector is of size 1x1296 for a single image.

3.4 Training - Mini-Batch Gradient Descent

Gradient descent is an optimization technique used for minimizing loss functions using their gradients. The procedure of repeatedly evaluating the gradient by decreasing loss and then performing a Parameter update is called Gradient Descent. In Mini-batch Gradient descent we compute gradient over small batches of training set in each iteration such as batch of 200 images. This is computationally less expensive than Stochastic Gradient descent in which gradient is computed over whole dataset on every iteration also less time consuming[8]. The samples of say 200 examples are selected randomly from the dataset in every iteration because it is proven that this performs better than going over whole dataset in order.

3.5 Testing

The final model is tested against the test dataset of F-MNIST. This data used for testing is never been seen by the model therefore this is the best way for evaluating the model. The results of various models can be seen ahead in results and accuracy part.

3.6 Workflow

In below figure 4 we have shown how to test a single image on our model. We have to first convert the image to standard format of our model ie Image size of 28 by 28 and greyscaled, this can be achieved by using the functions which are available in python. Then we need to extract features from that image using HOG and then we need to multiply Weight vector of SVM with the features to get the predictions. The class with maximum probability is the predicted class.

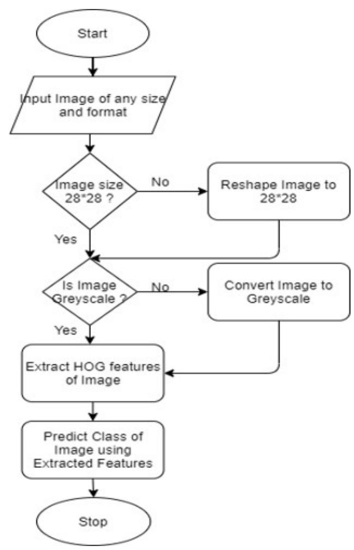

Fig. 3. Workflow Diagram for classifying an image

4 Results and Discussion

In this final part we apply the algorithm on every image and predict its label and check if the predicted label is right, in this way we calculate the accuracy of prediction on test dataset. the accuracy's of various algorithm's(Table 1). The first four algorithms refer to the literature survey and the next four are the algorithm's used by us. As it can be seen using SVM + HOG we have achieved highest accuracy which is 88.63.and second highest using Soft- max + HOG which is 88.05 percent .We have also tested SVM individually and Softmax individually which provide accuracies of 81 percent and 83 percent. Using HOG along with Softmax the accuracy increases by 5 percent which is very significant considering the size of the dataset. Using HOG with SVM the accuracy increases by 7 percent which is even bigger increase.

Table 1. Accuracy of various models

Algorithm	Accuracy
Random Forest Classifier	82
Decision Tree	79
SGD Classifier	81
Linear SVC	75
SVM	81
Softmax	83
Softmax + HOG	88.0
SVM + HOG	88.6

Rather than only checking how many values were predicted right and how many wrong for checking accuracy there is a better way which is Confusion Matrix. The confusion matrix is a matrix which gives the number of True Positive, True Negative, False Positive and False Negative. Where row 0 and column 1 represents the images whose actual label was 0 but were predicted as 1. It gives a better understanding of accuracy. We have also converted it into heatmap so that it is easier to read. The labels 0 to 9 are AS follows 0)T-shirt/top 1)Trouser 2)Pullover 3)Dress 4)Coat 5) Sandal 6)Shirt 7)Sneaker 8)Bag 9)Ankle boot

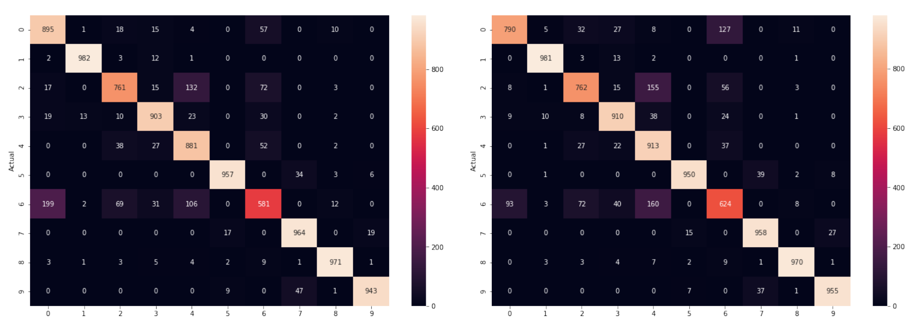

Fig. 4. Confusion Matrix of 1)SVM + HOG 2)Softmax + HOG

As it can be observed in the confusion matrix of SVM + HOG(Figure 5) ,the algorithm is confused a little bit about these types of labels 1) T-shirt – Shirt (199 images of Shirt wrongly predicted as T-shirt) 2) Pullover – Coat (132 images of Pullover wrongly predicted as Coat) 3) Shirt – Coat (106 images of Shirt wrongly predicted as Coat). Softmax + HOG(Figure 5) is confused about following labels 1) Shirt – Coat (160 images of Shirt wrongly predicted as Coat) 2) Pullover – Coat (155 im- ages of Pullover wrongly predicted as Coat) 3) T-shirt – Shirt (127 images of T-shirt wrongly predicted as Shirt). But these wrong predictions are still very less compared to the size of the dataset. Our model can be used anywhere where these types of predictions are required. For using our model the user just needs the weight vector found after training SVM and the code for converting the image in the standard format required for the classifier to run.

For demo purposes we have created a simple GUI using tkinter. In this GUI we can select any image from our computer and classify the image using the button provided. The GUI will output the label of the image in the output part. Below is a screenshot of the GUI working(Figure 7). In the GUI code all the steps have been implemented which are mentioned in (Figure 4). In the code SVM + HOG is used as the classifier which means that the weight vectors used for prediction belong SVM+ HOG training.

Fig. 5. Graphical User Interface(GUI)

5 Conclusion

We can conclude that we have provided an efficient way for apparel classification with an accuracy of 88.6 percent using HOG+SVM. Our model can be used in various applications which are listed in the paper. Our model can be integrated in any system by following the workflow diagram. These weights can also be used in neural network for transfer learning because the model is pretrained. In future, many modifications and improvements can be proposed on the pre-processing part and feature extraction and more combinations of features can be explored.

6 Acknowledgment

We express our deep gratitude to our project guide Prof. Vijayalakshmi Kadroli for providing timely assistant to our query and guidance that she gave owning to her experience in this field for past many year. She had indeed been a lighthouse for us in this journey

7 References

(1). Lawgali, A. (2016). Recognition of Handwritten Digits using Histogram of Oriented Gradients. International Journal of Advances Research in Science, Engineering and Technology, 3, 2359-2363.

(2). J Weston and C Watkins. Support vector machines for multi-class pattern recognition. In Proc. 7th European Symposium on Artificial Neural Networks, 1999.

(3). Xiao, H., Rasul, K., Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.

(4). N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

(5). Ebrahimzadeh, R., Jampour, M. (2014). Efficient handwritten digit recognition based on histogram of oriented gradients and svm. International Journal of Computer Applications, 104(9).

(6). Qian, Y., Xichang, W., Huaying, Z., Zhen, S., Jiang, L. (2013, November). Recognition method for handwritten digits based on improved chain code histogram feature. In 3rd International Conference on Multimedia Technology (ICMT-13).

(7). Khan, H. A. (2017). MCS HOG Features and SVM Based Handwritten Digit Recognition System. Journal of Intelligent Learning Systems and Applications,9(02).

(8). Li, Mu Zhang, Tong Chen, Yuqiang Smola, Alexander. (2014). Efficient mini-batch training for stochastic optimization. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 10.1145/2623330.2623612.