

Image Classification using HOG and LBP Feature Descriptors with SVM and CNN

Greeshma K V
 Asst. Professor on contract
 Department of Computer Science
 Carmel College, Mala

Dr. J. Viji Gipsy
 Assistant Professor
 Department of Computer Science
 PSGR Krishnammal College for Women, Coimbatore

Abstract— Image recognition and classification plays an important role in many applications, like driverless cars and online shopping. We present the classification of Fashion-MNIST (F-MNIST) dataset using two important classifiers SVM (Support Vector Machine) and CNN (Convolutional Neural Networks). In the first model two feature descriptors HOG (Histogram of Oriented Gradient) and Local Binary Pattern (LBP) with multiclass SVM. In this paper we explore the impact of various feature descriptors and classifiers on Fashion products classification tasks. We have used one of the most simple and effective single feature descriptor HOG. The multiclass SVM which is one of the best machine learning classifier algorithms is used in this method to train the images. In computer vision Convolutional Neural Networks (CNN or ConvNet) are the default deep learning model used for image classification problems. Selecting appropriate technique for feature extraction and choosing a best classifier algorithm remains a big challenging task for attaining good classification accuracy. However, the experimental results show that impressive results on this new benchmarking dataset F-MNIST.

Keywords—Convolutional Neural Networks, Fashion-MNIST, HOG features, Image Classification, LBP features, SVM Classifier

I. INTRODUCTION

In computer vision one of the most popular applications are Object Recognition or Object Classification. In object classification the main aim is to extract features from the images and classify it into right classes using any one of the classifiers or classification methods. For classifications of images or patterns the best classification methods are CNN and multiclass SVM. HOG and LBP are efficient gradient based feature descriptors for data discrimination and its performance is excellent comparing with other feature sets. This work classified the fashion products in Fashion-MNIST dataset using combined features of HOG and LBP with multiclass SVM classifier and CNN features with SVM. Fashion-MNIST (F-MNIST) is a dataset of 70000 fashion articles developed by Zalando Research Company. Figure 1 shows some images in F-MNIST.

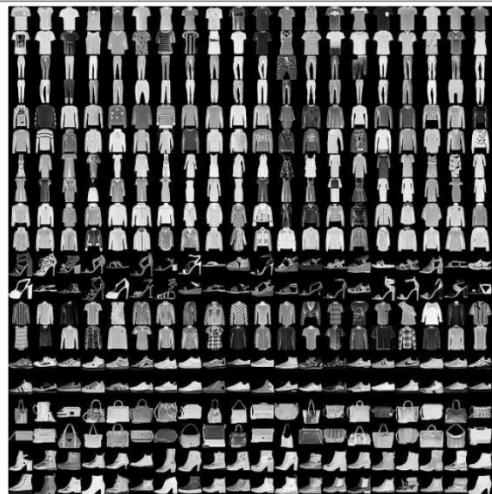
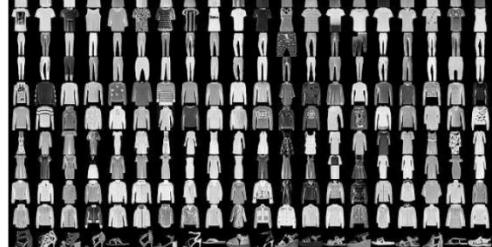
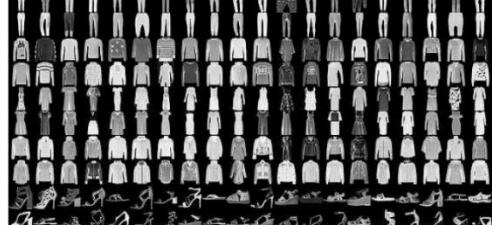
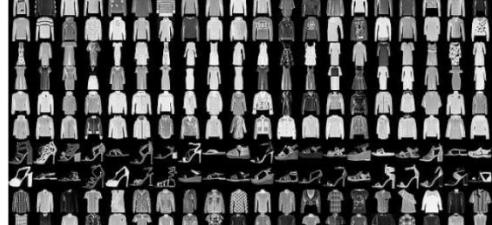
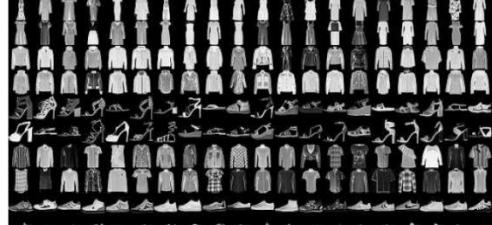
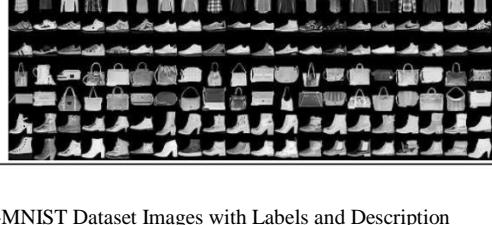
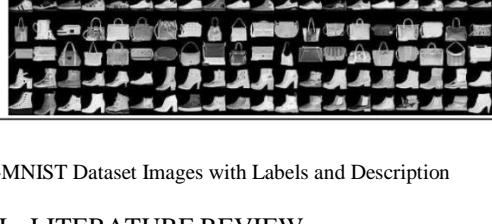
Label	Description	Examples
0	T-Shirt/Top	
1	Trouser	
2	Pullover	
3	Dress	
4	Coat	
5	Sandals	
6	Shirt	
7	Sneaker	
8	Bag	
9	Ankle boots	

Fig. 1. Fashion-MNIST Dataset Images with Labels and Description

II. LITERATURE REVIEW

In image classification different methods are used such as methods based on low-level image feature representation which consider image as a collection of low-level characteristics like texture, shape, size, colour, etc. and methods based on mid-level visual feature constructions for image classification tasks. Nowadays, usage of deep neural networks and neural-networks to obtain image representation is trending. Such architectures allows us to extract features from a specified layer of trained neural network and then use extracted feature maps as a numeric image representation. There are a large number of publications related to the image processing with various ensemble methods. Image classification in the fashion domain has numerous benefits and applications and has various research works have been presented about it.

Fashion-MNIST dataset has been presented by Zalando Research (Xiao et al., 2017) [29]. F-MNIST is proposed to intend for a direct drop-in substitute for the classical MNIST handwritten digits dataset which has been considered as the benchmark for machine learning techniques, as it contains the same structure, image format and size of train and test set splits. The images in F-MNIST are transformed to a format which matches with original

MNIST, for making F-MNIST compatible with the ML packages which are efficient on working with the MNIST dataset. They have provided some results of classification in this paper to form a benchmark on this dataset. All algorithms presented on that were repeated five times by shuffling the training data and the mean of the accuracy on the test data were reported on it.

2.1 HOG using SVM

HOG was first introduced by Dalal and Triggs (2005) for the human detection and it is one of the most popular and successful feature descriptors in pattern recognition and CV. They practically presented that the grids of the HOG descriptor can remarkably surpass the existing feature sets for the human detection. In one of the research paper presented by Ebrahimzadeh and Jampour szeshows that they have achieved very high accuracy on HDR (Automatic Handwritten Digits Recognition) using this efficient HOG descriptor with multiclass SVM. A feature-based approach is proposed by them in which the data is processed using HOG. HOG is the gradient-based descriptor and it is more efficient descriptor for the handwritten digits. And the classifier had been used is the linear SVM which has good results than RBF, polynomial and sigmoid kernels.

III. PROPOSED METHODOLOGY

A. Feature Extraction

The various features of the images are extracted in this phase and then they have used with SVM for classification of fashion objects in F-MNIST dataset. In advance of training a classifier and evaluating the test, a preprocessing task is introduced to decrease noise artifacts produced while collecting samples of images. For training the classifiers by applying pre-processing, it provides better feature vectors. Preprocessing is very much important task because its efficient functioning reduces the misclassification and improves the recognition rate [8]. Herein HOG based feature extraction scheme for recognizing fashion products is used for the proposed work. Every fashion article image of dimension 28x28 is used to extract HOG feature.

B. Histogram of Oriented Gradients (HOG)

One of the simple and effective feature extraction methods is HOG feature descriptor. It is a fast and efficient feature descriptor in compare to the SIFT and LBP due to the simple computations, it has been also shown that HOG features are successful descriptor for detection. Mainly it is used for object detection in image processing and computer vision. Using HOG the shape and appearance of the image can be described. It divides the image into small cells like 4-by-4 which is used in this work and computes the edge directions. For improving the accuracy the histograms can be normalized.

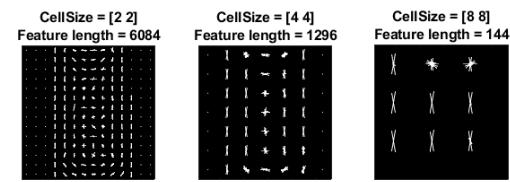
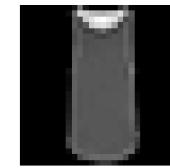


Fig. 2. HOG features of an image with different CellSize

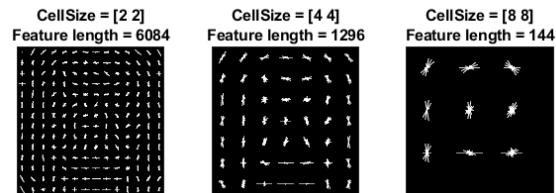


Fig. 3. Extracted features of an image in Fashion-MNIST

In Figure 2 extracted HOG features of one image using three different cell sizes are shown. In this figure the visualization of cell size [2 2], [4 4] and [8 8] are shown. From that it is clearly understood that the cell size [2 2] contains more shape information than the cell size of [8 8] in their visualization. But in the latter case the dimensionality of feature vector using HOG increases comparing with the former. A good choice is the [4 4] cell size. By using this size the numbers of dimensions are limited and this helps to speed up the training process. Also it contains enough information to visualize the fashion image shape. For identifying the suitable parameter setting configuration of HOG parameters more training and testing processes using the classifier has to be performed.

C. Local Binary Pattern (LBP)

Local Binary Patterns (LBPs) converts a grayscale image at pixel level to a matrix of integer numbers. This matrix of labels describes the original image. It computes the local representation of the texture. It is a visual descriptor used in CV for classification. This model is proposed in 1990 and first described in 1994. When combining with the HOG feature descriptors it significantly improves the performance. LBP feature descriptor is a powerful feature used for texture classification.

D. Convolutional Neural Network (CNN)

Among various deep learning architectures, ConvNets stands out for its unprecedented performance on computer vision. ConvNet is an Artificial Neural Network inspired by biological visual cortex and been successfully applied to image processing tasks. A special kind of artificial neural network is ConvNet which contains at least one convolutional layer. A typical ConvNet takes an input image, pass it through a set of layers convolution, nonlinear activation, pooling (downsampling) and fully connected, and retrieve an output of classification labels. This output of this CNN layer is an activation map.

The first ConvNets architecture of the model defined in this paper consists of 2 convolution layers succeeded by activation, pooling, fully connected and softmax layers respectively. Multiple filters are used at each ConvNet layer, for various types of feature extraction. In our first ConvNet layer 32 numbers of filters of the dimension (3, 3) is given and in the second layer 64 filters of (3, 3) is applied. In the second ConvNet model 4 convolution layers followed by Batch Normalization, relu, maxpooling and dropout. First two convolution layers contain 32 numbers of filters and next two with 64 filters. Each filter has the dimension of (3, 3).

E. Support Vector Machine (SVM)

In machine learning one of the most common and successful classifier in supervised learning is SVM which can be used for classification and regression tasks [6]. Supporting Vector Machine has been successfully applied in the field of pattern recognitions, like face recognition, text recognition and so on. It shows good performance in applications [8]. So this part we utilize SVM to train and test. This paper employed a multiclass SVM classifier as a classification tool of HOG feature space developed for a complete dataset of fashion images from F-MNIST database. The HOG feature of dimension 1x1296 for each individual fashion object have been arranged in the row wise to prepare complete feature space.

IV. EXPERIMENTAL RESULTS

A. Fashion-MNIST Dataset

F-MNIST dataset is a collection of fashion objects in grayscale. It contains 4 files including the labels and images which are again subdivided into sets of training and test. The labels and images in training set consists of 60000 numbers and in the test set, it is 10000. F-MNIST contains 10 classes of images and the labels and description of each class is given in Fig. 1.

TABLE I. ACCURACY OF F-MNIST IN LITERATURES

Method	Accuracy
Linear SVC [4]	83.6
HOG + SVM [6]	86.53

Method	Accuracy
SVC C=10; kernel : rbf [4]	89.70
EDEN [5]	90.60

SVC: Support Vector Classifier; HOG: Histogram of Oriented Gradient; SVM: Support Vector Machine; EDEN: Evolutionary Deep Networks

TABLE II. ACCURACY OF F-MNIST USING DIFFERENT FEATURES AND CLASSIFIERS

Method	Accuracy
HOG + LBP SVM	87.4
CNN	91.15
CNN FC SVM	91.59

Comparing with the accuracy results on F-MNIST dataset test data results with various models in literature as shown in Table I, the CNN FC SVM model shows better accuracy results of 91.59%.

V. CONCLUSION

In general, proposed work presents an efficient system for the effective and accurate classification and recognition of the fashion products images. After successful implementation of the proposed fashion articles classification system using HOG and LBP feature space and multiclass SVM classifier, it has shown that the proposed system provides relatively good fashion object classification efficiency as compared to available literature works. In future, many modifications and improvements can be proposed on the preprocessing part and feature extraction and more combinations of features can be explored. We may modify the feature extraction and classification using many other techniques and can produce outstanding performance on fashion image classification. We can explore the other feature types for training the classifiers and analyze the effects of other machine learning algorithms for classifying fashion images.

REFERENCES

- [1] Ebrahimbzadeh, R., & Jampour, M. (2014). Efficient handwritten digit recognition based on histogram of oriented gradients and svm. International Journal of Computer Applications, 104(9).
- [2] 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.
- [3] Khan, H. A. (2017). MCS HOG Features and SVM Based Handwritten Digit Recognition System. Journal of Intelligent Learning Systems and Applications, 9(02), 21.
- [4] Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.
- [5] Dufourq, E., & Bassett, B. A. (2017, November). EDEN: Evolutionary deep networks for efficient machine learning. In Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech), 2017 (pp. 110-115). IEEE.

- [6] Greeshma K. V., & Sreekumar K. Fashion-MNIST Classification Based on HOG Feature Descriptor Using SVM, IJITEE.
- [7] Bhatnagar, S., Ghosal, D., & Kolekar, M. H. (2017, December). Classification of fashion article images using convolutional neural networks. In Image Information Processing (ICIIP), 2017 Fourth International Conference on (pp. 1-6). IEEE.
- [8] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.
- [9] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27.
- [10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).