Advances in Technology Innovation, vol. 2, no. 4, 2017, pp. 119 - 125

A CNN Based Approach for Garments Texture Design Classification

S. M. Sofiqul Islam, Emon Kumar Dey*, Md. Nurul Ahad Tawhid, B. M. Mainul Hossain

Institute of Information Technology, University of Dhaka, Bangladesh.

Received 03 October 2016; received in revised form 04 February 2017; accepted 08 February 2017

Abstract

Identifying garments texture design automatically for recommending the fashion trends is important nowadays because of the rapid growth of online shopping. By learning the properties of images efficiently, a machine can give better accuracy of classification. Several Hand-Engineered feature coding exists for identifying garments design classes. Recently, Deep Convolutional Neural Networks (CNNs) have shown better performances for different object recognition. Deep CNN uses multiple levels of representation and abstraction that helps a machine to understand the types of data more accurately. In this paper, a CNN model for identifying garments design classes has been proposed. Experimental results on two different datasets show better results than existing two well-known CNN models (AlexNet and VGGNet) and some state-of-the-art Hand-Engineered feature extraction methods.

Keywords: CNN, deep learning, AlexNet, VGGNet, texture descriptor, garment categories, garment trend identification, design classification for garments

1. Introduction

Online shopping is popular nowadays. Customer select products from the web pages according to their choice and that can help to predict the direction of trends. If a retailer knows popular design styles of clothing products, it can increase the production of those styles to achieve more profit. Therefore, if a system can classify the garments products according to different style, texture, size etc., it can automatically suggest different products to the customers based on their choices. The system proposed in this paper, can classify clothes according to textures.

Effective design classification based on textures, local spatial variations of intensity or colour in images has been an important topic of interest in the past decades. A successful classification, detection or segmentation requires an efficient description of image textures. To fulfil this purpose, lots of well-known Hand-Engineered feature extraction methods such as CENsus Transform hiSTogram (CENTRIST) [1], Local Binary Pattern (LBP) [2], Histogram of Oriented Gradient (HOG) [3] etc., are exist. LBP gains popularity because of their computational simplicities and better accuracies. But, it is very sensitive to uniform and near uniform region. LTP [4], Completed Local Binary Pattern (CLBP) [5] can handle this issue more accurately. Between these two methods, CLBP is better choice because this method is rotation invariant. CENTRIST [1] has gain popularity by incorporating Spatial Pyramid (SP) structure. But, most recently Completed CENTRIST (cCENTRIST) and Ternary CENTRIST (tCENTRIST) [6] gained high accuracies for garments design classification. Although several Hand-Engineered feature extraction approaches exist for garments design classification, deep learning is rarely used in this field. Our goal is to apply appropriate deep learning model to measure the performance of garments design identification based on textures.

In recent year, deep learning has become popular in the field of machine learning and computer vision. Using large architectures with numerous features, many deep learning models achieve high performance in the field of object detection, text classification, image classification, face verification, gender classification, scene-classification, digits and traffic signs recognition, etc. Some of the available deep learning models are AlexNet [7], VGGNet [9], Berkeley-trained models [10], Places-CNN model [8], Places-CNDS models on Scene Recognition [11], Models for Age and Gender Classification [12], GoogLeNet model [13], etc. These methods have achieved dramatic improvements and attracted considerable interest in both the academic and industrial communities. In general, deep learning algorithms attempt to learn hierarchical features, corresponding to different levels of abstraction. Each of these models concerned about some specific issues: preventing over-fitting, connection of nodes between adjacent layers, large learning capacity, etc. Several factors need to be considered for working with deep learning network such as availability of large training set, powerful GPU for training and testing, better model regularization strategies, the amount of training time that one can tolerate, etc.

The major contributions of this paper are as follows.

(1) In this paper, a brief review on existing well known Hand-Engineered feature extraction methods for garments design class identification has been conducted.

(2) This research has applied some existing Deep Convolutional Neural Network models for classifying the clothing products on some datasets and compared the results with several state-of-the-art Hand-Engineered feature extraction methods.

(3) A new Deep Convolutional Neural Network model has been proposed for classifying garments design class. This proposed model is applied on two different datasets and has found a remarkable output.

The rest of the paper is structured as follows. Section 2 and Section 3 describe the background studies and the methodology respectively. Section 4 has presented the experimental results and finally Section 5 concluded the overall work with necessary explanation.

2. Background studies

In this section, some existing garments clothing segmentation and classification strategies have been described. Some existing deep learning models; that have

*Corresponding author. Email: emonkd@iit.du.ac.bd

been used for several applications in computer vision are also narrated in this section.

2.1. Garment product segmentation and identification

Yamaguchi et al. [14] proposed a method for clothing parsing. For this work, they created Fashionista dataset consisting of 158,235 images. From this dataset, they selected 685 images for training and testing their system. They identified 14 different parts of a body and different clothing regions. In [15], they deal with clothing parsing problem using retrieval based approach. Their proposed approach focused on pre-trained global clothing models, local clothing models, and transferred parse. Authors found that their proposed final parse achieve 84.68% parsing accuracy. Menfredi et al. [16] proposed a new approach for automatic garments segmentation and classification. They classified garments into nine different classes such as skirts, shirt, dresses, etc. For this work, authors used a projection histogram for extracting few specific garments. They divided the whole image into 117 cells and group them into 3*3 cells. They computed HOG features [17] from each cell and the orientations are grouped into nine bins. They used multiclass linear support vector for training. Serra et al. [18] did similar type of work, where authors used conditional random field (CRF) for divided outfits.

Vittayakorn et al. [19] used five different features such as color, texture, shape, parse and style descriptor to identify three different visual trends, namely floral print, pastel color and neon color from runway to street fashion. However, using more color and design classes would be more beneficial in this field. Kalantidis et al. [20] proposed a system to identify the relevant product where they firstly estimated the pose of a person from an input image and then segmented the clothing area such as shirt, tops, jeans, etc. Finally, they applied an image retrieval technique which is 50 times faster than [14] for identifying similar clothes for each class.

Gallagher et al. [21] used grab cut algorithm for identifying a person by segmenting the clothing parts. Bourdev et al. [22] proposed a new method for detecting some attributes and type of cloths from an input image. Here attributes are gender, hair style and types of clothes such as t-shirts, pants, jeans, and shorts etc. For this work, they created a dataset consisting of 8000 people images with annotation.

2.2. Texture based classification

Nowadays, garments design classification based on texture has become more popular and there are several existing well known methods such as Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), Features Wavelets transform, Noise Adaptive Binary Pattern (NABP), Gabor filters, Scale-Invariant Feature Transform (SIFT) etc. Recently, LBP has become popular because of its computational simplicity. LBP was proposed for describing the local structure of an image and it has been used in several areas such as facial image analysis, including face detection, face recognition and facial expression analysis, demographic (gender, race, age, etc.) classification, moving object detection, etc. However, LBP is very sensitive in uniform and near uniform regions. In the last few years, lots of researches have been done by modification of LBP to improve the performance. Such as derivative-based LBP, dominant LBP, Rotation Invariant, center-symmetric LBP, etc. Tan and Triggs [4] proposed a new texture based method Local Ternary Patterns (LTP), which can tolerate noises up to a certain level. They used a fixed threshold (±5), for making LTP more discriminant and less sensitive to noise in a uniform region. There are also several other methods that can handle noises in different application areas, such as the methods described by Jun et al. [25]. They proposed Local Gradient Pattern (LGP) for texture based face detection. This method is a variant of LBP and uses adaptive threshold for code generation. Guo et al. [5] proposed Completed Local Binary Pattern (CLBP), which incorporates sign, magnitude and center pixel information. This method is rotation invariant and capable of handling the fluctuation of intensity. Wu et al. [1] proposed CENsus Transform histogram (CENTRIST) which is very similar to LBP and mainly work as a visual descriptor for recognizing scene categories. CENTRIST proposes a spatial representation based on a Spatial Pyramid Matching Scheme (SPM) [26] to capture global structure from images. CENTRIST uses total 31 blocks to avoid the artefacts. Dey et al. [6] proposed two new descriptors for garments design class identification namely Completed CENTRIST (cCENTRIST) and Ternary CENTRIST (tCENTRIST). These descriptors are based on Completed Local Binary Pattern (CLBP), Local Ternary Pattern (LTP) and CENsus TRanformed hISTogram (CENTRIST). Authors applied these two descriptors on two different publically available databases and achieve nearly about 3% more accuracy than the existing state-of-the art methods.

2.3. Deep learning

This sub-section, will describe some deep learning techniques for garments design classification. Deep Network learn features automatically from large number of unlabelled data, hence more useful hidden discriminative features are extracted. It has achieved popularity in classic problems, such as speech recognition, object recognition and detection, natural language processing, etc. Convolutional Neural Networks (CNN) is now being used in several image, pattern and signal processing researches. Liu et al. [33] introduced AU-aware Deep Networks (AUDN) by constructing a deep architecture for facial expression recognition. For extracting high level features from each AU-aware receptive fields (AURF), they used restricted Boltzmann machine (RBMs). Later, this technique was applied on three expression database namely CK+, MMI and SFEW. Results achieved from this technique were better or at least competitive. However, this method fails when several kinds of challenging images (e.g., the subjects have higher expression non-uniformity, most of them have moustache and wear accessories such as glasses) are appeared.

Krizhevsky et al. [7] proposed a new CNN architecture which achieved top-1 and top-5 error rates of 37.5% and 17.0% on the test data. However, there is still an open issue that, if a single convolutional layer is removed, Network's performance is degraded. Here, authors did not use any unsupervised pre- training data to simplify this work but it could be more helpful if the computational power and size of the Network were increased. Dey et al. [6] used deep learning model in texture based garments design classification. In their experiment, using Berkeley-trained model [10], they obtained

73.54% accuracy in Clothing Attribute dataset. However, they claimed that accuracy might be improved by changing layers and other related issues.

Zhoub et al. [8] proposed a technique which extracted the difference between the density and diversity of image datasets. Here, authors used CNN to learn deep features for scene recognition tasks. For their dataset VGG S-16 models achieved 88.8% accuracy in top-5 val/test. However, there exist some difficulties such as the variability in camera poses, decoration styles or the objects that appear in the scene. Lao et al. [27] used Convolutional Neural Network for fashion class identification. Authors divided their work into four parts those are multiclass classification of clothing type; Clothing Attribute classification; clothing retrieval of nearest neighbours; and clothing object detection. For this work they used Apparel Classification with Style (ACS), Clothing Attribute (CA) and Colourful-Fashion (CF) datasets and found 50.2% and 74.5% accuracy for clothing style classification and Clothing Attribute datasets. Hu et al. [28] used deep convolutional neural networks for high-resolution remote sensing (HRRS) scene classification. For this work, they proposed two models for extracting CNN features from different layers. Authors also used convolutional feature coding scheme for aggregating the dense convolutional features into a global representation. Their proposed two models achieved remarkable performance and improved the state-of-the-art by a significant margin.

For garments design class identification many approaches have been proposed. But, there are only a few works that have been conducted based on deep learning. This research has experimented different deep learning methods for identifying different garments design class based on textures.

3. Methodology

This section describes the methodology for identifying the garments design classes. Basic steps of the procedure are shown in Fig. 1. Input images are firstly segmented and classified into several classes based on their texture design. After that, these images are separated for training, validation and testing from each of the class. Proposed model is then applied alongside with two well-known deep Convolutional Neural Network (CNN) models AlexNet and VGG_S in two different garment datasets for the purpose of training and testing. Finally, the accuracy of proposed system is compared with the existing models. We have also compared the results with traditional state-of-the-arts Hand-Engineered feature extraction method. AlexNet and VGG_S have been chosen in this work because of their computational simplicity and better performance in several areas. They work well on unsupervised dataset. These two models can handle over-fitting problem when working with large dataset by using data augmentation technique. Besides, these two models use a recently-developed regularization method called "Dropout" that is proven to be very effective. These two models gained significant results in challenging benchmarks on image recognition and object detection. Brief descriptions about these two models alongside our proposed model are described in the following sub-sections.

Fig. 1 Basic steps of our working procedure

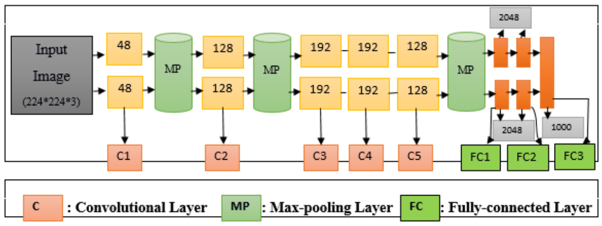

Fig. 2 The full architecture of AlexNet Model

3.1. AlexNet model

AlexNet model was proposed by Krizhevsky et al. [7]. There are three types of layer in a deep convolution neural network; such as Convolution layer, Pooling layer and Fully-Connected (FC) layers. Full architecture of AlexNet model was created by combining these three layers. In this architecture, there are total eight learned layers: five convolutional layers and three fully connected layers. Convolution layer is the core building block and each of those convolution layer consists of some learnable filters. Filters size are different from one another. Full AlexNet architectural model is shown in Fig. 2. First convolution layer takes the input images by resizing each of the images into 224×224 with 96 kernels. The second layer takes the input from first convolution layer with 256 kernels after passing through a pooling layer. Pooling layer operates independently and reduce the amount of parameters and computation in the network. Hence, control the over-fitting problems. In this architecture, the third, fourth and fifth layers are connected to one another without any connection of pooling layers. The third layer consists of 384 kernels which takes input from the output of second layer. The fourth layer has 384 and fifth layer contains 256 kernels. Each of last three fully connected layers contains 4096 neurons. The output of the last fully connected layer is sent as input to a 1000 way softmax layer which produces a distribution over the 1000 class labels. Here, multinomial logistic regression is also used for maximizing the training cases.

3.2. VGGNet model

Chatfield et al. [9], based on Caffe toolkit proposed three different architectures of deep CNN models: VGG_F, VGG_M and VGG_S; each of which explores a different speed/accuracy trade-off:

(1) VGG_F: This CNN architecture is almost similar to AlexNet. But VGG_F contains smaller number of filters and small stride in some convolutional layers.

(2) VGG_M: It is a medium size CNN which is very similar proposed by Zeiler et al. [30]. The 1st convolution layer of this network has smaller stride and pooling layer. 4th convolution layer use smaller numbers of filters for balancing the computational speed.

(3) VGG_S: This architecture is relatively slow than VGG_F and VGG_M and it is a simplified version of accurate model in the Over-Feat framework which has six convolutional layers. Fig. 3 shows the full architecture of VGG_S model. It has taken the first five layers from the original model and has a smaller number of filters in 5th layer. It has large pooling size in 1st and 5th convolutional layer than VGG_M. This model has been used to evaluate the garments design

class identification. As depicted in Fig. 3, this VGG_S model contains five convolution layers with smaller number of filters in the 5th layer and three fully connected layers. There are another two models based on VGGNet namely VGG-VD16 and VGG-VD19. Between AlexNet and VGG_S models, the main difference is that VGG_S model has small stride in some convolutional layers and pooling size is large attached with the 1st and 5th convolutional layer. Here, fully-connected layers 6 and 7 are regularized using Dropout and the last layer acts as a multi-way soft-max classifier.

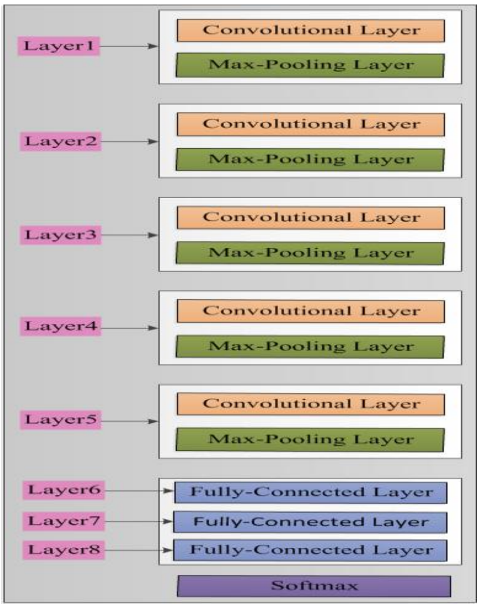

Fig. 3 The full architecture of VGG_S Model

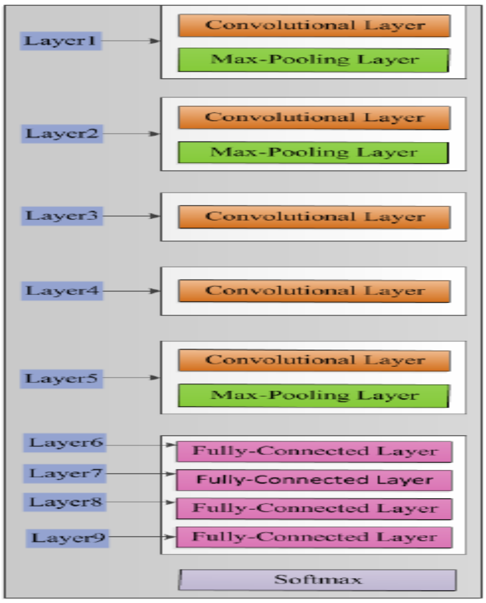

Fig. 4 The full architecture of our proposed model

3.3. Proposed transferred CNN

For classifying garments design class, a new scenario has been proposed in this paper based on AlexNet, to observe the performance and effectiveness of deep features by total nine learned layers. Among these layers five of these layers are Convolutional layer and remaining four are Fully Connected layers. Like AlexNet, first convolution layer of proposed model takes the input images by filtering each of the images into 224×224 size with 96 kernels. The second layer takes the input from first convolution layer after passing through a pooling layer. Pooling layers are added after first, second and fifth convolution layer like AlexNet. A new Fully Connected layer (FC3) which takes input from the output of second Fully Connected layer (FC2) has been added in this proposed model. Output of the last layer (FC4) is connected to a softmax layer for classifying the categories. The proposed model used data augmentation technique to reduce overfitting in the training stage. Because, recent works show that data augmentation also helps to improve classification performance [7]. The full architecture of the proposed model is shown in Fig. 4.

3.4. Datasets

Fig. 5 Example of clothing attribute dataset: column 1 to 6 represents example of floral, graphics, plaid, solid color, spotted and stripe respectively

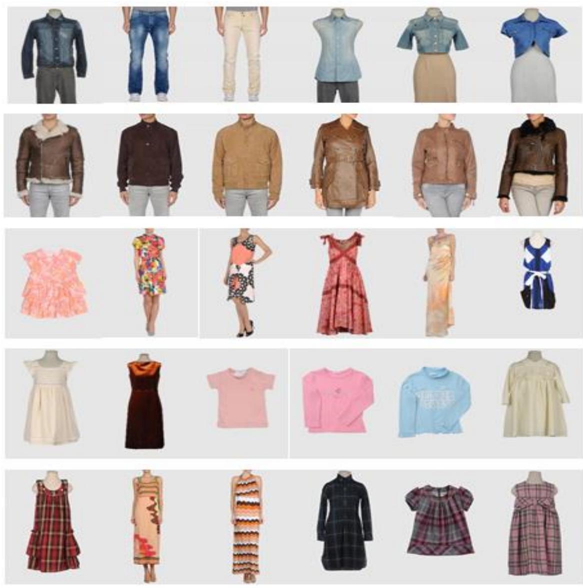

Fig. 6 Example images from fashion dataset: Each of the row represents jeans, leather, print, single color and stripe category respectively

Two publicly available datasets: Fashion [31] and Clothing Attribute datasets (CAD) [32] that was originally created for Garment Product Recognition have been considered for this research. From Fashion dataset, 5400 images are manually selected and categorized into five design classes, namely "Single color" (2440 images), "Print" (1141 images), "Stripe" (565 images), "Jeans" (614 images) and "Leather" (640 images). Again from Clothing Attribute dataset; 1575 images and manually selected and categorized into six different categories; after segmenting garments area from the original dataset. The categories are "Floral" (69 images), "Graphics" (110 images), "Plaid" (105 images), "Spotted" (100 images), "Striped" (140 images) and "Solid" pattern (1051images). Original CAD contain 1856 different images with 26 ground truth clothing attributes such as necktie, color, pattern etc. Fig. 5 and Fig. 6 show some sample images from "Fashion" and "Clothing Attribute" datasets used in our work.

Table 1 describes proper training and validation samples about Clothing Attribute and Fashion datasets. For Clothing Attribute dataset, different training and validation samples has been used; such as 10, 20, 30 images per class for training and validation, and rest of the images for testing to identify the classification results. In Fashion dataset, 60, 100, 200 and 300 images are used for training and 10, 10, 20 and 30 images for validation and rest of the images for testing respectively.

Table 1 Dataset used for experiments sample with different training and validation samples

Databases	Clothing Attribute Dataset (CAD)	Fashion Dataset
Classes	6	5
Total samples	1575	5400
Training sample/class	I)10 II) 20 III) 30	I) 60
		II) 100
		III) 200
		IV) 300
Validation sample/class	I) 10 II) 20 III) 30	I) 10
		II) 10
		III) 20
		IV) 30

4. Experimental result

This section describes the experimental detail and divided into two sub-sections. First sub-section discusses about the implementation environment and next one describes the results.

4.1. Implementation Environment

Experimentation environment for this research has been set by following a straightforward process. We fine-tuned the CaffeNet [29] model and use Ubuntu 12.4 operating system. This research considered high speed GPU for making the computation faster. Because CPU is nearly ten times slower than GPU for working with large datasets and complex CNN. NVIDIA GEFORCE GTX 950 4GB GPU and Intel core i7 processor has been used for faster training and testing.

4.2. Experimental result and discussion

This research mainly experimented on two existing deep convolutional neural network models alongside with the proposed model on Fashion dataset and Clothing Attribute dataset. Performance of the proposed deep learning model has been compared with the existing models and also with some existing well-known Hand-Engineering feature extraction approaches for garment design class identification. Different training, validation and testing sample from two different datasets have been used and shown in Table 1. The training and testing results of AlexNet, VGG_S and proposed model are provided in Table 2, Table 3, Table 4, and Table 5. These accuracies are calculated based on the training, validation samples/class used for each dataset. From Table 3 and Fig. 7 it can be found that in most of the cases VGG_S performs better than AlexNet model.

Table 2 Recognition rate (%) in training phase of CAD

Dataset	**Models**	**Training Sample**	**Validation Sample**	**Results**
CAD with 6 classes	AlexNet	10	10	75.1
		20	20	75.6
		30	30	75.5
	VGG_S	10	10	76.2
		20	20	76.5
		30	30	76.6
	Proposed Model	10	10	77.2
		20	20	77.3
		30	30	77.7

Table 3 Recognition rate (%) in testing phase of CAD

Dataset	**Models**	**Training Sample**	**Validation Sample**	**Results**
CAD with 6 classes	AlexNet	10	10	75.3
		20	20	75.5
		30	30	75.6
	VGG_S	10	10	76.5
		20	20	76.4
		30	30	76.8
	Proposed Model	10	10	77.1
		20	20	77.4
		30	30	77.8

Table 4 Recognition rate (%) in training phase of fashion dataset

Dataset	**Models**	**Training Sample**	**Validation Sample**	**Results**
Fashion Dataset with 5 classes	AlexNet	60	10	74.3
		100	10	75.9
		200	20	78.1
		300	30	81.5
	VGG_S	60	10	75.3
		100	10	76.8
		200	20	78.6
		300	30	82.7
	Proposed Model	60	10	76.6
		100	10	78.1
		200	20	81.1
		300	30	84.1

Table 5 Recognition rate (%) in testing phase of fashion dataset

Dataset	**Models**	**Training Sample**	**Validation Sample**	**Results**
Fashion Dataset with 5 classes	AlexNet	60	10	74.8
		100	10	76.6
		200	20	79.1
		300	30	81.8
	VGG_S	60	10	76.1
		100	10	77.3
		200	20	80.8
		300	30	82.9
	Proposed Model	60	10	76.7
		100	10	78.1
		200	20	82.7
		300	30	84.5

Using Clothing Attribute dataset, AlexNet and VGG_S model of CNN shows maximum 75.6% and 76.8% accuracies respectively while our proposed model of CNN achieved 77.8% accuracy. On the other hand, using Fashion dataset with 5 different classes, 81.8% accuracy has been achieved using AlexNet and 82.9% accuracy using VGG_S respectively and our proposed model achieved 84.5% accuracy. From Table 3 and Table 5, it is clear that more training sample increase the accuracy. Table 6 and Table 7 describe the experimental results using seven different Hand-Engineered feature extraction methods which are HOG, GIST, LGP, CENTRIST, tCENTRIST, cCENTRIST, and NABP on Clothing Attribute Dataset and Fashion Dataset. For these methods Support Vector Machine (SVM) was used for classification purpose.

Table 6 Experimental results of different methods for clothing attribute dataset

Method	Accuracy
HOG	63.76%
GIST	72.31%
LGP	65.55%
CENTRIST	71.97%
tCENTRIST	74.48%
cCENTRIST	74.97%
NABP	74.18%
Berkeley	73.54%
AlexNet (30)	75.6%
VGG_S (30)	76.8%
Proposed Model	77.8%

Table 7 Experimental results of different methods for fashion dataset

Method	Accuracy
HOG	79.15%
GIST	81.67%
LGP	79.79%
CENTRIST	79.72%
tCENTRIST	84.07%
cCENTRIST	84.23%
NABP	83.22%
AlexNet	81.8%
VGG_S	82.9%
Proposed Model	84.5%

Table 6 also shows the result of three deep learning models Berkeley, AlexNet, VGG_S along with our proposed model for Clothing Attribute dataset. From this table, it is clear that performance of different deep learning models are better than any Hand-Engineering feature extraction method for Clothing Attribute Dataset.

Table 7 shows that, for Fashion dataset our proposed method performs better. Though AlexNet and VGG_S show slightly less accuracy than tCENTRIST, cCENTRIST and NABP.

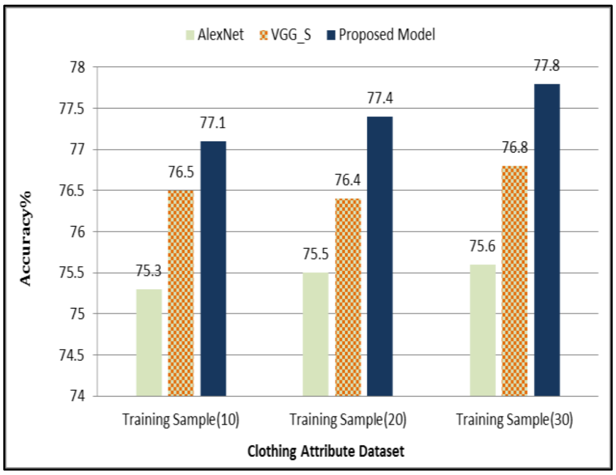

Fig. 7 Comparison between AlexNet, VGG S and our proposed models for clothing attribute dataset

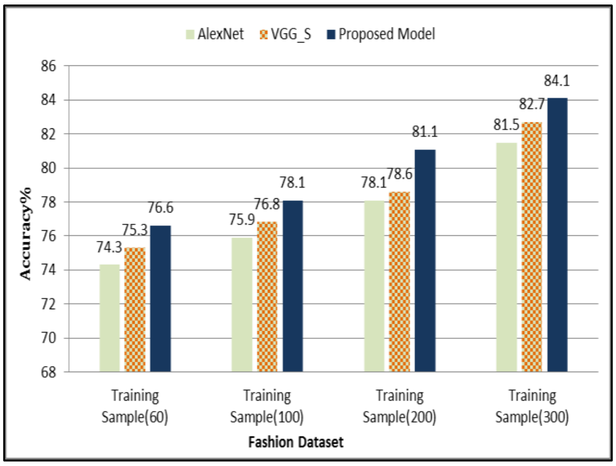

Fig. 8 Comparison between AlexNet, VGG S and our proposed models for fashion dataset

5. Conclusion

In this paper, some deep CNN models for identifying garments design class along with our proposed CNN have been used and also the results are compared with several Hand-Engineered feature extraction methods.

Using two different datasets, this proposed Deep Convolutional Neural Network with five convolutional layers and four fully connected layers shows better performance than some existing deep convolutional model as well as several Hand-Engineered feature extraction methods. FC layers and Convolutional layers used in a Deep CNN represent the features more elaborately, which are stronger than any of Hand-Engineered feature extraction techniques.

77.8% accuracy has been achieved on Clothing Attribute Dataset with 6 different classes and 84.5% accuracy on Fashion dataset containing 5 texture design categories using the proposed model. When a database contains more generic properties for every class, then a deep network can extract the generic features easily and accurately. It is mentioned earlier that the used datasets were manually categorized in different clothing product classes and used only a few numbers of classes. For this reason, the classes contain less generic properties most of the time. Additional FC layer used in the proposed model helps the model to understand the features from these datasets more accurately.

This research work will help other future researchers for choosing appropriate deep learning model for garments texture design classification. In future, we will try to improve the results by adopting more sophisticated strategies.

Acknowledgement

This research is funded by Bangladesh University Grants Commission, No: Reg/Prosha-3/2016/46889.

References

[1] J. Wu and J. M. Rehg, “Centrist: a visual descriptor for scene categorization,” IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 8, pp. 1489-1501, August 2011.

[2] T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[3] O. L. Junior, D. Delgado, V. Gonçalves and U. Nunes, "Trainable classifier-fusion schemes: an application to pedestrian detection," Proc. 12th International IEEE Conference on Intelligent Transportation Systems (ITCS 09), IEEE Press, pp. 1-6, 2009.

[4] T. X. Yang and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE transactions on image processing, vol. 19, no. 6, pp. 1635-1650, 2010.

[5] Z. Guo, L. Zhang and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1657-1663, 2010.

[6] E. K. Dey, M. N. A. Tawhid and M. Shoyaib, "An automated system for garment texture design class identification," Computers, vol. 4, no. 3, pp. 265-282, 2015.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Proc. Advances in neural information processing systems, pp. 1097-1105, 2012.

[8] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba and A. Oliva, "Learning deep features for scene recognition using places database," Proc. Advances in neural information processing systems, pp. 487-495, 2014.

[9] K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets," arXiv preprint arXiv: 1405.3531, 2014.

[10] P. Heit, "The berkeley model," Health education, vol. 8, no. 1, pp. 2-3, 1977.

[11] L. Wang, C. Y. Lee, Z. Tu and S. Lazebnik, "Training deeper convolutional networks with deep supervision," arXiv preprint arXiv: 1505.02496, 2015.

[12] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34-42, 2015.

[13] Z. Ge, C. McCool and P. Corke, "Content specific feature learning for fine-grained plant classification," Proc. CLEF (Working notes), 2015.

[14] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz and T. L. Berg, "Parsing clothing in fashion photographs," Proc. Computer Vision and Pattern Recognition (CVPR), pp. 3570-3577, 2012.

[15] K. Yamaguchi, M. H. Kiapour and T. L. Berg, "Paper doll parsing: retrieving similar styles to parse clothing items," Proc. of the IEEE International Conference on Computer Vision, pp. 3519-3526, 2013.

[16] C. Shan, S. Gong and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.

[17] X. Feng, A. Hadid and M. Pietikäinen, "A coarse-to-fine classification scheme for facial expression recognition," Proc. International Conference Image Analysis and Recognition, Springer Berlin Heidelberg, pp. 668-675, 2004.

[18] E. Simo-Serra, S. Fidler, F. Moreno-Noguer and R. Urtasun, "A high performance CRF model for clothes parsing," Proc. Asian conference on computer vision, Springer International Publishing, pp. 64-81, 2014.

[19] S. Vittayakorn, K. Yamaguchi, A. C. Berg and T. L. Berg, "Runway to realway: visual analysis of fashion," Proc. IEEE Winter Conference on Applications of Computer Vision, IEEE Press, pp. 951-958, 2015.

[20] Y. Kalantidis, L. Kennedy and L. J. Li, "Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos," Proc. of the 3rd ACM conference on International conference on multimedia retrieval, ACM, pp. 105-112, 2013.

[21] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," Proc. Computer Vision and Pattern Recognition (CVPR 2008), IEEE Press, pp. 1-8, 2008.

[22] L. Bourdev, S. Maji and J. Malik, "Describing people: a poselet-based approach to attribute classification," Proc. International Conference on Computer Vision, IEEE Press, pp. 1543-1550, 2011.

[23] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," Pattern recognition letters, vol. 24, no. 9, pp. 1513-1521, 2003.

[24] M. M. Rahman, S. Rahman, M. Kamal, E. K. Dey, M. A. A. Wadud and M. Shoyaib, "Noise adaptive binary pattern for face image analysis," Proc. 18th International Conference on Computer and Information Technology (ICCIT), IEEE Press, pp. 390-395, 2015.

[25] B. Jun, I. Choi and D. Kim, "Local transform features and hybridization for accurate face and human detection," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 6, pp. 1423-1436, 2013.

[26] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE Press, pp. 2169-2178, 2006.

[27] B. Lao and K. Jagadeesh, "Convolutional neural networks for fashion classification and object detection," http://cs231n.stanford.edu/reports/BLAO_KJAG_CS231N_FinalPaperFashionClassification.pdf, June 26, 2016.

[28] F. Hu, G. S. Xia, J. Hu and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," Remote Sensing, vol. 7, no. 11, pp. 14680-14707, 2015.

[29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Caffe: convolutional architecture for fast feature embedding," Proc. 22nd ACM international conference on Multimedia, ACM, pp. 675-678, 2014.

[30] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," Proc. European Conference on Computer Vision, Springer International Publishing, pp. 818-833, 2014.

[31] M. Manfredi, C. Grana, S. Calderara and R. Cucchiara, "A complete system for garment segmentation and color classification," Machine Vision and Applications, vol. 25, no. 4, pp. 955-969, 2014.

[32] H. Chen, A. Gallagher and B. Girod, "Describing clothing by semantic attributes," Proc. European Conference on Computer Vision, Springer Berlin Heidelberg, pp. 609-623, 2012.

[33] M. Liu, S. Li, S. Shan and X. Chen, "Au-aware deep networks for facial expression recognition," Proc. 10th Automatic Face and Gesture Recognition (FG), IEEE Press, pp. 1-6, 2013.