MDPI

Article

Garments Texture Design Class Identification Using Deep Convolutional Neural Network

S.M. Sofiqul Islam [1,†,‡]*, Emon Kumar Dey [2,‡], Md. Nurul Ahad Tawhid [3,‡] and B. M. Mainul Hossain [4,‡]

[1]   Affiliation 1; bit0310@iit.du.ac.bd
[2]   Affiliation 2; emonkd@iit.du.ac.bd
[3]   Affiliation 3; tawhid@iit.du.ac.bd
[4]   Affiliation 4; raju@du.ac.bd
*   Correspondence: bit0310@iit.du.ac.bd; Tel.: +880-152-140-9419
†   Current address:Institute of Information Technology, University of Dhaka, Dhaka 1000, Bangladesh;
‡   These authors contributed equally to this work.

1 **Abstract:** Automatic garments design class identification for recommending the fashion trends is
2 important nowadays because of the rapid growth of online shopping. By learning the properties
3 of images efficiently, a machine can give better accuracy of classification. Several methods, based
4 on Hand-Engineered feature coding exist for identifying garments design classes. But, most of the
5 time, those methods do not help to achieve better results. Recently, Deep Convolutional Neural
6 Networks (CNNs) have shown better performances for different object recognition. Deep CNN
7 uses multiple levels of representation and abstraction that helps a machine to understand the types
8 of data (images, sound, and text) more accurately. In this paper, we have applied deep CNN for
9 identifying garments design classes. To evaluate the performances, we used two well-known CNN
10 models AlexNet and VGGNet on two different datasets. We also propose a new CNN model based
11 on AlexNet and found better results than existing state-of-the-art by a significant margin.

12 **Keywords:** CNN; Deep Learning; AlexNet; VGGNet; Texture Descriptor; Garment Categories;
13 Garment Trend Identification; Design Classification for Garments.

1. Introduction

15 Forecasting fashion trends have a great business value in fashion industry. A fashion forecaster
16 predicts the colours, fabrics and styles for the upcoming seasons. One of the main resources of
17 identifying trends is to analyse the choices of the customers. Online shopping is popular nowadays.
18 Customer select products from the web pages according to their choice and that can help to predict
19 the direction of trends. If a retailer knows popular design styles of clothing products, it can increase
20 the production of those styles to achieve more profit. Therefore, if a system can classify the garments
21 products according to different style, texture, size etc., it can automatically suggest different products
22 to the customers based on their choices. In this paper, we have developed a system which can classify
23 clothes according to textures.
24 Effective design classification based on textures, local spatial variations of intensity or colour
25 in images has been an important topic of interest in the past decades. A successful classification,
26 detection or segmentation requires an efficient description of image textures. To fulfill this
27 purpose, lots of well-known Hand-Engineered feature extraction methods such as CENsus Transform
28 hiSTogram (CENTRIST) [1], Local Binary Pattern (LBP) [2], Gabor [3], Histogram of Oriented
29 Gradient (HOG) [4], GIST [5] etc., are exist. LBP gain popularity because of their computational

simplicities and better accuracies. But, it is very sensitive to uniform and near uniform region. LTP [6], Completed Local Binary Pattern (CLBP) [7] can handle this issue more accurately. Between these two methods, CLBP is better choice because this method is rotation invariant [8]. CENTRIST [1] has gain popularity by incorporating Spatial Pyramid (SP) structure. But, most recently Completed CENTRIST (cCENTRIST) and Ternary CENTRIST (tCENTRIST) [8] gained high accuracies for garments design classification. Although several Hand-Engineered feature extraction approaches exist for garments design classification, deep learning [9] is rarely used in this field. Our goal is to apply appropriate deep learning model to measure the performance in 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 [10], text classification [11], image classification [12], face verification [13], gender classification [14], scene-classification [15], digits and traffic signs recognition [16], etc. Some of the available deep learning models are AlexNet [12], VGGNet [17], Berkeley-trained models [18], Places-CNN model [15], Fully Convolutional Semantic Segmentation Models (FCN-Xs) [19], CNN Models for Salient Object Subitizing [20], Places CNDS models on Scene Recognition [21], Models for Age and Gender Classification [22], GoogLeNet model [23], etc. These methods have achieved dramatic improvements beyond the state-of-the-art records in such broad domains, and they have 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 we can tolerate, etc.

In this research, we apply two deep learning models namely AlexNet and VGG_S alongside our proposed model for garments design classification and compare them with some well-known Hand-Engineered methods. The major contributions of this paper are as follows.

1. We conduct a brief review on existing well known Hand-Engineered feature extraction methods which are used in garments design classification.

2. We apply some existing Deep Convolutional Network for classifying design class of garments clothing products on some datasets and make a comparison with several state-of-the-art Hand-Engineered feature extraction methods.

3. We propose a new Deep Convolutional Network model for classifying garments design class and apply it on two different datasets which gives a remarkable output.

The rest of the paper is structured as follows. Section 2 discusses the background studies, Section 3 discusses the working method, Section 4 presents the experimental results and Section 5 concludes the work.

2. Background Studies

In this section, we discuss about some existing garments clothing segmentation and classification strategies with existing descriptors. We also describe some deep learning models; that are used for several applications in computer vision.

2.1. Garment Product Segmentation and Identification

Yamaguchi et al. [24] proposed a method for clothing parsing. For this work, they create Fashionista dataset, that consist of 158,235 images. From this, they selected 685 images for training and testing. They identified 14 different parts of a body and different clothing regions. For pose detection they used a method described in [25]. In [26], they deal with clothing parsing problem using retrieval based approach. Their proposed approach focus on pre-trained global clothing models, local clothing models, and transferred parse. Authors found out that their proposed final parse achieve

84.68% parsing accuracy. Menfredi et al. [27] proposed a new approach for automatic garments segmentation and classification. Here, authors classified garments into nine different classes such as skirts, shirt, dresses, etc. For this work, authors used a projection histogram for extracted few specific garments. Since, they divided the whole image into 117 cells and group them into 3×3 cells. For computing HOG features [28] from each cell the orientations are grouped into nine bins. For concatenating the projection histogram and HOG features they used multiclass linear support vector for training. Serra et al. [29] did similar type of work, where authors used conditional random field (CRF) for divided outfits.

Vittayakorn et al. [30] 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. [31] proposed a method for identify relevant products. From an input image first they estimate the pose of the person, then segment the clothing area such as shirt, tops, jeans, etc. Finally they applied an image retrieval technique which is 50 times faster than [24] for identifing similar clothes for each class.

Gallagher et al. [32] used grab cut algorithm [33] for clothing part segmentation for identifying a person. Authors successfully extracted the region of interest (ROI) which is only torso region; using this method. Bourdev et al. [34] proposed a new method for attributes and type of cloth identification 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 annotated attributes.

2.2. Texture Based Classification

Nowadays garments design classification based on texture become more popular and there are several existing well known methods such as Histogram of Oriented Gradients (HOG) [4], Local Binary Pattern (LBP) features [2], Wavelets transform [35], Noise Adaptive Binary Pattren (NABP) [36], Gabor filters [3], Scale-invariant feature transform (SIFT) [37]. Recently, LBP become popular which was proposed for describing the local structure of an image because of computational simplicity. It has been used in several areas such as facial image analysis, including face detection [38], face recognition and facial expression analysis [39]; demographic (gender, race, age, etc.) classification [40], [41]; moving object detection [42], etc. However, LBP is very sensitive in uniform and near uniform regions. In the last few years, lots of researches are done by modification of LBP such as derivative-based LBP [43], dominant LBP [44], Rotation Invariant [45], center-symmetric LBP [46], etc. to improve the performance.

Tan and Triggs [6] proposed a new texture based method Local Ternary Patterns (LTP), which can tolerance noises up to a certain level. They used a fixed threshold (± 5), for made LTP more discriminant and less sensitive to noise in a uniform region. Beside this there are several methods that also handle noises in different application areas, such as the methods described by Jun et al. [47]. 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. [7] proposed a new method Completed Local Binary Pattern (CLBP), which incorporates sign, magnitude and center pixel information. This method also 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) [48] to capture global structure from images. Which is a collection of orderless feature histograms computed over cells. CENTRIST uses total 31 blocks to avoid the artefacts. All blocks are belong to 3 level, here one block from level 0, five blocks from level 1, 25 blocks come from level 2.

Dey et al. [8] proposed two new descriptors for garments design class identification namely Completed CENTRIST (cCENTRIST) and Ternary CENTRIST (tCENTRIST). These descriptors are

based 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. In above, we have discussed about some supervised methods for feature extraction applied on various garments design classification areas.

2.3. Deep Learning

In this sub-section, we describe some deep learning techniques for garments design classification. Deep learning is a research trend in machine learning researchers nowadays. These methods learn features automatically from large number of unlabelled data hence more useful hidden discriminative features are extracted. Deep learning as unsupervised feature learning methods achieved popularity in classic problems, such as speech recognition, object recognition and detection, natural language processing, etc. Convolutional Neural Networks (CNN) are now being used in several image, pattern and signal processing researches. Liu et al. [13] 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), authors used restricted Boltzmann machine (RBMs). Later, this technique was applied on three expression database namely CK+, MMI and SFEW, and found that better or at least competitive results are achieved. However, this method fails when several kinds of challenging images (eg., the subjects have higher expression non-uniformity, most of them have moustache and wear accessories such as glasses) are appeared.

Krizhevsky et al. [12] 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 was increased. Dey et al. [8] used deep learning model in texture based garments design classification. In their experiment, using Berkeley-trained model [18], 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. [15] 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 this 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. [49] 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 neighbors; and clothing object detection. For this work they used Apparel Classification with Style (ACS), Clothing Attribute (CA) and Colorful-Fashion (CF) datasets and find out 50.2% and 74.5% accuracy for clothing style classification and Clothing Attribute datasets. Hu et al. [50] 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 use convolutional feature coding scheme for aggregating the dense convolutional features into a global representation. Their proposed two models achieve remarkable performance and improve the state-of-the-art by a significant margin.

For garments design class identification many approaches have been proposed for solving the problem in different applications. But, there are only a few works that have been conducted based on deep learning. In this paper, We experiment different deep learning methods for identifying different garments design class based on textures.
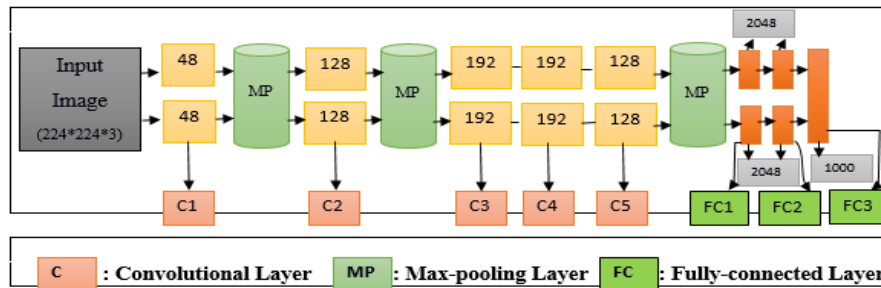
Figure 1. Basic steps of our working procedure

Figure 2. The full architecture of AlexNet Model

3. Working Procedure

This section describes the procedure we used for classifying garments design class. Basic steps of our working procedure is shown in Figure 1. Input images are firstly segmented and classified into several classes based on their texture design. After that, we separate these images for training, testing and validation. Then, we apply our proposed model alongside two well-known deep Convolutional Neural Network (CNN) models AlexNet and VGG_S in two different garment datasets, for the purpose of both training and testing. Finally, the accuracy of proposed system based on AlexNet, VGG_S and our new proposed models are compared. We also compare the results with traditional state-of-the-arts Hand-Engineered feature extraction method of machine learning for classifying garments design class.

We use CNN models AlexNet and VGG_S because of computational simplicity, better performance in several areas and 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. Also, these two models gain 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.

3.1. AlexNet Model

Krizhevsky et al. [12] proposed a deep convolutional neural network model, which is known as AlexNet model. There are three types of layer such as Convolution layer, Pooling layer and Fully-Connected (FC) layers. Full architecture of this model is 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 and filters size is different from one another. Full AlexNet model architecture shown in Figure 2.

First convolution layer takes the input images by resizing each of the images into 224×224 size with 96 kernels. The second layer takes the input from first convolution layer with 256 kernels after passing through a pooling layer. Pooling layer operate 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
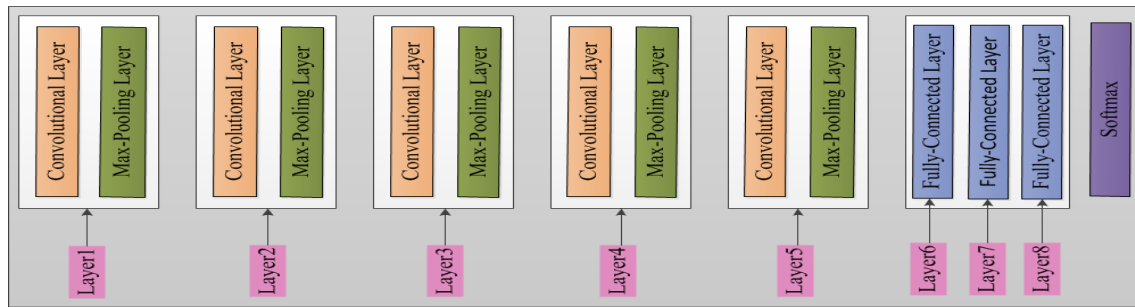
Figure 3. The full architecture of VGG_S Model

of pooling layer. 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 layer 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. [17], based on Caffe toolkit [51] 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 contain 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. [52]. First convolution layer 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 [53] which has six convolutional layers. VGG_S model 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. We use this model to evaluate the garments design class identification. As depicted in Figure 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. We use VGG_S model to classify the garments design class in our work.

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 [62] and the last layer acts as a multi-way soft-max classifier.

3.3. Proposed Transferred CNN

For classifying garments design class, we propose a new scenario based on AlexNet to observe the performance and effectiveness of deep features by total nine learned layers, among which five are Convolutional layers and four are Fully Connected layers. Like AlexNet, our model first convolution layer 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. We add pooling layers after first, second and fifth convolution layer like AlexNet. We also added a new Fully Connected layer (FC3) which takes input from the output of second Fully Connected layer (FC2). Output of the last layer (FC4) is connected to a softmax layer for classifying the categories. In our model, we use data augmentation technique to reduce overfitting in the training stage. Because,
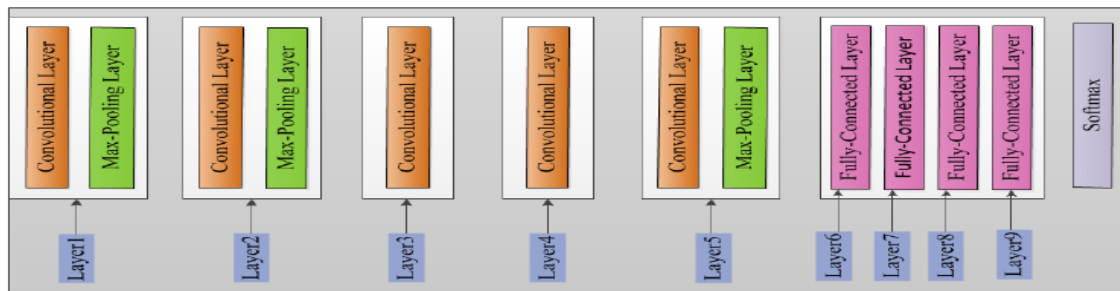
Figure 4. The full architecture of our Proposed model

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	5413
Training sample/class	I) 10	I) 60
	II) 20	II) 100
	III) 30	III) 200
		IV) 300
Validation sample/class	I) 10	I) 10
	II) 20	II) 10
	III) 30	III) 20
		IV) 30

239   recent works show that data augmentation also helps to improve classification performance [12]. The
240   full architecture of our Proposed model is shown in Figure 4.

241   *3.4. Dataset*

242   We have considered two publicly available datasets Fashion [54] and Clothing Attribute datasets
243   (CAD) [55] that was originally created for Garment Product Recognition. From Fashion dataset, firstly
244   we manually take 5400 images and categorized them into five design classes, namely "Single color"
245   (2440 images), "Print" (1141 images), "Strip" (565 images), "Jeans" (614 images) and "Leather" (640
246   images). From Clothing Attribute dataset; we take 1575 images and manually categorized those into
247   six different categories after extracting garments area from the original dataset. The categories are
248   "Floral" (69 images), "Graphics" (110 images), "Plaid" (105images), "Spotted" (100 images), "Striped"
249   (140 images) and "Solid" pattern (1051images). Figure 5 and Figure 6 show some sample images from
250   "Fashion" and "Clothing Attribute" datasets used in our work.
251   Table 1 represents proper training and validation samples about Clothing Attribute and Fashion
252   datasets. For Clothing Attribute dataset, we use different training and validation sample such as
253   10, 20, 30 images per class and rest of the images for testing to identify the classification results. In
254   Fashion dataset, we use 60, 100, 200 and 300 images for training and 10, 10, 20 and 30 images for
255   validation and rest of the images for testing.

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

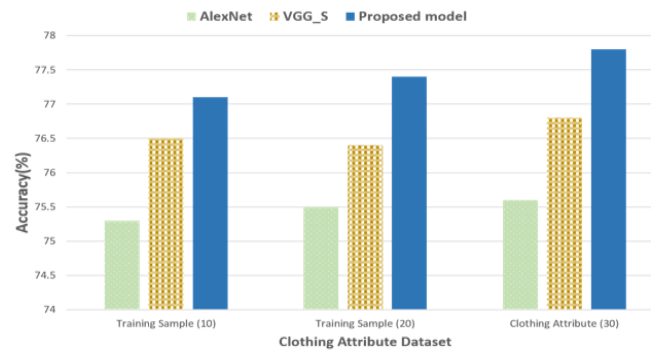| Dataset	CAD with 6 classes								
Models	AlexNet			VGG_S			Proposed Model		
Training Sample	10	20	30	10	20	30	10	20	30
Validation Sample	10	20	30	10	20	30	10	20	30
Results	75.1	75.6	75.5	76.2	76.5	76.6	77.2	77.3	77.7

Figure 5. Some example images from Fashion dataset: Each of the row represents jeans, leather, print, single color and stripe category respectively.

Figure 6. Example of Clothing Attribute dataset: Column 1 to 6 represents example of Floral, Graphics, Plaid, Solid Color, Spotted and Striped garments respectively.

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

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

Figure 7. Comparison between AlexNet, VGG S and our proposed models for Clothing Attribute Dataset.

Table 4. Recognition rate (%) in training phase of Fashion Dataset

Dataset	Fashion Dataset with 6 classes											
Models	AlexNet				VGG_S				Proposed Model			
Training Sample	60	100	200	300	60	100	200	300	60	100	200	300
Validation Sample	10	10	20	30	10	10	20	30	10	10	20	30
Results	74.3	75.9	78.1	81.5	75.3	76.8	78.6	82.7	76.6	78.1	81.1	84.1

Table 5. Recognition rate (%) in testing phase of Fashion Dataset

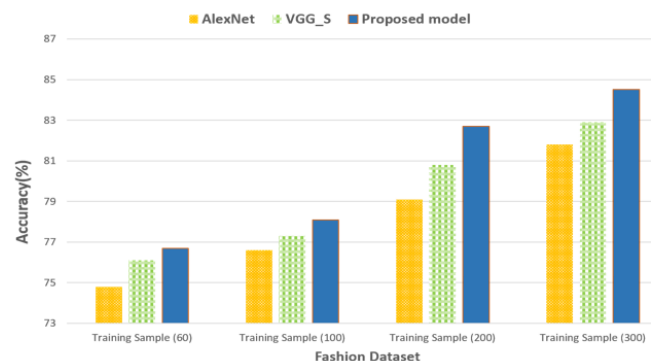| Dataset	Fashion Dataset with 6 classes											
Models	AlexNet				VGG_S				Proposed Model			
Training Sample	60	100	200	300	60	100	200	300	60	100	200	300
Validation Sample	10	10	20	30	10	10	20	30	10	10	20	30
Results	74.8	76.6	79.1	81.8	76.1	77.3	80.8	82.9	76.7	78.1	82.7	84.5

Figure 8. Comparison between AlexNet, VGG S and our proposed models for Fashion Dataset.

Table 6. Experimental results of different methods for Clothing Attribute dataset.

Method	Accuracy
HOG	63.01%
GIST	71.78%
LGP	64.95%
CENTRIST	71.27%
tCENTRIST	73.98%
cCENTRIST	74.07%
NABP	74.18%
Berkeley	73.54%
AlexNet (30)	75.6%
VGG_S (30)	76.8%
Proposed Model	**77.8%**

4. Experimental Result

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

4.1. Implementation Environment

For this work, we set our experimentation environment by following a straightforward process. We fine-tune the CaffeNet [51] model and use Ubuntu 12.4 operating system. We consider high speed GPU for this work, because working with large datasets and complex DNN, CPU is nearly ten times slower than GPU. We use NVIDIA GEFORCE GTX 950 4GB GPU and Intel core i7 processor for faster training and testing.

4.2. Experimental Result and Discussion

We use two deep convolutional neural network models alongside our proposed model on Fashion and Clothing Attribute datasets for comparing the performance of these models with some existing well-known Hand-Engineering feature extraction approaches for garment design classification. In this work, we use different training, validation and testing sample from these two datasets, that shown in Table 1. The training and testing accuracies of AlexNet, VGG_S and proposed model for these two datasets 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 Figure 7 we find that in most of the cases VGG_S performs better than AlexNet model.

For Clothing Attribute dataset, we achieve maximum 75.6% accuracy on AlexNet and 76.8% accuracy on VGG_S, but, our proposed model achieve 77.8% accuracy. For Fashion dataset with 5 classes Table 5 and Figure 8 shows that we have achieved 81.8% accuracy on AlexNet and 82.9% accuracy on VGG_S respectively and our proposed model achieve 84.5% accuracy. From these two tables, it can be concluded that using more training sample, we can produce more accuracy.

Table 6 shows the experimental results using seven different Hand-Engineered feature extraction methods which are (HOG, GIST, LGP, CENTRIST, tCENTRIST, cCENTRIST, and NABP). For these methods Support Vector Machine (SVM) is used for classification purpose. 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 model is better than any Hand-Engineering feature extraction method.

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

Table 7. Experimental results of different methods for Fashion dataset.

Dataset	Accuracy
HOG	78.75%
GIST	81.22%
LGP	79.34%
CENTRIST	79.02%
tCENTRIST	83.67%
cCENTRIST	83.73%
NABP	83.22%
AlexNet	81.8%
VGG_S	82.9%
Proposed Model	**84.5%**

5. Conclusions

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

Using two different datasets we show that a Deep Convolutional Neural Network with five convolutional layers and four fully connected layers show 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.

We achieve 77.8% accuracy on Clothing Attribute Dataset with 6 different classes and 84.5% accuracy on Fashion dataset containing 5 texture design categories using our proposed model. When a database contains more generic properties for every class, then a Deep network can extract the generic features easily and accurately. We mentioned that we manually categorized the clothing products and use only a few numbers of classes. For this reason, the classes contain less generic properties most of the time. Additional FC layer helps this model to understand the features from these datasets more accurately.

We hope that, this research work will help other future researchers for choosing appropriate deep learning model for garments texture design classification. In future, we plan to improve the results by adopting more sophisticated strategies in this field.

Author Contributions:

All authors contributed equally to this work, and have read and approved the final manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

References

1. Wu, J.; Rehg, J.M. CENTRIST: A visual descriptor for scene categorization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2011**, *33*, 1489–1501.
2. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2002**, *24*, 971–987.
3. Arivazhagan, S.; Ganesan, L.; Priyal, S.P. Texture classification using Gabor wavelets based rotation invariant features. *Pattern recognition letters* **2006**, *27*, 1976–1982.
4. Junior, O.L.; Delgado, D.; Gonçalves, V.; Nunes, U. Trainable classifier-fusion schemes: an application to pedestrian detection. Intelligent Transportation Systems, 2009, Vol. 2.
5. Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision* **2001**, *42*, 145–175.

6. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on* **2010**, *19*, 1635–1650.

7. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on* **2010**, *19*, 1657–1663.

8. Dey, E.K.; Tawhid, M.N.A.; Shoyaib, M. An Automated System for Garment Texture Design Class Identification. *Computers* **2015**, *4*, 265–282.

9. Lee, H.; Pham, P.; Largman, Y.; Ng, A.Y. Unsupervised feature learning for audio classification using convolutional deep belief networks. Advances in neural information processing systems, 2009, pp. 1096–1104.

10. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

11. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent Convolutional Neural Networks for Text Classification. AAAI, 2015, pp. 2267–2273.

12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.

13. Liu, M.; Li, S.; Shan, S.; Chen, X. Au-aware deep networks for facial expression recognition. Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, 2013, pp. 1–6.

14. Mäkinen, E.; Raisamo, R. Evaluation of gender classification methods with automatically detected and aligned faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2008**, *30*, 541–547.

15. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning deep features for scene recognition using places database. Advances in neural information processing systems, 2014, pp. 487–495.

16. Ciresan, D.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3642–3649.

17. Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* **2014**.

18. Heit, P. The Berkeley Model. *Health education* **1977**, *8*, 2–3.

19. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

20. Zhang, J.; Ma, S.; Sameki, M.; Sclaroff, S.; Betke, M.; Lin, Z.; Shen, X.; Price, B.; Mech, R. Salient object subitizing. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4045–4054.

21. Wang, L.; Lee, C.Y.; Tu, Z.; Lazebnik, S. Training deeper convolutional networks with deep supervision. *arXiv preprint arXiv:1505.02496* **2015**.

22. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 34–42.

23. Ge, Z.; Mccool, C.; Corke, P. Content specific feature learning for fine-grained plant classification. Working notes of CLEF 2015 conference, 2015.

24. Yamaguchi, K.; Kiapour, M.H.; Ortiz, L.E.; Berg, T.L. Parsing clothing in fashion photographs. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3570–3577.

25. Yang, Y.; Ramanan, D. Articulated pose estimation with flexible mixtures-of-parts. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 1385–1392.

26. Yamaguchi, K.; Kiapour, M.; Berg, T. Paper doll parsing: Retrieving similar styles to parse clothing items. Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3519–3526.

27. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing* **2009**, *27*, 803–816.

28. Feng, X.; Hadid, A.; Pietikäinen, M. A coarse-to-fine classification scheme for facial expression recognition. In *Image Analysis and Recognition*; Springer, 2004; pp. 668–675.

29. Simo-Serra, E.; Fidler, S.; Moreno-Noguer, F.; Urtasun, R. A high performance CRF model for clothes parsing. In *Computer Vision–ACCV 2014*; Springer, 2014; pp. 64–81.

30. Vittayakorn, S.; Yamaguchi, K.; Berg, A.C.; Berg, T.L. Runway to realway: Visual analysis of fashion. Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on. IEEE, 2015, pp. 951–958.

31. Kalantidis, Y.; Kennedy, L.; Li, L.J. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. Proceedings of the 3rd ACM conference on International conference on multimedia retrieval. ACM, 2013, pp. 105–112.

32. Gallagher, A.C.; Chen, T. Clothing cosegmentation for recognizing people. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008, pp. 1–8.

33. Rother, C.; Kolmogorov, V.; Blake, A. Grabcut: Interactive foreground extraction using iterated graph cuts. ACM transactions on graphics (TOG). ACM, 2004, Vol. 23, pp. 309–314.

34. Bourdev, L.; Maji, S.; Malik, J. Describing people: A poselet-based approach to attribute classification. Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 1543–1550.

35. Arivazhagan, S.; Ganesan, L. Texture classification using wavelet transform. *Pattern recognition letters* **2003**, *24*, 1513–1521.

36. Rahman, M.M.; Rahman, S.; Kamal, M.; Dey, E.K.; Abdullah-Al-Wadud, M.; Shoyaib, M. Noise Adaptive Binary Pattern for Face Image Analysis. Computer and Information Technology (ICCIT), 2015 18th International Conference on. IEEE, 2015.

37. Cheung, W.; Hamarneh, G. N-sift: N-dimensional scale invariant feature transform for matching medical images. Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on. IEEE, 2007, pp. 720–723.

38. Hadid, A.; Pietikäinen, M.; Ahonen, T. A discriminative feature space for detecting and recognizing faces. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, Vol. 2, pp. II–797.

39. Shoyaib, M.; Abdullah-Al-Wadud, M.; Youl, J.M.; Alam, M.M.; Chae, O. Facial expression recognition based on a weighted local binary pattern. 2010 13th International Conference on Computer and Information Technology (ICCIT), 2010.

40. Rahman, M.M.; Rahman, S.; Dey, E.K.; Shoyaib, M. A Gender Recognition Approach with an Embedded Preprocessing. *International Journal of Information Technology and Computer Science (IJITCS)* **2015**, *7*, 19.

41. Dey, E.K.; Khan, M.; Ali, M.H. *Computer Vision Based Gender Detection from Facial Image*; LAP LAMBERT Academic Publishing, 2013.

42. Heikkilä, M.; Pietikäinen, M.; Heikkilä, J. A texture-based method for detecting moving objects. BMVC, 2004, pp. 1–10.

43. Huang, X.; Li, S.Z.; Wang, Y. Shape localization based on statistical method using extended local binary pattern. Image and Graphics (ICIG'04), Third International Conference on. IEEE, 2004, pp. 184–187.

44. Liao, S.; Law, M.W.; Chung, A. Dominant local binary patterns for texture classification. *Image Processing, IEEE Transactions on* **2009**, *18*, 1107–1118.

45. Guo, Z.; Zhang, L.; Zhang, D. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern recognition* **2010**, *43*, 706–719.

46. Heikkilä, M.; Pietikäinen, M.; Schmid, C. Description of interest regions with local binary patterns. *Pattern recognition* **2009**, *42*, 425–436.

47. Jun, B.; Choi, I.; Kim, D. Local transform features and hybridization for accurate face and human detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2013**, *35*, 1423–1436.

48. Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, Vol. 2, pp. 2169–2178.

49. Lao, B.; Jagadeesh, K. Convolutional Neural Networks for Fashion Classification and Object Detection.

50. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing* **2015**, *7*, 14680–14707.

51. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the ACM International Conference on Multimedia. ACM, 2014, pp. 675–678.

52. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Computer vision–ECCV 2014*; Springer, 2014; pp. 818–833.

53. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229* **2013**.

54. Manfredi, M.; Grana, C.; Calderara, S.; Cucchiara, R. A complete system for garment segmentation and color classification. *Machine Vision and Applications* **2014**, *25*, 955–969.

55. Chen, H.; Gallagher, A.; Girod, B. Describing clothing by semantic attributes. In *Computer Vision–ECCV 2012*; Springer, 2012; pp. 609–623.