# LOCAL BINARY PATTERN NETWORK : A DEEP LEARNING APPROACH FOR FACE RECOGNITION

*Meng Xi*[1], *Liang Chen*[1], *Desanka Polajnar*[1], *and Weiyang Tong*[2]

[1]Department of Computer Science, University of Northern British Columbia, Canada
[2]Department of Mechanical and Aerospace Engineering, Syracuse University, USA

## ABSTRACT

Deep learning is well known as a method to extract hierarchical representations of data. In this paper a novel unsupervised deep learning based methodology, named Local Binary Pattern Network (LBPNet), is proposed to efficiently extract and compare high-level over-complete features in multilayer hierarchy. The LBPNet retains the same topology of Convolutional Neural Network (CNN) - one of the most well studied deep learning architectures - whereas the trainable kernels are replaced by the off-the-shelf computer vision descriptor (i.e., LBP). This enables the LBPNet to achieve a high recognition accuracy without requiring any costly model learning approach on massive data. Through extensive numerical experiments using the public benchmarks (i.e., FERET and LFW), LBPNet has shown that it is comparable to other unsupervised methods.

***Index Terms***— Deep learning, Local Binary Pattern, PCA, Convolutional Neural Network

## 1. INTRODUCTION

Deep learning is one brunch of machine learning which tends to extract high-level abstractions or representations of data through multiple processing layers [1]. A hierarchical architecture can be formed through sequential connections between multiple layers. The latter layers in the hierarchy extract a higher level of abstractions from the lower level ones of the earlier layers. In addition, features are extracted in a heavily overlapped manner. This means that a low-level feature can contribute to multiple high-level features in the later layer.

Convolutional Neural Network (CNN) is one of the most commonly studied deep learning architectures, which can be viewed as a variant of multilayer perceptron (MLP) neural network. CNN obtains the facial discriminative representations from a set of hierarchically connected and trainable convolutional kernels [2, 1]. Comparing with other regular face recognition methods, training CNN is troublesome. Difficulties in CNN generally come from two folds: (i) the learning approach itself is computational expensive due to a large amount of parameters in sequentially connected multiple layers, which makes the convergence undesirably time-consuming; (ii) overfitting is more likely to occur due to the existence of thousands of parameters in this model. The former issue is primarily solved using powerful computers and leveraging hardware accelerating techniques (e.g., GPU computing). To tackle the latter issue, in the case of face recognition, many state-of-the-art systems leverage massive external data to learn their networks. However, we believe that these are just workarounds by utilizing more computing resource rather than final solutions. Considering the complexities of CNN are mainly attributed to its trainable kernels, the question we want to address here is the possibility to replace the convolutional kernels with off-the-shelf computer vision descriptors; such a framework can be capable for the high-level feature extraction on dense data with only a few of adjustable parameters. This framework requires no additional data and can help avoid the costly training process.

In this paper, a deep network based on Local Binary Pattern (LBP) descriptor is proposed, which is named as Local Binary Pattern Network (LBPNet). Two filters are used in LBPNet, which are based on LBP and Principle Component Analysis (PCA) techniques, respectively. The over-complete patch-based features are extracted hierarchically using these two filters. After feature extraction, the LBPNet employs a simple network to measure the similarity of the extracted features. Major characteristics of the proposed LBPNet are summarized in the following:

**Feature extraction in dense grid:** Both of the two filters are replicated densely in layers.

**Multilayer architecture:** The representations are extracted hierarchically: the latter layer extracts a higher level of abstractions from the lower ones of the earlier layer.

**Partially connected layer:** Filters only compute based on the selected subset of the inputs from the earlier layer.

**Multi-scale analysis:** Filters with different parameters are used in each of the layers to capture multi-scale statistics.

This work is supported by NSERC Discovery Grant (No. 261403-2011 RGPIN).

**Unsupervised learning:** Since both LBP and PCA are unsupervised learning algorithms, LBNet is capable to perform unsupervised learning on data.

Since LBPNet contains all the fundamental characteristics of deep learning architecture, it can be classified as a simplified deep network with *hand-craft* filters. The proposed LBPNet retains the key features of CNN but simplifies the model to avoid the costly training approach, which makes LBPNet distinguished from the regular CNN architectures and significantly outperform the original LBP approach.

## 2. RELATED WORKS

In the context of face recognition, CNN gradually extracts a high-level discriminative facial descriptor using multiple trainable feature extractors. In the literature, many attempts have been made to fold shallow face recognition algorithms into similar deep architectures [3, 4].

One key stage used in many state-of-art methods is to extract over-complete features. The over-complete features can be obtained in many different ways: by image pyramid [5]; by dense grid [6]; by obtaining multi-scale representations [7]. The extracted features are then compressed by feature selection or subspace projection.

The patch-based system measures the similarity using a divide-and-conquer strategy. The naive algorithm is to partition the image into non-overlapped patches (with equal sizes) and sum up all the local similarities. [8, 9] enhanced this algorithm by allowing patches shift in a small range to tackle the misalignment issue. [10] suggested a component-level face alignment algorithm where the patches are centred at important components of the faces (e.g., eyes, nose, and mouth).

## 3. PROPOSED ALGORITHM

This section describes a powerful deep learning architecture called LBPNet, which provides a novel tool for face recognition.

The architecture of LBPNet can be divided into two parts: (i) deep network for feature extraction, and (ii) regular network for classification. Two layers, respectively using LBP and PCA filters, hierarchically connected in the deep network part to extract high-level over-complete representations of the images (Figure 1). Two such networks are connected with the classification networks, allowing taking two images as the input. Hence, the similarity measurement can be performed based on the extracted features. Details of each layer are described in the following two subsections.

### 3.1. LBP Filter Layer

In the LBP filter layer, filters are based on LBP operator described in [11]. The LBP operator, $LBP^{u2}_{P,R}$, labels each pixel

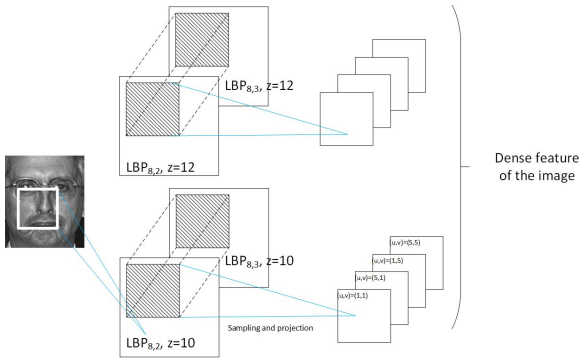


**Fig. 1**. The deep network part of LBPNet for feature extraction.

$g_c$ in the image by thresholding its $P$ surrounding points $g_p$ ($p \in [1, P]$) and then stacking labels $l_p$ (defined in Eq. 1) into one binary string

$$l_p = \begin{cases} 1, \text{when } g_p > g_c \\ 0, \text{when } g_p \leq g_c \end{cases} \quad (1)$$

The points are sampled from a circle with a radius of $R$, whose centre is at $g_c$. In addition, we use the unique pattern (denoted by $u2$ in the operator) to encode the labels. The feature generated in the filter is formulated as

$$h_{P,R,z}(i) = (\sum B(LBP^{u2}_{P,r}(x, y) = i))^{-\frac{1}{2}} \quad (2)$$

where $z$ is the filter size, and $B(v)$ is 1 when $v$ is true; 0 otherwise. Note that here the square root of the LBP histograms is used to increase the discrimination ability [12]. By replicating the kernel, a 3-dimensional feature cube can be generated from the image.

To capture multi-scale representations of the image, the computation is repeated in the LBP filter layer using multiple kernels subject to different combinations of LBP radius, $r$ and filter size, $z$. The features obtained in this layer represent the multi-scale LBP histogram features of the image. When considering them as one feature vector, it is of high dimensionality, which can be over 10 millions in our experiments.

### 3.2. PCA Filter Layer

The PCA filter reduces the dimensionality of input features while improving their discrimination abilities. To start with, the input features are sampled and concatenated into the vector $p_z$, which is given by

$$\begin{aligned} p_z = &[h_{r_1,z}(u, v), h_{r_1,z}(u + s, v), \dots \\ &h_{r_k,z}(u + n \times s, v + n \times s)], n = \lfloor M/s \rfloor \end{aligned} \quad (3)$$

where the filter size is $M \times M$, $s$ is the sampling stride, $h_{r,z}(u, v)$ is the feature vector located at the $u$-th column and the $v$-th row of the feature cube generated by the LBP filter

with sample radius $r$ and size $z$. $(u, v)$ denotes the starting point of sampling. Considering the feature extraction is in a dense grid in the earlier layer, the features are highly redundant —in general, two neighbourhood features can share up to 90% of the same LBP labels. Therefore, features are sampled to reduce the resulting vector length while preserving the critical discriminative information.

Here, the PCA filter only computes based on the features that are generated by LBP filters of the same size. This resembles the partial connections between convolutional layers in CNN. This is to help simplify the computation and increase the discrimination ability.

After obtaining the concatenated vector, the dimensionality by PCA projection can then be reduced.. In general, for a given matrix $A$ with zero mean, PCA seeks a transformation matrix, $W$, which minimizes the reconstruction error, $||A - W^T A||$. The solution is known as the matrix constructed by the first $n$ eigenvectors of the covariance matrix $C = A^T A$. In our case, the input matrix, $A$, is formed by

$$A = [q_1, q_2, \ldots, q_n]^T \tag{4}$$

where $q_i$ is defined as

$$q_i = p_i - \frac{1}{J} \sum_{j=1}^{J} p_j \tag{5}$$

In Equ 5 is the total number of features in the training set. In this method the extracted feature ($p_i$) yields another vector ($q_i$) by subtracting the mean vector of the entire training set from itself.

In the context of face recognition, high variability of images generally corresponds to the changes of illumination, facial expression, etc. Such impact of high variability can be reduced by downweighting the high variance directions whereas increasing the weak ones, since the discriminative information is uniformly distributed over all directions of the data [13]. Here, we normalize all the eigenvectors by whitening, the transformation matrix is then expressed as

$$W = [\lambda_1^{-\frac{1}{2}} x_1, \lambda_2^{-\frac{1}{2}} x_2, \ldots \lambda_n^{-\frac{1}{2}} x_n] \tag{6}$$

where $\lambda_i$ is the eigenvalue of the corresponding eigenvector $x_i$. The output feature is then extracted by

$$\sigma_i = W_i^T q_i \tag{7}$$

where $W_i$ is the whitened PCA transformation matrix. Similar to the first layer, multi-scale representation can be obtained using multiple filters with different parameters. Here we vary the starting point, $(i, j)$, and leave other parameters unchanged. Figure 2 shows an example of 4 filters with different starting points. The diagram of the deep part of LBP-Net consisting of the first two layers is presented in Figure 1. The outputs of the deep network represent patch-based overcomplete features of the image.

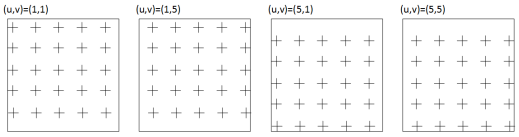


**Fig. 2**. 4 different PCA filters. Only vary the starting point of sampling and keep other parameters unchanged.

### 3.3. Similarity Measurement

Two deep networks are connected to accept two images as the input. The extracted features in the upper layer consist of two subsets from each image, respectively. The regional similarity scores, $\delta_i$, are computed pairwisely between two corresponding features. Here we use angle-based measure, cosine similarity, which is formulated as

$$\delta_i = \frac{\sigma_i \cdot \sigma\prime_i}{\|\sigma_i\| \, \|\sigma\prime_i\|} \tag{8}$$

where $\sigma_i, \sigma\prime_i$ are two features from upper layers, respectively. The output of these layers represents the regional similarity of two faces in specific scale.

It is assumed that the similarity of different pairs of features contribute equally to the final score. The output is then computed by averaging all regional similarities.

## 4. EXPERIMENTS

Numerical experiments are conducted in this section to validate the proposed framework using the public benchmarks. We are aware of some other CNN-based works (e.g. FaceNet [14]) which achieve superior results; however, they used a less restricted setting: external data are allowed for training. To fairly evaluate our method, we only compare LBP-Net to other unsupervised learning methods which are single descriptor based and do not require additional training data.

### 4.1. Experiment on Face Identification: FERET

To evaluate the capability of LBPNet on face identification, the well-known Face Recognition Technology (FERET) [15] dataset was used. This dataset contains controlled images of $1,196$ individuals, sorted in one gallery set and 4 probe sets: (i) Fb set, taken in the same condition but with different facial expressions, (ii) Fc set, taken in different light conditions, (iii) Dup-I set, taken between one minute and 1031 days after the gallery set, (iv) Dup-II set, a subset of Dup-I taken at least 18 months later.

The network was trained on a Linux node (16 CPU cores and 32GB of RAM) for 4 hours, and the classification was finished in less than a minute. The centre region of image was cropped in a size of $150 \times 130$ to exclude most background region. The images were also preprocessed following the suggested steps by Tan *et al.* [16]. Three different LBP

operators, $\{LBP^{u2}_{2,8}, LBP^{u2}_{3,8}, LBP^{u2}_{4,8}\}$, and two different filter sizes, $z = \{11, 12\}$, were used in the first layer. For the second layer, the filter size was 110, sampling stride in the filter and stride of the filter were $s_1 = z$, $s_2 = 10$, respectively, the starting point of sampling was at $i, j \in \{1, z/2\}$ to uniformly sample all features in the filter region, and the first 2000 dimensions of PCA were selected to retain most of the discrimination information.

As the results of the LBPNet and other stat-of-the-art methods shown in Table 1, LBPNet achieved comparable (in Fb), same good (in Fc) or better (in Dup-I and Dup-II) results.

**Table 1**. Comparison using the FERET probe sets.

| Methods | Fb | Fc | Dup-I | Dup-II |
|---|---|---|---|---|
| LBP [11] | 0.93 | 0.51 | 0.61 | 0.50 |
| LBP Template [9] | 0.989 | 0.928 | 0.760 | 0.634 |
| LGBPWP [17] | 0.981 | 0.989 | 0.838 | 0.816 |
| G-LQP [13] | 0.999 | 1.00 | 0.932 | 0.910 |
| **LBPNet** | **0.996** | **1.00** | **0.942** | **0.936** |

### 4.2. Experiment on Face Verification: LFW

The *de-facto* evaluation benchmark, Labeled Faces in the Wild [18, 19] dataset, was used to evaluate proposed framework for face verification. LFW contains 13,233 images of faces of 5,749 individuals taken in unconstrained conditions. The View 2 of the dataset was selected, which contains 6000 pairs of faces, and the evaluated system confirms or rejects two faces having the same identity. The standard 10-fold cross validation was implemented as specified in [18].

The experiment on LFW was conducted under unsupervised protocol. The model was trained without prior label information or external data. The LFW-a version of the dataset was used, in which images are aligned by commercial software. The centre area of $170 \times 100$ of image was cropped. Three different LBP operators, $\{LBP^{u2}_{1,8}, LBP^{u2}_{2,8}, LBP^{u2}_{3,8}\}$, and seven different filter sizes, $z = \{10, 12, 14, 16, 18, 20, 22\}$ were used in the LBP layer; in the PCA filter layer, the filter size was 80, and the first 500 dimensions of PCA were selected; the rest of the parameters were set the same as in the FEREST experiment.

The training and the classification were finished in 200 hours and 10 minutes, respectively. Comparison of our results with other baselines and state-of-the-art results are reported in Table 2, and the corresponding ROC curves are shown in Figure 3. The AUC of LBPNet is 0.9404 and is ranks the third among all. However, this value is extremely close to the first (0.9428) and the second (0.9405) bests.

**Table 2**. Comparison using the LFW dataset view 2 under unsupervised protocol

| Methods | AUC |
|---|---|
| SD-MATCHES [20] | 0.5407 |
| H-XS-40 [20] | 0.7547 |
| GJD-BC-100 [20] | 0.7392 |
| LARK [21] | 0.7830 |
| MRF-MLBP [22] | 0.8994 |
| Pose Adaptive Filter [23] | 0.9405 |
| Spartans [24] | 0.9428 |
| **LBPNet** | **0.9404** |

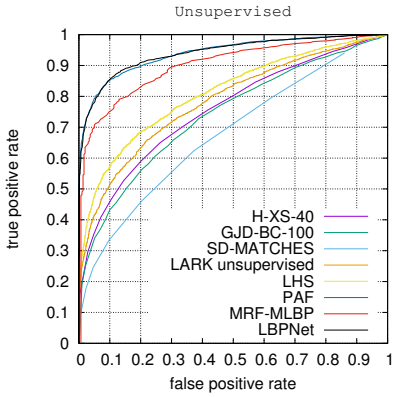


**Fig. 3**. ROC curves by various methods using LFW dataset view 2 under unsupervised protocol

## 5. CONCLUSION

In this paper, a novel tool for face recognition named Local Binary Pattern Network (LBPNet) was proposed. The key ideas in LBPNet were inspired by the successful LBP descriptor and the deep learning architecture. LBPNet retains a similar topology of CNN while avoiding the costly model learning on massive data by replacing its convolutional kernels with off-the-shelf computer vision descriptors.

Extensive experiments were conducted using two public benchmarks (i.e., FERET and LFW) to evaluate the proposed LBPNet. The results showed that LBPNet achieved promising performance in these benchmarks compared with other state-of-the-art methods.

## 6. REFERENCES

[1] Li Deng and Dong Yu, "Deep learning: methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[2] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document

recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, "Deep fisher networks for large-scale image classification," in *Advances in neural information processing systems*, 2013, pp. 163–171.

[4] Zhenyao Zhu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning identity-preserving face space," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 113–120.

[5] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3025–3032.

[6] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher Vector Faces in the Wild," in *British Machine Vision Conference*, 2013.

[7] Chi-Ho Chan, Josef Kittler, and Kieron Messer, *Multi-scale local binary pattern histograms for face recognition*, Springer, 2007.

[8] Liang Chen and Naoyuki Tokuda, "A unified framework for improving the accuracy of all holistic face identification algorithms," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 107–122, 2010.

[9] Liang Chen, Ling Yan, Yonghuai Liu, Lixin Gao, and Xiaoqin Zhang, "Displacement template with divide-&-conquer algorithm for significantly improving descriptor based face recognition approaches," in *Computer Vision–ECCV 2012*, pp. 214–227. Springer, 2012.

[10] Zhimin Cao, Qi Yin, Xiaoou Tang, and Jian Sun, "Face recognition with learning-based descriptor," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2707–2714.

[11] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen, "Face recognition with local binary patterns," *Computer vision-eccv 2004*, pp. 469–481, 2004.

[12] Hieu V Nguyen and Li Bai, "Cosine similarity metric learning for face verification," in *Computer Vision–ACCV 2010*, pp. 709–720. Springer, 2011.

[13] Sibt Ul Hussain, Thibault Napoléon, and Fréderic Jurie, "Face recognition using local quantized patterns," in *British Machive Vision Conference*, 2012, pp. 11–pages.

[14] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," *arXiv preprint arXiv:1503.03832*, 2015.

[15] P Jonathon Phillips, Hyeonjoon Moon, Syed Rizvi, Patrick J Rauss, et al., "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[16] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[17] Hieu V Nguyen, Li Bai, and Linlin Shen, "Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person," in *Advances in Biometrics*, pp. 269–278. Springer, 2009.

[18] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[19] Gary B. Huang Erik Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Tech. Rep. UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.

[20] Javier Ruiz-del Solar, Rodrigo Verschae, and Mauricio Correa, "Recognition of faces in unconstrained environments: a comparative study," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1, 2009.

[21] Hae Jong Seo and Peyman Milanfar, "Face verification using the lark representation," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1275–1286, 2011.

[22] Shervin Rahimzadeh Arashloo and Josef Kittler, "Efficient processing of mrfs for unconstrained-pose face recognition," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[23] Dong Yi, Zhen Lei, and Stan Z Li, "Towards pose robust face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3539–3545.

[24] Felix Juefei-Xu, Khoa Luu, and Marios Savvides, "Spartans: Single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios," 2015.