Real-Time Image Classification Using LBP and Ensembles of ELM

Stevica Cvetković, Boban Rajković, Saša V. Nikolić

Abstract: This paper presents a method for real-time image classification which combines Local Binary Pattern (LBP) descriptor and ensembles of Extreme Learning Machines (ELM). We start by extraction of a multi-channel LBP descriptor over separate color channels and several scales of the input image. The descriptor is characterized by compactness and robustness to illumination and resolution changes. Image category prediction is done using recently introduced specific single layer neural network called Extreme Learning Machine (ELM). To overcome main disadvantages of the ELM, instability and non-optimal output, we combined multiple ELMs into an ensemble using appropriate score aggregation strategy. Our evaluation on a standard benchmark dataset consisting of a thousand images from ten categories, has shown high accuracy of results while executing in real-time during tests.

Keywords: image classification, local binary patterns, neural networks, extreme learning machines

1 Introduction

Automatic image classification based on visual content is an important problem in computer vision. The goal of an image classification system is to assign a category with the most similar visual content to the given input image. In general, there are two main steps of an image classification system. The first step is to define a robust and compact image description (feature) which includes sufficient information for the forthcoming classification step. Visual similarity between images is measured using extracted descriptors, where the choice of the descriptor essentially affects the overall performance of the classification system. There is a wide set of visual descriptors available in the literature [14, 3]. As one of the most successful, Local Binary Pattern (LBP) descriptor has gain large popularity due to: robustness to resolution and lighting changes, low computational complexity, and compact representation [10, 11, 15]. In this work we will further improve its robustness by including color and multi-resolution information. Concretely we will extend LBP by extracting it over

separate color channels and multiple image scales, what will led to improved classification results. The second phase of the system is to predict the category of the input image using appropriate classifier. It is the crucial to choose an appropriate machine learning technique to be applied for classification of descriptors. Support Vector Machine (SVM) is the most widely used machine learning technique for image classification purposes [8, 2]. Despite its popularity it was found that SVM has to be combined with nonlinear kernels to achieve highly accurate results [8]. However, nonlinear SVM introduces significant computational costs during training and testing steps, what makes this approach difficult to apply for real-time applications. As an alternative to the commonly used SVM, we will investigate application of Extreme Learning Machines (ELM) [7, 6, 5, 12] for image classification. ELM is an emerging machine learning technique applied to single hidden layer feed-forward neural networks (SLFN). It overcomes an important drawback of traditional neural networks - their slow learning speed. It increases training speed by randomly assigning weights and biases to the hidden layer, instead of iteratively adjusting its parameters by gradient-based methods. As well as minimizing training error, ELM finds smallest norm of output weights and hence have better generalization capabilities then gradient based training algorithms, such as backpropagation. Furthermore, it can naturally handle multi-class classification problem using the architecture of multiple output nodes equal to the number of classes. However, random assignment of parameters introduces suboptimal input weights and biases into hidden layer that may result in instable and non-optimal output. A natural way to overcome this drawback is to combine several ELM-s into an ensemble. The main advantage of ensembles is that combined outputs from many diverse learners can increase generalization capabilities of a single model used in ensemble. For aggregating multiple ELMs into a single classification system, the most common method is voting based aggregation [6, 9]. Recently we introduced Average Score Aggregation [4] which shows improved classification accuracy without additional complexity. In this paper we will apply Average Score Aggregation to the ELM ensembles for image classification. In the rest of the paper we first describe the process of extended LBP descriptor extraction using multiple color channels and several image resolutions. Then we give an overview of ELM algorithm followed by description of ELM ensembles for multi-class image classification. Finally, experimental evaluation and conclusion are presented.

2 Local Binary Patterns - LBP

Local Binary Pattern (LBP) is a popular visual descriptor that captures local appearance around a pixel. LBP descriptor of the complete image is then formed as a histogram of quantized LBP values computed for every pixel of the image. It was introduced in [11] for the texture classification problem, and extended to general neighborhood sizes and rotation invariance in [10]. Since then, LBP has been extended and applied to variety of applications [15]. For a given image I, the local LBP descriptor centered at pixel $I(x, y)$ is an array of 8 bits, with one bit encoding each of the pixels in the 3×3 neighborhood (Figure 1). Each

neighbor bit is set to 0 or 1, depending on whether the intensity of the corresponding pixel is greater than the intensity of the central pixel. To form a binary array, neighbors are scanned in anti-clockwise order, starting from the one to the right at position $I(x+1,y)$.

178	230	229
129	**215**	197
212	250	212

0	1	1
0		0
0	1	0

01100010

a)   b)   c)

Fig. 1. Example of a LBP extraction process for the central pixel of intensity 215: a) pixel intensities, b) thresholded difference, c) LBP.

When using 3×3 neighborhood, there are 256 possible basic LBP patterns. This can be further reduced into a smaller number of 58 rotation-invariant patterns, by using an extension from [10]. The extension was inspired by the fact that some binary patterns occur more frequently than others. To describe the complete image, the quantized LBP patterns of every image pixel are aggregated into a histogram. The image could be additionally divided into blocks, with a histogram computed for every block and concatenated to form the final descriptor. We used one image block per image, i.e. a global histogram of LBP codes computed for the complete image. In order to form the final multi-channel LBP descriptor which exploits color information of the image, we extracted LBP descriptor over all color channels. Since default RGB color space shows large correlation among color channels, conventional color spaces such as $YCbCr$ could be more effective for image classification. Therefore we propose to extract a LBP histogram for every color channel of the $YCbCr$ color space. To include image details at multiple scales, we extracted LBP histograms over the original image and several times resized image. Resizing is done to the half width and height of the original image using bicubic interpolation method. Multi-resolutional descriptor is formed by concatenation of the previously extracted descriptors at several scales. In the concrete implementation we used 3 color channels at 3 scales, forming a final $3 \times 3 \times 58 = 522$ dimensional LBP image descriptor.

3   Ensembles of Extreme Learning Machines - ELM

Let us define N training examples as (x_j, y_j), where $x_j = [x_{j1}, x_{j2}, ..., x_{jn}]^T \in \mathbf{R}^n$ denotes j-th training instance of dimension n and $y_j = [y_{j1}, y_{j2}, ..., y_{jm}]^T \in \mathbf{R}^n$ represents j-th training label of dimension m, where m is the number of classes. LBP image descriptor from the previous section will further be denoted as x_j, while y_j will denote m dimensional vector of binary class labels with value 1 denoting membership to the class. Single hidden layer feed-forward neural networks (SLFN) with activation function $g(x)$ and L hidden neurons could be defined as:

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{f}_j \, , j = 1, ..., N \tag{1}$$

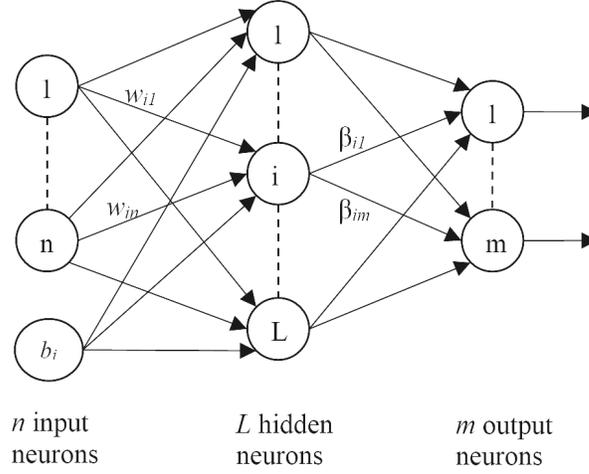n input neurons L hidden neurons m output neurons

Fig. 2. Structure of an ELM network.

where $wi = [w_{i1}, w_{i2}, ..., w_{in}]^T$ denotes the vector of weights which connects the ith hidden neuron and all input neurons, $i = [_i1, _i2, ..., _im]^T$ is the weight vector which connects i-th hidden neuron and all output neurons, and b_i is the bias of the i-th hidden neuron. By ELM theory [7], w_i and b_i can be assigned in advance randomly and independently, without apriori knowledge of the input data. The ELM network structure is presented in Figure 2.

SLFN in (1) should satisfy $\sum_{i=1}^{L} \|\mathbf{f}_i - \mathbf{y}_i\| = 0$, i.e. there exist β_i, $\mathbf{w}_i$ and b_i such that:

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{y}_j, j = 1, ..., N \tag{2}$$

If we denote as $\mathbf{H}$ a hidden layer output matrix of the ELM, then the i-th column of $\mathbf{H}$ represents the i-th hidden neurons output vector regard to inputs $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$:

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & ... & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ ... & ... & ... \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & ... & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix} \tag{3}$$

and

$$\beta = \begin{bmatrix} \beta_1^T \\ . \\ . \\ . \\ \beta_L^T \end{bmatrix} \tag{4}$$

and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ . \\ . \\ . \\ . \\ \mathbf{y}_N^T \end{bmatrix} \tag{5}$$

Then the equivalent matrix form of (2) can be represented as:

$$\mathbf{H}\beta = \mathbf{Y} \tag{6}$$

The output weights are then computed by finding the unique smallest norm least-squares solution of the linear system (5) as:

$$\beta = \mathbf{H}^\dagger \mathbf{Y} \tag{7}$$

where $\mathbf{H}^\dagger$ represent the Moore-Penrose generalized inverse of the $\mathbf{H}$.

3.1 Ensemble of ELMs (E-ELM) using Average Score Aggregation

Ensembles of ELMs are introduced to overcome known drawbacks of ELM caused by the randomness of input weights and biases [1, 9]. In order to improve accuracy of an ensemble, we proposed to directly accumulate output scores of the ELMs into a final decision [4], instead of accumulating their binary votes. The proposed algorithm, denoted as Average Score Aggregation (ASA), is described below.

Given a number of classes m, the sigmoid activation function $g(x)$, the hidden neuron number L, and the number of ELMs in ensemble Q (where every independent ELM use the same L and $g(x)$):

Algorithm 1 E-ELM training phase:

Input: Training set $S = \{(\mathbf{x}_j, \mathbf{y}_j) | \mathbf{x}_j \in \mathbf{R}^n, \mathbf{y}_j \in \mathbf{R}^m, j = 1, ..., N\}$ with N instances of n-dimensional descriptors.
for $k = 1$ to k **do**
    Train ELM(k) and compute β_k using the complete $\mathbf{S}$, according to (7).
end for

---

4 Experimental Evaluation

To test the proposed method for image classification, we used publicly available Corel1000 dataset [13]. It consists of 1000 images classified into following 10 categories: Africa people, Beach, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food.

Algorithm 2 E-ELM testing phase:

Input: Test set $S' = \{\mathbf{x}_j | \mathbf{x}_j \in \mathbf{R}^n, j = 1, ..., N'\}$ with N' instances of n-dimensional descriptors.

for $k = 1$ to Q **do**

    Obtain output matrix $Y_{test}(k) = H_{test}(k) \times \beta_k$ of dimensionality $N' \times m$.

end for

Sum up all Q output matrices $Y_{test} = \sum_{k=1}^{Q} \mathbf{Y}_{test}(k)$.

For every test instance (i.e. Y_{test} row), compute the class label as the index of the maximal value in the row.

The dataset is characterized by large variation of visual content inside a category, what makes this dataset close to the real world image classification scenario.

For the tests, we implemented a method in MATLAB and used it to measure the classification accuracy and speed. Final LBP image descriptor is formed by concatenation of LBP histograms of the three color channels at 3 scales, resulting in a 522 dimensional descriptor. To achieve correctness of results, tests were repeated for 50 times over random partitions per category, where we randomly selected 50 images for training and other 50 for testing. Classification accuracy is measured by varying number of hidden neurons per ELM for a fixed number of ELMs in ensemble (k = 15). We used ASA strategy [4] as aggregation method. Results are presented in Figure 3. It can be observed that noticeable accuracy improvements are achieved up to 1000 neurons, afterwards improvements are negligible. Increasing the number of neurons constantly improves classification results at the cost of increased training/testing time. Top achieved results are near to 89% for k=15 ELMs in an ensemble. For practical applications, one should experiment with values larger than 1000 neurons.
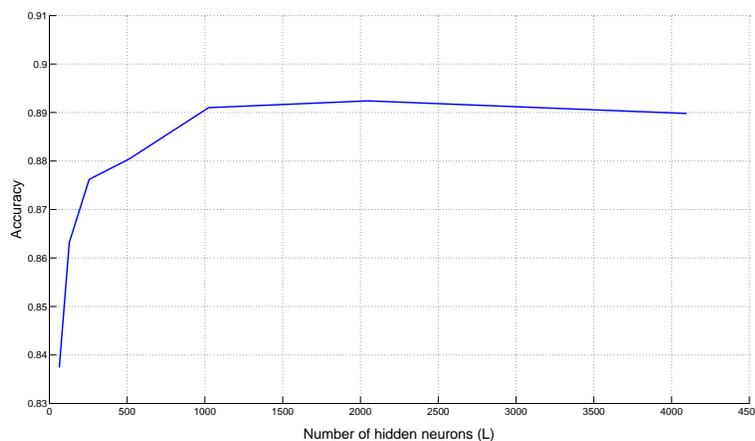

Fig. 3. Classification accuracy depending on the number of hidden neurons in ELM, for a fixed number of ELMs in ensemble (k = 15).

We further measured average training and testing time on an Intel Core i7 3.5GHz computer. Results are presented in Figure 4 and Figure 5. Training time when using 1000 neurons is about 1 second (Figure 4), while test image classification is done in less than a second. These results demonstrate high performances of the method in terms of training and testing speed on this dataset.
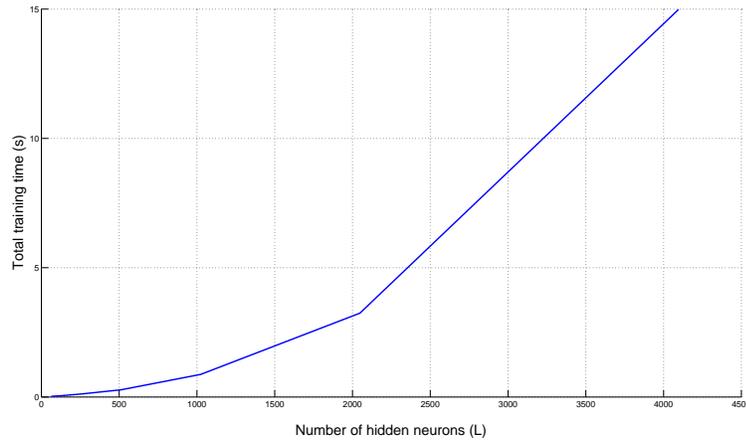

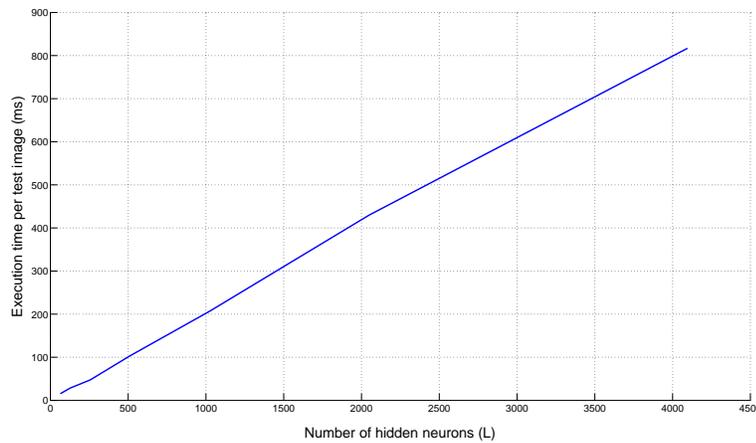Fig. 4. Total training time of an ELM ensemble for all 500 training images.

Fig. 5. Average test time per image (in milliseconds).

In order to compare results of the ELM with other common classification techniques, we measured accuracy of the Linear SVM and kernelized RBF SVM [6], on the same dataset. Linear SMV accuracy was 87.81%, while RBF SVM reached 88.8%. It can be noted that ELM outperforms Linear SVM in terms of accuracy, having the similar algorithm complexity. On the other side, ELM reaches results comparable to the kernelized SVM,

while operating significantly faster during the training.

5 Conclusion

The paper presents results of our research in the field of automatic image classification. Colorized multi-scale LBP descriptor is combined with fast and powerful ELM ensemble classifier. Achieved classification accuracy of 89% is satisfied result for rapid image classification. It can be concluded that combination of LBP descriptor with ELM classifiers is a reasonable choice for image classification applications. ELM classifier could be used as an alternative to the commonly used SVM. Our plan is to investigate performance of other types of image descriptors combined with ELM classifier, particularly fusion of color and texture descriptors.

References

[1] Jiuwen Cao, Zhiping Lin, Guang-Bin Huang, and Nan Liu. Voting based extreme learning machine. *Information Sciences*, 185(1):66 – 77, 2012.

[2] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.

[3] Stevica Cvetkovic, Sasa V. Nikolic, and Slobodan Ilic. Effective combining of color and texture descriptors for indoor-outdoor image classification. *Facta Universitatis Series: Electronics and Energetics*, 27(3):399 – 410, 2014.

[4] Stevica Cvetkovic, Milos B. Stojanovic, and Sasa V. Nikolic. Multi-channel descriptors and ensemble of extreme learning machines for classification of remote sensing images. *Signal Processing: Image Communication*, 39, Part A:111 – 120, 2015.

[5] Gao Huang, Guang-Bin Huang, Shiji Song, and Keyou You. Trends in extreme learning machines: A review. *Neural Networks*, 61:32 – 48, 2015.

[6] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. Extreme learning machine for regression and multiclass classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(2):513–529, April 2012.

[7] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489 – 501, 2006. Neural Networks Selected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04) 7th Brazilian Symposium on Neural Networks.

[8] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178, 2006.

[9] Nan Liu and Han Wang. Ensemble based extreme learning machine. *Signal Processing Letters, IEEE*, 17(8):754–757, Aug 2010.

[10] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, July 2002.

[11] Timo Ojala, Matti Pietikinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.

[12] J. Tang, C. Deng, and G.-B. Huang. Extreme learning machine for multilayer perceptron. *Neural Networks and Learning Systems, IEEE Transactions on*, PP(99):1–1, 2015.

[13] J.Z. Wang, Jia Li, and G. Wiederhold. Simplicity: semantics-sensitive integrated matching for picture libraries. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(9):947–963, Sep 2001.

[14] Xin Zhang, Yee-Hong Yang, Zhiguang Han, Hui Wang, and Chao Gao. Object class detection: A survey. *ACM Comput. Surv.*, 46(1):10:1–10:53, July 2013.

[15] Yang Zhao. Theories and applications of lbp: A survey. In De-Shuang Huang, Yong Gan, Phalguni Gupta, and M.Michael Gromiha, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, volume 6839 of *Lecture Notes in Computer Science*, pages 112–120. Springer Berlin Heidelberg, 2012.