International Journal of Engineering & Technology, 7 (4.39) (2018) 655-659

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Object Recognition with Improved Features Extracted from Deep Convolution Networks

K Manohar[1*], S Irfan[2], K Sravani[3]

[1,2]Department of Electrical and Computer Engineering, Wollo University, Ethiopia
[3]Department of Computer Science and Engineering, Khammam Institute of Technology and science
**E-Mail: manohar646@gmail.com*

Abstract

Object identification is a method for identify an exact object in an image. Object identification algorithm depends on matching, learning, and pattern recognition using appearance based feature technique. Object recognition has various method to detect objects it's include feature extraction and machine learning methods, deep learning search as CNN. The deep learning convolution neural network (CNN) has been proved to be very efficient in feature extraction. CNN is compressed of one or more convolution layers and then follow by one or more totally connected layers. In Image types, the work with classifiers aims at explore the most appropriate types for high level deep features. The feature extracted from the image plays a significant role in image type. It is the process of retrieving the main data from the raw data, its finding the set of parameters that recognize the object uniquely. In feature extraction every nature is represent by a feature vector it's become identity. In this work the features extracted from CNN applied as input to train machine learning classifiers and perform image classification. A systematic comparison between various classifiers is made for object recognition.

Keywords: *Feature extraction, neural network, object recognition, Machine learning*

1. Introduction

The first intention of trying to "understand the scene" is one of the base ideas in computer vision that lead to a continuous increase in the need to apprehend the high-level context in images regarding object recognition and image classification. By becoming a fundamental visual expertise that Computer Vision systems require, the field has rapidly grown. Images have become ubiquitous in a variety of fields as so many people and systems extract vast amounts of information from imagery. Information that can be vital in areas such as robotics, hospitals, self-driving cars, surveillance or building 3D representations of objects. While each of the above-mentioned applications differs by numerous factors, they share the common process of correctly annotating an image with one or a probability of labels that correlates to a series of classes or categories. This procedure is known as image classification and, combined with machine learning, it has suit an main do research area in the field, and on account of the focus on the understanding of what an image is representative of. The complex process of identifying the type of materials in diverse tasks linked to image-based scene perspectives has taken advantage of the combination of machine learning techniques applied to the up-to-date development of neural networks. This outlines the challenging problem of material classification due to the variety of the definite features of materials.

The state of the-art solutions rely massively on the attention that Computer Vision systems have received, which led to a series of algorithms being developed and images being collected in datasets. People are able to recognize the environment they are in as well as the various objects in their everyday life no matter the influence on the item's features or if their view is obstructed, as this is one of the very first skills we learn from the moment we are born. Computers, on the other hand, require effort and powerful computation and complex algorithms to attempt to recognize correctly patterns and regions where a possible object might be. Object detection and recognition are two main ways that have been implemented over multiple decades that are at the centre of Computer Vision systems at the moment. These approaches are presented with challenges such as scale, occlusion, view point, illumination or background clutter, all issues that have been attempted as research topics that provided functionality that led to the introduction of Neural Networks and Convolution Neural Networks (CNN). The newly added functionality is composed of distinct types of layers that consist of many parameters that are able to figure out the features present in a given image. These architectures have since been built on and a more complex structure with hidden non-linear layers between the input and output layers of a CNN has been identified as Deep Convolution Neural Network (DCNN). In Computer Vision systems, datasets are divided into two main categories: a training dataset used for training the algorithm learn to perform its desired task and a testing dataset that the algorithm is tested.

Extraction of discriminative features starting with information pictures may be a standout amongst those A large portion testing errands previously, object distinguishment frameworks. Considerably exert need pointed in deciding ideal characteristic sets to a particular task, dependent upon the qualities of Questions should a chance to be perceived Also classifiers will make utilized. A large number for these offers generated all the altogether guaranteeing outcomes. However, because of those vagueness Furthermore absence of all task-independent guidelines for ideal characteristic selection, the methodology from claiming information arrangement need been as of late overwhelmed Eventually Tom's perusing Different methodologies utilizing neural networks. That significant point of these neural system methodologies may be that throughout the preparation transform those organize self-determines the ideal situated about features from those information. Those hindrance is that extensive preparation information sets might make obliged Also In this way those preparation transform Might make exceptionally long. Neural networks have been indicated will furnish fantastic execution on numerous picture orders.

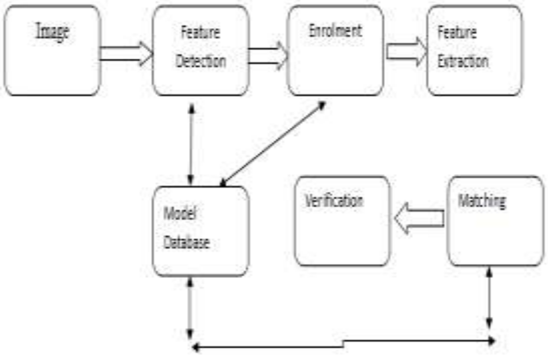


Figure 1: Components of object recognition

2. Object Recognition Algorithms

In this chapter, famous algorithms from the most promising approaches are demonstrated. Short descriptions and the general function of SIFT and SURF, which are examples for the feature-based approach, PCA and LDA, which are appearance-based methods, and convolution neural networks are displayed.

2.1 SIFT - Scale-Invariant Feature Transform

Those scale-invariant feature transform (SIFT) may be an algorithm over workstation dream on recognize and more portray nearby offers over pictures. That calculation might have been protected for patented in Canada by the perusing those school for British Columbia and distributed eventually Tom's perusing David Lowe over 1999. Filter way focuses about Questions are principal concentrated from a set from claiming reference pictures Furthermore put away On a database. An article may be distinguished to another picture Eventually Tom's perusing separately contrasting each characteristic from those new picture should this database What's more finding hopeful matching Characteristics In light of Euclidean separation about their characteristic vectors. From those full situated from claiming matches, subsets for magic focuses that concur on the article and its location, scale, Furthermore introduction in the new picture need aid distinguished should channel out handy matches. Those determination of reliable groups is performed quickly Eventually Tom's perusing utilizing an proficient hash table execution of the summed up Hough convert. Each bunch of 3 alternately All the more Characteristics that consent on an article and its pose is afterward liable on further nitty gritty model confirmation and hence outliers would dispose of. At last the likelihood that An specific situated about Characteristics demonstrates the vicinity about a item may be computed, provided for those exactness from claiming fit What's more amount for possible false matches. Object matches that every last bit these tests can be identified as right with more certainty.

2.2 SURF - Speeded-Up Robust Features

The SURF technique will be a quick Also strong calculation for local, similitude invariant representational Also correlation about pictures. Also with a significant number different neighborhood descriptor built approaches, investment focuses of a provided for image is characterized as notable features starting with a scale-invariant representational. Such a multiple-scale dissection will be Gave toward that convolution of the introductory picture for discrete kernels during a few scales (box filters). The second step comprises to fabricating introduction invariant descriptors; Eventually Tom's perusing utilizing nearby gradient facts (intensity and orientation). The principle investment of the surf methodology lies on its quick calculation for operators utilizing box filters, along these lines empowering ongoing requisition for example, following and more object distinguished.

2.3 PCA - Principle Component Analysis

Principle component analysis (PCA) is a procedure of statistics which reduces the dimensionality of a collection of observed data. In general, PCA is a orthogonal linear transformation which transforms data to a new coordinate system. The first coordinate equals to the direction of the greatest variance of the data, this is called first principle components. The second principle component equals to the second greatest variance and so on.

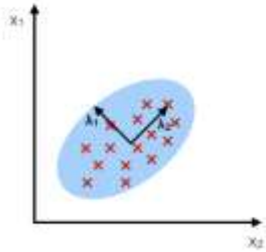


Figure .2: Collection of data with two principle components λ1 and λ2.

In Fig. 1.3, the two axis λ1 and λ2 represent the two principle components of the observed data, where λ2 is the first principle component. In PCA for object recognition mostly the first principle component is needed as it still contains most information about the data. The data is projected on the first principle component in order to maintain maximum variance.

$$w(1) = argmax_{||w||=1}\left\{||Xw||^2\right\} = argmax\{\frac{w^T X^T w}{w^T w}\} \qquad (1)$$

The weights of the first principle component can be calculated using Eq. 1, which is in matrix format, where X is a m ∗ n matrix containing m n-dimensional samples.PCA is a popular technique for pattern and object recognition, however, it is not suitable for classification as separation does not work well. In the next section, an alternative method is proposed.

2.4 LDA - Linear Discriminator Analysis

Another famous appearance-based algorithm is called linear discriminate analysis (LDA). It is a method used in statistics for dimensionality reduction or classification. Each class is represented by mean μi and the same covariance Σ. The algorithm minimizes the intra-class variance Σ, while the inter-class variance Σb is maximized. With C as the number of classes and μ as the mean of the class means one obtains:

$$\sum b = \frac{1}{C}\sum_{i=1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T \qquad (1)$$

the class separation S in direction #»w is calculated:

$$S = \frac{\sum b\overrightarrow{w}^{T}\overrightarrow{w}}{\overrightarrow{w}^{T}\sum\overrightarrow{w}} \qquad (2)$$

A projection is good, if it separates classes well like it is shown in Figure 1.3 The projection on the y-axis yields a bad projection as the two classes are not separable anymore, while a projection in x-axis direction provides a good result.

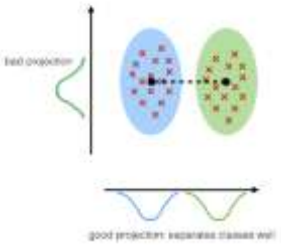


Figure.3: a projection can lead to good or bad class separation.

As means and covariance are not always known, different approaches have to be applied to still be able to use LDA. Maximum likelihood estimation or maximum a posterior estimation can help.

2.6 Characterization of General Object Recognition Strategies

Appearance-based Method

The concept of the appearance-based object recognition strategy is presented. Appearance-based methods are popular for face or hand-writing recognition. For this strategy, a set of reference training images, which are highly correlated, is needed. For example, 100 images of faces and a set of images containing background or random objects. This dataset is compressed using dimensionality reduction techniques to obtain a lower dimension subspace, also called eigen space. Parts of the new input images are projected on the eigen space and then correspondence is examined. More details about appearance-based methods is when describing the two famous algorithms PCA and LDA.

Feature-based Method

The next strategy is called feature-based, because algorithms recognize objects based on specific features. Features are supposed to be characteristic for each object; often one object is not only described by one attribute but multiple features. Colors, contour lines, geometric forms or edges (gradient of pixel intensities) are popular choices. As already mentioned, object features can have many faces, but simplified spoken, they all can be divided in just two categories. Features and their descriptors can be either found considering the whole image (global feature) or after observing just small parts of the image (local feature).An histogram of the pixel intensity or color are simple examples for global features. It is not always reasonable to compare the whole image, as already slight changes in illumination, position (occlusion) or rotation lead to significant differences and a correct recognition is not possible anymore. [GL] Descriptors of local features are more robust against these problems and therefore algorithms with local features often outperform global feature-based methods. In general concept of local feature-based algorithms can be seen. Two small patches are compared and not the whole image, these patches may be rotated and normalized first in order to achieve higher accordance. This approach has lead to much progress on research in object recognition.

Pattern Matching

Methods of pattern matching, or sometimes called template matching, are often used because of their simplicity. \The computation is quite easy: In above equation, the squared differences between an image patch I and a template M are summed pixel wise. A threshold has to be provided in order to let the algorithm decide whether a template matched and an object was recognized.

$$r(x,y) = \sum_{i\in M}\sum_{j\in M}(I(x+i, y+j) - M(i,j))^2 \qquad (1)$$

The result can be adjusted in order to stabilize against small distortion and light changes. Here n is the number of pixels.

$$r = \frac{\sum IM - \sum I * \sum M}{\sqrt{(n\sum I^2 - (\sum I)^2)(n\sum M^2 - (\sum M)^2)}} \qquad (2)$$

One famous application of template matching is traffic sign recognition, small parts of the input image are tried to be matched with a database full of different images of traffic signs. As this approach has lots of disadvantages such as problems with occlusion, rotation, and scaling, different illuminations and so on, it will not be given further attention in this work.

Artificial Neural Networks

This networks models motivated by biological neural network, such a model consists of several layers, as it can be seen in below fig, in which each layer is composed of a certain number of neurons. An input and an output layer is the minimum amount of layers a network can have, but normally hidden layer are included to be able to learn more complex things such as object recognition.

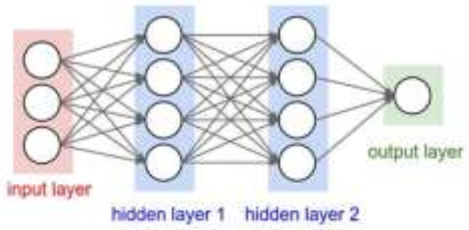


Figure.4: Neural Networks

All neurons from one layer are connected to all neurons from the next layer and therefore create a huge network with millions of parameters. All of these connections have a weight which is updated during learning phase. Neurons are activated

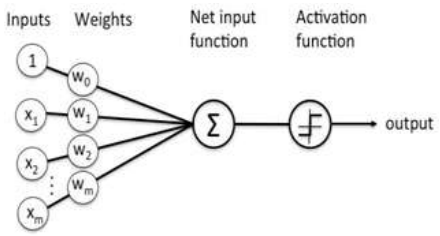


Figure .4.1: A neural network containing one input layer, two hidden layer and one output layer.

A neural network containing one input layer, two hidden layer and one output layer. There are different types of networks such as feed-forward, recurrent with different number and types of hidden layers, while the input (e.g. number of pixels) and output (number of classes) layer are fixed. Later, convolution neural networks and their hidden layers. New inputs go through the same way, some neurons might be activated based on the trained network and finally, this leads to the most suitable classification.

Deep learning

Deep learning will be and only state-of-the-symbolization frameworks over Different disciplines, especially workstation dream automatic speech recognition (ASR). Comes about ahead regularly utilized assessment sets for example, such that TIMIT (ASR) and MNIST (image classification), and additionally a reach for large-vocabulary discourse different assignments have relentlessly enhanced. Convolution neural networks (CNNs) were superseded to ASR, to LSTM yet all they are a greater amount effective On PC dream. Developments clinched alongside equipment enabled those replenished investment. On 2009, Nvidia might have been included over the thing that might have been known as those “big bang” for profound learning, as deep-learning neural networks were prepared for Nvidia graphics transforming units (GPUs). That year, Google cerebrum utilized Nvidia GPUs to make skilled DNNs. same time there, ng confirmed that GPUs Might increment those pace from claiming deep-learning frameworks toward over 100 times. Over particular, GPUs are well-suited to the matrix/vector math included on machine learning.

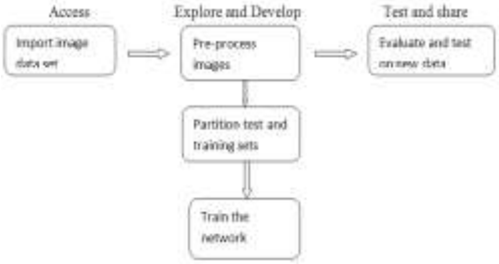


Figure 4.2: Deep learning workflow

Features extracted from CNN applied as input to train machine learning classifier. Systematic comparison between various classifiers is made for object recognition. we will present a study of Nearest Neighbor (NN).

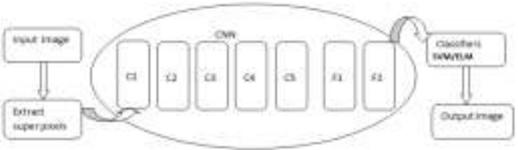


Figure 4.3: Block diagram

Convolution Neural Networks (CNN) are composed of several inter-connected hidden layers which process and transforms inputs to outputs. They are inspired from the biological structure of the visual cortex (the part of the brain responsible for sight). Convolution layers map inputs to certain neurons in different regions Convolution layers account for non-linearity in the model. Pooling layers down sample the data to reduce the number of inputs to the next layer. The final layer is fully connected, mapping to each input into single output.

Extreme Learning Machine

Extreme learning machine (ELM)as a sort from claiming summed up single-hidden layer feed-forward networks (SLFNs) need exhibited its useful generalization execution for amazing quick Taking in speed done a number benchmark What's more true requisitions. These paper further investigations the execution of elm and its variants clinched alongside object distinguishment utilizing two diverse characteristic extraction routines. The principal technique extracts composition features, power features from histogram Also features starting with two sorts from claiming shade space: HSV & RGB. Those second system extracts state offers In light of radon change. The order exhibitions from claiming elm Also its variants need aid compared with those execution from claiming support vector Machines (SVMs). As checked Eventually Tom's perusing Recreation results, elm accomplishes better testing precision for a significant part preparing time once greater part instances over SVM for both characteristic extraction strategies.

Given a dataset X=[x1,x2,⋯,xN]∈ℜ× of N samples with label T =[t1,t2,⋯,tn]∈ℜ×, where d is the dimension of sample and c is the number of classes. Note that if (i=1,⋯n,) belongs to the k-th class, the k-th position of (i=1,⋯n,) is set as 1, and -1 otherwise. The hidden layer output matrix H with L hidden neurons can be computed as

$$H = \begin{bmatrix} h(w_1^T x_1 + b_1) & h(w_2^T x_1 + b_2) \dots & h(w_L^T x_1 + b_L) \\ \vdots & \vdots & \vdots \\ h(w_1^T x_N + b_1) & h(w_2^T x_N + b_2) \dots & h(w_L^T x_N + b_L) \end{bmatrix}$$

where $h(\cdot)$ is the activation function of hidden layer, = [w1,⋯wl,]€R and =[b1,⋯bl,]∈ℜ $\min_{\beta \epsilon R^{L\times c}} \frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{i=1}^{N}\|\varepsilon_i\|^2$

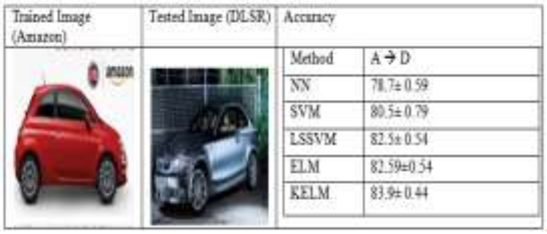
$$s.t.\, h(x_i)\beta = t_i^T - \varepsilon_i^T, i = 1, \dots, N \leftrightarrow HB = T^t - \varepsilon^T$$

Where $\Omega \epsilon R^{NXN}$ it denotes the output weights between hidden layer & output layer

3. Results

Shown the below table Compressions ELM and SVM performance

Table.1: Accuracy Results



Trained Image (Amazon)	Tested Image (DLSR)	Accuracy	
		Method	A → D
		NN	78.7± 0.59
		SVM	80.5± 0.79
		LSSVM	82.5± 0.54
		ELM	82.59±0.54
		KELM	83.9± 0.44

This experimental analysis performs the ELMs output from SVMs in cross domain recognition. ELMs output forms SVM based various different settings. The kernel ELM shown below table.

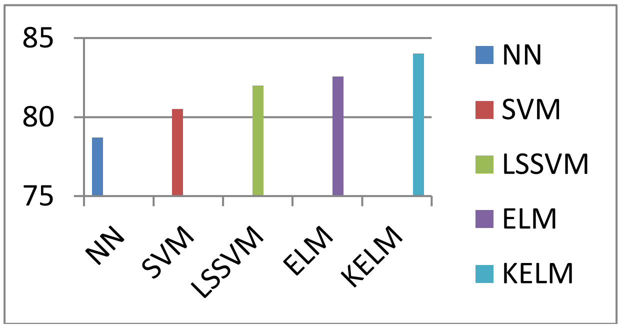


Figure.5: Experimental Results Chart

Comparative Study of Object Recognition methods:
The Comparative Study of Various methods and its accuracy rate, Time efficiency and user rate are given below

Table.2: Comparative Study of Object Recognition methods

Methods	**Accuracy Rate**	**Time Efficiency**	**User Rate**
Deep Learning	High (85%)	High	80%
Affine scale invariant feature transform (ASIFT)	Moderate (76%)		60%
Background subtraction	Moderate	Moderate	40%
Optical Flow	Moderate	High	20%
Frame Differencing	High	Low to Moderate	30%

Figure.6: Experimental Simulation

4. Conclusion

By using object recognition technique easily identify the object a present image or a video. There are a number of techniques and methods are applied for the required result. As human can easily identify the object and recognize, but machine it is not an easy task, hence the work can be done with the help of the model of Artificial intelligence. Convolution neural net will be a prevalent profound taking in techno babble to current visual distinguishment assignments. Similar to every last bit profound taking in techniques, CNN is thick, as reliant on those extent What's more personal satisfaction of the preparing information. Provided for a great ready dataset, CNNs need aid fit of surpassing people during visual distinguishment errands. However, they are at present not hearty on visual artifacts for example, such that glare Furthermore noise, which people have the capacity with adapt.

References

[1] G.B. Huang, H. Lee, and E. Learned-Miller, “Learning hierarchical representations for face verification with convolutional deep belief networks,” in Proc. IEEE Int’l Computer Vision and Pattern Recognition, pp. 2518-2525, 2012.

[2] A. Krizhevsky, I. Sutskever, G.E. Hinton, “Image Net classification with deep convolutional neural networks,” NIPS, 2012

[3] A. Karpathy , G. Toderici, S. Shetty , and T. Leung, “Large-Scale Video Classification with Convolutional Neural Networks,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 1725-1732, 2014.

[4] A.S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 512-519, 2014.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Accurate Object Detection and Semantic Segmentation,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 580-587, 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” arXiv: 1406.4729.

[7] Y. Sun, X. Wang, and X. Tang, “Hybrid Deep Learning for Face Verification,” in Proc. IEEE Int’l Conf. Computer Vision, 2013.

[8] Y. Taigman, M. Yang, M.A. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in Proc. IEEE Int’l Computer Vision and Pattern Recognition, 2014.

[9] E. Zhou, Z. Cao, and Q. Yin, “Naïve-Deep Face Recognition: Touching the Limit of LFW Benchmark or Not?” arXiv: 1501.04690, 2015.

[10] A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” NIPS, 2012.

[11] D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 3642-3649, 2012.

[12] A. Karpathy, G. Toderici, S. Shetty, and T. Leung, “Large-Scale Video Classification with Convolutional Neural Networks,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 1725-1732, 2014.

[13] A.S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 512-519, 2014.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Accurate Object Detection and Semantic Segmentation,” in Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, pp. 580-587, 2014. [15] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” arXiv: 1406.4729.

[15] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, “What is the best multi-stage architecture for object recognition?” ICCV, pp. 2146-2153, 2009.

[16] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, “DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition,” arXiv: 1310.1531, 2013.

[17] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Information Theory, vol. 13, no. 1, pp. 21-27, 1967.

[18] V. Vapnik, “Statistical learning theory,” John Wiley: New York, 1998.

[19] J.A.K. Suykens and J. Vandewalle, “Least Squares Support Vector Machine Classifiers,” Neural Processing Letters, vol. 9, no. 3, pp. 293-300, 1999.