INSPECTING THE EFFICIENCY OF DEEP NEURAL NETWORKS FOR CLASSIFICATION OF IMAGES WITH NOISE

[1]SHAIFY KANSAL, [2]SHIVANI GOEL, [3]JHILIK BHATTACHARYA

[1]Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India-147004
[2]Department of Computer Science and Engineering, School of Engineering and Applied Sciences, Bennett University, Greater Noida, India-201310
E-mail: shaify.kansal@thapar.edu

Abstract - Noise is the undesirable data in digital images which can produce undesirable impacts like obscured objects. It becomes necessary to apply filters for denoising the images. There are many deep learning techniques for image classification. The study in this paper aims to challenge the performance of deep learning in the presence of noise in the images. The experiments are carried out on thepopular dataset- Caltech 101. The images are corrupted using Gaussian noise. Denoising is performed using combination of Median and Gaussian filters. Feature extraction methods considered are edge-detection, wavelet transformation, their concatenation, LBP and Gabor filter. Classifiers such as kNearest Neighbours, Support Vector Machine, Neural Network and Deep neural network with autoencoders are used. The experimental outcomes demonstrate that though noise can upset the execution of image mining task in terms of accuracy,deep neural networks with autoencoders has reported higher accuracy of the classification of images with noise as well as without noise.

Index Terms - Deep neural networks, Support Vector Machine, Noise,Image classification

I. INTRODUCTION

Millions of images are captured and stored with the frequent use of digital cameras and smartphones. The image retrieval applications such as online shopping, biometrics and crime prevention persuade the need of efficient retrieval of images. It is difficult to extract patterns from a large dataset of images and retrieve only the useful images. Browsing the database manually to hunt for identical images in the database is not a feasible approach. Content based image retrieval(CBIR) is used for getting significant images from a vast image database.

The main steps in CBIR include image retrieval and classification. Noise can cause damage to an image and cause difficulty in retrieval. So, the image recovery is done with respect to the content i.e. the numerical qualities of the pixels of an image. In general, the execution of content based image recovery depends on features, feature extraction procedures, closeness measures and the range of database [1]. In this paper, a study has been proposed to check the effects of noise on image classification using various machine learning classifiers. DNN has shown the least variation in accuracy of classification in the presence of noise.

The rest of the paper is structured as follows: literature work is reviewed in Section 2. The flow of proposed work is given in section 3. In Section 4, the methodology followed is discussed. The experimental results are presented and discussed in Section 5. The conclusions are drawn in last section.

II. RELATED WORK

Early phases of the image classification pipeline i.e. acquisition and pre-processing of images can specifically impact classification performance [2]. Authors using different techniques for feature extraction and image classification may or may not consider presence of noise in images. A content-based information retrieval method presented by Vijendran and Kumar[3]proved that discrete wavelet transform (DWT) followed by histogram of obedient gradient approach is very accurate for feature extraction. Iterative DWT and sparse representation was used efficiently in a content-based image recovery framework [4]. Noise was not considered in these frameworks. In a framework for imbalanced classification proposed by He et al. [5], the dataset taken was noisy and a fusion model K-nearest Gaussian was found to be the best for classification with an accuracy of 82%. Teoh and Ibrahim [6] presented a survey on frameworks using median filters for reducing impulse noise from grayscale digital images. According to the authors, there were many versions of median filter like adaptive median filter, median filter with fuzzy logic, directional median filter, switching median filter along with standard, weighted, iterative and recursive median filters. These were found to be producing best results when used in combination. Various classifiers used successfully are SVM, NN, DNN. Thing et al. [7] presented a framework for classification of noisy logo used in fraud detection. SVM has been used as multi class classifier while principal component analysis (PCA) has been used as binary classifier. An edge based heuristic method is used for extracting the features. The success rate is found to be different for different datasets ranging from 81% to 86%.As deep learning is in use recently, Zhang et al. [8] proposed an image classification model and utilized the DCNN for preparing and integrating the image features into a conclusion to-end learning structure in person-re-

identification in surveillance without considering the noise. DCNN based component learning has been displayed by Xu et al. [9] to group the Epithelial (EP) and stromal (ST) locales from digitized tumor Tissue Microarrays. The correlation of DCNN based models have been finished with three handcraft feature extraction based strategies: color feature based approach, LBP and contrast measures. The outcomes demonstrated that the DCNN based models always outpaced handmade components based models. The noise present in case of medical images was not considered. Xiao et al. [10] trained CNN with supervised datasets for various computer vision problems for some clean labels and many noisy labelsto efficiently classify the noisy labels(78%).

Many researchers have used CNN on Caltech101 dataset. With 1-NN and Gaussian processes, an accuracy upto 96% is achieved by Kapoor et al. in 2010[11].Donahue et al. (2014) have obtained an accuracy of 86.7% when SVM is used with CNN[12]. Kamarudin et al. [13]used NaïveBayesUpdateable classifier, Random Tree, Bayes Net and IBk an extension of nearest neighbour algorithm on Caltech101 dataset images 12. The highest accuracy of 99% was provided by random tree with feature selection Correlation-based Feature Selection SubsetEvaluator. Wang et al. [14]used CNN with gnostic fields for object recognition and achieved accuracyupto 52%.Bărar et al. [15]obtained an accuracy of 77% with CNN and Aspect ratio Classifier (AR) as compared to SVM and AR during image recognition.

III. PROPOSED WORK

Presence of noise in an image affects the efficiency of all these steps during image mining. The proposed work here has two objectives: First is to study the effectiveness of filters for denoising the images. Second is to verify whether the presence of noise in image affects the accuracy of classification by deep learning.

IV. METHODOLOGY

In this section, the steps followed for implementing the proposed work are discussed. The image classification results are obtained from the Caltech-101 datasetthat is subject to denoisingand feature extraction. Different types of noise, denoising filters and feature extraction techniques are elaborated in sections A, B and C respectively. The classifiers used in framework are explained in section D.

A. Corrupting the image

Adigital image is not a genuine projection of a real domain into a two-dimensional plane. The function of a pre-processing filter is to expel the noise to the degree that it is conceivable without losing the significant data from the image. Scientifically this can be communicated as described in (1):

$$A(x, y) = \hat{A}(x, y) + \Phi(x, y) \quad (1)$$

where Φ is assumed to be the noise component. A(x,y) and $\hat{A}$(x,y) are the original and noisy images respectively. In this research work, Gaussian noise is added to the images. Gaussian noise is brought about by normal sources.For example, warm vibration of atoms and noise that occurs in warm objects due to radiation of their discrete nature. It is additive in nature and follows Gaussian distribution (2):

$$P(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

where z {\displaystyle z}zz represents the grey level, σ{\displaystyle \sigma } denotes the standard deviation and μ is {\displaystyle \mu }the mean value.

B. Denoising

To remove these types of noises, different filters and their combinations can be applied[16]. These are discussed in the following section.

Mean filter

It substitutes each center value in a window or local neighbourhood by the average (mean) of all the pixel values of that window. Mean filter can be demonstrated by (3):

$$h(i,j) = \frac{1}{M} \sum_{(k,l) \in N} f(k,l) \quad (3)$$

where, M is the number of pixels in the surrounding N, h(i,j) and f(k,l) are the old and new image pixels respectively.

Median filter

It substitutes each center value in a n x n window by the median of all the pixel values of that window. This filter can be best defined by (4):

$$z(i, j) = \text{median}(z_p \mid p = 1,\ldots, n \times n) \quad (4)$$

where z_p for p = 1,…, n x n are the pixel values in the neighbourhood centered on (i, j).

C. Feature Extraction

The feature-vectors of RGB intensity images for Wang and Caltech 101 datasets are obtained through edge-detection, wavelet transformation, concatenation of edge-detection and wavelet transformation, LBP and Gabor Filter methods. The edge–detection and wavelet transformation have been used widely because most of the density of the images are around the edges of the images. Feature extraction methods (FEM) used in this paper are:

a) Kirsch operator on grayscale images of the datasets.
b) Kirsch operator on RGB images of the datasets.

c) Single level discrete 2-dimensional wavelength transform, 'dwt2' on grayscale images of the datasets.
d) Single level discrete 2-dimensional wavelength transform, 'dwt2' wavelet transformation on RGB images of the datasets.
e) Single level discrete 2-dimensional wavelength transform, 'dwt2' wavelet transformation on all the three components of RGB images of the datasets.

D. The concatenation of methods (B) and (E).

The probability histogram value is then calculated for each resultant image for above mentioned transformations. The probability histogram is then divided into 10 bins and for each bin the statistical features such as standard deviation, skewness and Kurtosis are calculated. The resultant feature vector is obtained by combining the statistical features of each image.

a) In case of LBP feature extraction method, each RGB image of the datasets is converted into grayscale image, which is then subdivided into 16 images. On each of the 16 subdivided images, 'dwt2' wavelet transformation is applied. Finally, LBP is then applied on the lowpass filtered image obtained after wavelet transformation.
b) The Gabor filter is applied on the 'dwt2' wavelet grayscale image of the datasets.

E. Feature Classification

The classification methods used for the experimental work are kNN, SVM, NN and DNN with autoencoders. NN comprise of three layers namely- input layer, hidden layer and output layer. Supposea$_1$,a$_2$, a$_3$, a$_4$…..a$_k$ are the inputs to the neuron. These inputs are the feature vectors. w_1, w_2, w_3, w_4…...w_k are the weights on each input. Output of the neuron is calculated using the function,

$$a = f (\sum X_i W_j)+b$$

f is known as activation function whereas b is the bias.Various activation functions can be used for NN such as Sigmoid, Tanh, ReLU, Leaky ReLu and Maxout[17]. Maxout has the advantage that for every single neuron, it doubles the number of parameters which leads to a hike in total number of parameters.In this research work, Maxout activation function with 20 numbers of neurons has been used. When number of hidden layers is more in a NN, the results are more precise and accurate[18]. Layers are built with nodes. In this research work, DNN with autoencoders is usedwhich is aNNusing an unsupervised learning algorithm and backpropagation method.It is used to set the target values to be equal to the inputs to the network.

V. EXPERIMENTAL RESULTS

The experiment is performed by applying all the above methods on two datasets.

A. Datasets and experimentation details

The proposed method is applied on 9,146 images of Caltech 101 dataset [19] of size 384 × 256. The images of the datasets have been changed to size 256 × 256, in JPG format. For Caltech101 dataset, 101 classes are there comprising of different number of images for each class. The images in the same category are considered as similar images. The experiments are carried out on the image dataset comprising of original database images(ODI), filtered original images(FOI), noisy images(NI) and filtered noisy images(FNI) using statistical features content for the sake of comparison. Various classifiers such as kNN, SVM,NN and DNNusing autoencodersare evaluated on the basis of accuracy for image classification. The experiments are carried out on Intel® Core™ i7-5500U CPU, 2.40 GHz processor with 12.0 GB RAM using MATLAB R2015a.

B. Indexing Results of Feature Extraction

Various feature extraction methods used are A-H as described in section 4(C) and classifiers used are as described in section 4(D). These are used onODI, FOI, NI and FNI to predict the accuracy. Gaussian Noise is added to the dataset images and the NI and ODI are filtered with the combination of Median and Gaussian filters. The experimental results for different methods stated above are mentioned in Table 1. From the above results, it is found that concatenation of Kirsch operator on RGB images and 'Db2' wavelet on all the three components of RGB images is the best feature extraction method and DNN comes out to be the most accurate classifier. The graphical representation is shown in Fig. 1.

DISCUSSION

In this paper, the effectiveness of feature extraction methods along with image classifiers like kNN, SVM, NN and DNN with autoencoders is studied. Different methods are thoroughly evaluated on popular Caltech 101 dataset [20]. The images are corrupted with Gaussian Noise and filtered with the combination of Median and Gaussian filters. The experimental results presented that concatenation of Kirsch operator and Db2 wavelet on three different components of RGB intensity images with DNN is more accurate and superior method for effective image classification than the other state-of–the-art image classification methods. The degradation due to noise in image classification is least in DNN i.e. from 99% to 95% only. The process of denoising of images by application of median and Gaussian filters increase the accuracy of image classification procedure which is again highest in case of DNN, i.e.

from 91% to 97% using 'dwt2' wavelet transformation on RGB images.

CONCLUSION

From the experimental results, it is found that DNN with autoencoders shows more accurate results than the other classifiers (upto 99%) for images with noise and without noise. The presence of noise lowers the accuracy for image classification as the noise makes it tough for the classifier to separate the classes. As it is clear from the results, the accuracy of DNN is reduced by only 4% with noisy images which is the lowest among all the other classifiers used.As a future direction, use of wavelet based denoising techniques for denoising the images can be tested.

REFERENCES

[1] J.K. Dash, S. Mukhopadhyay, and R.D. Gupta,"Content-based image retrieval using fuzzy classmembership and rules based on classifier confidence",IET Image Process., Vol. 9, no. 9, pp. 836-848. 2015.

[2] M. Ponti, T.S. Nazaré, and G.S.Thumé,"Image quantization as a dimensionality reduction procedure in color and texture feature extraction", Neurocomputing,Vol. 173, pp. 385–396. 2016.

[3] A.S. Vijendran, and S.V. Kumar,"A new content based image retrieval system by HOG of wavelet sub bands", Int. J. of Engg. Res. & Techno., Vol. 3, no. 3, pp. 1476-1480.2014.

[4] S. Mohamadzadeh, and H. Farsi, "Content-based image retrieval system via sparse representation", IET Comput. Vis., Vol. 10, no. 1, pp. 95-102. 2016.

[5] M. He, J.D. Weir, T. Wu, "K Nearest Gaussian-A model fusion based framework for imbalanced classification with noisy dataset", Artificial Intelli. Res., Vol. 4, (2), pp. 126-135.2015.

[6] S.H. Teoh, andH. Ibrahim, "Median Filtering Frameworks for Reducing Impulse Noise from Grayscale Digital Images: A Literature Survey", Int. J. of Future Comput. andCommun., Vol. 1, no. 4, pp. 323-326. 2012.

[7] V.L.L. Thing, W. Lim, J. Zeng, "Design of a Modular Framework for Noisy Logo Classification in Fraud Detection", Proc. Int. Conf. on Security Tech., pp. 53-64. 2011.

[8] R. Zhang, L. Ling, R. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification", IEEE Trans. Image Process., Vol. 24, (12), pp. 4766-4779. 2015.

[9] Xu, X. Luo, G. Wang, H .Gilmore, A. Madabhushi., "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images", Neurocomputing, Vol. 191, pp. 214-223. 2016.

[10] T. Xiao,T. Xia,Y. Yang,"Learning from Massive Noisy Labeled Data for Image Classification", Proc. IEEE Int. Conf. Comput. Vision and Pattern Recog., Boston, pp. 2691-2699. June 2015.

[11] Kapoor, K. Grauman, R.Urtasun,"Gaussian Processes for Object Categorization", Int. J. Comput. Vis.,Vol. 88, no. 2, pp. 169 – 188. 2010.

[12] J.Donahue, Y. Jiaet al.,"DeCAF: a deep convolutional activation feature for generic visual recognition", Proc. 31st Int. Conf. on Machine Learning, Beijing, China, Vol. 32, pp. 647-655. Jun. 2014.

[13] N.S. Kamarudin, M. Makhtar, and S.A. Fadzli.,"Comparison of image classification techniques using Caltech 101 dataset", J. of Theoretical and Applied Infor. Tech., 2015, Vol. 71,no. 1, pp. 79-86. 2015.

[14] P. Wang, G. Cottrell, and C. Kanan, "Modeling the Object Recognition Pathway: A Deep Hierarchical Model Using Gnostic Fields", Proc. CogSci.,pp. 2601-2606.

[15] Bărar, V. Neagoe, and N. Sebe,"Image Recognition with Deep Learning Techniques", Recent Adv. in Image, Audio and Signal Proc., pp. 126- 132. 2017.

[16] M. Tayel, A.Gamal, and H. Shawky,"Denoising of Stego-images for different noise models", IEEE Int. Conf. Ad8vanc.Commun. Tech. (ICACT), Pyeongchang, South Korea, pp. 65-69. Jul. 2015.

[17] N.A. Ibraheem, andR.Z. Khan,"Vision Based Gesture Recognition Using Neural Networks Approaches: A Review", IJHCI Int. J. Hum. Comput.Interact.,Vol. 3, no. 1, pp. 1-14. 2012.

[18] Y. Guo, Y. Liu, A.Oerlemanset al.,"Deep learning for visual understanding: A review", Neurocomputing, Vol. 187, pp. 27-48. 2016.

[19] "Caltech 101- Computational Vision", https://www.vision.caltech.edu/Image_Datasets/Caltech101/, accessed 25 October 2017

[20] Y. Mistry, D.T. Ingole, and M.D. Ingole, "Content based image retrieval using hybrid features and various distance metric", J. Elect. Syst. and Informat.Tech., 2017.

Table I. Accuracy of classification (%) of feature extraction methods with classifiers for Caltech 101 dataset

Classifier	SVM				kNN				Neural Network				DCNN			
FEM	NI	ODI	FNI	FOI	NI	ODI	FNI	FOI	NI	ODI	FNI	FOI	NI	ODI	FNI	FOI
A	61	64	66	67	63	67	69	71	67	69	71	74	74	77	79	82
B	64	70	72	72	69	72	72	74	76	80	78	81	90	95	96	97
C	61	63	65	68	64	66	68	69	64	67	66	69	82	84	86	89
D	65	67	68	69	61	68	71	74	77	80	79	84	91	96	97	98
E	63	65	67	69	66	69	72	72	67	70	70	72	92	94	95	95
F	70	72	74	75	71	74	76	77	82	86	84	87	95	98	98	99
G	65	66	69	70	64	67	68	69	65	70	68	72	82	87	89	94
H	68	70	72	72	66	72	74	76	81	84	83	85	92	97	97	98

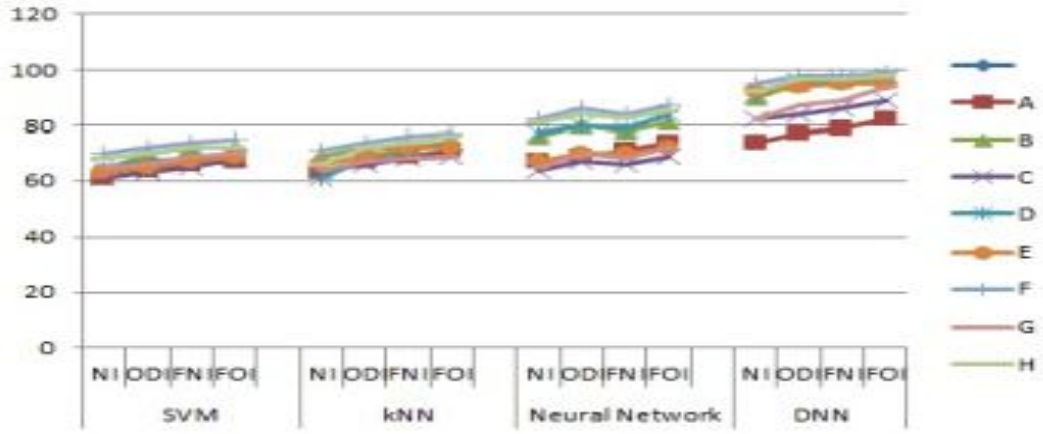


Fig.2.Accuracy of classification (%) for Caltech 101 dataset

★★★