Design of Objects Detection System using Deep Learning Algorithms

Taoufik Saidani[1,2] and Yahia Fahem Said[2]

[1]Department of Computer Sciences, Faculty of Computing & Information Technology, Northern Border University, Rafha, Saudi Arabia

[2]Laboratory of Electronics and Microelectronics (EµE), Faculty of Sciences of Monastir, University of Monastir, TUNISIA

Summary

Classification of objects in visual-images is a hot open research problem in the area of computer vision, especially in modern Internet of Things (IoT) systems. In this paper,a robust and reliable object taxonomy purposed framework is presented. The proposed model functionalize the research future targeted, deep convolutional neural network components to construct a Robust Object Recognition Network (RORNet). The RORNet consists of four convolutional layers, four Relu/Leaky Relu activation layers, three max-pooling layers and only two fully connected layers for extracting expected input image features. To speedup the training process, we used non-saturated neurons with a very efficient Graphics Processing Unit (GPU) coding for the convolution operation. To minimize the over-fitting issue in the fully-connected layers, we fictionalize a recently-developed regularization approach “dropout” with a dropping probability of 55%. The experimental and simulations results show that proposed RORNet framework has a high potential capability in the recognition of unseen images.

Key words:

Deep Learning; Object Recognition; Transfer Learning; RORNet; CIFAR10; CIFAR100; PASCIFAR.

1. Introduction

One of the basic goals of computer vision is to understand visual scenes. Understanding the scene includes many tasks such identifying what objects provide, settling objects in two and three dimensions, characterizing and marking objects, characterizing relationships between objects and providing an important description of the scene. Deep learning draw attention the researchers from different directions in image processing and computer vision. Due to continuous research in artificial intelligence (AI), in order to develop the techniques of AI the researchers are performing significant effort to find new methodologies to make the performance of AI more reliable [1–5]. For example, automatic system learning, without being explicitly programmed, automatically classifies and discloses data patterns for other purposes [6–8]. It has many applications and applications in various areas of security issues to detect diseases in the field of medicine and the computer vision community [9–12]. Deep learning is one of the important machine learning techniques, refers to deep linear groups of several simple layers or functions defined by variables. The structure, especially the composition of layers or functions, defines a boundary function with dozens of parameters that must be improved in order to reduce "loss" or objective function through some form of gradient gradient based on artificial neural networks (ANNs) [13–15]. In general, deep learning is the representation of multi-layered hierarchical data in the form of ANNs with more than two layers. It has the ability to create new features from a limited set of features in training data without human intervention [16,17].

ANN has been applied to classification, approximation, clustering and recognition issues in many applications. The performance of ANNs improved drastically when learned based on the principles of deep learning models [7,8,11,18]. Deep learning is the newborn generation of machine learning. The Convolutional Neural Network (CNN) is a very effective deep learning approach based on artificial neural networks that has attracted the attention of many scientists because of its similarity with the biological systems [19–24]. The Deep Convolutional Neural Network (DCNN) is a more effective technique than CNN and shows promising performance in visual image classification. It has many interconnected components and requires large datasets to train networks with a depth of zero. The common layers of the deep learning approach are as follows: data separation (training, test and validation sets), random sampling during training, loading and collection of image data samples, data augmentation, DCNN, rapid computational structure for optimization and inference, evaluation metrics performance during training and inference. In the literature, many DCNN models appeared, such as GoogleNet [22], AlexNet [23], ResNet [24], VGGNet [25] and others. However, most, if not all, computations are overheads because of the large number of layers in the training and the update weights in the network. In addition, they have difficulties in training the first few layers in the transfer of learning efficiently. This paper proposed the creation of a new robust network for identifying objects, called RORNet.

Manuscript received March 5, 2020
Manuscript revised March 20, 2020

2.Related Work

Small-sample UAV recognition requires an effective detecting and recognition method. When identifying a UAV target using the backward propagation (BP) neural network, fully connected neurons of BP neural network and the high-dimensional input features will generate too many weights for training, induce complex network structure, and poor recognition performance. In this paper, a novel recognition method based on non-negative matrix factorization (NMF) with sparseness constraint feature dimension reduction and BP neural network is proposed for the above difficulties. The Edge boxes are used for candidate regions and Log-Gabor features are extracted in candidate target regions. In order to avoid the complexity of the matrix operation with the high-dimensional Log-Gabor features, preprocessing for feature reduction by down sampling is adopted, which makes the NMF fast and the feature discriminative. The classifier is trained by neural network with the feature of dimension reduction [24]. Starting with LeNet-5 [25], convolutional neural networks (CNN) typically have a standard structure stacked convolutional layers followed by one or more fully-connected layers. Variants of this basic design are prevalent in image classification literature and have achieved the best results so far in MNIST and CIFAR, most notably the ImageNet classification challenge [15,16]. For larger datasets such as Imagenet, the last trend was to increase the number of layers and layer size [17], while using dropout [18] to address the problem of overfitting. Recently, a large number of state-of-the-art object detectors rely on the structure of DCNN to overcome the traditional CNN weakness such as the class imbalance and overfitting. The main contributions of this work are as follows: A very effective way to classify visual images was provided on a DCNN basis. Adapts to well-known DCNN concepts for robust RORNet construction. The proposed RORNet structure consists of three convolutional layers, four Relu activation layers, three max-pooling layers, and two fully connected layers to extract the filter input image features. In order to make the training phase more rapid, we used unsaturated neurons with a highly effective graphics processing unit (GPU) for the convolution process. The over-fitting problem in the fully connected layers is reduced by "dropout" with a 55% dropping probability. Experimental modeling simulation showed that the proposed model has high classification accuracy. The proposed RORNet framework is highly capable of identifying test images.

The paper is organized as follows: In the section 3, the proposed deep model named RORNet is presented. In the section 4, we discussed the materials and methods utilized. In section the experimental settings are presented, while the results are presented in the 6 section.

3.The Proposed Deep Model

Here, we propose an effective framework of deep learning to recognize visual images. The structure of the proposed topology contains Convolution, ReLU, Cross Channel Normalization, Pooling, Dropout of different blend connections.RORNet consists of 19 layers of convolutional, some of which are followed by max-poolingl layers, and the layers are fully-connected with the softmax finally two megabits. In order to speed up the training phase, we used unsaturated neurons and performed a highly effective GPU for the convolution process. To get rid of the over-fitting in the fully connected layers, we adapted the newly developed method of regularization called "dropout" which proved very effective. We use the Stratified Sampling Division (SSDiv [22]) To divide the data into separate subgroups of the design / test. Detailed steps for the proposed framework are announced in the algorithm 1:

Algorithm 1 Describes the pseudo code steps of proposed RORNet/CIFAR100 image classification framework.

Require: An image data-store $\mathcal{I} = \{I_1, I_2, \ldots, I_n\}$.
Ensure: A label set $\mathcal{L}$ for detected objects in each visual images $\mathcal{I}' \notin \mathcal{I}$.

Training phase:

1. SSDiv split $\mathcal{I}$ into Design$_{Set}$/Test$_{Set}$; $\mathcal{I} = (D_{Set} \bigcup T_{Set})$.
2. SSDiv split D_{Set} into Training/Validation; $D_{Set} = (Tr_{Set} \bigcup Val_{Set})$.
3. Train(RORNet,Tr_{Set}, Val_{Set}).
4. Stop training when predefined stopping criteria is satisfied.
5. **Return:** Trained RORNet.

Test phase:

1. Evaluate RORNet($\mathcal{I}'$) metrics listed in table 1.
2. Extract the 2-megabits of Test$_{Set}$ features $Test_{\mathcal{F}}$ of $\mathcal{I}'$ at FC$_2$ layer.
3. Train SVM, KNN, NB and DT classifiers Using $test\mathcal{F}$.
4. Compute T_{Set} classification evaluation measures listed in table 1.
5. **Stop.**

4.Materials and Methods

4.1 CIFAR100

The CIFAR-100 dataset consists of 100 classes with 600 images each. There are 500 training images and 100 test images per class. 100 classes were grouped into the CIFAR-100 in 20 super classes. Each image comes with a "fine" label (the category to which it belongs) and a "coarse" label (the superclass to which it belongs). When the class name is plural, the labels are directed not to reject images that show multiple instances of the object. The CIFAR-100 class structure is shown in [25]. Each image contains 3 RGB color channels and 32x32 pixel dimensions for the total size of each input 3 x 32 x 32 = 3072.

4.2 Object Classification

The object classification task requires binary labels to indicate whether the objects are in an image. Early data sets of this type contain images that contain a single object with blank backgrounds, such as COIL household objects [19].Caltech 101 [20] and Caltech 256 [21] marked the transition to more realistic object images retrieved from the Internet, with the number of object categories also increasing to 101 and 256 respectively. In this paper, the CIFAR-100 dataset is presented. Once the training set is obtained, a classification algorithm is applied to extract the knowledge base, which is necessary to make a decision in an unknown state. Based on knowledge, smart decision is taken as output and return to knowledge base at the same time, generalizing the way inspectors perform tasks. The hard part of computationally classification is to urge the classifier to determine the ideal values for any parameters. Classifiers can provide a simple answer to yes or no, and can also estimate the probability that an object belongs to each class of candidate.

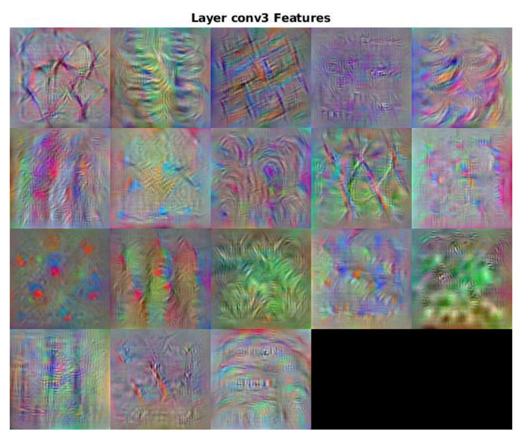

Fig. 1 Visualize the DNN activations of First Convolutional Layer at a specific selected channel, input image came from CIFAR100 database

5. Experiment settings

5.1 Features Classification

As an alternation for softmax rating class computed at layer 17, we extracted the train/test features of 2-megabits at the fully connected layer FC2. Extraction features are used to train a variety of basic classifiers such as K-Nearest Neighbor (K-NN),Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayesian (NB). We used the default Matlab implementation [22] for these classifiers.

5.2 Experimental Evaluation Metrics

To evaluate the proposed RORNet on the CIFAR100 dataset, the overall classification model is evaluated using common measures such as Sensitivity, Specificity, Accuracy, Precision, Recall, G-mean and F-Measure . Table 1 shows the set of evaluation metrics for the general performance of the proposed model. Accuracy of the recorded training phase coupled with satisfactory validation accuracy. A deep-dream visualization of images of the best channels learned by the RORNet of Blenheim dog breeds is presented at different hierarchical levels are shown in Figures 3,7.

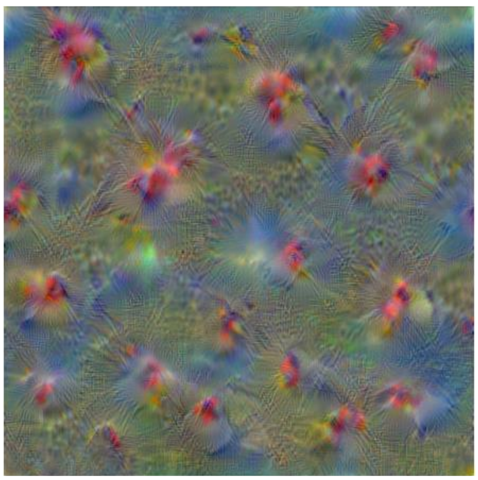

Fig. 2 Visualize All DNN activations of First Convolutional Layer, input image came from CIFAR100 database

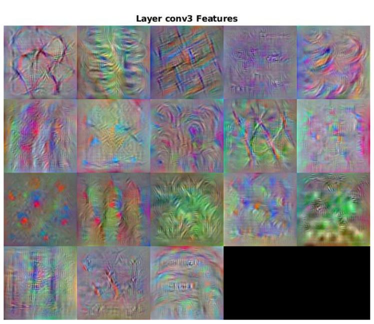

Fig. 3 Visualize the DNN activations of First Convolutional Layer at the strongest activation channel, input image came from CIFAR100 database

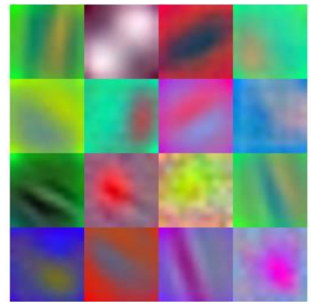

Fig. 4 Visualize All DNN activations of First Convolutional Layer, input image came from CIFAR100 database

Fig. 5 Visualize All DNN activations of First Convolutional Layer, input image came from CIFAR100 database

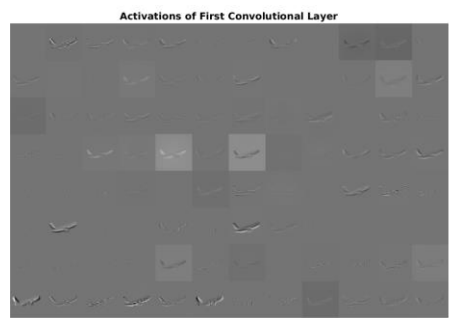

Fig. 6 Visualize All DNN activations of First Convolutional Layer, input image came from CIFAR100 database

Fig. 7 Visualize All DNN activations of First Convolutional Layer, input image came from CIFAR100

6. Results and Discussion

In this work we proposed an efficient Robust Object Recognition Network of category convolutional network. The proposed approach adapted Convolution, Cross-Channel, Max Pooling,Fully Connected, and Softmax as components of CNN procedures to classify the ImageNet dog breeds images. The trained RORNet performance is evaluated via a subset of evaluation metrics (see section 5.2). Apply deep network activation to extract application features on input layer images. Using 2Mbps extracted activation features, we are training a set of features classifiers (see section 5.1), using the training set then the classifiers were evaluated using the test set. Results reported in Table 1 shows evidence of the durability and reliability of the proposed model as a model with DNN layers. Our future expanded work on RORNet is to provide a full demonstration of the use of the proposed model in the full ImageNet database.

Table 1: Set of evaluation metrics for overall performance of the RORNet model

Classifier	RORNet	SVM	K-NN,(k=1)	Naïve Bayesian	Decision Tree
Accuracy	0.9814	0.9949	0.9786	0.9772	0.9246
Sensitivity	1	1	1	1	1
Specificity	0.9809	0.8784	0.9568	0.8700	0.8955
Precision	0.9870	0.9243	0.9012	0.9788	0.9489
Recall	1	1	1	1	1
F1-Measure	0.9538	0.9744	0.9234	0.9222	0.9533
G-Mean	0.9352	0.9501	0.9234	0.9788	0.9340

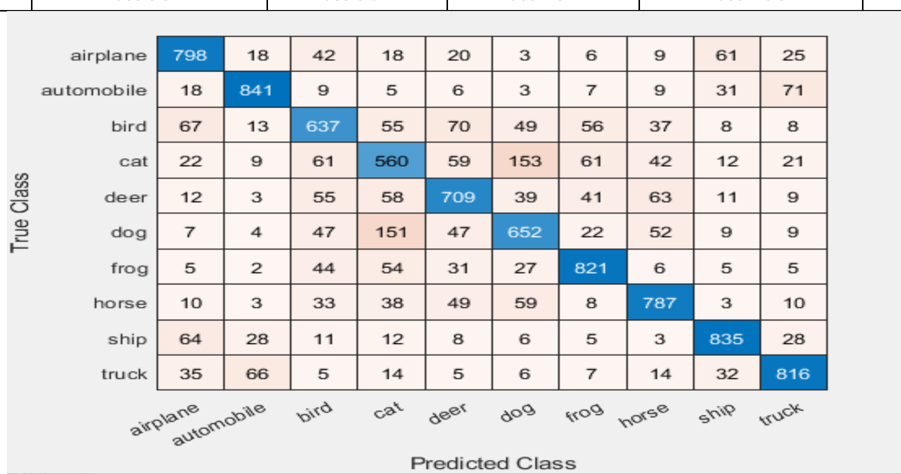

Fig. 8 Confusion Matrix of RORNet Network:

7. Conclusion

We have used RORNet network for the object detection we have tested our network for detection of stop sign from images ,Network got the accuracy around 98 % .Further images are tested ,results are as score value 1 ,score value values from 0 to 1 represents the confidence level of detection of particular object. We got approximate score value of test images near to 0.9 ,it means network is performing good.

Acknowledgments

The authors wish to acknowledge the approval and the support of this research study by the grant N_. CIT-2017-1-8-F-7432 from the Deanship of the Scientific Research in Northern Border University,Arar, KSA.

References

[1]. Huang, Y. Advances in Artificial Neural Networks Methodological Development and Application. Algorithms 2009, 2, 973–1007.

[2]. H. Ding, j. Wu, X.L. Evolving neural network using hybrid genetic algorithm and simulated annealing for rainfall runoff forecasting. Advances in Swarm Intelligence 2011, 7331, 444–451.

[3]. M. Zidan, A.-H. Abdel-Aty, M.E.s.M.F.Y.A.S.H.E.M.A.A. Quantum Classification Algorithm Based on Competitive Learning Neural Network and Entanglement Measure. Appl. Sci. 2019, 9, 1277.
[4]. El-Rahiem, B.A.; Ahmed, M.A.O.; Reyad, O.; El-Rahaman, H.A.; Amin, M.; El-Samie, F.A. An Efficient Deep Convolutional Neural Network for Visual Image Classification. The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019); Springer International Publishing: Cham, 2020; pp. 23–31.
[5].
[6]. Ahmed, M.A.O.; Didaci, L.; Lavi, B.; Fumera, G. Using Diversity for Classifier Ensemble Pruning: An Empirical Investigation. Theoretical and Applied Informatics 2018, 29, 25–39. doi:10.20904/291-2025.
[7]. Shiddieqy, H.A.; Hariadi, F.I.; Adiono, T. Implementation of deep-learning based image classification on single board computer. Electronics and Smart Devices (ISESD), 2017 International Symposium on. IEEE,2017, pp. 133–137.
[8]. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing 2018, 108, 33–47.
[9]. Gibson, E.; Li, W.; Sudre, C.; Fidon, L.; Shakir, D.I.; Wang, G.; Rosen, Z.E.; Gray, R.; Doel, T.; Hu, Y.; Whyntie, T.; Nachevc, P.; Modat, M.; Barratta, D.C.; Ourselin, S.; Cardoso, M.J.; Vercauteren, T. NiftyNet: a deep-learning platform for medical imaging. Computer Methods and Programs in Biomedicine 2018, 158, 113–122.
[10]. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors 2015, 2015.
[11]. El-Sayed, M.A.; Estaitia, Y.A.; Khafagy, M.A. Automated Edge Detection Using Convolutional Neural Network. International Journal of Advanced Computer Science and Applications 2013, 4, 10–20. doi:10.14569/IJACSA.2013.041003.
[12]. Lavi, B.; Ahmed, M.A.O. Interactive Fuzzy Cellular Automata for Fast Person Re-Identification. The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018); Springer International Publishing: Cham, 2018; pp. 147–157.
[13]. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. nature 2015, 521, 436.= Li, J.; Zhang, B.; Lu, G.; Zhang, D. Generative multi-view and multi-feature learning for classification. Information Fusion 2019, 45, 215–226.
[14]. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; others. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 2015, 115, 211–252.
[15]. Ahmed, M.A.O. Trained Neural Networks Ensembles Weight Connections Analysis. International Conference on Advanced Machine Learning Technologies and Applications. Springer, 2018, pp. 242–251.
[16]. Radovic, M.; Adarkwa, O.; Wang, Q. Object recognition in aerial images using convolutional neural networks. Journal of Imaging 2017, 3, 21.
[17]. Ren, Y.; Zhu, C.; Xiao, S. Small object detection in optical remote sensing images via modified faster R-CNN. Applied Sciences 2018, 8, 813.
[18]. C, S.; Jia Y, L.; et al.. Going deeper with convolutions. IEEE conference on computer vision and pattern recognition. Boston, USA. IEEE, 2015, pp. 1–9.
[19]. A, K.; I, S.; GE., H. ImageNet classification with deep convolutional neural networks. Advances in NeuralInformation Processing Systems. California, USA, 2012, pp. 1097–105.
[20]. K, H.; X, Z.; S, R.; et al.. Deep residual learning for image recognition. IEEE conference on computer vision and pattern recognition. Las Vegas, USA. IEEE, 2016, pp. 770–778.
[21]. K, S.; A., Z. Very deep convolutional networks for large-scale image recognition. Link: https://arxiv.org/pdf/1409.1556v6.pdf. 2015/2017.
[22]. Lei, S.; Zhang, B.; Wang, Y.; Dong, B.; Li, X.; Xiao, F. Object Recognition Using Non-Negative Matrix Factorization with Sparseness Constraint and Neural Network. Information 2019, 10. doi:10.3390/info10020037.
[23]. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.
[24]. 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.
[25]. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 2012. Version September 10, 2019 submitted to J. Imaging 9 of 9
[26]. Griffin, G.; Holub, A.; Perona, P. Caltech-256 object category dataset 2007. Statistics, M.; Toolbox, M.L. The mathworks 2018.

Taoufik Saidani received the M.S. and Ph.D degrees in Computer science and engineering from Faculty of Science of Monastir, Tunisia in 2007 and 2014, respectively. His major research interests include VLSI and embedded system in video and image compression. His current research interests include digital image processing, Artificial Intelligence and Deep Learning and hardware-software co-design for rapid prototyping in real time video processing.

Yahia Fahem Said received the M.S. and Ph.D degrees in Computer science and engineering from Faculty of Science of Monastir, Tunisia in 2007 and 2014, respectively. His major research interests include VLSI and embedded system in video and image compression. His current research interests include digital image processing, Artificial Intelligence and Deep Learning and hardware-software co-design for rapid prototyping in real time video processing.