ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

Survey of Object Detection using Deep Neural Networks

Mrs. Swetha M S[1], Ms. Veena M Shellikeri[2] , Mr. Muneshwara M S[3], Dr. Thungamani M[4]

Assistant Professor, Dept. of ISE, BMSIT & Management, Bangalore, India[1]

Student, Dept. of ISE, BMSIT & Management, Bangalore, India[2]

Assistant Professor, Dept. of CSE, BMSIT & Management, Bangalore, India[3]

Assistant Professor, Dept. of CSE, GKVK, Bangalore, India[4]

Abstract: Object detection using deep neural network especially convolution neural networks. Object detection was previously done using only conventional deep convolution neural network whereas using regional based convolution network [3] increases the accuracy and also decreases the time required to complete the program. The dataset used is PASCAL VOC 2012 which contains 20 labels. The dataset is very popular in image recognition, object detection and other image processing problems. Supervised learning is also possible in implementing the problem using Decision trees or more likely SVM. But neural network work best in image processing because they can handle images well.

Keywords: Object Detection; Neural Network, Artificial Neural Network (ANN), Feed-forward networks, Feedbacks networks

Fig 1: Object Detection using Deep Neural Network

1. INTRODUCTION

Object detection is detecting a specific object from an image of multiple and complex lines and shapes. Object detection is used in face detection, object tracking, image retrieval, automated parking systems [12]. The number of the applications is increasing in number. The main use of object detection is image classification or more precisely image retrieval. For understanding the convolution neural network, deep neural network is important. Papers in deep neural network are studied to understand the concepts of convolution neural network. NIPS paper on Regional based convolution neural network is also referred for further comparison [3]. Object detection is being used in various other fields like defense, architecture etc.

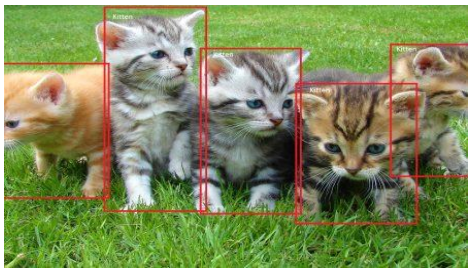


Fig 2: End to End object detection

1.1 Deep Neural Networks: A deep neural network is a neural networks with a certain level of complexity, a neural network with more than two layers. Deep neural networks use sophisticated mathematical modeling to process data in complex ways.A neural network is a technology built to simulate the activity of the human brain specifically pattern recognition and the passage of input through various layers of simulated neural connections. The phrases “deep learning “ is also used to describe these deep neural networks, as deep learning represents a specific form of machine learning

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

where technologies using aspects of artificial intelligence seek to classify and order information in ways that go beyond simple input/output protocols.

1.2 Artificial Neural Network: An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones.

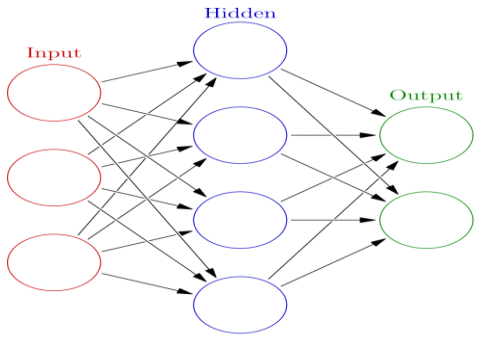


Fig 3: Artificial Neural Network

Types of Neural networks

1. Feed-forward networks: Feed-forward ANNs (figure 1) allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organisation is also referred to as bottom-up or top-down.

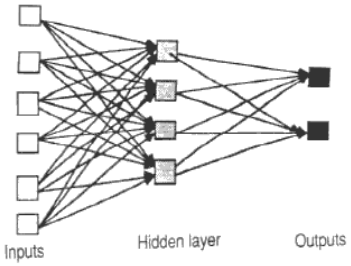


Fig 3: Top down approach of ANN

2. Feedback networks: Feedbacks networks can have signals travelling in Feed-forward ANNs allow signals to both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent, although the latter term is often used to denote feedback connections in single-layer organisations

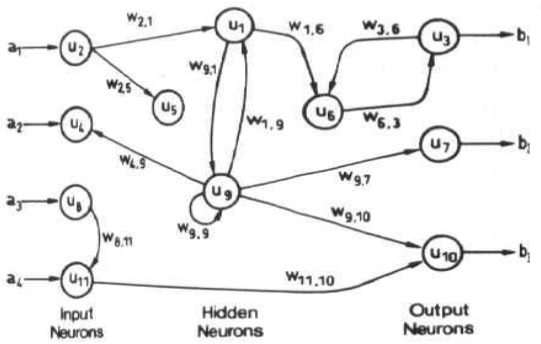


Fig 4: Equilibrium of ANN

3. Network layers: The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

- The activity of the input units represents the raw information that is fed into the network.
- The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.
- The behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

This simple type of network is interesting because the hidden units are free to construct their own representations of the input. The weights between the input and hidden units determine when each hidden unit is active, and so by modifying these weights, a hidden unit can choose what it represents.

1.4 Perceptrons

The most influential work on neural nets in the 60's went under the heading of 'perceptrons' a term coined by Frank Rosenblatt. The perceptron (figure 4.4) turns out to be an MCP model (neuron with weighted inputs) with some additional, fixed, pre-processing. Units labelled A1, A2, Aj , Ap are called association units and their task is to extract specific, localised featured from the input images. Perceptrons mimic the basic idea behind the mammalian visual system. They were mainly used in pattern recognition even though their capabilities extended a lot more.

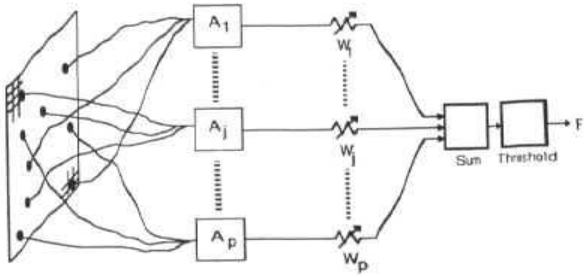


Fig 5: Perceptrons of ANN

2. LITERATURE SURVEY

1 .Focal Loss for Dense Object Detection

The highest accuracy object detectors to date are based on a two-stage approach popularized by R-CNN, where a classifier is applied to a sparse set of candidate object locations. In contrast, one-stage detectors that are applied over a regular, dense sampling of possible object locations have the potential to be faster and simpler, but have trailed the accuracy of two-stage detectors thus far. In this paper, we investigate why this is the case. We discover that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause. We propose to address this class imbalance by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. Our novel Focal Loss focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training. To evaluate the effectiveness of our loss, we design and train a simple dense detector we call Retina Net. Our results show that when trained with the focal loss, Retina Net is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors. This paper pushes the envelope further: we present a one stage object detector that, for the first time, matches the state-of-the-art COCO AP of more complex two-stage de-tectors, such as the Feature Pyramid Network (FPN).or Mask R-CNN [14] variants of Faster R-CNN [28]. To achieve this result, we identify class imbalance during training as the main obstacle impeding one-stage detector from achieving state-of-the-art accuracy and propose a new loss function that eliminates this barrier.

2. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World

The paper explores domain randomization, a simple but promising method for addressing the reality gap. Instead of training a model on a single simulated environment, we randomize the simulator to expose the model to a wide range of environments at training time. The purpose of this work is to test the following hypothesis: if the variability in simulation is significant enough, models trained in simulation will generalize to the real world with no additional training. This paper explores domain randomization, a simple technique for training models on simulated images that transfer to real images by randomizing rendering in the simulator. With enough variability in the simulator, the real world may appear to the model as just another variation. We focus on the task of object localization, which is a stepping stone to general robotic manipulation skills. We find that it is possible to train a real-world object detector that is accurate to 1:5 cm and robust to distracters and partial occlusions using only data from a simulator with non-realistic random textures. To demonstrate the capabilities of our detectors, we show they can be used to perform grasping in a cluttered environment. To our knowledge, this is the first successful transfer of a deep neural network trained only on simulated RGB images extend to any large number of transformations without specialized designs.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

3. Aggregated Residual Transformations for Deep Neural Networks

We present simple, highly modularized network architecture for image classification. Our network is constructed by repeating a building block that aggregates a set of transformations with the same topology. Our simple design results in a homogeneous, multi-branch architecture that has only a few hyper parameters to set. This strategy exposes a new dimension, which we call "cardinality" (the size of the set of transformations), as an essential factor in addition to the dimensions of depth and dataset, we empirically show that even under the restricted condition of maintaining complexity, increasing cardinality is able to improve classification accuracy. In this paper, we present a simple architecture which adopts VGG/ResNets' strategy of repeating layers, while exploiting the split-transform-merge strategy in an easy, extensible way. A module in our network performs a set of transformations, each on a low-dimensional whose outputs are aggregated by summation. We pursuit a simple realization of this idea — the transformations to be aggregated are all of the same topology. This design allows us to extend to any large number of transformations without specialized designs.

4. Application of Deep Learning in Object Detection

This paper deals with the field of computer vision, mainly for the application of deep learning in object detection task. On the one hand, there is a simple summary of the datasets and deep learning algorithms commonly used in computer vision. On the other hand, a new dataset is built according to those commonly used datasets, and choose one of the network called faster r-cnn to work on this new dataset. Through the experiment to strengthen the understanding of these networks, and through the analysis of the results learn the importance of deep learning technology, and the importance of the dataset for deep learning. Object detection as one of the important applications in the field of computer vision has been the focus of research, and convolution neural network has made great progress in object detection. Object detection is developing from the single object recognition to the multi-object recognition. The meaning of the first is just from an image to identify a single object, it can be said that it is a problem of classification, and the meaning of the later is not only can identify all the objects in an image, including the exact location of the objects. Deep learning has formed a mainstream object recognition algorithm based on RCNN and this algorithm is refreshing the higher accuracy in a number of famous datasets. In this paper, we first summarize the some algorithms related to deep learning for object detection, and then apply one of the algorithms to a new dataset to verify its wide applicability.

5. Rapid Object Detection Using a Boosted Cascade of Simple Features

This paper describes a machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. This work is distinguished by three key contributions. The first is the introduction of a new image representation called the Integral Image which allows the features used by our detector to be computed very quickly. The second is a learning algorithm, based on Ada Boost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers [6]. The third contribution is a method for combining increasingly more complex classifiers in a cascade which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The cascade can be viewed as an object specific focus-of-attention mechanism which unlike previous approaches provides statistical guarantees that discarded regions are unlikely to contain the object of interest. In the domain of face detection the system yields detection rates comparable to the best previous systems. Used in real-time applications, the detector runs at 15 frames per second without resorting to image differencing or skin color detection.

6. Orientation Robust Object Detection in Aerial Images using Deep Convolutional Neural Network

Detecting objects in aerial images is challenged by variance of object colors, aspect ratios, cluttered backgrounds, and in particular, undetermined orientations. In this paper, we propose to use Deep Convolutional Neural Network (DCNN) features from combined layers to perform orientation robust aerial object detection. We explore the inherent characteristics of DCNN as well as relate the extracted features to the principle of disentangling feature learning. An image segmentation based approach is used to localize ROIs of various aspect ratios, and ROIs are further classified into positives or negatives using an SVM classifier trained on DCNN features. With experiments on two datasets collected from Google Earth, we demonstrate that the proposed aerial object detection approach is simple but effective.
The remainder of the paper is organized as follows. The orientation robust feature extraction procedure and the aerial object detection approach are presents experimental results concludes the paper with discussion of future works.

3. PROPOSED SYSTEM

The proposed system involves the optimization of two networks, Yolov2 and SSD. Yolov2 and SSD will be used for object detection in images and also for real-time object detection. These two network models will be optimized in several ways simultaneously. One of the most important way is using an effective training datasets because the quality of the model depends on the quality of the dataset being used. This can be done by getting more data, inventing/creating

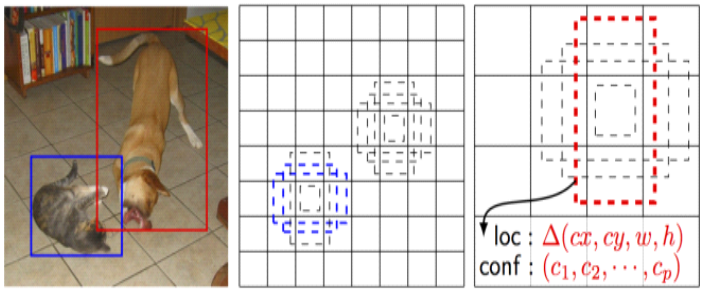
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

more data, re-scaling the data, transforming the data or by feature selection. Further, these models will be optimized by tuning the algorithms.



(a) Image with GT boxes (b) 8 × 8 feature map (c) 4 × 4 feature map

Fig 6: YOLO & SSD

4. METHODOLOGY

4.1 Training methodology: We trained the classifier on a data set comprising of _ 10 million crops overlapping some object with at least 0:5 Jaccard overlap similarity. The crops are labeled with one of the 20 VOC object classes. 20 million negative crops that have at most 0:2 Jaccard similarity with any of the object boxes. These crops are labeled with the special "background" class label.

4.2 Evaluation methodology: In the first round, the localizer model is applied to the maximum center square crop in the image. The crop is resized to the network input size which is 220 _ 220. A single pass through this network gives us up to hundred candidate boxes. After a non-maximum-suppression with overlap threshold 0:5, the top 10 highest scoring detections are kept and were classified by the 21-way classifier model ina separate passes through the network. The final detection score is the product of the localizer score for the given box multiplied by the score of the classifier evaluated on the maximum square region around the crop. These scores are passed to the evaluation and were used for computing the precision recall curves.

5. APPLICATIONS OF NEURAL NETWORKS

5.1 Neural Networks in Practice

Given this description of neural networks and how they work, what real world applications are they suited for? Neural networks have broad applicability to real world business problems. In fact, they have already been successfully applied in many industries.

5.2 Neural networks in medicine

Artificial Neural Networks (ANN) are currently a 'hot' research area in medicine and it is believed that they will receive extensive application to biomedical systems in the next few years. At the moment, the research is mostly on modelling parts of the human body and recognising diseases from various scans (e.g. cardiograms, CAT scans, ultrasonic scans, etc.).

5.2.1 Modelling and Diagnosing the Cardiovascular System: Neural Networks are used experimentally to model the human cardiovascular system. Diagnosis can be achieved by building a model of the cardiovascular system of an individual and comparing it with the real time physiological measurements taken from the patient. If this routine is carried out regularly, potential harmful medical conditions can be detected at an early stage and thus make the process of combating the disease much easier.

5.2 .2 Electronic noses: ANNs are used experimentally to implement electronic noses. Electronic noses have several potential applications in telemedicine. Telemedicine is the practice of medicine over long distances via a communication link. The electronic nose would identify odours in the remote surgical environment. These identified odours would then be electronically transmitted to another site where an door generation system would recreate them. Because the sense of smell can be an important sense to the surgeon, telesmell would enhance telepresent surgery.

5.2.3 Instant Physician

An application developed in the mid-1980s called the "instant physician" trained an auto associative memory neural network to store a large number of medical records, each of which includes information on symptoms, diagnosis, and treatment for a particular case. After training, the net can be presented with input consisting of a set of symptoms; it will then find the full stored pattern that represents the "best" diagnosis and treatment.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

5.3 Neural Networks in business

Business is a diverted field with several general areas of specialisation such as accounting or financial analysis. Almost any neural network application would fit into one business area or financial analysis. There is some potential for using neural networks for business purposes, including resource allocation and scheduling. There is also a strong potential for using neural networks for database mining that is, searching for patterns implicit within the explicitly stored information in databases. Most of the funded work in this area is classified as proprietary. Thus, it is not possible to report on the full extent of the work going on. Most work is applying neural networks, such as the Hopfield-Tank network for optimization and scheduling. system which increase the profitability of the existing model up to 27%. The HNC neural systems were also applied to mortgage screening. A neural network automated mortgage insurance underwritting system was developed by the Nestor Company. This system was trained with 5048 applications of which 2597 were certified. The data related to property and borrower qualifications. In a conservative mode the system agreed on the underwritters on 97% of the cases. In the liberal model the system agreed 84% of the cases.

6. RESULT

This result was gained from applying the algorithms the dataset contains 20 labels having minimum 1000 images. The images are in the .jpeg format. The size of the images ranges from 300x300 to 500x500. The images have to be converted into the desired format to comply with the net. The label varies from object such as table, aero-plane, and car to person, dog, cat, etc.

CONCLUSION

From the above result, one can conclude that the region based convolution neural network is more optimized at a very basic level. It is in dispute whether it can be said as the best form of solution to the problem or not. This result is valid International Conference on Intelligent Computing and Control Systems only in certain parameter. Another researcher can engender new parameters and would achieve less error rates than this but one cannot argue that the RCNN is better than the other neural net.

REFERENCES

[1]. Haigang Zhu1, Xiaogang Chen1, Weiqun Dai1, Kun Fu2, Qixiang Ye1, Jianbin Jiao1_1"ORIENTATION ROBUST OBJECT DETECTION IN AERIAL IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK "University of Chinese Academy of Sciences, Beijing, China © 2015 IEEE
[2]. Huieun Kim, Youngwan Lee, Byeounghak Yim, Eunsoo Park, Hakil Kim"On-road object detection using Deep Neural Network" INHA University Computer Vision Laboratory Incheon, South Korea ©2016 IEEE.
[3]. Malay Shah, Institute of Technology:Object Detection Using Deep Nerual Neworks" Nirma University Ahmedabad, India©2017 IEEE
[4]. Elham Etemad, Qigang Gao, "Object Localization By Optimizing Convolutional Neural Network Detection Score Using Generic Edge Features" Dalhousie University Faculty of Computer Science Halifax, Canada @IEEE2017
[5]. Güner Alpayd_n,Ph.D. "An Adaptive Deep Neural Network For Detection, Recognition Of Objects With Long Range Auto Surveillance" MAY Cyber Technology Ankara, Turkey©2018 IEEE
[6]. Shaukat Hayat, She Kun, Zuo Tengtao, Yue Yu, Tianyi Tu, Yantong Du" A Deep Learning Framework Using Convolutional Neural Network for Multi-class Object Recognition" University of Electronic Science and Technology of China Chengdu, China©2018 IEEE
[7]. A Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification using deep neural networks," University of Toronto part of Advances in Neural information processing systems 25 (NIPS 2012).
[8]. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, " Going deeper with convolution," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp.
[9]. Piotr Doll´ar, Ron Appel, Serge Belongie, and Pietro Perona, "Fast feature pyramids for object detection," Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1532–1545, 2014.
[10]. Martinez-Martin, E. and A.P.d. Pobil, Object Detection and Recognition for Assistive Robots: Experimentation and Implementation. IEEE Robotics & Automation Magazine, 2017.
[11]. Zhang, Y., H. Wang, and F. Xu. Object detection and recognition of intelligent service robot based on deep learning. in 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). 2017.
[12]. Turaga, P., et al., Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video technology,2008.
[13]. Bhatnagar, S., et al., IITP at SemEval-2017 Task 5: An Ensemble of Deep Learning and Feature Based Models for Financial Sentiment Analysis. 2017.
[14]. Szegedy, C., S. Ioffe, and V. Vanhoucke, Inception-v4, Inception- ResNet and the Impact of Residual Connections on Learning. 2016.
[15]. Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
[16]. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, "Decaf:A deep convolutional activation feature for generic visual recognition," arXiv preprint arXiv:1310.1531,2013.
[17]. Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," Journal of Machine Learning Research, vol. 9, no. 2579-2605, pp. 85, 2008.
[18]. Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision, vol. 88, no. 2, pp. 303–338, 2010.
[19]. Piotr Doll´ar, Ron Appel, Serge Belongie, and Pietro Perona,"Fast feature pyramids for object detection," Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1532–1545, 2014.
[20]. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2004.