IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

Object Detection in Military & Space Image by Deep Learning with Convolutional Neural Network

Chitra J. Patil[1*], Swati V. Shinde[2]

[1,2]Information & Technology Dept. Pimpri Chinchwad College of Engineering Sector No. 26, Pradhikaran, Nigdi, Pune, Maharashtra, 411044, India

[*]Corresponding Author: chitra.patil023@gmail.com.

Available online at: www.ijcseonline.org

Accepted: 25/Nov/2018, Published: 30/Nov/2018

Abstract— In recent years, the various Deep Learning architectures have been applied in fields such as speech recognition, natural language processing, and many more classification tasks, where they have usually undergoing the traditional methods. The motivation for such an idea is inspired by the fact that the human brain is organized in a deep architecture, with a given input percept represented at several levels of abstraction. Previous research has problems like – complex unstructured data of satellite images to be observed in short period of time & result is misjudged. So, it is difficult to obtain accurate result immediately. Therefore, proposed paper addresses need of Convolutional Neural Network (CNN) for automatic object detection in military & space image. As per study of CNN we conclude as it is for accurate classification of object from image.

Keywords—Convolutional Neural Network (CNN); Features; Kernels; Pooling; Rectified Linear Unit (ReLU), etc.

I. INTRODUCTION

1. *Deep Learning (DL)*

Classification of data is one of the most frequent decision making tasks performed by human [1]. Image Classification is the process to classify the entire pixels of an image into different possible feature classes as per their attribute values [2]. Deep learning allows computational models that nothing but representations of data with several levels of abstraction. DL solve complex structure in large data sets by implementing the back propagation algorithm. Vital role of DL is show how a machine changes its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Machine Learning utilized in many aspects like web search, e-commerce website, and identification of objects in image, translate speech into text, match news items and many more use. Increasingly, these applications make use of deep learning. The hyperline segment layer of fuzzy hyperline segment neural network (FHLSNN) consists of number of n- dimensional hyperline segments [3].

Deep learning has many advantages compared to shallow learning: network structure of deep learning is the best simulation of the human cerebral cortex. Deep learning, a branch of Machine Learning where neural networks are constructed with more than one hidden layer called as Deep learning Networks [4]. Deep structure neural network can better represent the hyper variable functions and other complex high-dimensional functions; when the larger deep network structure express a function, sometimes can significantly reduce the computational complexity [5]. The key aspect of deep learning is that layers of features they are learned from data using a general-purpose. Learning procedure which are not designed by human engineers. It is computer sytems inspired by the biological neural network that constitute working structure of neurons. The deep architecture allows the system to learn to represent features by themselves based on the nature of the data, rather than the subjective nature of human perception. This deep architecture has been shown to achieve state of the art in many computer vision tasks with little effort in tuning the model including text recognition , object detection , object recognition , face recognition , scene parsing/labeling [6] . Different deep-learning model with ANN are as :

- Feed-forward neural network
- Recurrent neural network
- Multi-layer perceptrons (MLP)
- Recursive neural networks
- Convolutional deep belief networks
- Convolutional neural networks

Here in this paper we studied about CNN among all above.

2. Convolutional Neural Network

Convolution Neural Networks or covnets are neural networks. CNN is combination of math and biology. CNN used in situation where data can be expressed as like map. CNN takes image as input in array of number form &

performed series of operations to that array and at end return probability that object in image belongs to particular class of object. CNN contains layers like –
1. Convolution Layer
2. ReLU (Rectified Linear Unit)
3. Pooling layer:
4. Fully Connected Layers:

1. Convolutional Layer – For generating features map here convolution layer take array of weight (array of no.) as input & taking dot product of array and pixel value of image & this operation called convolution. The convolutional and fully- connected layers have parameters but pooling and non-linearity layers don't have parameters. The CNN has an excellent performance in machine learning problems. Specially the applications that deal with image data, such as largest image classification data set (Image Net), computer vision, and in natural language processing (NLP) and the results achieved were very amazing [7].

2. ReLU – It increases non-linear properties of decision function & overall network without affecting receptive field of convolution layer. Neural networks with rectified linear unit (ReLU) non-linearities have been highly successful for computer vision tasks and proved faster to train than standard sigmoid units, sometimes also improving discriminative performance [8].

3. Pooling layer – Pooling reduces size of feature map, thus simplifying computation in later layer. Applying an additional pooling operation on the local features extracted from multiple image regions can significantly boost classification performance [9]. Many pooling function used like L2 pooling , Max Pooling , Average pooling, sum-pooling or Fisher vector based pooling.

4. Fully Connected Layer – a fully-connected layered model that enables global reasoning about the complicated segmentations of real objects. Optimization with fully-connected graphical models is challenging [10]. It gives final decision like probability that object in photo or image.

Consider following five papers which are written by different authors having same objective & work with satellite image by CNN. Identifying military vehicle from social media image [11]. To do so, a vehicle recognition system use three classes to training purpose like BMP, T-72, other vehicle, the author discussed that with help of data augmentation and transfer learning performs experiment with accurate result. The results show that transfer learning outperforms data augmentation, achieving an average accuracy of 95.18% using 10-fold cross validation, while also generalizing well on a separate testing set consisting of social media content.
For Distinguishing Cloud and Snow in Satellite Images via fully CNN in pixel level [12]. System designed to identifying cloud & snow from the multispectral satellite images designed fully convolutional network with learn deep patterns for cloud and snow detection. Then, to integrate the low-level spatial information and high-level semantic information simultaneously for a multiscale prediction strategy was implemented. Finally, a new and challenging cloud and snow data set was labeled manually for training purpose. For the Detection of Informal Settlements in VHR Images [13]. Here author suggest fully convolutional networks (FCNs) by pixel wise image labeling by automatically. Informal settlements or slums are proliferating in developing countries and their detection and classification provides vital information for decision making and planning urban upgrading processes. Deep FCNs can learn a hierarchy of features associated to increasing levels of abstraction, from raw pixel values to edges and corners up to complex spatial patterns.

For a analysis of satellite images for disaster detection [14], here author implements advance deep learning techniques by implementing CNN. Proposed neural network consists of 3 convolutional layers, followed by max-pooling layers after each convolutional layer, and 2 fully connected layers. Each disaster's training data set consists of 30000~40000 patches and all patches are trained automatically in CNN to extract region where disaster occurred instantaneously. The results reveal accuracy of 80%~90% for disaster detection.

In this paper space target recognition system used for space traffic controls by deep learning, author investigate different single and hybrid data augmentation methods for both training and testing images [15]. Thus propose system follow a data augmentation-based on deep learning approach to recognize space-debris & spacecrafts. Experimental performed on 400 synthetic space target images rendered by the Systems Tool Kit (STK). The paper is organized as follows. Next section introduced about different techniques for object detection using CNN as Residual neural network, VGG network module, Convolution with dilated kernel, Based architecture of CNN, Deep convolutional neural network model. Section 3 introduced comparative study of five papers, section 4 concludes this paper with study of implementing CNN by different techniques.

II. DIFFERENT TECHNIQUES FOR OBJECT DETECTION USING CNN

A. Residual Neural Network [11]

Here author proposed residual neural network For recognizing military vehicles in social media images using deep learning. Two different options were explored for selecting and training a neural network for the recognition task:
(1) Training a network from scratch using data augmentation to increase the volume of data.
(2) Training a network without data augmentation, but opting to use transfer learning instead.

In this case, transfer learning involved using a neural network pre-trained on data from a different domain to extract features from the data set and feeding these features a small, fully-connected neural network .The first option, training a neural network from scratch, involved using a CNN. This architecture, hereafter referred to as ConvNet, features two convolutional blocks, each consisting of two layers with 32 and 64 feature maps, respectively, followed by a maxpooling operation. The feature maps are then flattened before a fully-connected layer with 256 nodes, which is followed by a Softmax activation that outputs a probability distribution over the three classes, as shown in Figure no. 2. To combat overfitting, Dropout is used to randomly disconnect neurons in the fullyconnected layer to prevent the network from seeing the same input twice. The second option, transfer learning, used a deep residual network (ResNet). The difference between the convolutional and residual neural networks is that residual networks adds the original layer input to the layer output by skipping a connection. In Figure no. 3 illustrates the elaboration of a residual network, the residual block, which accepts the activation from the previous layer as an input. This input is then given to a convolutional layer, followed by batch normalization (BN) and a rectified linear unit (ReLU) activation. The number of feature maps learned in each convolutional layer. Finally, the original activation is then added to the output ahead of another ReLU activation, whose output given as the input for the next residual block.

ResNets stack several residual blocks before performing average pooling, which outputs a 2048-dimensional feature vector that is then fed to a fully-connected layer.

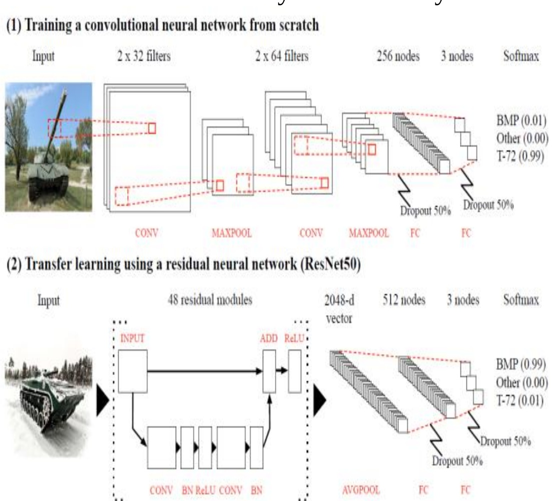

Fig.1. Convolutional Neural Network & Residual Neural Network

B. Visual Geometric Group (VGG) Network Module [12]

Author proposed network structure to differentiate snow & cloud. According to structure of VGG modules as shows in figure no. 3. It is well known that the low-level feature retains much spatial information, while the high-level feature holds much semantic information. To exploit the low-level feature and high-level feature simultaneously, further apply a multiscale prediction module from intermediate layers to supplement low-level localization information. Such a multiscale intermediate module improves pixel-level accuracy significantly in the cloud/snow detection task. In summary, the main contributions of the proposed approach are highlighted as follows.

1) A simple yet effective fully convolutional deep model was proposed to classify the cloud area and snow area in pixel level simultaneously. The proposed model can be trained from scratch in an end-to-end manner without a pertaining model and any hand-tuned parameters.

2) A multiscale learning module was proposed to exploit the low-level spatial information and high-level semantic information simultaneously, which can achieve better localization accuracy than the model with only high level information.

C. Convolution with Dilated Kernel [13]

Author proposed Detection of Informal Settlements in VHR Images. The main building blocks of network are convolutional layers. They compute the convolution of the input image with a bank of filter kernels and add a bias term like following figure . The learnable weights of the filter bank are represented by a 4-D array **w** belongs to $R^{H*W*D*K}$, where H and W are the height and width of the kernel, D is the number of input feature channels, and K is the number of

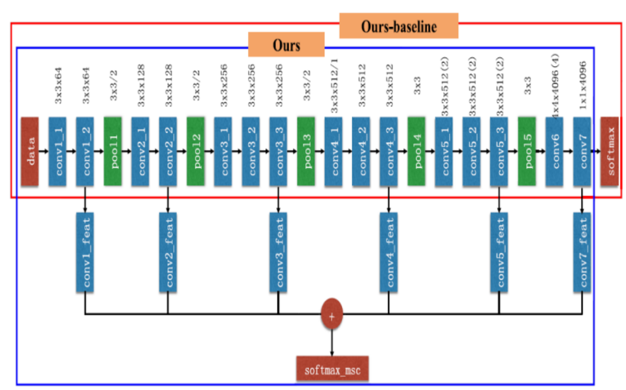

Fig.2 Visual Geometric Group Network Architecture

filters. Additional parameters that control the size of the output are the stride and zero padding. The stride s is the spatial interval between convolution centers and defines the downsampling factor (s = 1 means no downsampling). Padding consists in adding zeros around the border of the input image before applying the filter. The parameter p defines the number of pixels used for zero padding. Given an input image $\mathbf{x}$ of size $M \times N \times D$, the tth channel of the output feature map $\mathbf{y}$ is given by

$$yqrt = bt + \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{d=1}^{D} wijdt \bullet Xs(q-1)$$
$$+ i - p, s(r-1) + j - p, d$$

Assumin equal strides *s* in both spatial dimensions. The size of the output feature map equals.
M' × N' × K = L((M − H + 2 p)/s) + 1_ × L((N − W + 2 p)/s) + 1 × K

In order to capture long-distance dependency, one should adopt kernels of large spatial support *H ×W*. Using large filters, however, increases the number of parameters, making the training more difficult and decreasing the generalization capability of the network. To alleviate this problem, several techniques adopt downsampling in the convolutional layers, which then requires an upsampling strategy. Then convolutions with dilated kernels (DKs), or dilated convolutions, instead of downsampling. This allows us an exponential expansion of the receptive field without increasing the number of learnable parameters per layer. DKs are obtained by inserting zeros between filter elements, effectively enlarging the spatial support of the filter without increasing the number of elements.

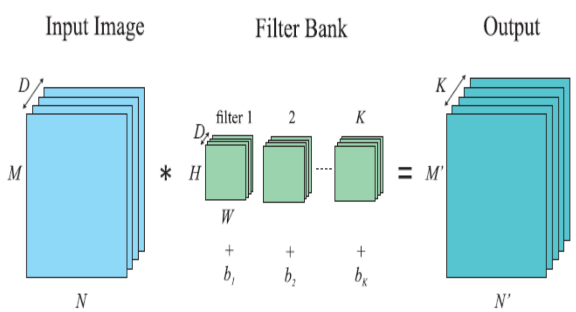

Fig.3. Convolution of image with Filter Bank

D. Base Architecture of CNN [14]
Author suggest a system to Disaster Detection by deep learning. A convolutional layer has a number of filters and convolves them on an input image for extracting features. A pooling layer applies subsampling to the output of the next lower layer for achieving translational invariance as shown following Fig. No.5.

First, RGB channels of pre-disaster and post-disaster (6 channels) are merged into 1 image. Then, raster scan is conducted to this image by sliding over 16 pixels to obtain best predictions value of disaster occurrence. Based on the knowledge obtained in train phase, the highest predictions value with label 1 only will be extracted and 32x32 pixel sized rectangle will be drawn. The drawn region refers to disaster region.

In order to evaluate the accuracy of disaster region, output in Result is compared with ground truth images by undergoing raster scan on a region of interest of 32x32 pixel. Accuracy is calculated based on precision, recall and f-measure with following formulae.

1. Precision = TP / (FP+TP) (3)
2. Recall = TP/ (FN+TP) (4)
3. F-Measure = 2*[Precision*recall/Precision + recall] (5)

Where as ,

TP = True Positive (2)
FP = False Positive
FN = False Negative

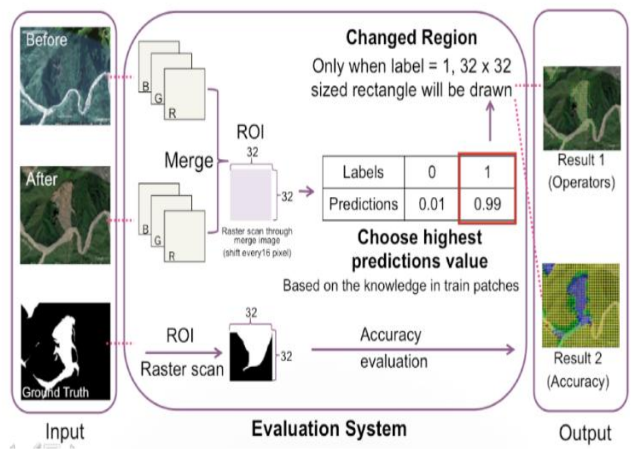

Fig.4. Base architecture of Convolutional Neural Network

E. Deep Convolutional Neural Network (DCNN) Model [15]
Here author proposed space target recognition by deep learning. The DCNN model is constructed based on the the well-known LeNet-5 mode. As shown in Fig. 7, this DCNN consists of nine layers, including three convolutional layers, three spatial pooling layers and three fully connected layers. In each convolutional layer, a feature map is obtained by convolving the input image or severals input feature maps with a linear filter, adding a bias term and then applying a non-linear activation function, shown as follows

$$X_j^{\,l} = f(\,B^l_j\ \ down\ (X^{l-1}{}_i)) \qquad (6)$$

Each convolution layer is followed by a spatial pooling layer, which is a form of down-sampling, where down (.) denotes the down-sampling operator B^l_j and b^l_j stand for multiplicative bias and additive bias respectively. Pooling layers can reduce the computation complexity and provide robustness to translation invariance. Each of the first two fully-connected layers consists of 64 neurons. The dropout technique has been adopted to prevent over-fitting. The last fully-connected layer is a 4-dimensional Softmax layer, which produces the likelihood of an image belonging to each of those four classes.

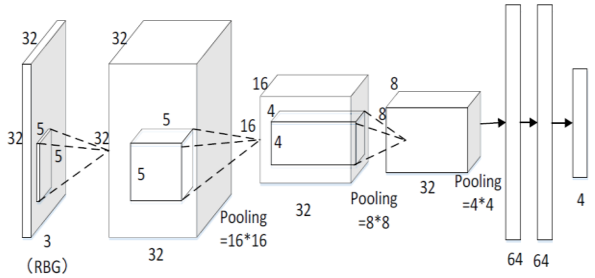

Fig. 5. Architecture of Nine Layer CNN

III. COMPARATIVE STUDY OF FIVE PAPERS

Here we studied 5 papers about their datasets, techniques, number of patches and also their application, numbers of layers used. Following table shows comparative study of different five papers. So by comparing this papers we can judge accuracy of each papers and finds how each techniques will accurate for application.

Table 1 Comparison of Techniques

	Tech. 1	**Tech. 2**	**Tech. 3**	**Tech. 4**	**Tech. 5**
Used Methods	Deep Residual Network	VGG Module	Dilated Convolutional	Deep Learning with CNN	Deep Learning with CNN Data Augme-ntation
Input	Image of Vehicle	Multisp-e-ctral Satellite Image Gaofen #1	VHR Satellite Image	Arial Satellite Image	Synthetic Satellite Image
No. of Layers	6	21	20	8	9
Accuracy	95%	91.4%	-	90%	95.6%
Training Datasets (No. of patch)	6555	50	3000	30000-40000	400

IV. CONCLUSION

The proposed study focus on various layers of Convolution Neural Network and also how deal satellite image data with Deep Learning. From studied five papers we conclude that Deep learning with Convolutional Neural Network classified objects accurate by pixel-wise image labelling than traditional hand-craft features extraction method. Also studied about a general framework of Deep Learning for remote sensing data is provided. CNN is superior because of pooling layers, rectified linear unit layer, convolution layer, deconvolution layer, fully connected layer. Also studied about Different techniques like Data Augmentation & Transfer learning plays vital role in accurate object identification.

REFERENCES

[1] S. V. Shinde and U. V. Kulkarni, "Exctracting Classification Rules from Modified Fuzzy Min-Max Neural Network for Data with Mixed Attributes," Elsevier Journal -Applied Soft Computing Journal, pp. 364, 2016.

[2] Asha Rathee, "Satellite Image Classification using Ant Colony Optimization and Neural Network", in International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 1, Issue 3 , pp. 76, 2017.

[3] S. V. Shinde and U. V. Kulkarni, "Extended Fuzzy Hyperline-Segment Neural Network with Classification Rule Extraction", **Article** in Press, Neurocomputing, April 2017.

[4] Parminder Kaur Birdi, Karbhari Kale, "Deep Convolutional Neural Network Models for Land Use and Land Cover Identification Using Dataset Created From LISS-IV Satellite Images", in International Journal of Scientific Research in Computer Science, Engineering and Information Technology,Vol. 3, Issue 3, pp. 2123, 2018.

[5] Kuang ping, cao wei-na, wu qiao. "Preview on structures and algorithms of deep learning". In IEEE, pp. 176, 2014.

[6] Nguyen Kien, Fookes Clinton, Sridharan Sridha. "Improving deep convolutional neural networks with unsupervised feature learning". In IEEE 978-1-4799-8339-1/15/2015.

[7] Albawi saad ."Understanding of a Convolutional Neural Network". In IEEE 978-1-5386-1949-0/17/ 2017.

[8] Dahl George E., Sainath Tara N., Hinton Geoffrey E." Improving deep neural networks for lvcsr using rectified linear units and dropout". In IEEE 978-1-4799-0356-6/13, pp. 8609, 2013.

[9] Liu Lingqiao, Shen Chunhua, Hengel Anton van den."Cross-Convolutional-Layer Pooling for Image Recognition". In IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 11,pp. 2305, november 2017.

[10] Sun Deqing, Wulff Jonas, Sudderth Erik B., Pfister Hanspeter, Black Michael J." A Fully-Connected Layered Model of Foreground and Background Flow". In IEEE Conference on Computer Vision and Pattern Recognition, pp. 2451, 2013.

[11] Hiippala Tuomo. Recognizing military vehicles in social media images using deep learning. In IEEE, 978-1-5090-6727-5/17/$31.00, PP. 60-62 2017.

[12] Zhan Yongjie, Wang Jian, Shi Jianping, Cheng Guangliang, Yao Lele, And Sun Weidong. "Distinguishing Cloud And Snow In Satellite Images Via Deep Convolutional Network". In IEEE Geoscience And Remote Sensing Letters, 1545-598x, pp.3-4 2017.

[13] Persello Claudio." Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images". In IEEE geoscience and remote sensing letters, vol. 14, no. 12, pp 2327-2328, december 2017.

[14] Siti Nor Khuzaimah Binti Amit, Shiraishi Soma, Inoshita Tetsuo, Aoki Yoshimitsu. "Analysis Of Satellite Images For Disaster Detection". In IEEE, 978-1-5090-3332-4/16/$31.00, pp. 5190-5191, 2016.

[15] Zeng Haoyue, Xia Yong. "Space Target Recognition based on Deep Learning". In 20th International Conference on Information Fusion Xi'an, China - July 10-13, 2017.

Authors Profile

Chitra J. Patil pursuing ME (Information & Tech.) from Savitribai Phule, Pune university. She has completed BE(Computer Engg.) from North Maharashtra University in 2012. She has 4.5 years of teaching experience. She has Published 4 papers in national & international journals. Her master level project "Extraction of Leaf Features for Plant Classification using Deep Learning Approach." got AYUSH Student fellowship program from Pune University.

Dr. Swati V. Shinde has completed Ph.D. in Computer Science and engineering from Swami Ramanand Teertha Marathwada University, Nanded, ME degree in Computer engineering from Bharti Vidyapeeth, Pune, in 2006 and BE(Computer Science & Engineering) degree from Swami Ramanand Teertha Marathwada University, Nanded in 2001. She has total 16 years of teaching experience and currently she is working as a Professor in Pimpri Chinchwad College of Engineering, Pune. Her research interests include Data Mining, Machine Learning, Soft Computing, Artificial Neural Network and Fuzzy Logic. She has published 53 research papers in reputed conferences and journals. Two of these are Published in the prestigious Sciencedirect- Elsevier journals. She has filed three research patents. She also has received the research grant of SPPU University, Pune. She is conferred with the "Dr. APJ Abdul kalam Women Achievers Award" by IItech Banglore. She received the three Best Paper Awards in IEEE Conferences. She got awarded by Indo Global Engineering Excellence Award by Indo Global Chamber of Commerce Industries and agriculture. She has guided more than 40 UG and PG Projects in the different domains of Computer Science and Engineering and Information Technology.

