**A Review on Land Use Land Cover Classification of Satellite Images using Deep Learning Approach**

*Archana Hebbar KV, Harshitha D N, Nagraj Bhat, Jayalakshmi, Pooja*

*Shri Madhwa Vadiraja Institute of Technology and Management, Bantkal, Udupi-574115*

***Abstract-Land cover and land use and its change is the most important, as well as the most widely researched topic in remote sensing. Land cover and land use have been used extensively to derive human and climate activities. The deep neural network today is playing a very important role in the classification of images, which is required in some popular applications like Urban planning. This paper focuses mainly on a deep learning approach i.e. especially the deep convolution neural network (DCNN) for the classification of photogenic images i.e., acquired from mostly satellites. So here in this paper, the different architectures or neural networks of DCNN like Alexnet, VGG, Cascaded cross channelPooling, etc.., how these architectures work better in the classification of satellite images are discussed. The result of this architectures is compared with other classification algorithms like Support vector machine and maximum likelihood classification. One advantageous result that is found from this study is that some of the architectures like cross channel pooling and average pooling with DCNN can automatically construct the training dataset and classify images. And finally, the accuracy is observed between the different architectures of DCNN compared and the accuracy of some of this architecture is compared with SVM, MLC, and RF.***

I.Introduction

A Satellite Image is a picture of the earth taken using artificial satellites. Those images obtained using these satellites can be an image of lots of disturbances i.e it can be light images, water vapor images, infrared images, etc[1]. We know that satellite images provide high spectral and multispectral images of different ranges. So multispectral images generally refer to 3-10 bands. Each band is obtained using sensing radiometers. Each band has a spatial reduction of about 30 meters. Whereas high spectral satellite images are considered to be one the best for the classification purpose because it consists of much narrow-band than multispectral images and it gives the best results and reduces misclassification. It gives a better capability to see from human eyes but the disadvantage is that it increases complexity. Nowadays, large amounts of high-resolution

remote-sensing images are acquired daily. However, the satellite image classification is requested for several applications like modern town planning, agriculture, and environmental monitoring.

Deep Learning is a quite popular technique when it comes to classification. Deep Learning is a subset of Machine Learning. So when it comes to deep learning, deep learning layers depend on the different layers of Artificial Neural Network, whereas the Machine Learning always needs the structured data so in deep learning the most popular Convolution Neural Network is mainly based on the multilayer interconnected channels which enables the data to be learned more that is a high capacity of learning the features and classification of objects from data are quick[1]. Deep learning works better for millions of data at a time and output classification results of deep learning can be score classification, classified element, free text, sound, etc. This is also one of the most important advantages of deep learning over Machine Learning and the most widely used technique. This study will be explaining the different classification methods like Deep convolution neural network, Support vector machine and Maximum likelihood classification methods. And finally, we are going to conclude here that DCNN is one of the most widely used methods for classification.SVM and MLC are good for the classification, regression of objects but they are considered to be good only for a small amount of data[7,8]. When the complexity of the data increases methodology and the architecture also change and DCNN has a number of architectures like Resnet, Googlenet which work on these and give better classification images based on classes considered. And finally, we see that gives the best classification results.

II. Data and Preprocessing

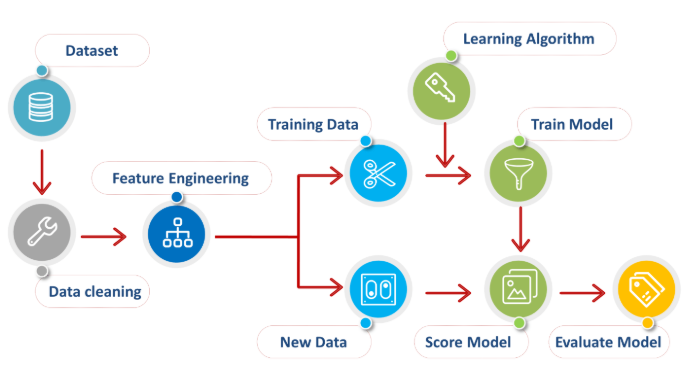
Datasets are something that is most necessary when it comes to classification. The availability of the datasets is itself one of the crisis. Here we are going to see the datasets that are used in this whole study. And all these datasets are preprocessed and well trained and each object is considered into different classes[2]. The most commonly considered class in this study is the cultivated land, wetland, building land and grassland and water bodies. Based on these classes the accuracy of different algorithms is calculated.

ImageNet is an open-source repository of images consisting of 1000 classes and over 1.5 million images. It is a very huge dataset. To introduce a new own network like Resnet, InceptionNet and all kinds of the advanced network then for this network we have to use these datasets that are available from ImageNet. So the image is one of the powerful repository database, so while constructing a new network then that specific network should be trained on imageNet. ImageNet is so capable because it is trained over 1.5 million images.

Now to classify the classes we need the datasets so in this paper there are a number of means mentioned to from which datasets are obtained like LandSat-8 –this is the satellite launched by NASA for the Earth resources. This satellite actually carries two sensors i.e, OLI and TIRS[11]. This satellite mainly uses multispectral images

Other sets of datasets are obtained from NWPU-RESISC45. These are one that was created by NWPU. Here the entire dataset contains around 31500 high-resolution images by Google Earth of 45 different classes.[12]. Some of the datasets are obtained from IARPA Functional Map of the World (fMoW), these datasets are classified into 63 classes. Another important and standard organization that provides dataset is UC-Mercede datasets, these are 21 class Landuse datasets which are mainly used for research purpose and it consists of around 100 images per class. Another important satellite which provides datasets for free is Sentinel-2 [13] ,it provides multispectral high-resolution time-series datasets of landcover maps globally. So these are some of the means to obtain the datasets.as per the review.

Once we have data it is necessary that before its feed into some model for classification or to extract the information it has to pre-processed mainly for radiation calibration and atmospheric correction[9]. Remote sensing data suffer from a variety of radiometric and geometric errors. These errors harm the accuracy of information content and hence reduce the utility of data. Image pre-processing involves the removal of degradation and noise. So finally the aim of any study is to get the best classification result with high accuracy so as to remove the different kinds of noise in an image. Pre-processing plays the first and most important role in the classification of images, because the clean data obtained after preprocessing is used for the classification purpose.

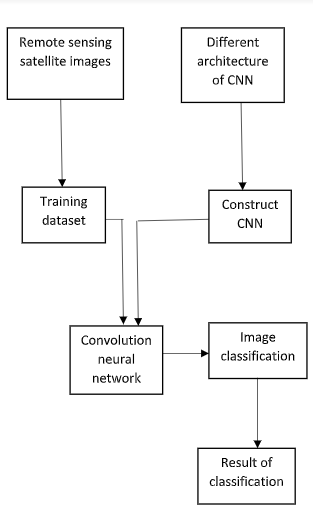


**Figure 1:** Preprocessing of data

III. Methodology

*A. General Process*

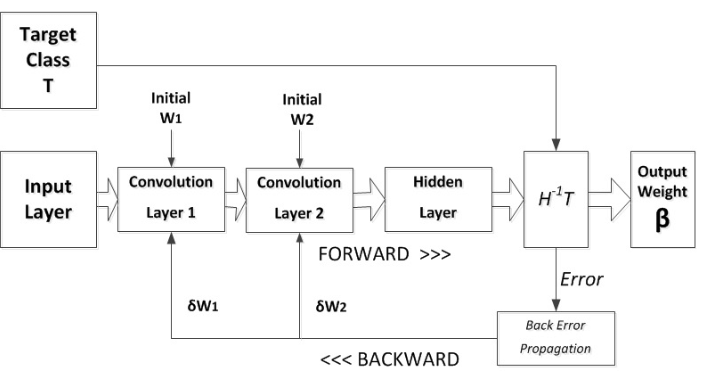
In this one thing is clearly understood any classification techniques have to undergo certain steps to find the result so here we see that basically some of the common steps that are usually followed for classification of an image. Dataset availability and Preprocessing: before starting we have to see that dataset is available based on the needs of the experiment. The DCNN Model requires the amount of dataset for obtaining accurate information. Building a deep CNN model: so this model is built with different methods like cascade pooling and average pooling etc..So based on these trained datasets and classified objects the final accuracy is obtained.



**Figure 2:** Flow chart of the general process for classification

*B. Convolution Neural network*

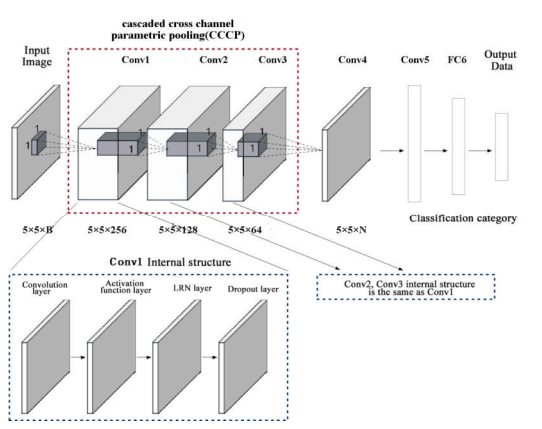
CNN is a feed-forward network that is generally used to analyze visual Image by processing data with grid-like topology. In Convnet every mage is represented in the form of arrays and pixels. As the name itself says convolution is nothing but coil or a twist. So here it sees everything as an object in an image. Basically, there are three layers of the input layer, the hidden layer, and the output layer. Pixels of images are accepted by the input layer in the form of arrays. Next, the hidden layer performs feature extraction by calculation and manipulation. It arranges pixels in a number of ways so that it becomes easy and readable for neural networks. The hidden layer is a convolution layer or pooling layer or Relu layer, it uses a matrix filter and performs convolution operation to detect a pattern in the image. Relu activation function applied on the convolution layer to get rectified feature mapping of image[3]. Pooling is used for another kind of detection and pooling information. Finally, the output layer identifies the object in an image.



**Figure 3:** General process of CNN

*C. Cascaded cross channel Parametric Pooling*

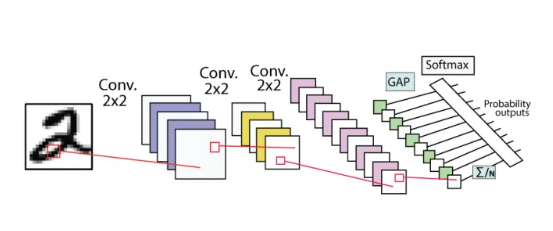
Cascades Cross channel a method for reducing the size of state and number of parameters needed to have a given number of filters in the model Say let us say with an example suppose you have a linear module, with 50 outputs and you want to have 5 outputs. You use a cross-channel pooling to downsample the number of channels. The number of units in a maxout layer is not limited, while in a cross channel pooling layer it cannot be above the number of inputs. That is exactly what cross-channel pooling is because the pooling also considers other feature maps or channels on an equivalent level during the pooling process.



**Figure 4:** Cascaded Average Channel parametric Pooling

*D. Global Average Pooling*

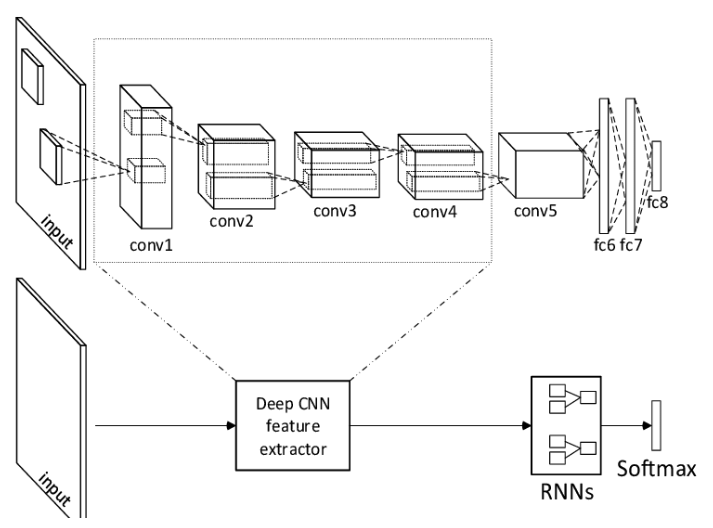
The main function of the Global Average Pooling layer is the pooling of the average value of the single feature map and the average value is taken as the probability value of each feature[6]. When the input layer is passed for Global Average Pooling operation it calculates the average value of each and every single map and returns the average into the output node. So this node at the output layer will contain the average value of the feature map. In the Global Average Pooling, the number of nodes in the input layer is the same as the output layer. This method drastically reduces dimensions, it needs no parameter to train and inherently reduces overfitting.



**Figure 5:** Global Average Channel Pooling

*E. Alex Net*

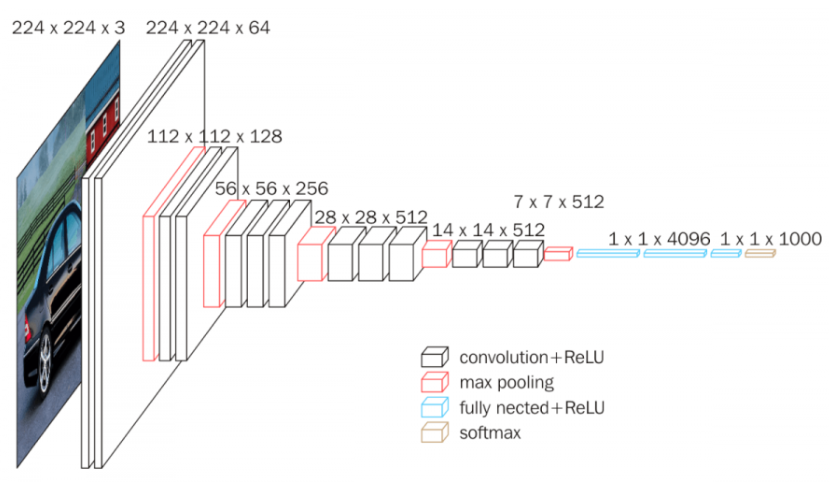
One of the most important architecture. In 2012 AlexNet was introduced by 18 researchers. It is one of the first DCNN for ImageNet. It almost reduced around the top 5 errors from 26% to 18%. It consists of 60 million parameters, It actually trains on 2 GPUs on 6 days. At that time but classification result. Internal representation generated by Alex Net was studied and utilized as a base for live scale image retrieval. Alex Net is the underlying base model for kind of other CNN architecture till now. Alex Net consists of 8 hidden layers, 5 convolution layers used for feature extraction and a fully connected layer. It also attached Relu activation after every convolution and fully connected layer. The number of the layer increases layer by layer.



**Figure 6:** Global Average Channel Pooling

*F. VGG*

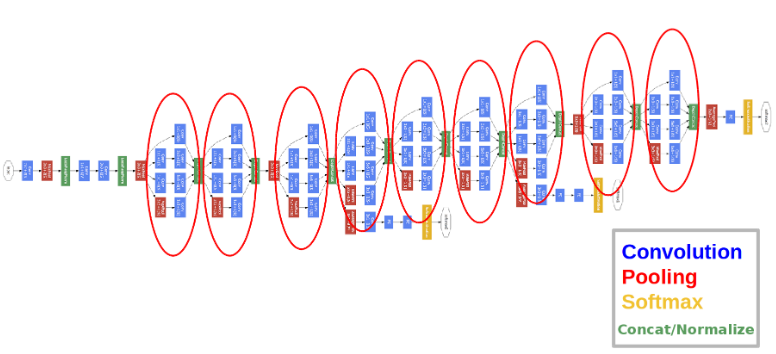
It was introduced in 2014, It is similar to Alexnet. It has only 3x3 convolution but has more number of filters. It reduced the Imagenet top 5 errors i.e, around 80% this can be done by a single model It has around 138 million parameters and trains 4 GPUs for 2-3 weeks. VGG training is similar to AlexNet with additional multi-scale cropping. It has 3x3 layers because stacked convolution has a large receptive field. Decision functions become more complicated due to  3 non-linearity. By introducing more convolution or pooling layer we can reduce error.



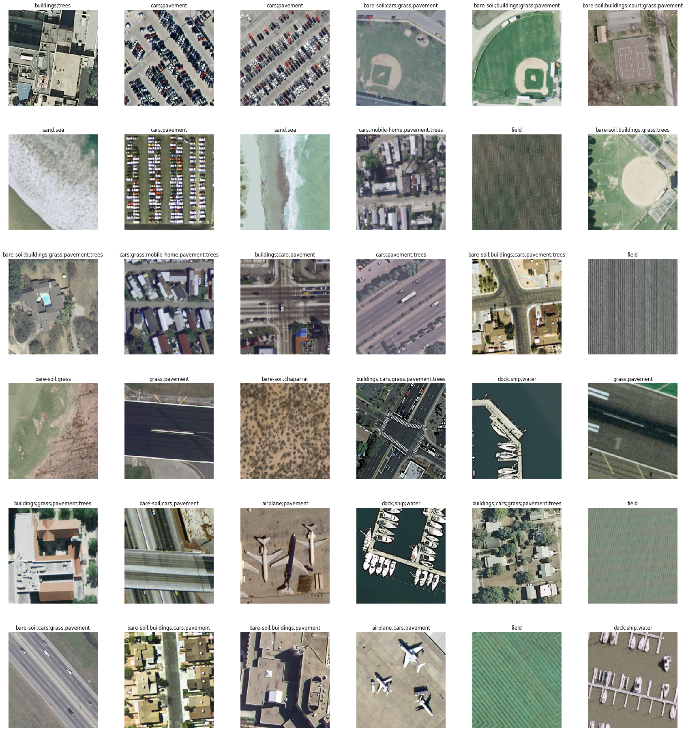
**Figure 6:** VGG

*G. Inception*

It was first introduced in 2015. Inception architecture is associated with GoogleNet family i.e, GoogleNet is nothing but InceptionV1.InceptionV1 was developed in 2014 by Google. It is one that is used with a large dataset in the real world. The main reason Google has to process millions of images so Google came up with this architecture. It is a deeper model and reduced the Imagenet top 5 errors around 5-6%. It has around 25 million parameters and trains on 8 GPUs for 2 weeks[7]. Inception blocks are basic building blocks for the Inception model. A deeper model probably better results in recognition. Depth of learned representation reflects the hierarchy of features at various levels of abstraction.



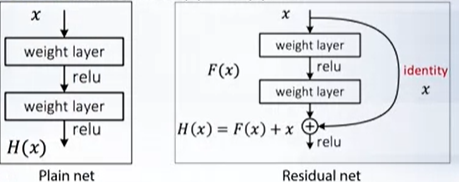
**Figure 7:** Inception



**Figure 8:**Datasets available from different satellites.

*H. ResNet*

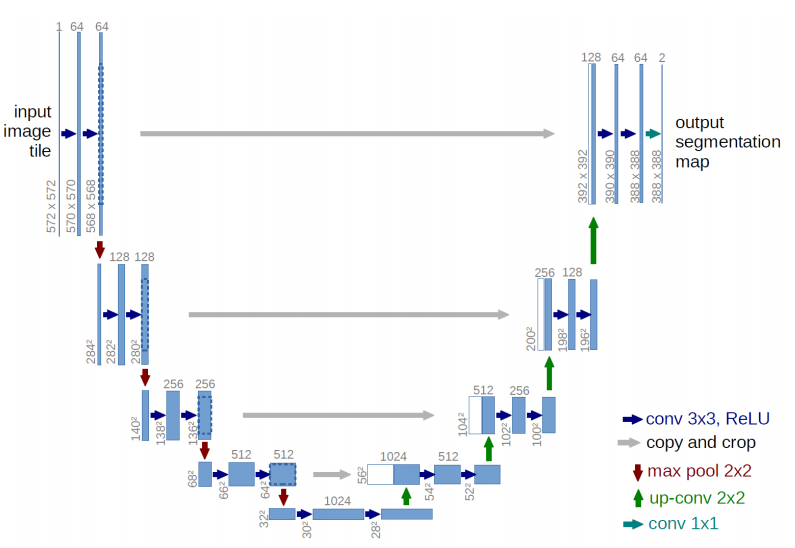
Residual Network: If we want a network to fit some mapping from the pixel space to the space of labels. Say instead of hopping each few stack layers it can be directly implemented at desired underlying mapping explicitly i.e, this layer to residual mapping instead of just using Resnet. We can incorporate residual connections into the more sophisticated architecture that is building a Resnet with Inception architecture. This Inception Resnet will achieve good performance at relatively low computation cost.



**Figure 9:**ResNet and InceptionResNet

*I. U-Net*

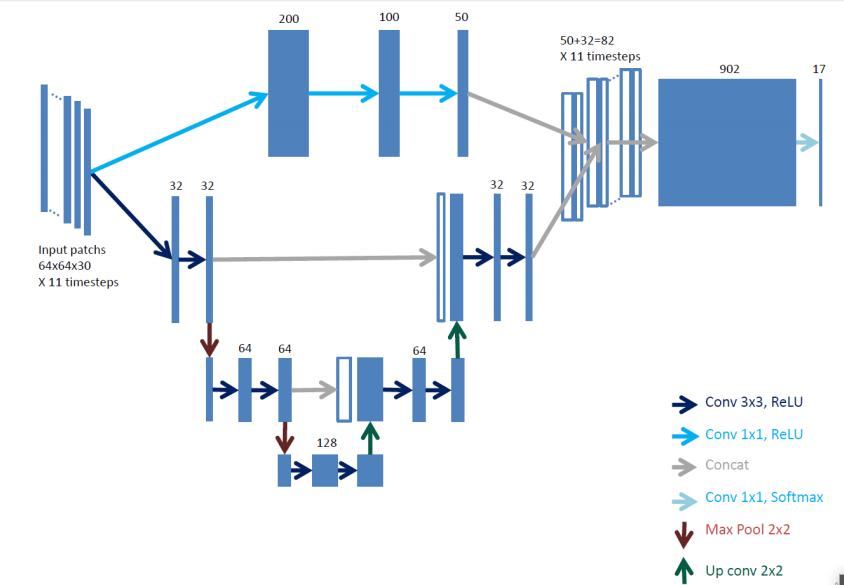
U-Net architecture is one of the classical architecture i.e. [14] fully convolutional U-Net architecture. This is mainly used for semantic segmentation. It is initially applied to biomedical images. It is one which will be able to gather context in the encoding phase and accurately localize feature in the decoding phase. It contains symmetric downsampling and upsampling paths i.e 5 steps of 2 convolution layer. On the downsampling side, the network takes as input an image patch with C channels of size N × N × C. For each step, a convolution layer, followed by a ReLU activation function and a subsampling by maximum pooling are used. s. On the upsampling side, skip connections allow the transfer of raw information from layers on the downsampling side. The upsampling path is implemented through the concatenation of the output of convolution layers on the downsampling side to features produced on upsampled input.



**Figure 10:**Original U-Net architecture

*J. Fine-Grained U-Net*

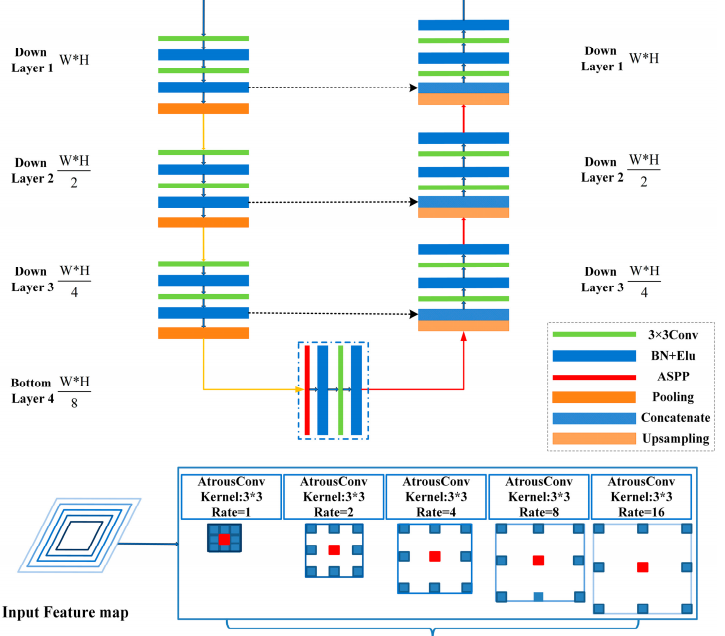
Fine-Grained U-Net architecture is one that is used for the classification of images. This architecture is inspired by the U-Net architecture. This is mainly implemented from the following adaptations: The receptive field and the number of steps in the network are adapted to our remote sensing data resolution and number of dimensions. 2. A pixel-wise fully connected path is added to improve boundary delineation in the produced maps. 3. The temporal component of the data is dealt with by replication of the network on subsets of the time series.



**Figure 11:**Fine Grained U-Net architecture

*K. ASPP-UNet*

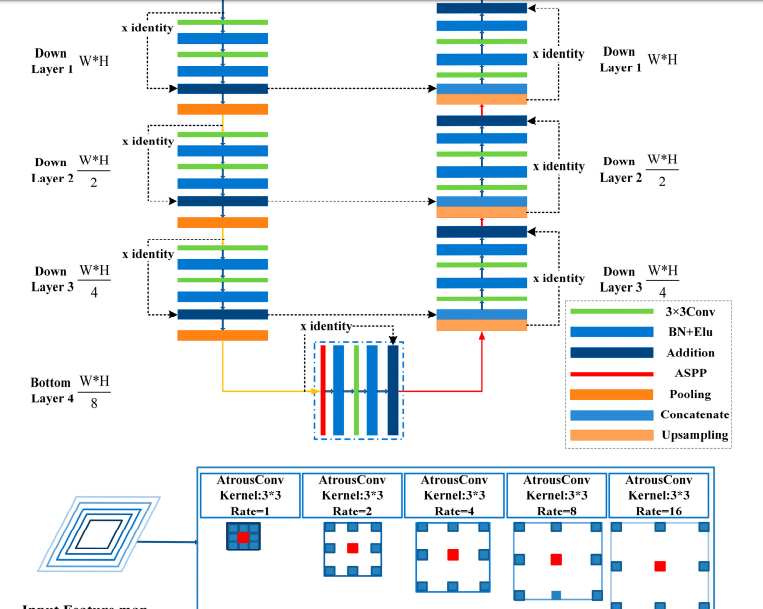
In one of the paper for urban cover classificationof images ASPP net is used .Here the architetur is designed with atronus convolution for dense feature extraction with variant felid of view.Here ASPP implements atronous convolution layer with different sampling rate in parallel manner.Then this multi scale feature are used to generate a final feature map[15]. Based on the U-Net model the ASPP Net model was propsed. Each down layer contains two sequential unpadded 3 × 3 convolutions. Each convolution operation is followed by an element-wise activation function. The number of feature maps is doubled after the previous convolution.



**Figure 12:**ASPP U-Net architecture

*L. ResASPP-UNet*

Recent researches have found that classification accuracy really don’t increase with increase in convolution layers especially for smaal training examples[16].So the ASPP-Unet model is alterd by adding identity mapping from the input of each layer to the output of the same layer. Thus, each layer forms a residual block. Within each block, the layer structure remains the same as ASPP-Unet except for the shortcut connection part, which performs an identity mapping from the shallower layer and then addition to the current layer. This new architecture is named ResASPP-Unet model.



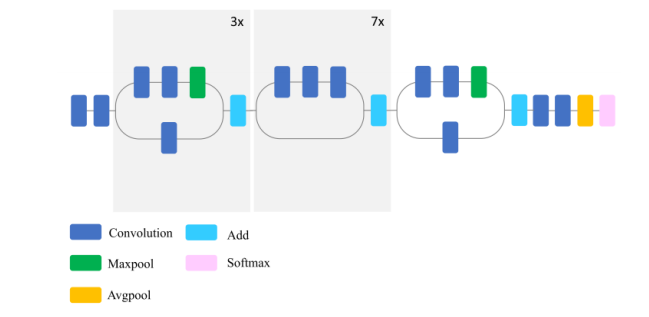
**Figure 13:**ResASPP U-Net architecture

*M. Conditional Random Field*

A conditional random field (CRF) is an undirected graphical model that is used to probabilistically predict pixel labels based on weak supervision, which could be manual label classification outputs from discrete regions of an image based on outputs from a trained DCNN[2]. The trained DCNN is used to classify small spatially distributed regions of pixels in a sample image, which is used in conjunction with the same CRF method used for label image creation to estimate a class for every pixel in the image.

*N. Xception*

Xecption is one of the architcute like Inception, but inception module is substituted with depth-wise separable convolutional layers Xception’s architecture is constructed based on a linear stack of a depth-wise separable convolution layer i.e., 36 convolutional layers with linear residual connections. There are two important convolutional layers in this configuration: A depth-wise convolutional layer [17], where a spatial convolution is carried out independently in each channel of input data, and a pointwise convolutional layer: where a 1 × 1 convolutional layer maps the output channels to a new channel space using a depth-wise convolution.



**Figure 13:**Xception architecture

**IV. Discussion**

Deep learning and its most important approach for classification of satellite images that is the deep convolution neural network and its various types of architecture were disused here is output obtained from those techniques for classification of images.

Firstly architecture of DCNN that is the global average pooling cascaded cross channel pooling was used for classification of land course mapping of Qinhuangdao provision of china which could automatically construct the training dataset and classify images and also this technique performed the classification of both multispectral and hyperspectral images[1]. And also the provided but classification result in that area. Accuracy was about 82.0% and also observed that accuracy was 5% and 14% higher when compared to SVM and MLC respectively.

Next, the DCNN was applied to the landscape images obtained by satellite images for the classification using the condition random field[2]. Here the classes that were considered were desert, forest, island, water bodies, etc. Based on these classes the classification was performed and finally, the result was that CNN training and testing data using minimal manual superstition and the most important result that was obtained is that DCNN prediction of a small region of images using CRF could give a highly accurate pixel-level classification of images[3]. DCNN with different architecture like RENET, VGG, Inception V3, was used in this paper. They used the dataset from IARPA which consists of around 63 classes. So they tried to classify the images based on the classes of the dataset obtained i.e. from IARPA[15], python libraries, Keras and TensorFlow was used for the

classification using for DCNN architecture like inception-V3, Xception, VGG and finally classify the classes with an accuracy of 83%.

From this review, most of the classification is carried out with different CNN architecture this mainly the Alextnet architecture is used accuracy and compared with the other architecture like google net and Caffnet and as used the UC-Merced land use dataset[6]. Google gave an accuracy of 97% and Caffnet around94% whereas Alextnet had an accuracy of 94% but it is observed that Alextnet was faster than google net.

There are other methods in deep learning like supervised, unsupervised learning, in this, the object-oriented method with DCNN for land use classification (COCNN) is used for classifying 10 land-use types of Fuxian lake is classified. Firstly segmentation is done next feature object is constructed on the basis of OOS the CNN model is built for classification purpose it is observed that the COCNN gives higher accuracy about 8.98% more than CNN that is the accuracy is around 96.2% so this method solves the problem of inaccurate classification of typical feature and also the classification accuracy better than CNN.

Another study on the classification of wetland classes of Canada is done using most known methods deep neural networks i.e. DenseNet21, InceptionV3, VGG16, VGG19, Xception, ReNet50, Inception ResNetV2.That is compared with ML methods SVM and Random Forest to test the efficiency of wetland. The accuracy obtained with Inception ResNet V2, Xception, ReNet50 was around 96.17%, 93.57% and 94.81% whereas with SVM and RF[9], it was 74.89%, 70.08%. this shows how DL works better than traditional classification methods. Another approach for classification of land use is studied using fully conventional neural network architecture.that is the replacement of random forest classifier and this approach gives the best classification result with respect to pixel-wise and gives different variability across different landscapes. The use of another type of architecture for classification of urban land covers on high resolution.

The satellite image is done using ASPP-Unet and ResASPP-Unet.ASPP allow spatial pyramid pooling which is used for extra high-level feature and give high-level output .these methods were built for classification of world view 2 and 3 satellite images of region around 85.2% for WV2 images and 83.2% for WV3 images accuracy was obtained using this ASPP-Unet model[10]. This study gave a very good classification result. On their approach of classification on land swap images collected from amazon with CNN model produced accuracy was around 84% and this approach mainly worked for multilabel classification of images. Another approach of multilabel classification on the amazon forest was performed mainly to

identify the illegal mining operation.

Here they used to build the CNN model obtained about 0.76 and built the same model with Resnet50 to get the higher focus about 0.91 and this model could identify the weather condition, natural terrain feature and also the manmade improvements. From this, we are able to make out illegal mining in the Amazon forest.

So this approach of classification of satellite images has a wide range of applications like urban planning, identification of illegal activities etc.in different sectors.

We could also observe from this study that though there are a number of architectures of most popular CNN the different architecture of CNN like Resnet, google net, Alextnet will give a higher accuracy rate compared to other ML models like SVM, RF, etc.

V.Conclusion

This paper gives an introduction about deep learning and ita approaches for classification also about the satellite images and satellite that provide the satellite sensing images and its types. Briefly, the deep neural network architectures are explained most of the popular architecture of DCNN are explained and also how they are helpful in image classification. All the papers are reviewed and observed that the deep neural architectures like Resnet, Alexnet etc.. give the best classification result and also the best accuracy when compared to Machine Learning methods. And about 94.08% accuracy is obtained with DCNN architectures. Using some of the methods like global average pooling, the cascaded cross channel will automatically construct the training datasets. And some of the most useful application with this classification of images is urban planning, identification of illegal mining activities in the Amazon forest etc..This method also reduces the manual supervision. So this paper finally gives a review of using this technique with these different architectures and to better understand which methods give good classification results and accuracy.

REFERENCE

[1]. Hu, Yunfeng, Qianli Zhang, Yunzhi Zhang, and Huimin Yan. "A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China." *Remote Sensing* 10, no. 12 (2018): 2053.

[2]. Buscombe, Daniel, and Andrew C. Ritchie. "Landscape classification with deep neural networks." *Geosciences* 8, no. 7 (2018): 244.

[3]. Pritt, Mark, and Gary Chern. "Satellite image classification with deep learning." In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1-7. IEEE, 2017.

[4]. Shafaey, Mayar A., et al. "Deep Learning for Satellite Image Classification." *International Conference on Advanced Intelligent Systems and Informatics*. Springer, Cham, 2018.

[5]. Jin, Baoxuan, Peng Ye, Xueying Zhang, Weiwei Song, and Shihua Li. "Object-oriented method combined with deep convolutional neural networks for land-use-type classification of remote sensing images." *Journal of the Indian Society of Remote Sensing* 47, no. 6 (2019): 951-965.

[6]. Mahdianpari, Masoud, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery." *Remote Sensing* 10, no. 7 (2018): 1119.

[7]. Stoian, Andrei, Vincent Poulain, Jordi Inglada, Victor Poughon, and Dawa Derksen. "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems." *Remote Sensing* 11, no. 17 (2019): 1986.

[8]. Zhang, Pengbin, Yinghai Ke, Zhenxin Zhang, Mingli Wang, Peng Li, and Shuangyue Zhang. "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery." *Sensors* 18, no. 11 (2018): 3717.

[9]. Shendryk, Yuri, Yannik Rist, Catherine Ticehurst, and Peter Thorburn. "Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 157 (2019): 124-136.

[10]. Gardner, Daniel, and David Nichols. "Multi-label Classification of Satellite Images with Deep Learning." (2017): 2017.

[11]. Roy, D.P.; Wulder, M.; Loveland, T.R.; Woodcock, C.; Allen, R.; Anderson, M.; Helder, D.; Irons, J.; Johnson, D.; Kennedy, R. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. 2014, 145, 154–172.

[12]. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: benchmark and state of the art. P. IEEE 583 2017, 105, 1865-1883

[13].Hagolle, O.; Sylvander, S.; Huc, M.; Claverie, M.; Clesse, D.; Dechoz, C.; Lonjou, V.; Poulain, V. SPOT4 (Take5): Simulation of Sentinel-2 Time Series on 45 Large sites. Remote Sens. 2015, 7, 12242–12264.

[14]. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

[15]. 6. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv, 2018; arXiv:1802.02611.

[16].Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. IEEE Trans. Geosci. Remote Sens. 2018, 56, 847–858

[17].Sifre, L.; Mallat, S. Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1233–1240.