Project Report

Name : Shriya Rai | Email : [rai00016@umn.edu](mailto:rai00016@umn.edu) | Student ID : 5462955

# Introduction

Land use classification is defined as a classification technique which provides information on land cover and the types of human activity involved in land use. It may also facilitate the assessment of environmental impacts on and potential or alternative uses of land.[2] Classifying and mapping land cover is an integral step in understanding the Earth's biophysical systems. Data on the area and distribution of wildlife habitat, for example, is useful in managing and mitigating development impacts on protected and endangered species. Similarly, information on the type, area, and configuration of buildings, roads, and other impervious land-cover facilitates the modeling of storm-water runoff and watershed hydraulics and hydrology.[3]

Different land-cover surfaces reflect sunlight in distinct wavelengths of the electromagnetic spectrum. These spectral signatures, when recorded as remotely sensed images, permit the classification and mapping of individual land-cover types. Sources of data for the classification and mapping of land cover and land use include aerial photographs, satellite imagery, and maps of assorted social and cultural attributes (e.g., property boundaries, population density).[3] In this project, two such remote sensing imagery datasets have been utilised - UC Merced[1] and NWPU-RESISC45[2]. These datasets are curated such that the images have been labelled based on their land use, and these labelled classes are balanced. UC Merced dataset has images spanning 21 land use classes containing a variety of spectral signatures. Similarly, the NWPU-RESISC45 dataset contains 45 land use classes. Further details for each of the dataset is provided in the Dataset section of the report.

Remarkable progress has been made in satellite imagery classification in recent times. The most notable acceleration came with the advent of convolutional network architectures and their success in the field of image recognition. However, deep networks require large amounts of training data to mitigate the bias problem and overcome overfitting, which in turn negatively impacts the classification accuracy on the test data. This proves to be a hindrance due to the lack of larger, high-quality satellite imagery datasets. As a possible solution to this issue, one can leverage the principle of transfer learning in tandem with convolutional networks. Transfer learning applys the knowledge gained while solving one problem to a different but related problem. In this project, transfer learning utilises the knowledge gained by solving the problem of Visual Object Recognition to solve the related problem of Land Use Classification. The varied sizes of the two datasets utilised helps demonstrate the effectiveness of the proposed approach to mitigate the shortage of polished labelled satellite imagery data.

The motivation for tackling land use classification is that it has multiple diversified real-world applications, as stated above. The following expands on a few of these applications:

1. Resource planning: Resource planning deals with the allocation and utilization of resources to achieve maximum efficiency. Management of land is largely dependent on natural resources and land

segregation for various activities. Soil quality, water availability, biodiversity and population density in an area are key factors for the planning of resources. Modern approaches, such as land use classification, can be extended to real-time classification not only to determine the appropriate land use type but also to provide decision makers with sustainable land resource management strategies that improve productivity and sustainability. Land use classification will assist the decision makers in determining and putting into practice the best land use management options for sustaining production.

1. Infrastructure planning: With the increase in population and the advent of modern architecture, infrastructure planning is highly important in developing nations. Land use classification helps in determining the areas associated with residential, water, agriculture, etc., which gives the planners needed guidance when planning the various aspects of a city. A well-planned infrastructure strengthens the sustainability and livability of our cities and communities and helps in creating a flourishing environment. Land use classification can also help in the planning of factory installations to minimize the damage caused by hazardous waste emissions from the same.
2. Disaster Management: Large scale natural disasters cause a considerable damage to property and lives, which leaves a financial debt in the economy. Disaster managers employ mitigation strategies and work towards minimizing the economic damage of disasters. Land use classification can help disaster managers in both pre- and post-planning. This includes the location and temporal information of the land before and after the disaster. Land use classification can help in building structures in areas where the damage can be minimized or mitigated.

The primary objective of this project is to classify land use by utilising satellite images for various

land use types. The satellite images are taken from the UC Merced[1] and NWPU-RESISC45[2] land use dataset, which are balanced, labelled and publicly available datasets. The problem is solved by leveraging state-of-the-art CNN architectures and transfer learning techniques. Secondary objectives are to 1) demonstrate how the size of the dataset affects the classification accuracy of the CNN models and 2) investigate the classification accuracy for CNN models trained from scratch versus pre-trained CNN models in context with the same dataset. The project also hopes to analyze whether there needs to be a trade-off between classification accuracy and computational efficiency by reusing pre-existing knowledge.

# Related Work

There have been enormous strides in solving aerial imagery and land use classification based problems in the last few years. The approaches range from simpler image classification solutions, which are based on local descriptors computed by utilising scale-invariant feature transform (SIFT)[5] or histogram of gradients(HOG)[6], to a slightly more complex bag of visual words approach[7] used in combination with a Support Vector Machine or K-Means based classifier. In recent times, with ever-expanding data and the availability of more efficient hardware, numerous convolutional neural network architectures[8] have also been employed to aid aerial imagery classification.

In the paper, “Remote Sensing Image Scene Classification: Benchmark and State of the Art”[9], authors propose a large-scale land use aerial imagery dataset - “NWPU-RESISC45” along with providing a comprehensive review of the recent progress in remote sensing image scene classification. Paper illustrates the advantage of NWPU-RESISC45 dataset over the existing datasets. They investigate how well the current state-of-the-art scene classification methods perform on the NWPU-RESISC45 dataset. It is observed that fine-tuned versions of AlexNet, VGGNet-16, and GoogLeNet models, which were

pre-trained on ImageNet[10] dataset perform well with accuracies - 85.16%, 90.36% and 86.02% .

Pre-trained networks have been explored in the context of remote sensing imagery in the past [11][12]. In Marmanis et al. [11], the potential of using large pre-trained neural networks have been investigated for classifying remote sensing aerial images into a large set of diverse land use classes. The paper demonstrates promising results over the UC Merced[1] benchmark. In Castelluccio et al., the authors address the remote sensing scene classification task by employing two CNN architectures, GoogLeNet[15] and CaffeNet[17], which have been considered with three design modalities. Two of these modalities(fine-tuning and feature vector) are based on transfer learning technique. Experiments were conducted on two datasets, UC Merced[1] and Brazilian Coffee Scenes[16], which have quite different properties. The latter dataset helped understand the behavior of the proposed approach when pre-training and target data differ significantly. The results of the experiments reiterate that training a deep CNN from scratch is not always advisable with the limited-size datasets currently available in this field.

# Datasets

NWPU-RESISC45 and UC Merced datasets are the two publicly available land use datasets utilized to assess the performance of the proposed approach.

The NWPU-RESISC45 dataset is a publicly available benchmark for Remote Sensing Image Scene Classification, created by Northwestern Polytechnical University. This dataset contains 31,500 images, covering 45 scene classes with 700 images in each class. These 45 scene classes include airplane, airport, baseball diamond, basketball court, beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium residential, mobile home park, mountain, overpass, palace, parking lot, railway, railway station, rectangular farmland, river, roundabout, runway, sea ice, ship, iceberg, sparse residential, stadium, storage tank, tennis court, terrace, thermal power station, and wetland.[1]

The UC Merced dataset consists of 21 land use classes containing a variety of spatial patterns, some with texture and color homogeneity and others with heterogeneous presentations. The data set is compiled from a manual selection of 100 images per class, with each RGB image being approximately 256 by 256 pixels. The 21 land use types include agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium-density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court.[2]

The UC Merced dataset is similar to NWPU-RESISC45, as both are land use datasets. However, it is also a considerably smaller dataset compared to NWPU-RESISC45, and training a neural network using UC Merced would be less time consuming and simpler. On the other hand, for the same reason, models trained on UC Merced would have lower accuracy as they would be prone to overfitting, especially when trained from scratch on the training set.

Images in both the datasets have been resized to 224x224x 3 for GoogLeNet and VGG based models while images have been resized to 299x299x3 for InceptionResnet models. These CNN architectures are discussed in the Approach section below. For training and testing models on NWPU-RESISC45 dataset, only the images belonging to the 19 common classes between the two

datasets have been used. Furthermore, the extracted datasets have been split per class into 80:10:10 ratio for training, validation and test, respectively.

# Approach

Convolutional Neural Network architectures are employed for land use classification: 1) by training from scratch on the two datasets and 2) by leveraging transfer learning for training purposes.

Transfer learning is based on the idea of using a pre-trained CNN and repurposing it to the task of interest. The CNN models used here have been pre-trained on the ImageNet dataset. ImageNet is an image dataset organized according to the WordNet hierarchy. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet; the majority of them are nouns (80,000+). ImageNet provides, on average, 1000 images to illustrate each synset. Images of each concept are quality-controlled and

human-annotated[10]. The output of the penultimate layer of a pre-trained network which is used as a feature vector for classification. Features extracted from the lower layers of the CNN will be generalizable to other classification problems. Making use of CNNs trained on the Imagenet dataset would make sense for land use data as optical remote sensing images have strong low-level similarities with general-purpose optical images. Transfer learning also helps in reducing model training time and lends higher generalizability to the model predictions, especially when the training data may be small.[11][12]

GoogLeNet is a convolution network that utilises “inception modules”. These modules reduce the complexity of the 3-D filters of conventional architectures by means of a prior depth reduction phase. Due to reduced complexity, multiple filters can be used in parallel at different resolutions. To improve the effectiveness of the gradient backpropagation, given the depth of the network, GoogLeNet employs auxiliary classifiers connected to intermediate layers. Thus, the advantages of this architecture are twofold: 1) different sized filters can be used at each layer, therefore, retaining spatial information, and

1. the number of free parameters of the network is significantly reduced, making it less prone to overfitting and allowing it to be deeper [12].

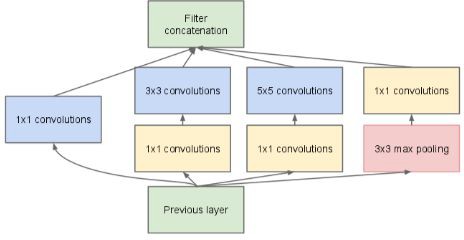


Fig 1. For the inception module, convolutions of different sizes allow the network to process features at different spatial scales. They are then aggregated and fed to the next layer. 1x1 convolutions are used for dimension reduction before the more expensive 3x3 and 5x5 convolution.[12]

VGGNet is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition”. The model achieves 92.7% top-5 test accuracy in the ImageNet classification problem.

The convolution layers in VGG use very small receptive fields (3x3) and have 3 fully connected layers at the end with 4096 channels in the first two layers and the third corresponding to the number of classes. The hidden layers in VGGNet use the Rectified Linear Unit (ReLU) as an activation function. Due to a larger number of trainable parameters, VGG has the potential to achieve higher accuracy.

However, it may also be prone to overfitting for the same reason [13].

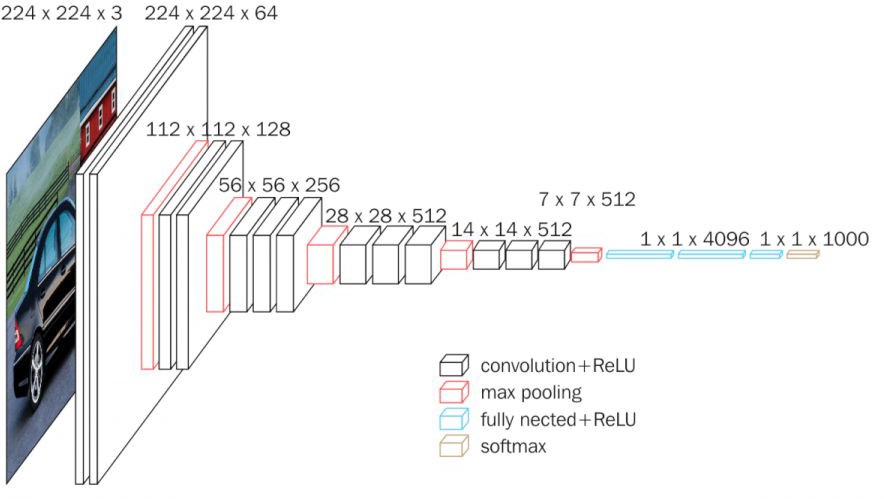
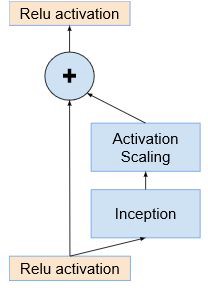


Fig 2.VGG16 Architecture [14]

InceptionResNet is a newer convolution network based on GoogLeNet and ResNet. The introduction of residual connections leads to dramatically improved training speed for the Inception architecture. The

foundation of residual connection is borrowed from ResNet architecture. Residual connections act as identity shortcut connections that skips one or more layers[15].



*Fig 3. The general schema for scaling combined Inception-resnet modules[15]*

# Result

Accuracy is utilised as the evaluation metric to judge the performance of the proposed CNN based models as both the datasets are perfectly balanced.

*Accuracy*

= *Number of correct predictions*

÷ *Total number of prediction*

The chosen CNN models are trained from scratch on the datasets and their performance evaluated. The same models utilise feature vectors for classification, which is generated via transfer learning on the ImageNet dataset.

The performance of the pretrained InceptionResNet model is the best for UC Merced dataset. All the pretrained CNN models trained perform better when compared to the performance of the same models trained from scratch. The results reaffirm the hypothesis that deep learning models, in conjunction with transfer learning, improve the prediction power of the models when the dataset utilised for training is small in size. It can also be observed that the model having the larger number of trainable parameters would perform worse on the test set due to overfitting on the training set. This is evident by the results as the VGGNet models have the highest number of trainable parameters and lack the benefit of skipping the layers which the InceptionResNet models can afford due to the presence of residual connections. The GoogLeNet models have fewer trainable parameters than the VGGNet models in spite of the fact that GoogLeNet models (22 layers) have more layers than the VGGNet models (16 layers).

However, the opposite can be observed from the accuracy results for the NWPU-RESISC45 dataset. The GoogLeNet model trained from scratch performs the best out of the lot. The InceptionResNet model trained from scratch proves slightly better than its pre-trained counterpart. Surprisingly, the VGGNet model, which had been trained from scratch, performs poorly relative to the VGGNet model trained by leveraging transfer learning. This may be due to the NWPU-RESISC45 dataset not being

large enough for training the high number of parameters of the former model.

|  |  |  |  |
| --- | --- | --- | --- |
| CNN Model | Design | Accuracy on UC Merced | Accuracy on NWPU-RESISC45 |
| GoogLeNet | From Scratch | 0.70 | **0.93** |
| Transfer Learning | 0.80 | 0.82 |
| VGGNet | From Scratch | 0.56 | 0.81 |
| Transfer Learning | 0.80 | 0.85 |
| InceptionResNet | From Scratch | 0.71 | 0.88 |
| Transfer Learning | **0.89** | 0.87 |

*Table 1. Classification accuracy of proposed CNN based solution on UC Merced and NWPU-RESISC45 test datasets. Best result in bold.*

The following images provide insight into the training and prediction of the best performing CNN models.

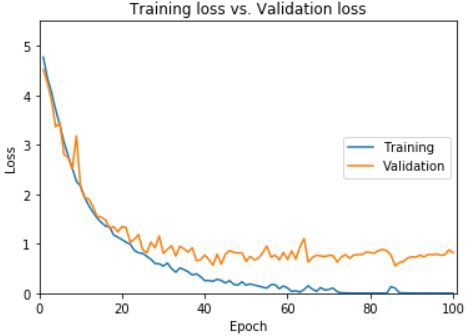
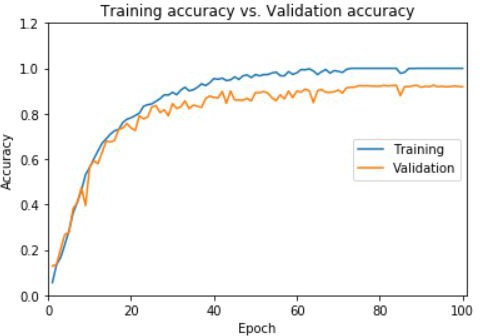


Fig 4 and 5. Plots shown here pertain to the training of GoogLeNet on NWPURESISC\_45 training set from scratch. The left image shows an accuracy plot as the model trains, and the right image shows a loss plot as the model trains.

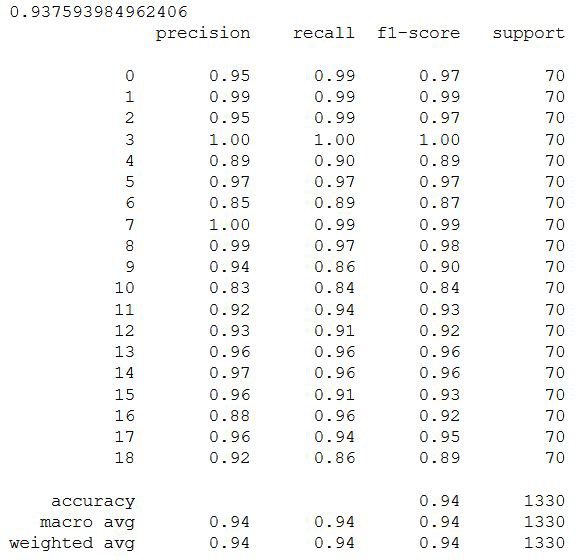


Fig 6.The image shows accuracy followed by the confusion matrix, which pertains to predictions made by GoogLeNet on NWPURESISC\_45 test set. The GoogLeNet model is trained from scratch on the NWPURESISC\_45 train set.

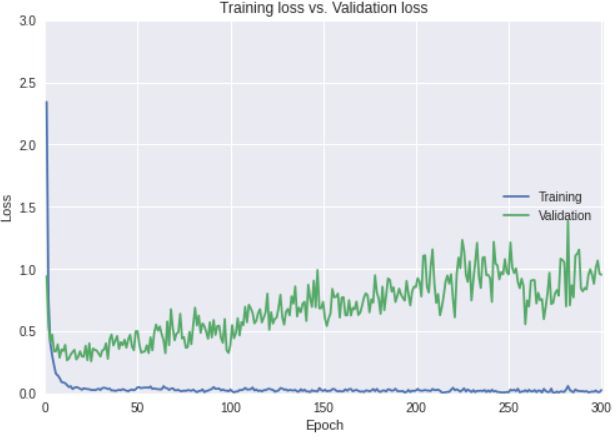
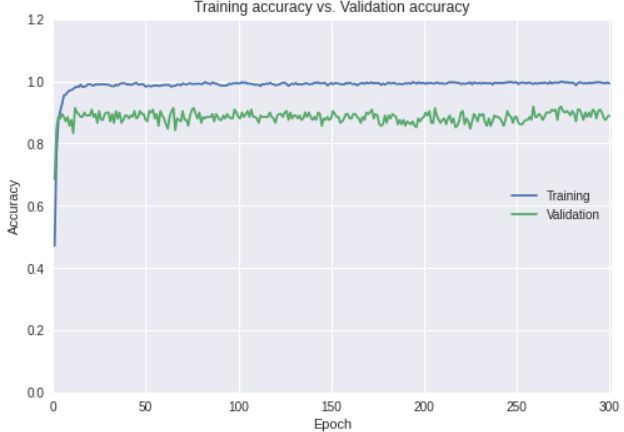


Fig 7 and 8. Plots shown here pertain to training of pretrained InceptionResNet on the UC Merced train set. The left image shows an accuracy plot as the model trains, and the right image shows a loss plot as the model trains.

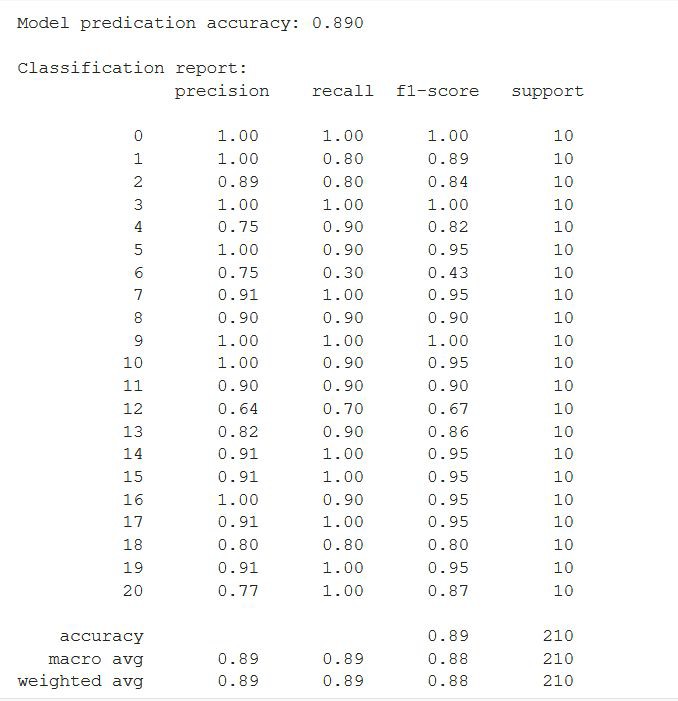


Fig 9.The image shows accuracy followed by the confusion matrix, which pertains to predictions made by the pretrained InceptionResnet on the UC Merced test set. The last layer of this InceptionResnet model is trained from on the UC Merced train set.

# Conclusion

In general, an increase in the size of the dataset improves the prediction accuracy of the deep learning model. Occam's razor comes into the picture here as predictive power cannot be improved by increasing the complexity of the model when the dataset is relatively smaller. Transfer Learning can be leveraged for training models in such cases to provide gains in accuracy by 1) increasing generalizability of the model and 2) mitigating overfitting on the training set. Additionally, training deep learning models from scratch is a good idea only when the size of the training set is adequately large.

# Future Work

The transfer learning-based CNN models used for classification on the UC Merced dataset can be pre-trained on the entire NWPU-RESISC45 dataset instead of the ImageNet dataset. The reasoning

behind the proposed approach is 1) a higher degree of similarity between the two land use datasets, 2) dissimilarity between aerial images belonging to UC Merced dataset and ground-level images of everyday items in the ImageNet dataset, and 3) sufficiently large size of NWPU-RESISC45. It is fair to argue that aerial images are unlike the images available in the ImageNet dataset, and they might have differing edges and low-level features.

# Code

The code for the project can be found at the link below:

*<https://github.com/shriya2909/NWPU-RESISC45-and-UC-Merced-Land-Use-Classification>*

# References

* 1. *NWPU-RESISC45 Dataset 2016:<http://www.escience.cn/people/JunweiHan/NWPU-RESISC45.html>*
  2. *UC merced land use dataset. UC Merced, Oct. 2010:*

*http://vision.ucmerced.edu/datasets/landuse.html*

* 1. *LAND USE CLASSIFICATION*

*<https://stats.oecd.org/glossary/detail.asp?ID=1502>*

* 1. *Land Use Classification:*

[*https://www.sciencedirect.com/topics/earth-and-planetary-sciences/land use-classification*](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/land-use-classification)

* 1. *D. Lowe, “Distinctive image features from scale-invariant keypoints keypoints”:* [*https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf*](%20https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf)
  2. *N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”:* [*https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf*](https://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf)
  3. *Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for classification”:*

[*https://dl.acm.org/doi/10.1145/1869790.1869829*](https://dl.acm.org/doi/10.1145/1869790.1869829)

* 1. *A.B. Penatti, K. Nogueira, and J.A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes do-mains?”:* [*http://openaccess.thecvf.com/content\_cvpr\_workshops\_2015/W13/papers/Penatti\_Do\_Deep\_Feat ures\_2015\_CVPR\_paper.pdf*](http://openaccess.thecvf.com/content_cvpr_workshops_2015/W13/papers/Penatti_Do_Deep_Features_2015_CVPR_paper.pdf)
  2. *Gong Cheng, Junwei Han, Xiaoqiang Lu, “Remote Sensing Image Scene Classification: Benchmark and State of the Art”:* [*https://arxiv.org/ftp/arxiv/papers/1703/1703.00121.pdf*](https://arxiv.org/ftp/arxiv/papers/1703/1703.00121.pdf)
  3. *About ImageNet :*

[*http://image-net.org/about-overview*](http://image-net.org/about-overview)

* 1. *D. Marmanis, M. Datcu, T. Esch, and U. Stilla - “Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks” :* [*https://drive.google.com/file/d/1p7eob-qY5f9db0A7EFVmKE8BRz2KGofL/view*](https://drive.google.com/file/d/1p7eob-qY5f9db0A7EFVmKE8BRz2KGofL/view)
  2. *M. Castelluccio, G. Poggi, C. Sansone, L. Verdoliva - “Land Use Classification in Remote Sensing Images by Convolutional Neural Networks” :* [*https://arxiv.org/pdf/1508.00092.pdf*](https://arxiv.org/pdf/1508.00092.pdf)
  3. *Karen Simonyan, Andrew Zisserman - “ Very Deep Convolutional Neural Networks for Large Scale Image Recognition”:*

[*https://arxiv.org/pdf/1409.1556.pdf*](https://arxiv.org/pdf/1409.1556.pdf)

* 1. *VGG16 – Convolutional Network for Classification and Detection:*

[*https://neurohive.io/en/popular-networks/vgg16/*](https://neurohive.io/en/popular-networks/vgg16/)

* 1. *C. Szegedy, S Ioffe,V.Vanhoucke, A.Alemi - “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”:*

[*https://arxiv.org/pdf/1602.07261.pdf*](https://arxiv.org/pdf/1602.07261.pdf)

* 1. *Brazilian Coffee Scenes Dataset:* [*http://patreo.dcc.ufmg.br/downloads/brazilian-coffee-dataset/*](http://patreo.dcc.ufmg.br/downloads/brazilian-coffee-dataset/)
  2. *Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick,S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding”:* [*https://arxiv.org/pdf/1408.5093.pdf*](https://arxiv.org/pdf/1408.5093.pdf)