A Survey to study about different Convolutional Neural Network on Various Image Classifications

Raj Shah, Sagar Patil, Anish Malhotra, Ranjita Asati

Department of Computer Engineering, University of Mumbai, Atharva College of Engineering, Malad, Mumbai, India

Publication Info

Article history:
Received : 21 February 2020
Accepted : 15 May 2020

Keywords:
Convolutional Neural Network, image classification, UC MERCED Dataset

****Corresponding author:***
Raj C Shah
e-mail: *shahraj88794@gmail.com*

Abstract

Multi label Image classification using Convolutional Neural Network is yet very difficult when it comes to performing. However, Single Label Image Classification can be performed easily and promisingly. As there are many categories of objects in a real world image, it becomes difficult to label them under various categories and also because of the lack of multi-label training image and high complexity. This paper surveys different Convolutional Neural Network (CNN) using Single Label Image Classification on which Multi Label Image Classification can be performed with High Accuracy. We have also learnt different trained Convolutional Neural Network architecture using UC MERCED Dataset which is essayed in this paper.

1. INTRODUCTION

Recent studies have found different Image Classification techniques which has got momentum with Convolutional Neural Network architecture. The task of satellite Image classification becomes vital as it requires an image to be labelled in one or multiple categories according to the objects present in it. We know that in real world, various scenes consist of images generally of more than one class of different categories. Therefore, it becomes difficult to classify labels precisely. This is still a problem to interpret since the complexity in Multi label classification is high and also due to the reason, the assumption that foreground objects are roughly aligned, which is usually true for single-label images, does not always hold for multi-label images.

Single Label Image Classification is performed on different Convolutional Neural Network using pre-trained dataset. From the total of 21 semantics of the UC MERCED Dataset VHR scene of images depends on the overall semantic theme used for training. For example, Figure 1.delineates three scenes from the UC-Merced dataset from the *Baseball court* and *Medium residential* categories, *airport* respectively.

In single label classification, for the last layer either the soft max or sigmoid activation function is used depending on the number of classes.

2. REVIEW OF LITERATURE

There are fundamentally three domains in which the current undertakings are investigated: methods dependent on handmade highlights, systems based information driven highlights and procedures dependent on move learning.

Baseball court

Medium residential: Cars, Pavement, grass, trees

Airport: airplane, cars, grass

Fig 1: Single and Multi label images from UC MERCED Dataset

2.1. Feature engineering based classifiers

E Aptoula [1] concentrated on Content-based picture recovery (CBIR) utilizing worldwide picture descriptors. The creator guessed that worldwide picture descriptors are progressively effective in CBIR contrasted with neighbourhood descriptors. Consequences of applying worldwide morphological surface descriptors (roundabout covariance histogram (CCH), revolution invariant point triplets (RIT) and descriptors dependent on fourier force range of the semi level zone-based scale space (FPS)) on UC Merced land-use (UMLU) informational index are introduced. The creator has closed through experimentation that highlights of the picture removed utilizing CCH, RIT and FPS are progressively appropriate for CBIR contrasted with nearby descriptors on the UMLU informational collection [2]. L Gueguen [3] proposed a compound structure portrayal for picture characterization. It used neighbourhood highlights descriptors removed from the multi scale division of the first picture. The separated highlights are bunched into visual words with the kd-Tree calculation. The visual words are at long last assembled into circulations to depict the picture compound structures. N He [4] proposed a structure to join two low-level visual highlights viz. Gabor highlight and shading highlight for scene order. S Kumar et al. [5] proposed an equal design for extraction of morphological highlights, for example, roundabout covariance histogram (CCH), pivot invariant point triplets (RIT).

2.2. Unsupervised feature engineering based classifiers

J Fan et al. [6] contended that handmade element descriptors (nearby and worldwide) are less reasonable for picture acknowledgment contrasted with highlights gained from the picture information in an information driven way. The creators have favoured an unaided learning approach for taking in highlight descriptors from pictures utilizing multi-way meager coding design. F Luus et al. [7] utilize a profound convolutional layer neural system (DCNN) for picture order. The creators like to maintain a strategic distance from high quality highlights depending rather on DCNN for include assurance. R Stivaktakis et al. [8] proposed a multi-mark picture grouping design utilizing DCNN. To keep away from model over fitting, information growth methods, for example, pivot of the picture by various sums, picture re-scaling, flat and vertical flips, interpretations to the x and y-hub, and the expansion of clamor were utilized. H Wu et al. [9] proposed a half and half design called profound channel banks. It consolidated multicolumn stacked denoising scanty auto-encoder (SDSAE) and Fisher vector (FV) to consequently become familiar with the delegate and discriminative highlights in a various levelled way for LU scene arrangement in the UMLU informational index. F Zhang et al. [10] proposed an unaided element learning system utilizing the saliency recognition calculation to separate a delegate set of patches from the remarkable areas in the UMLU picture informational collection. These delegate patches were given as contribution to an inadequate auto-encoder to change over these picture patches into low-measurement highlight vectors. These element vectors were utilized to prepare a SVM for picture grouping. F Zhang et al. [11] proposed an inclination boosting arbitrary convolutional organize (GBRCN) system for scene order. It was a gathering system comprising of different single profound neural systems. J Bergado et al. [12] proposed a multi-goals convolutional organize, called FuseNet, and its intermittent variant, called ReuseNet, to perform picture combination, grouping, and guide regularization of a multi-goals VHR picture in a start to finish design. Z Gong et al. [13] present an organized measurement learning, a procedure that alters the double cross-entropy misfortune metric to permit it to separate the remote detecting picture scenes with the incredible similitude. The new measurement is joined into a DCNN model for picture grouping in UMLU informational index. C Cao et al. [14] concentrated on the assessment of eight moved CNN-put together models with respect to land-use characterization undertakings and utilization of the best performing moved CNN-based model as a classifier to order and guide the land-use. R Minetto et al. [15] proposed Hydra, a group of convolutional neural systems (CNN) for geo-spatial land order. The outfit of CNN is made from ResNet and DenseNet designs pre-trained on ImageNet informational collection. Extra layers of ResNet and DenseNet designs are added to make a start to finish profound learning pipeline for picture arrangement. H Parmar [16] proposed a multi-neighbourhood LBPs joined with closest neighbour classifier can accomplish a precision of 77.76% for picture grouping on UMLU informational index. In [17], K Karalas et al. use a CNN to recognize the various kinds of land covers by relegating at least one marks to watched phantom vectors of the multi-label picture pixels.

2.3. Transfer learning approach for classification

D Marmanis et al. [18] expanded the utilization of DCNN in picture arrangement by using a DCNN pre-trained on image-net informational index as opposed to preparing a DCNN without any preparation. The creators contended that the enormous pre-prepared profound convolutional neural system (DCNN) produced a lot of elevated level portrayals, which could be utilized for picture characterization in the following handling stage. G Scott et al. [19] favoured highlights took in by DCNN from the preparation information than handmade highlights. The creators explored different avenues regarding three engineering viz. CaffeNet, GoogLeNet, and ResNet50 for picture grouping on UMLU informational index. To improve the model execution, move learning and information expansion were utilized. Q Weng et al. [20] utilized pre-prepared CNN to take in profound and hearty highlights from the pictures of the UMLU informational index. The creators adjusted the CNN design by supplanting the completely associated layers of the CNN by the extraordinary learning machine classifier. Y Zhen et al. [21] likewise contended for pre-prepared profound neural systems for picture arrangement. The creators changed the standard GoogLeNet with a structure called 'Initiation'. The altered system diminished the parameters and prepared quicker contrasted with the first system. E Flores et al. [22] utilized a ResNet-50 DCNN design pre-prepared on the ImageNet dataset for include extraction from pictures. The educated parameters of the ResNet-50 were utilized to separate 2048 piece profound element vectors of each information picture. These component vectors were then used to order the picture. N Uba [23] used designs of AlexNet, CaffeNet and GoogleNet on ImageNet informational index and moved the leanings of the model from the ImageNet informational index to the UMLU informational index for picture grouping. M Castelluccio et al. [24] guessed that preparation CaffeNet and GoogLeNet without any preparation would not be prudent for the restricted estimated UC Merced land-use (UMLU) informational collection. The creators saw that cautious adjusting of the CaffeNet and GoogLeNet designs pre-prepared on ImageNet informational index, including a few layers of the engineering, if great outcomes, all in all. J Li et al. [25] used a class enactment map (CAM) encoded CNN model prepared utilizing unique RGB patches of ImageNet informational index and consideration map based class data. The parameters of the engineering are then utilized for grouping on the (UMLU) informational index.

3. COMPARISON OF DIFFERENT CNN ARCHITECTURE ON VARIOUS PARAMETERS.

Comparison graph is made on the data of different networks with the use of UC MERCED Dataset.

3.1. Quantitative Parameters

Y-axis range:0–65000000(1cm = 5000000)

- The average range of parameters in this bar graph is 25885339
- The Neural Network which has highest parameter is Resent 152
- The Neural Network which has lowest parameter is Mobilenet V2
- Resnet 50 and Resnet V2-50 have almost same parameters

Y-axisrange:0–40000 (1cm=2000)

- The average range of parameters in this bar graph is 26999 and closest one near it is Inception Resent.
- The Neural Networ which has the highest parameter is Resent 152.
- The Neural Networks which have lowest parameter are VGG-16, VGG-19 they also have the same parameter which is 8721.

Y-axis range:0–65000000(1cm = 5000000)

- The average range of parameters in this bar graph is 24401832
- The Neural Network which has highest parameter is Resent 152
- The Neural Network which has lowest parameter is Mobilenet V2
- Resnet 50 and Resnet V2-50, Resnet 101 and Resnet V2-101 have almost same parameters.

Fig.1, Fig.2, Fig.3 shows Quantitative parameters on graphical comparison to different Convolutional Neural Network. As we want to improve the quality of image classification, consideration of "Total parameters",

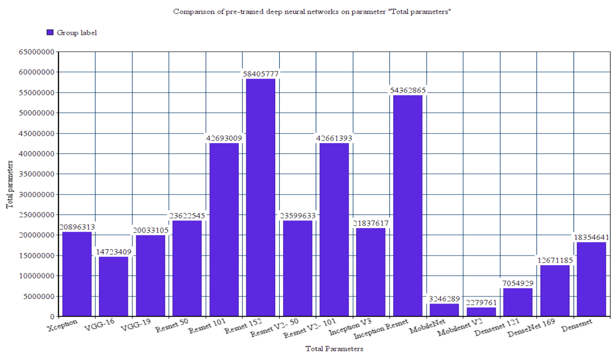


Fig. 2: Comparison of pre-trained deep neural networks on parameter "Total parameters"

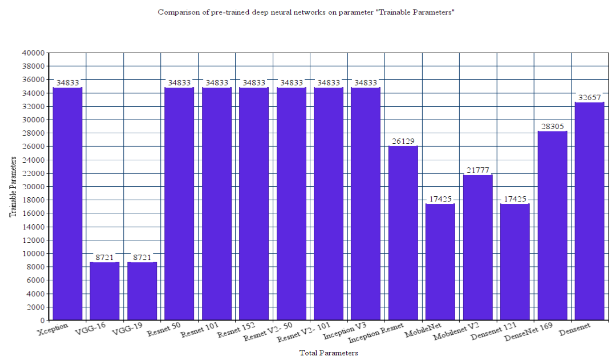


Fig. 3: Comparison of pre-trained deep neural networks on parameter "Trainable parameters"

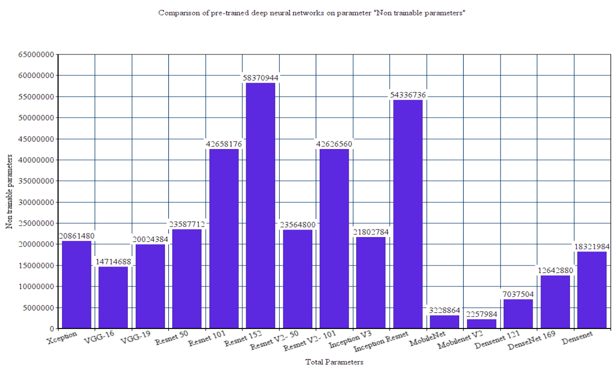


Fig.4: Comparison of pre-trained deep neural networks on parameter "Non Trainable parameters"

"Trainable parameters", "Non-trainable parameters" as a comparison parameter will not be of much help. You have the number of parameters increased or decreased quantitatively, image quality is never affected as it depends on qualitative parameters which is described below.

3.2. Qualitative parameters

Y-axis range: 0.00 – 1.05 (1cm = 0.05)

- The average range of parameters in this bar graph is 0.6818050667
- The Neural Network which has highest parameter is Resnet V2-50
- The Neural Network which has lowest parameter is Resnet 152.

Y-axis range: 0.00 – 1.00 (1cm = 0.05)

- The average range of parameters in this bar graph is 0.6490009334
- The Neural Network which has highest parameter is VGG-19
- The Neural Network which has lowest parameter is Resnet 50
- VGG-16 and Inception V3 have almost same parameters and are really close to average parameters.

Y-axis range: 0.00 – 1.00 (1cm = 0.05)

- The average range of parameters in this bar graph is 0.7241237334
- The Neural Network which has highest parameter is Resnet 50
- The Neural Network which has lowest parameter is VGG-19
- Inception Resent and Mobilenet V2 have almost same parameters.

Y-axis range: 0.00 – 1.10 (1cm = 0.05)

- The average range of parameters in this bar graph is 0.806667
- The Neural Network which has highest parameter is Resnet 50

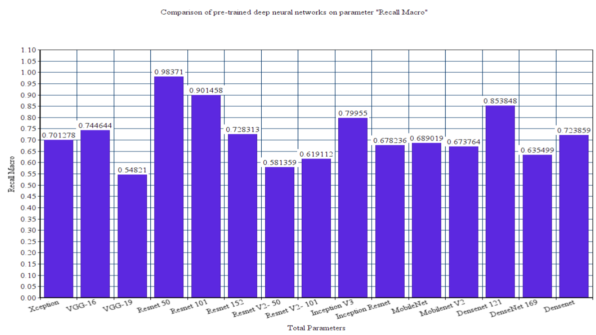


Fig.5: Comparison of pre-trained deep neural networks on parameter "Precision Macro"

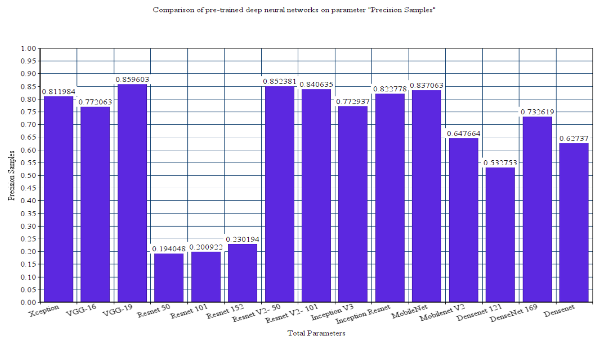


Fig.6: Comparison of pre-trained deep neural networks on parameter "Precision Sample"

- The Neural Network which has lowest parameter is Resnet V2-50
- MobileNet V2 and Inception V3 have almost same parameters and are really close to average.

Y-axis range: 0.00 – 0.85 (1cm = 0.05).

- The average range of parameters in this bar graph is 0.60704.
- The Neural Network which has highest parameter is Inception V3.
- The Neural Network which has lowest parameter is Resnet 152.
- VGG-19 and Mobilenet V2 have almost similar parameters.

Y-axis range: 0.00 – 0.90 (1cm = 0.05)

- The average range of parameters in this bar graph is 0.653583.
- The Neural Network which has highest parameter is VGG-16.
- The Neural Network which has lowest parameter is Resnet 50.
- Densenet 121 and Mobilenet V2 have almost same parameters and are really close to average parameters.

Fig. 5-11 shows qualitative parameters on comparison to

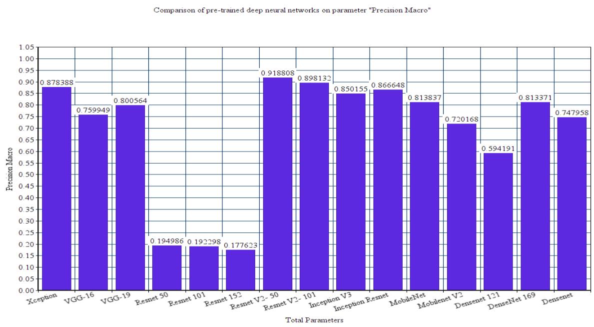


Fig.7: Comparison of pre-trained deep neural networks on parameter "Recall Macro"

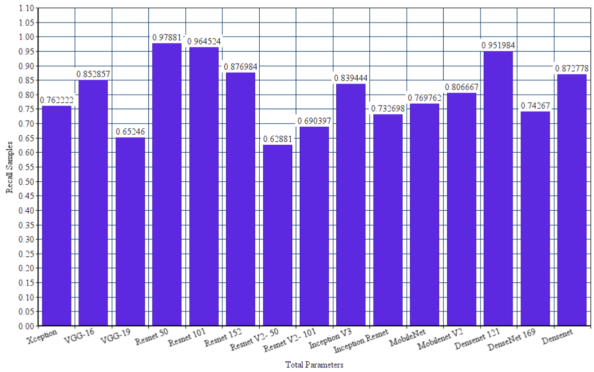


Fig.8: Comparison of pre-trained deep neural networks on parameter "Recall Samples"

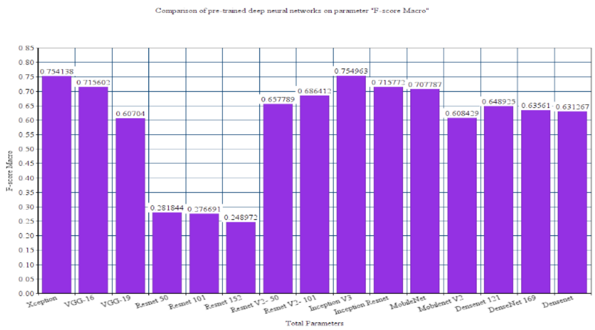


Fig.9: Comparison of pre-trained deep neural networks on parameter "F-score Macro".

various convolutional network models. These parameters help us effectively decide to what model is to be selected for multi label image classification that will give less error.

Y-axis range: 0.00- 0.18 (1cm= 0.01)

- The average range of parameters in this bar graph is 0.10.
- The Neural Network which has highest parameter is VGG-19.
- The Neural Network which has lowest parameter is Resnet 101 & Resnet 152.

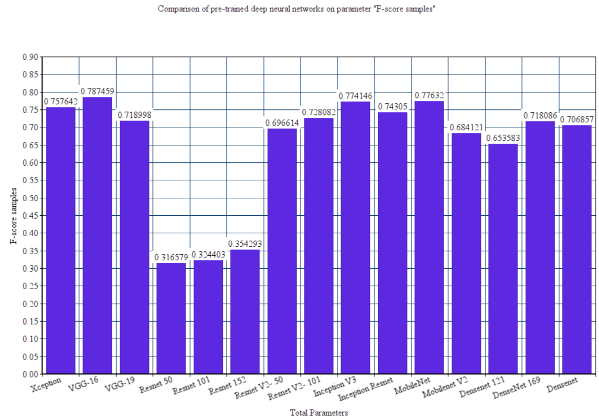


Fig.10: Comparison of pre-trained deep neural networks on parameter "F-score samples"

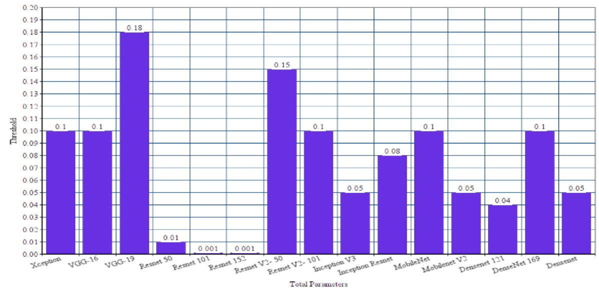


Fig.11: Comparison of pre-trained deep neural networks on parameter "Threshold"

4. RESULTS AND DISCUSSION (COMPARISON OUTCOME)

Above is the latex text for the comparison of different CNN models in terms of their accuracy. This text gives us proper idea of what model is to be selected for performing multi label classification. It has details of how accurate each model is, number of parameters used and the depth meaning the size of network use by the model and also space in MB. The text says, MobilenetV2 has the least size of 14 MB but when it comes to accuracy, it is lesser compared to other models. InceptionResNetV2 has big size of its architecture and also highest accuracy but is very complex. When it comes to ResNet models, they have big size as far as architecture is concerned but gives better accuracy compared to other models.

5. CONCLUSION

This work studies about performance analysis of high resolution satellite images on various Convolutional neural network (CNN). It simplifies approach to classify multi-label images using a trained CNN classifier. The training time for CNN is very fast. The main problem of classifying multi label images is that, if CNN is used for classifying

```
\begin{table}[]
\begin{tabular}{llllll}
Model              & Size     & Top-1 Accuracy  & Top-5 Accuracy  & Parameters    & Depth \\
Xception           & 88 MB    & 0.790           & 0.945           & 22,910,480    & 126   \\
VCG16              & 548 MB   & 0.713           & 0.901           & 138,357,544   & 23    \\
VCG19              & 549 MB   & 0.713           & 0.900           & 143,667,240   & 26    \\
ResNet50           & 98 MB    & 0.749           & 0.921           & 25,636,712    & -     \\
ResNet101          & 171 MB   & 0.764           & 0.928           & 44,707,176    & -     \\
ResNet152          & 232 MB   & 0.766           & 0.931           & 60,419,944    & -     \\
ResNet50V2         & 98 MB    & 0.760           & 0.930           & 25,613,800    & -     \\
ResNet101V2        & 171 MB   & 0.772           & 0.938           & 44,675,560    & -     \\
InceptionV3        & 92 MB    & 0.779           & 0.937           & 23,851,784    & 159   \\
InceptionResNetV2  & 215 MB   & 0.803           & 0.953           & 55,873,736    & 572   \\
MobileNet          & 16 MB    & 0.704           & 0.895           & 4,253,864     & 88    \\
MobileNetV2        & 14 MB    & 0.713           & 0.901           & 3,538,984     & 88    \\
DenseNet121        & 33 MB    & 0.750           & 0.923           & 8,062,504     & 121   \\
DenseNet169        & 57 MB    & 0.762           & 0.932           & 14,307,880    & 169   \\
DenseNet201        & 80 MB    & 0.773           & 0.936           & 20,242,984    & 201   \\
\end{tabular}
\end{table}
```

multiple labels then there is a chance that the network will be very complex and it will need a lot of time for training. Thus exploring easy and promising functionalities will also require a lot of time. And hence this study helps in knowing different CNN architecture using single label image classification and their comparison which makes multi label image classification even simpler with lesser errors and more accuracy. Depending on the comparison, Multi label image classification can be performed on the discussed CNN architecture depending on their advantages and disadvantages, giving rise to further exploration of image classification in near future.

6. ACKNOWLEDGEMENTS

We express our gratefulness to honorable Principal Mr. S. P. Kallurkar Sir for encouragement and facilities provided to us and we would like to express our deep sense of gratitude to Prof.SuvarnaPansambal, Head of Department of Computer Engineering, Atharva College of Engineering, Mumbai, Prof.RanjitaAsati for her generous support and suggestions.

7. REFERENCES

[1] ErchanAptoula, "Remote detecting picture recovery with worldwide morphological surface descriptors", IEEE exchanges on geoscience and remote detecting, 52(5):3023–3034 , 2013.

[2] Yi Yang and Shawn Newsam,"Sack of-visual-words and spatial augmentations for landuse grouping", In Proceedings of the eighteenth SIGSPATIAL worldwide meeting on progresses in geographic data frameworks, pages 270–279. ACM, 2010. QianWeng, Zhengyuan Mao, Jiawen Lin, and WenzhongGuo,"Land-use grouping by means of outrageous learning classifier dependent on profound convolutional highlights", IEEE Geoscience and Remote Sensing Letters, 14(5):704–708, 2017.

[3] [Lionel Gueguen,"Ordering compound structures in satellite pictures: A compacted portrayal for quick inquiries", IEEE Transactions on Geoscience and Remote Sensing, 53(4):1803–1818, 2014.

[4] Nanjun He, Leyuan Fang, Shutao Li, and Antonio J Plara, "Covariance grid based element combination for scene order", In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pages 3587–3590. IEEE, 2018.

[5] Sandeep Kumar, Swati Jain, and TanishZaveri,"Equal way to deal with speed up morphological element extraction of remote detecting pictures for cbir framework", In 2014 IEEE Geoscience and Remote Sensing Symposium, pages 2471–2474. IEEE, 2014.

[6] Jiayuan Fan, Tao Chen, and Shijian Lu,"Unaided component learning for landuse scene acknowledgment", IEEE Transactions on Geoscience and Remote Sensing, 55(4):2250–2261, 2017.

[7] Francois PS Luus, Brian P Salmon, Frans Van lair Bergh, and BodhaswarTikanathJugpershadMaharaj,"Multiview profound learning for land-use arrangement", IEEE Geoscience and Remote Sensing Letters, 12(12):2448–2452, 2015.

[8] Radamanthys Stivaktakis, GrigoriosTsagkatakis, and PanagiotisTsakalides,"Profound learning for multilabel land

spread scene classification utilizing information expansion", IEEE Geoscience and Remote Sensing Letters, 2019.

[9] Hang Wu, Baozhen Liu, Weihua Su, Wenchang Zhang, and Jinggong Sun,"Profound channel banks for land-use scene arrangement", IEEE Geoscience and Remote Sensing Letters, 13(12):1895–1899, 2016.

[10] Fan Zhang, Bo Du, and Liangpei Zhang,"Scene grouping by means of an angle boosting arbitrary convolutional arrange structure", IEEE Transactions on Geoscience and Remote Sensing, 54(3):1793–1802, 2015.

[11] Fan Zhang, Bo Du, and Liangpei Zhang,"Saliency-guided unaided element learning for scene arrangement", IEEE exchanges on Geoscience and Remote Sensing, 53(4):2175–2184, 2014.

[12] John Ray Bergado, Claudio Persello, and Alfred Stein, "Intermittent multiresolution convolutional systems for vhr picture characterization", IEEE exchanges on geoscience and remote detecting, (99):1–14, 2018.

[13] Zhiqiang Gong, Ping Zhong, Weidong Hu, and Yuming Hua,"Joint learning of the middle focuses and profound measurements for land-use grouping in remote detecting", Remote Sensing, 11(1):76, 2019.

[14] Cong Cao, SuzanaDragi'cevi'c, and Songnian Li,"Landuse change recognition with convolutional neural system strategies", Conditions, 6(2):25, 2019.

[15] Rodrigo Minetto, Maur'ıcio Pamplona Segundo, and Sudeep Sarkar, "Hydra: an outfit of convolutional neural systems for geospatial land characterization", IEEE Transactions on Geoscience and Remote Sensing, 2019.

[16] Harjot Singh Parmar, "Land use grouping utilizing multineighborhoodlbps",arXiv preprint arXiv:1902.03240, 2019.

[17] Konstantinos Karalas, Grigorios Tsagkatakis, Michael Zervakis, and Panagiotis Tsakalides,"Land order utilizing remotely detected information: Going multilabel", IEEE Transactions on Geoscience and Remote Sensing, 54(6):3548–3563, 2016.

[18] Dimitrios Marmanis, MihaiDatcu, Thomas Esch, and UweStilla,"Profound learning earth perception order utilizing imagenetpretrained systems", IEEE Geoscience and Remote Sensing Letters, 13(1):105–109, 2015.

[19] Grant J Scott, Matthew R England, William A Starms, Richard A Marcum, and Curt H Davis,"Preparing profound convolutional neural systems for land–spread grouping of high-goals symbolism", IEEE Geoscience and Remote Sensing Letters, 14(4):549–553, 2017.

[20] Jun Li, Daoyu Lin, Yang Wang, GuangluanXu, and Chibiao Ding, "Profound discriminative portrayal learning with consideration map for scene arrangement",arXiv preprint arXiv:1902.07967, 2019.

[21] Yumei Zhen, Huanyu Liu, Junbao Li, Cong Hu, and Jeng-Shyang Pan, "Remote detecting picture object acknowledgment dependent on convolutional neural system", In 2017 First International Conference on Electronics Instrumentation and Information Systems (EIIS), pages 1–4. IEEE, 2017.

[22] Eliezer Flores, MacielZortea, and Jacob Scharcanski,"Word references of profound highlights for land-use scene characterization of extremely high spatial goals pictures", Example Recognition, 89:32–44, 2019.

[23] Nagesh Kumar Uba,"Land use and land spread order utilizing profound learning systems",arXiv preprint arXiv:1905.00510, 2019.

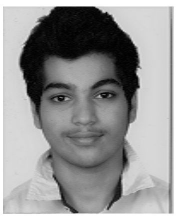
[24] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "Land use characterization in remote detecting pictures by convolutional neural systems",arXiv preprint arXiv:1508.00092, 2015.

[25] Liangpei Zhang, Lefei Zhang, and Bo Du. Profound learning for remote detecting information,"A specialized instructional exercise on the best in class", IEEE Geoscience and Remote Sensing Magazine, 4(2):22–40, 2016.

AUTHORS

Raj C Shah, Final Year BE Student, Department of Computer Engineering, Atharva College of Engineering.

SagarPatil, Final Year BE Student, Department of Computer Engineering, Atharva College of Engineering.



Anish Malhotra, Year BE Student, Department of Computer Engineering, Atharva College of Engineering.

RanjitaAsati, Assistant Professor, Department of Computer Engineering, Atharva College of Engineering.

