

Article

# Multi-Granularity Neural Network Encoding for Land Cover and Land Use Image Classification

WANG Gouyin<sup>1</sup>, MUSABE Jean Bosco<sup>1</sup>, and Hategekimana Yves<sup>2</sup>

- 1 Chongqing Key Laboratory of Computational Intelligence, Department of Computer Science and Technology, Chongqing University of Posts and Telecommunication, 2 Chongwen Rd, Huangjueya, Nan'an District, Chongqing, China, 400065; l201710001@stud.cqupt.edu.cn, wanggy.cq@cqupt.edu.cn(WG).
- 2 Rwanda Space Agency; Kigali City-Rwanda; yhatekekimana@space.gov.rw(YH)  
1 Correspondence: l201710001@stud.cqupt.edu.cn

**Abstract:** Deep learning classification is the state-of-the-art of machine learning approach. Earlier work proves that the deep convolutional neural network has successfully and brilliantly in different applications such as images or video data. Recognizing and clarifying the remote sensing aspect of the earth's surface and exploit land cover land use (LCLU). This article summarized the remote sensing emerging application and challenges for deep learning methods. We propose and examine four ways to learn efficient and effective CNNs to transfer image representation on the ImageNet dataset to recognize LCLU datasets. We use VGG16, Inception-ResNet-V2, Inception-V3, and DenseNet201 models pre-trained on the ImageNet dataset to extract features from the EAC dataset. We use pre-trained CNNs on ImageNet to extract features. The essential thing of our study is that we used principal component analysis (PCA) for feature selection to improve accuracy and speed up the model. We train our model by multi-layer perceptron (MLP) as a classifier. Lastly, we apply the multi-granularity encoding ensemble model. We achieve an overall accuracy of 92.3% for the nine-class classification problem. This work will help remote sensing scientists understand deep learning tools and apply them in large-scale remote sensing challenges.

**Keywords:** Multi-granularity neural networks encoding (MGNNE), multi-classification feature representation, Multi-layer perceptron (MLP), Principal component analysis (PCA), Remote Sensing Application.

---

## 1. Introduction

Land cover is the physical material of the earth's surface[1], has been using in many practical applications such as environmental protection[2], urban planning[3], economic resource management[4]. Land cover and land use (LCLU) are issues in remote sensing areas to organize the earth's land for actual utilization. In a modern economic growth, land cover and land use are broadly defined to include all the natural provides including minerals, forest products, water, and other land resources, and they can support economic development. By growing economies and improving living conditions, the accessibility to public services makes the main challenges effect of land cover, such as environmental degradation, climate change, deforestation, increase demand for public service, and pressure on public infrastructure. The massive amount of remote sensing datasets in the earth observation increased in the significant data era. To fight against the challenges of the enormous amount of datasets, the applications of the multi-granularity neural network encoding approach for LCLU classification using remote sensing imagery datasets have become increasing in recent. Deep learning methods have been made significant progress in computer vision and artificial intelligence for different tasks such as scene

classification[5], change detection[6], and semantic segmentation[7]. Based on sensing multi-granularity land use and land cover transformation of functional land structures and mapping land use pattern is caused not only by bottom-up residential influence and top-down influence from movement. The change of land use and land cover patterns is accelerating in different land support actions. The large-scale satellite imagery introduces new challenges to parse the earth through satellite images that need a contribution to deep learning[8], which includes different competition for land use applications[9]. Remote sensing data present some new challenges for deep understanding because satellite image analysis extending striking gaps such as data are passive multi-model from optical (multi-band and hyperspectral), Lidar [10]. Synthetic aperture radar (SAR) image segmentation, where the imaging geometries and content are entirely different, and unlike imaging modalities, missing of three dimensional (3-D), additional convolution of decision fusion[11]. Satellite imagery is measured by determining the complex relationship within and between spatial, spectral, and temporal dimensions. Classification of complex land cover mapping using multispectral for the wetland is not enough because it compares wetland vegetation. Many spatial and spectral classification experiments studies are based on the conditional random field model, Markov random field model[12], and experimentation of a composite kernel for hyperspectral image classification. Deep learning has improved computer vision tasks based on different machine learning tools[13] to learn satellite imagery to identify patterns in urban environments at a large scale and explain satellite dataset challenges[14]. Machine learning methods emphasize shallow convolution structures' limitations, such as support vector machines (SVM) and random forests (RF).

Current deep learning approaches such as Convolutional Neural Networks (CNNs)[15] by grouping several convolutional layers trained up to the small patch of local receptive range as the deepest input the hierarchical structure. In some image classification problems such as scene classification[16],[17],land change detection[18], object detection[19], and brain image analysis[20]. deep convolutional neural networks (DCNNs) have been in view powerful presentation by extracting multi-source and multi-level features with hierarchical representation[21],[22],[23],[24]. It used deep convolutional neural networks (DCNNs) methods to implement data-driven in remote sensing scene classification problems such as large intra-class variance[25]and information confused from large geographical locations [16].

This work establishes the Multi-granularity Neural Network encoding based on ImageNet feature extraction and multi-granularity transformation, which outputs images within different granularity levels. Our classification accuracy improved. We introduce Inception-ResNet-V2, InceptionV3, VGGNet, and DenseNet201, consisting of weighted layers in the feature learning-based representation approach of two awarding methods of the transfer pre-trained feature extraction and feature selection through PCA (principal component analysis). Furthermore, by the autoencoder, the characteristics approach is applied on the top of the representation feature learned from DCNNs pre-trained based features to improve speed up and performance. By grain data augmentations with the objective of the segmentation, the grain is due to the batching method. After multi-granularity feature extraction, MLP has been applied to classify final output and presents overall accuracy and

confusion matrix to distinguish which categories are intra-class diversity and information are confused.

The remainder of this paper is organized as follows. Section II presents the land cover (LCLU) application and its challenges. Section III offers a multi-granularity neural network based on feature extraction CNN methods for object recognition tasks to improve our classification accuracy, experiment, and results in Comparison. Section IV, we concluded, and further work was summarized.

## 2. Related Work

### 2.1 Remote Sensing Application

The idea behind this is just multi-class classification at the pixel level [26], we want to assign one of the ends, and these inputs might be represented, pixels might be represented either as a spectral reflectance curve, could be a temporal time series curve, or we often see images being represented a patch where we want to predict the central pixel. Land cover and land use classification mainly get the labels for these models from national databases such as the USGS to publish the national land cover datasets NLCD. The USDA produces a cropland data layer every year. These are used as the ground truth. Still, in many places that do not have a national program for collecting these data, we have researchers conducting field campaigns to acquire that data or make photo interpretations from very high-resolution images[27]. The most common types of machine learning methods that are used for this right now are decision tree (DT) and random forests (RF) for supervised classification[9]. Still, increasingly deep learning approaches are employed for land cover and land use classified. These are typically convolutional neural networks in 1-D, especially when looking at the [12]. Many recurrent neural networks (RNN), precisely long short term memory (LSTMs) heard earlier today, are used for classifying time series[28]. Image level change detection wants to see meaningful change in surface features that were often more object-based or deep learning approaches. We compare temporal images at the feature level rather than the pixel level. For example, we should apply an autoencoder neural network to extract the latent representing the most salient features from two different images and compare them in the feature space to see if something meaningful changes. To ignore irrelevant variations such as lighting conditions or miss registration.

### 2.2 Scene Classification

Image classification and machine learning, becoming more popular as images of higher resolution that we are collecting from these remote sensing satellites[29] where pixels may be used to represent an entire field or entire class types, we have more detailed images where we might be more interested in finding more neurons classes and these methods are enabled by deep learning, in particular convolutional neural networks (CNNs). One challenge in scene classification for remote sensing is that there is less labeled training data than pixel-level land cover classification. At the pixel level, we might have thousands of examples just in this one image. Still, for image classification, we have much fewer. So for that reason, many researchers are doing things like transfer learning, pre-training models

on larger image datasets, ImageNet, or even benchmark remote sensing datasets and transferring these models to fine-tune them for a different task[30]

### *2.3 Novelty or Anomaly Detection*

Another common area of machine learning for remote sensing is anomaly detection<sup>[25]</sup>. The unseen detection pattern, outliers defined as spatial may be morphological spectral outliers or temporal outliers. It is a kind of tough to do with machine learning typically, because looking for large-scale ways and not so much the things that we only see a few of and what anomaly detection is suitable for and either unsupervised methods or on class supervised process by using machine learning to characterize the regular type. Then we look at deviations from that common representation to identify anomalies. The majority of machine applications are for novelty detection are using read the [34] detected method that computes pixelized anomaly scores, that is, the Mahalanobis distance[9] between a single pixel and a background distribution, which could define the background to be a window around the pixel the rest of the image or even an entire dataset of typical images. The most common methods for anomaly detection in the machine learning literature are reconstruction-based methods where minimizing model reconstruction of not common examples such that this distance reconstruction error will be significant for the novel[34].

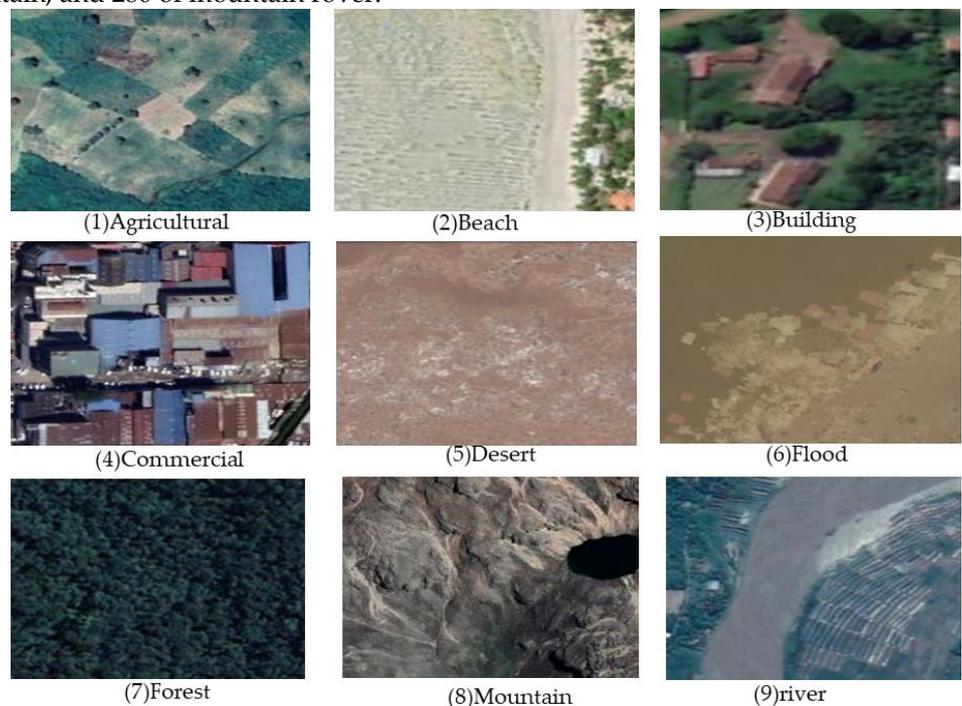
### *2.4 Estimation of Physical Quantities*

There are many applications for this on earth and another common application for regression or estimating physical quantities from remote sensing. An example very commonly used in agriculture for estimating yields[35] for different types of crops from remote sensing data such as maize yield being assessed. The second example is Greg Adler's lab[36], where they evaluate the growth carbon density directly from remote sensing data. These types of models are often combining multiple data sources to predict these variables. Most scientists use regression trees and peat world feed-forward neural networks and process models like Gaussian processes. However, increasingly again seeing convolutional neural networks (CNNs) and Recurrent Neural Networks (LSTMs) being used[2]. We summarized section two by saying remote sensing is non-trivial preprocessing and cleaning. Many issues represent time series or describe the data to account for all the variance across pixels in the same image, co-registration of ideas, aligning images from two different periods, or taking it to another time is a significant challenge. Second, labeling remote sensing images often requires domain expertise. For example, it might be self-evident for somebody on a mechanical truck to mark cat versus togs for 100% images. This is different for labeling things like impact craters or geologic features. Input remote sensing sources have different physical units. Lastly, different illumination conditions which are more of a problem on other planets where on earth, we kind of can get regular crossing times by choosing our orbits and imaging regularly, it is still a problem on land, and it is undoubtedly on other planets where you see how in these images showing the same site. After reviewing different

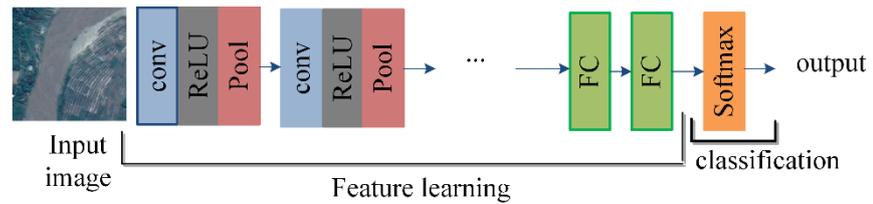
### 3. Methods and Materials

#### 3.1 Study Area and Data

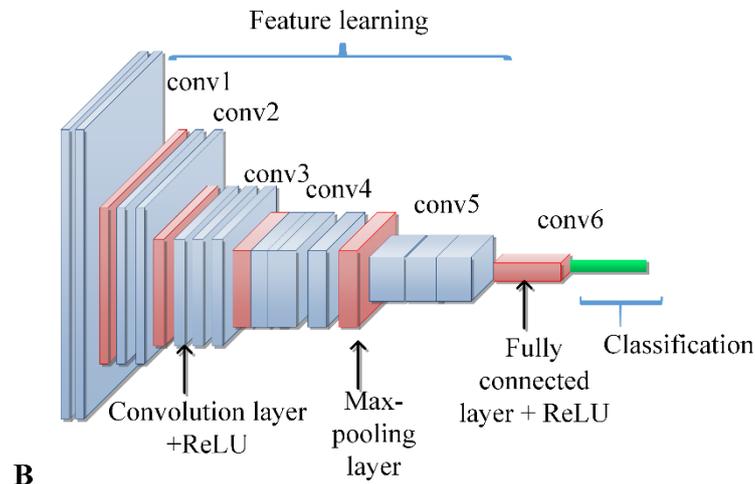
Dataset used in this experiment were collected from Easter Africa Community in six countries (Rwanda, Uganda, Kenya, Tanzania, South-Sudan, and Burundi) by using the google earth pro. Software connected with universal map of Google satellite map. After, downloading quality and higher resolution imagery from different area of EAC, we clean it and resizing 256×256 pixels in the RGB space by using python3.7. EAC-Dataset is a novel Land Use Data Set that contains 2110 images, we divide into nine object categories such as agricultural and forest, we list each scene category in as fellow agricultural, beach, buildings, commercial, desert, flood, forest, mountain, and river . First, we introduce the experimental dataset; second, we evaluate indexes and testing procedures by constructing an experimental dataset of high-resolution remote sensing images. Figure1 indicates different images located to each class whereby agriculture has 228 images, 249 of beach, 202 of building, 193 of commercial, 151 of a desert, 290 of a flood, 268 of a forest, 251 of a mountain, and 280 of mountain rover.



**Figure 1.** This is a sample image from the Easter Africa Countries (EAC.)



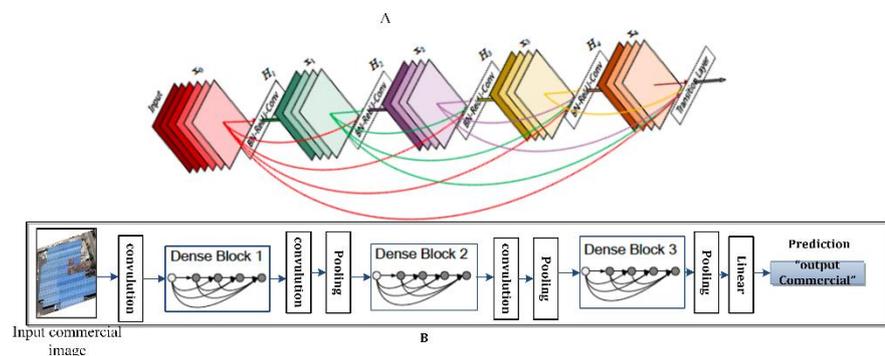
A



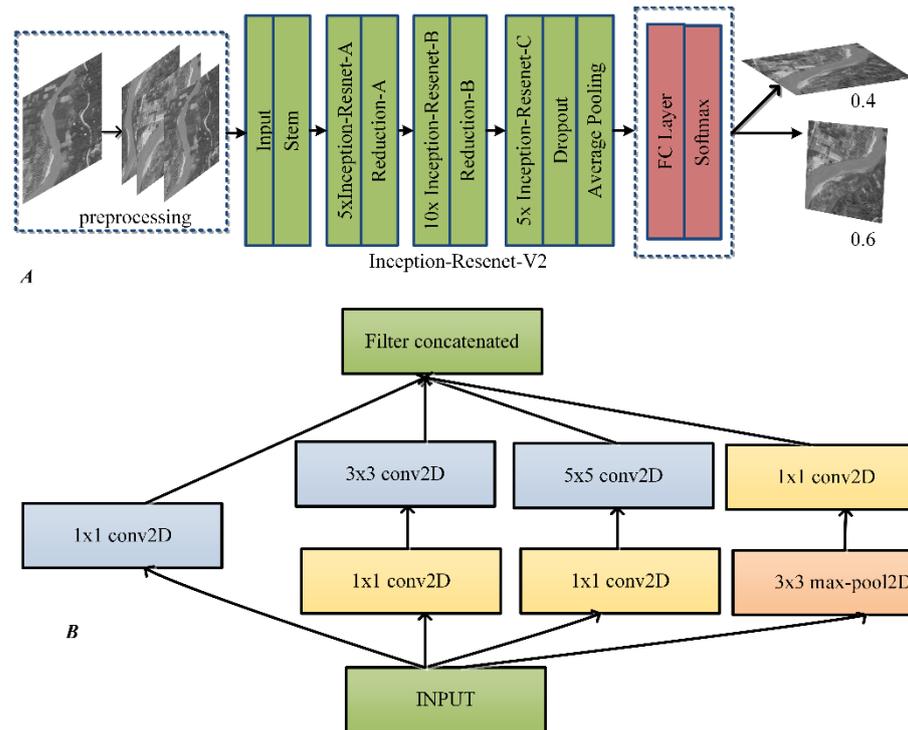
B

**Figure 2.** High-level granularity hierarchical representation, A) represent the most principle layers of CNN model while B) present VGG-16 architecture for feature learning method.

We provide a set of state-of-the-art deep learning models along with pre-trained weights on ImageNet. These pre-trained models can be used for image classification, feature extraction, and transfer learning. Our goal here we were clustering a subset of EAC dataset images. Our approach consists of two major components: using a pre-trained model VGG-16 to extract the feature of given image classes. We used PCA to speed up and define input features extracted, dense layer, and hidden layers. We used softmax to evaluate the performance accuracy classification method within the same process with inception-resentv2 and DenseNet201.

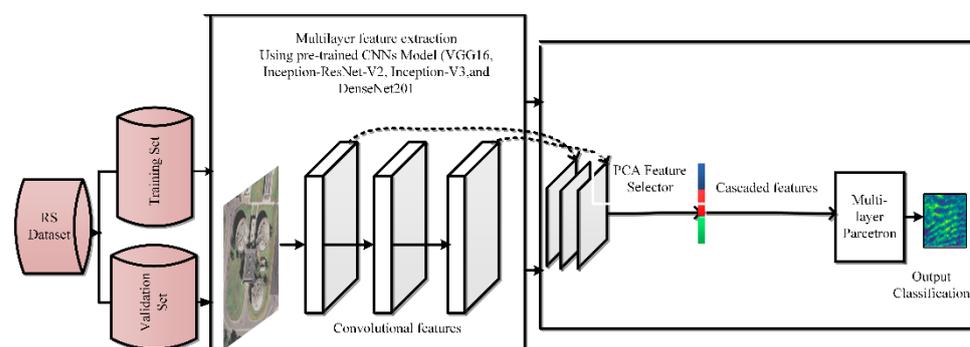


**Figure 3.** Dense Net architecture for the feature extraction, A) indicates five-layer Dense Net Block while B) is a Dense Net for three Dense Block. Transition layers concatenate preceding feature maps from each layer in a block.



**Figure4.** Structure of Inception-ResNetV2, the number of parameters increases in some layers.

As we see above figure4, the whole network expanded, this network of Inception-ResNet-V2 considerably deeper than the InceptionV3. The inception-ResNet-v2 framework is a convolutional neural architecture that constructs the inception-friendly architectures but incorporates residual connection, the filter concatenation stage of the inception architecture. It is more accurate than the previous state-of-the-art models. We use weights trained on ImageNet as feature extracted; after extracted features using Inception-Resner-v2 ImageNet weight, we reduce dimensional using PCA to increase speed and overcome computational cost. We applied MLP to define the final model effectively and efficiently for each concatenated feature selected.



**Figure5.** Our proposed multi-granularity feature extractor is for improving classification accuracy and speed up the computational model.

### 3.2 Feature Learning Based on Pre-trained weighs

The most important thing in computer vision is to take a model trained on a vast dataset and run it on your own, smaller dataset and extract the features that the model generates. Our first model for feature extraction, VGG, is a convolutional neural network model for image recognition proposed by [43], figure2 B) illustrates the architecture of VGG16, the input layer takes an image in the size of

224x224x3, and the output layer is a softmax prediction on nine classes based on our dataset. Based on ImageNet, the output layer is also softmax prediction on 1000 classes. From the input layer to the last max-pooling layer regarded feature extraction part of the model, the rest of the network is considered the classification part model. We applied the same steps to the other CNNs architectures.

### 3.3 *Principal Component Analysis (PCA)*

PCA is a non-parametric method of extracting relevant information from data by reducing complex data to a lower dimension. It is also a standard statistical technique that helps to find patterns in data of large sizes. It has application in image processing and objects recognition for subspace face recognition. It reduces the dimensionality of large data sets by transforming an extensive collection of variables into smaller ones containing most of the information in the large stage. To reduce the number of variables of a data set naturally comes at the expense of accuracy. Still, the trick in dimensionality reduction is to trade a little accuracy for simplicity. The smaller dataset is easier to explore and visualize and makes analysis data much easier and faster for machine learning algorithms without extraneous variables to process. It is an unsupervised learning method for dimensionality reduction. We dropped to 1000 dimensions, and we found that the dimensionality reduction score points, which shows that our PCA can reduce the dimensionality of the data while keeping the original characteristics of the data as much as possible. Moreover, lessen the subsequent clustering operation time between 15-22s. The effect of looking at the score is good.

### 3.4 *Implementation of Classifiers*

Our implementation is based on the Tensor Flow-GPU1.14.0 framework, Keras, python 3.7, and tested on a performance computer (Intel ® Core™ i5 CPU, 8GB RAM) equipped with an NVIDIA-GTX 1070 with 8GB of memory. First, we split our dataset randomly into two independent data subsets with 80% and 20% for training and testing. Then we normalized the dataset and applied one-hot encoding to facility multi-granularity neural network encoding. The trainable classifier block parameters are initialized randomly to random values. We used the rmsprop optimizer as a non-adaptive optimizer, optimizing the neural network for absolute accuracy or multi-class. An accuracy performance metric can be decisive when dealing with imbalanced data. In the step of training, the pre-trained feature extraction parameter becomes frozen to their initial values. To make a comparison, the number of neural in trainable layers and dropout values remain the same. After that, the feature extracted from CNNs was selected to add it between the feature extraction block and the classifier block of the convolutional neural network. The selection features using PCA (Principal Component Analysis), these selected features act as an input to the multi-layer perceptron (MLP) classifier. The final step was a multi-layer perceptron (MLC) neural network whereby is a class of feed-forward artificial neural consists of at least three layers of nodes, input layer, a hidden layer, and an output layer. It applies in different applications such as classification and regression problems. In MLP, we use three hidden layers, input, one dense, and output layer. Then we used an error backpropagation algorithm to optimize the network parameters with a learning rate of 0.3 on the training dataset. At the end of the process, after training of MLP, we get the confusion matrix for both training and test datasets, and we get the performance measures of overall accuracy. In this work, we learned about measuring different models are accuracy percentage of optimistic or correct predictions. The confusion matrix for the EAC dataset was split randomly into two independent data subsets with 80% and 20% for training and testing.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

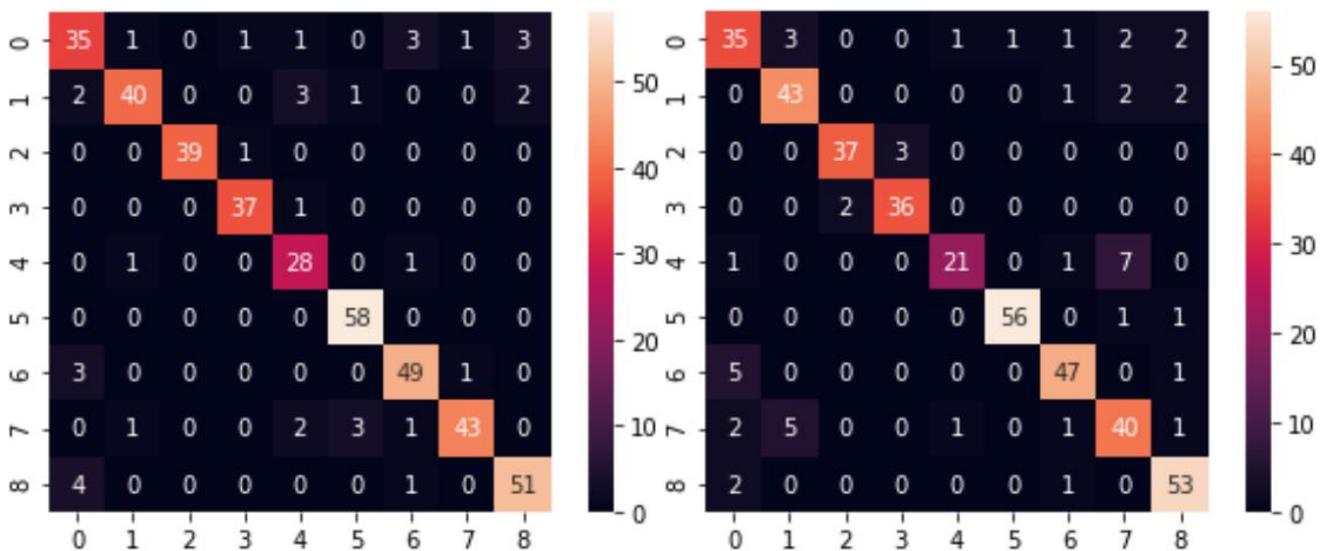
$$\text{FNR} = \frac{FN}{P} = \frac{FN}{FN + TP} \quad (4)$$

$$\text{FPR} = \frac{FP}{N} = \frac{FP}{FP + TN} \quad (5)$$

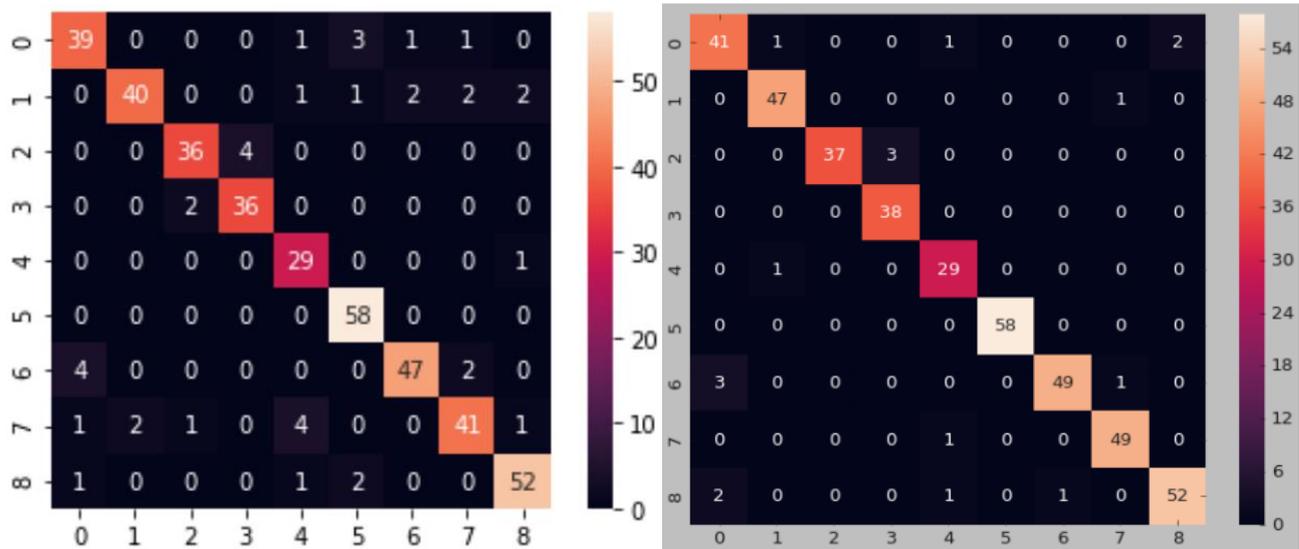
$$\text{cross-entropy}(x) = -\sum y \log(x) \quad (6)$$

$$\text{soft max} = \frac{\exp(x_i)}{\sum \exp(x)} \quad (7)$$

The loss could be calculated in many different ways: softmax and cross-entropy equation (6).



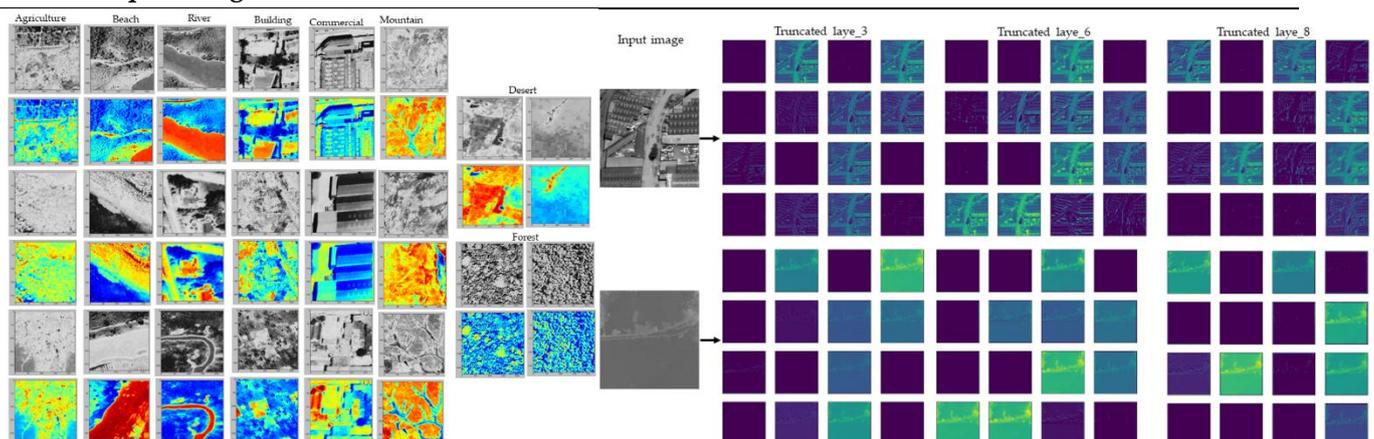
**Figure 6.** Two figures, left is confusion matrix for MLP classifier used VGG16 feature extracted and PCA as feature selection, right represent confusion matrix of Inception-ResNet-V2 pre-trained elements using PCA as feature selected to train MLP



**Figure 7.** Two figures left is the Confusion matrix of DenseNet201 pre-trained features using PCA as feature selected to train MLP. Right is the confusion matrix of VGG16-DenseNet201 pre-trained elements using PCA as a feature set to train MLP.

Table 1. Comparison accuracy for multi-class based EAC dataset

Model	Execution time with PCA/second	OA Acc.	Pre-trained param	MLP Param#
VGG16+PCA	11.6646	90.90%	14714688	424,969
Inception V3+PCA	12.4267	87.79%	21,802,784	424,969
Inception ResNet v2+PCA	24.3376	88.03%	54336736	424,969
DenseNet201+PCA	33.6673	90.43%	18,321,984	424,969
Ensemble	17.2245	90.88	14,848,329	424,969
VGG16+DenseNet201+PCA		92.30%	-	-



**Figure 7.** Sample of feature map visualizing of MG-ENN (VGG16) on EAC dataset. The first figure indicates the output of the ensemble encoding neural network for random images. The second figure demonstrates that two images are randomly selected during testing and used to generate feature maps, and the network is established from different granularity levels of networks.

#### 4. Discussion

The multi-granularity neural network encoding (MNNE) model comes from three concepts: sparse interaction, parameter sharing, and equivariant representation. Some ideas transform to the network configuration demonstrates in Figure 2A. The network may be viewed mainly of two feature learning processes: convolution, nonlinearity, and pooling stages incorporated to the fully connected stage to extract the features in deep. Our approach aims to study D-CNNs for classification imagery to address the gap of class similarity and dissimilarities between classes. The CNN has more powerful granular feature extraction to enhance the achievement of state-of-the-art deep learning methods. To build MNNE model training, we used four models to extract features such as pre-trained VGG16, InceptionResNetV2, Inception-V3, and DenseNet201 ImageNet weights to train and test parts. It explained the process of using pre-trained weights as feature extractors for both neural networks. After the feature is extracted, we use PCA (Principal component analysis) for dimensions reduction and feature selection. First, we verified the ideal number of PCA components not to lose much information, and we remind that the  $n_{\text{components}}$  must be lower than the number of features. Then we pick the optimal number of members; this is how many features we would have for our machine learning. After selecting features, we applied a multi-layer perceptron with three hidden dense layers, ReLU as activation and softmax. We used adaptive learning rate gradient decent RMSPROP and absolute accuracy for regularization within existing metrics learning to learn discriminative feature representations. The purpose of PCA is to reduce the number of features for speed training. It is a beneficial step for visualizing and processing high-dimensional datasets, still retaining as much of the variance in the dataset as possible.

We discussed and analyzed different contributions from ImageNet challenges in remote sensing tasks. The first critical aspect of classification for full convolution (FCN) encodes Net. Convolution neural network, which is maintained computational cost increase resolution of the intermediate feature map. This approach is smaller and can be added to any model, takes away for thought: parameterized everything including high-order hyper-parameters. After summarizing different remote sensing applications, we present our work using pre-trained models for feature extraction in image classification. We investigated the performance of VGG16 InceptionResNet-V2, Inception-V3, and DenseNet201 as feature extractors under internal classify validation using PCA for feature selection. Image classification is an exciting challenge. Lastly, we selected features and concatenating to feed multi-layer perceptron to improve final results.

#### 5. Conclusions

In this work, we investigate the emerging application of remote sensing. Also, we proposed four different approaches for the EAC dataset scene classification. For the extraction of features using pre-trained and PCA, we visualized and plotting of figure. The traditional and classical convolution neural network model fails to use multi-source satellite imagery. It has been used for RGB bands or small numbers of the dataset. The classic analyzing digital image is based on single granularity and with limited accuracy. Due to the granular computing with multiple delicate layers for brain big data processing and the data-driven granular cognitive computing, this framework supports the combination of multi-granularity and convolutional neural networks on distributed multi-granularity encoding module. Our results indicate that these methods could prove helpful to classify data of any domain with a smaller dataset or with a limited dataset. We suggested this design framework to increase the helpfulness. Further work, we would investigate multi-granularity cascading cross the channel parameters pooling for large remote sensing imagery.

**Author Contributions:** MJB and WG were involved in conceptualizing the problem and identifying the multi-granularity neural network encoding, designed the research methods, and worked on the entire framework of this research, data collection and comparison experiment. YH involved in the revision of the paper.

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 61772096), the Foundation for Innovative Research Groups of Natural Science Foundation of Chongqing (No. cstc2019jcyj-cxttX0002).

**Acknowledgment:** The authors would like to thank the handling editor and anonymous reviewers for their reading and remarks.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] C. Yang, F. Rottensteiner, and C. Heipke, "CLASSIFICATION of LAND COVER and LAND USE BASED on CONVOLUTIONAL NEURAL NETWORKS," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 4, no. 3, pp. 251–258, 2018, DOI: 10.5194/isprs-annals-IV-3-251-2018.
- [2] J. M. Deines, S. Wang, and D. B. Lobell, "Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt," *Environ. Res. Lett.*, vol. 14, no. 12, 2019, DOI: 10.1088/1748-9326/ab503b.
- [3] Y. Qian and Z. Wu, "Study on urban expansion using the spatial and temporal dynamic changes in the impervious surface in Nanjing," *Sustain.*, vol. 11, no. 3, 2019, DOI: 10.3390/su11030933.
- [4] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, 2020, DOI: 10.3390/RS12152495.
- [5] G. Chen *et al.*, "Training small networks for the scene classification of remote sensing images via knowledge distillation," *Remote Sens.*, vol. 10, no. 5, 2018, DOI: 10.3390/rs10050719.
- [6] D. A. J. Quispe and J. Sulla-Torres, "Automatic building change detection on aerial images using convolutional neural networks and handcrafted features," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 6, pp. 679–684, 2020, DOI: 10.14569/IJACSA.2020.0110683.
- [7] M. Alam, J. F. Wang, C. Guangpei, L. Yunrong, and Y. Chen, "Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images," *Mob. Networks Appl.*, 2021, DOI: 10.1007/s11036-020-01703-3.
- [8] M. Rezaee, M. Mahdianpari, Y. Zhang, and B. Salehi, "Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 9, pp. 3030–3039, 2018, DOI: 10.1109/JSTARS.2018.2846178.
- [9] S. Talukdar *et al.*, "Land-use land-cover classification by machine learning classifiers for satellite observations-A review," *Remote Sens.*, vol. 12, no. 7, 2020, DOI: 10.3390/rs12071135.
- [10] C. Szigarski *et al.*, "Potential Improvements," pp. 1–15, 2018, DOI: 10.3390/rs10111776.
- [11] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional CNN framework for land cover/land-use change detection in multi-source remote sensing datasets," *Remote Sens.*, vol. 12, no. 12, 2020, DOI: 10.3390/rs12122010.
- [12] S. Zhang *et al.*, "EMMCNN: An ETPS-based multi-scale and multi-feature method using CNN for high spatial resolution image land-cover classification," *Remote Sens.*, vol. 12, no. 1, pp. 1–33, 2020, DOI: 10.3390/RS12010066.
- [13] K. Tong, Y. Wu, and F. Zhou, "Recent advances in small object detection based on deep learning: A review," *Image Vis. Comput.*, vol. 97, p. 103910, 2020, DOI: 10.1016/j.imavis.2020.103910.
- [14] A. Albert, J. Kaur, and M. C. Gonzalez, "Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1296, pp. 1357–1366, 2017, DOI: 10.1145/3097983.3098070.
- [15] J. Lee, D. Han, M. Shin, J. Im, J. Lee, and L. J. Quackenbush, "Different spectral-domain transformation for land cover classification using convolutional neural networks with multi-temporal satellite imagery," *Remote Sens.*, vol. 12, no. 7, 2020, DOI: 10.3390/rs12071097.
- [16] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, no. April, pp. 166–177, 2019, doi: 10.1016/j.isprsjprs.2019.04.015.
- [17] B. Petrovska, E. Zdravevski, P. Lameski, R. Corizzo, I. Štajduhar, and J. Lerga, "Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification," *Sensors (Switzerland)*, vol. 20, no. 14, pp. 1–22, 2020, DOI: 10.3390/s20143906.
- [18] D. Verma and A. Jana, "LULC classification methodology based on simple Convolutional Neural Network to map complex urban forms at finer scale: Evidence from Mumbai," *arXiv*, pp. 1–28, 2019.

- [19] A. Varghese, J. Gubbi, A. Ramaswamy, and P. Balamuralidhar, "ChangeNet: A deep learning architecture for visual change detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11130 LNCS, pp. 129–145, 2019, DOI: 10.1007/978-3-030-11012-3\_10.
- [20] J. Bernal *et al.*, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review," *Artif. Intell. Med.*, 2018, DOI: 10.1016/j.artmed.2018.08.008.
- [21] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience Remote Sens.*, vol. 57, no. 1, pp. 1–20, 2020, DOI: 10.1080/15481603.2019.1650447.
- [22] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-CapsNet," *Remote Sens.*, vol. 11, no. 5, 2019, DOI: 10.3390/rs11050494.
- [23] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, no. March 2018, pp. 430–443, 2019, DOI: 10.1016/j.rse.2018.11.032.
- [24] Y. Hu, Q. Zhang, Y. Zhang, and H. Yan, "A Deep Convolution Neural Network Method for Land Cover Mapping: A Case Study of Qinhuangdao, China," *Remote Sens.*, vol. 10, no. 12, p. 2053, 2018, DOI: 10.3390/rs10122053.
- [25] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing Images: A survey and a new benchmark," *arXiv*, 2019.
- [26] G. Wang, A. Li, G. He, J. Liu, Z. Zhang, and M. Wang, "Classification of High Spatial Resolution Remote Sensing Images Based on Decision Fusion," *J. Adv. Inf. Technol.*, no. January, pp. 42–46, 2017, DOI: 10.12720/jait.8.1.42-46.
- [27] D. Marcos, M. Volpi, B. Kellenberger, and D. Tuia, "Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 96–107, 2018, DOI: 10.1016/j.isprsjprs.2018.01.021.
- [28] Z. Zhang, G. Vosselman, M. Gerke, D. Tuia, and M. Y. Yang, "Change detection between multimodal remote sensing data using siamese CNN," *arXiv*, pp. 1–17, 2018.
- [29] C. Tao, W. Lu, J. Qi, and H. Wang, "for Scene Classification," pp. 1–5, 2020.
- [30] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, 2019, DOI: 10.1109/TGRS.2018.2864987.
- [31] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, 2019, DOI: 10.3390/rs11131578.
- [32] A. R. D. Putri, P. Sidiropoulos, and J. P. Muller, "Anomaly detection performance comparison on anomaly-detection based change detection on martian image pairs," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 2/W13, pp. 1437–1441, 2019, DOI: 10.5194/isprs-archives-XLII-2-W13-1437-2019.
- [33] M. Campos-Taberner *et al.*, "Understanding deep learning in land use classification based on Sentinel-2 time series," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020, DOI: 10.1038/s41598-020-74215-5.
- [34] H. R. Kerner *et al.*, *Comparison of novelty detection methods for multispectral images in rover-based planetary exploration missions*, vol. 34, no. 6. Springer US, 2020.
- [35] K. Bashir, M. Rehman, and M. Bari, "Detection and Classification of Rice Diseases: An Automated Approach Using Textural Features," *Mehran Univ. Res. J. Eng. Technol.*, vol. 38, no. 1, pp. 239–250, 2019, DOI: 10.22581/muet1982.1901.20.
- [36] V. Kumar Chavan, M. Sreenivasulu, and B. R. V. Rao, "Optical applications of copper doped lithium phosphate glasses," *Optik (Stuttg.)*, vol. 221, no. June, p. 165291, 2020, DOI: 10.1016/j.ijleo.2020.165291.
- [37] C. Zhang, P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu, "A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high-resolution remote sensing images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 88, no. February, p. 102086, 2020, DOI: 10.1016/j.jag.2020.102086.
- [38] R. Li *et al.*, "DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2018, DOI: 10.1109/JSTARS.2018.2833382.
- [39] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network," *arXiv*, pp. 1–27, 2020.

- [40] X. Li *et al.*, "A Remote-Sensing Image Pan-Sharpener Method Based on Multi-Scale Channel Attention Residual Network," *IEEE Access*, vol. 8, pp. 27163–27177, 2020, DOI: 10.1109/ACCESS.2020.2971502.
- [41] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 148–164, 2018, DOI: 10.1016/j.isprsjprs.2017.12.012.
- [42] M. Vakalopoulou, S. Christodoulidis, M. Sahasrabudhe, S. Mougiakakou, and N. Paragios, "Image Registration of Satellite Imagery with Deep Convolutional Neural Networks," *Int. Geosci. Remote Sens. Symp.*, pp. 4939–4942, 2019, DOI: 10.1109/IGARSS.2019.8898220.
- [43] K. He, "CVPR 2017 Tutorial Learning Deep Features for Visual Recognition Outline • Introduction • Convolutional Neural Networks : Recap • LeNet, AlexNet, VGG, GoogleNet ; Batch Norm."