MGFEEN: A Multi-Granularity Feature Encoding Ensemble Network for Remote Sensing Image Classification

Musabe Jean Bosco, Rutarindwa Jean Pierre, Mohammed Saleh Ali Muthanna, Kwizera Jean Pierre, Ammar Muthanna, Ahmed A. Abd El-Latif

***Abstract*—Deep convolutional neural networks (DCNNs) have emerged as powerful tools in diverse remote sensing domains, but their optimization remains challenging due to their complex nature and the large number of parameters involved. Researchers have been exploring more sophisticated methodologies to improve image classification accuracy. In this paper, we introduce a multi-granularity feature encoding ensemble network (MGFEEN) that is designed to fine-tune features at different levels of granularity. The network is trained in a two-step process: first, the output of granularity level i is used as the input for the next level; then, a fully connected layer is added to the pre-trained network to advance to the next level.**

The effectiveness of the MGFEEN's feature extraction is evaluated by feeding the globally extracted features to a softmax classifier for classification. By applying ensemble learning principles, our proposed MGFEEN achieves more accurate final predictions. We evaluate our model on three widely recognized benchmark datasets: UC-Merced, SIRIWHU, and EAC-Dataset. Notably, on the EAC-Dataset, our results show a significant 0.54% improvement in accuracy over a single-training-network setup, resulting in an impressive 98.70% accuracy level.

***Index Terms*—Multi-Granularity feature representation, convolution neural network, feature ensemble network, remote sensing image classification.**

I. Introduction

The authentic supervision of remotely captured imagery and its subsequent analysis holds immense importance within the realm of environmental and natural resources management. This significance is evident in applications such as monitoring agricultural regions through the analysis of Sentinel-2 image time series [1], planning for water resource development[2], tracking land changes [3], and detecting alterations in high-resolution imagery[4]. The classification of Land Cover and Land Use (LCLU) poses a present-day challenge in the field of remote sensing, aiming to regulate the responsible utilization of Earth's land. While land cover and land use are distinct concepts, there is a growing tendency to use them interchangeably in the context of remote sensing satellite data. Remote sensing technology has a multitude of pivotal applications, including semantic segmentation[5], scene classification in Remote Sensing (RS) [6], change detection using Remote Sensing (RS) techniques[7], and object detection in Remote Sensing (RS) scenarios[8]. Among the emerging applications mentioned above, one of the prominent areas of interest is RS scene classification. This involves categorizing remote-sensing scene images into distinct dissimilarity and similarity classes. Recent research has shown a concentrated focus on various computer vision tasks utilizing deep convolutional neural networks (DCNNs)[9], [10], [11], [12]. There are some image classification tasks such as classification[13], object detection[14], change detection[15]and medical image classification [16], [17], and automatic ship detection[18]. Recognize that the image classification process consists of the following steps: (a) image preprocessing, (b) feature extraction, and (c) classifier selection and designing. Practically, feature extraction plays an essential important role in the whole process. The image features mostly contain image color features, image color space, texture, shape features, texture features, and spatial relational ship features.

Cultural features and color attributes constitute overarching characteristics that portray the landscape of the scene, pertaining to the precise location within the image. However, they do not fully capture the essential attributes of the objects present. The shape attribute predominantly delineates the outline of the image. Spatial relationship traits entail spatial positioning, while relative relationship traits pertain to distinct elements within the image. DCNNs have demonstrated their optimal efficacy through the extraction of multiple features within a hierarchical, finely-grained representation[19], [20], [21], [22], [23]. At its core, DCNNs transform images into probabilities for global feature categorization.

Over the last decade, computer vision has significantly harnessed the capabilities of convolutional neural networks (CNNs) [24]to solve challenges and learn hierarchical feature representation to classifier fine and coarse for a massive dataset, depending on handcrafted features. For systems to be manageable, resilient, low-cost, and harmonious, researchers have been looking into multi-granularity feature extraction, a

Musabe Jean Bosco, Jean Pierre Rutarindwa, and Kwizera Jean Pierre are with Kigali Independent University (ULK), School of Science and Technology, Department of Computer Science and Technology,2280 KIGALI, deanfstkigali@ulk.ac.rw, hodcskigali@ulk.ac.rw, and kwijpeter01@gmail.com

Mohammed Saleh Ali Muthanna M.Muthanna is with the Institute of Computer Technologies and Information Security, Southern Federal University, 347922 Taganrog, Russia (e-mail: muthanna@sfedu.ru)

Ammar Muthanna is with the Peoples' Friendship University of Russia (RUDN University) 6 Miklukho-Maklaya, 117198 Moscow, Russia; (email: muthanna.asa@spbgut.ru)

Ahmed A. Abd El-Latif is with the EIAS Data Science Lab, College of Computer and Information Sciences, and with Center of Excellence in Quantum and Intelligent Computing, Prince Sultan University, Riyadh 11586, Saudi Arabia, and with the Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, 32511, Egypt (email: a.rahiem@gmail.com; aabdellatif@psu.edu.sa)

Corresponding Author: Ahmed A. Abd El-Latif

high multilevel feature that involves making use of tolerance for imprecision, incomplete, uncertain, and massive information. Multiple granular layers for brain big data processing [25] introduced multi-granularity information knowledge representation and [26] learning decision-making and problem-solving. High semantic level features [27] can roughly localize, detect, and classify the objects [28]. Very Fine Spatial Remote Sensing [29], land cover challenges played a contribution for deep learning [30].As an example, the analysis of satellite images presents distinct challenges that give rise to complex and novel scientific inquiries. Certain remote sensing datasets encompass multimodal information derived from optical sources, such as Lidar [31] and synthetic aperture radar (SAR) sensors. However, applying DCNNs directly to remote sensing (RS) scene classification introduces the following challenges: The initial concern arises from intra-class variations due to the diverse resolutions present in RS images. The second challenge stems from the complexity of representing information at a granular level over vast land areas, resulting in the representation of both similarity and dissimilarity aspects. In order to address these challenges, we propose a concise yet comprehensive theoretical framework that guides our approach.

Fine-grained features, when combined with global features, enhance RS scene classification. For instance, in RS images, we can identify a 'train station' by detecting a train or recognize an 'airport' upon sighting an airplane. Furthermore, RS imagery inherently encapsulates latent semantic and structural information, even in the absence of detailed annotations like bounding boxes or pixel-level annotations. Given these challenges, we suggest investigating the use of a Multi-granularity Feature Encoding Ensemble Network (MGFEEN) for the purpose of improved Remote Sensing Image Classification. Our approach delves into multiple levels of granularity in order to enhance accuracy and mitigate the challenges arising from intra-class variations related to similarity and dissimilarity. Our primary focus lies in capturing aggregated features that encompass structural information. At each granularity level, we introduce an ensemble module that combines distinct high-level multi-granularity segments featuring comparable receptive fields yet varying resolutions. In summation, the key points of our work are as follows:

- To facilitate the automated extraction of advanced features, a Multi-granularity Feature Encoding framework has been devised. Nevertheless, a more comprehensive and abstract depiction of the image can be acquired by assimilating these attributes.
- "Multi-granularity Feature Encoding Ensemble module" is introduced to efficiently decompose the optimization process of DCNNs in remote sensing scene categorization.

II. Learning Multi-Granularity levels in Remote Sensing Image

Multiple levels of granularity analysis exist, ranging from broader to more detailed and vice versa. The primary goal of comprehending objects at various levels in remote sensing images involves analyzing an image systematically. The significance of these levels is determined by criteria such as purpose consistency or the degree of similarity and difference in sparse features. This importance is established based on the choice of an objective function. The specifics of segmentation might vary based on the chosen standards and segmentation objectives, particularly when defining information units and their interconnections to create an assessment framework. Nevertheless, the connections identified through diverse standard criteria could be further subdivided into smaller units, referred to as information granules[32] is a collection of entities with a common element, i.e., Entities comprised of elements aggregated due to their resemblance, functional closeness, temporal similarity, and distinct adjacency are subsequently handled as a solitary processing unit with semantic significance. This mirrors the inherent limitations in the human capacity to manage and retain information effectively. **Scenario1:** Let thing object detection, for example, bounding box $\mathbf{b} = (b_x, b_y, b_w, b_h)$ hold four coordinates of a granular image patch$x$. The task of the bounding box is to regress a candidate bounding box b into the bounding target $\mathbf{g}$ using the regression function $f(x, \mathbf{b})$. To generate a reversal invariant to scale and location from the training sample $(\mathbf{g}_i, \mathbf{b}_i)$, we minimize the bounding box L_1 loss function $L_{loc}(f(x_i, b_i), g_i)$ in order to perform the operation of distance vector$\Delta = (\delta_w, \delta_y, \delta_x, \delta_h)$. is defined by:

$$\delta_w = \log(g_w/b_w) \tag{1}$$

$$\delta_x = (g_x - b_x)/b_w \tag{2}$$

$$\delta_h = \log(g_h/b_h) \tag{3}$$

$$\delta_y = (g_y - b_y)/b_h \tag{4}$$

Regression tasks usually involve substantially fewer elements compared to classification tasks. This contributes to improving the efficiency of the multi-granularity approach. Learning Δ is generally standardised by its mean and variance $\delta'_x = (\delta_x - \mu_x)/\delta_x$, in a single regression step f is incomplete for accurate localization. In different iterative of the post-processing step, we can apply

$$f'(x, \mathbf{b}) = f \circ f \circ f \circ .....f(x, \mathbf{b}) \tag{5}$$

The intersection over union can be utilized to assess both negative and positive metrics in evaluating the quality of detection. If $I \circ U$ it is greater than a specified threshold u , the patch evaluates it as an instance for the class.

$$y = \begin{cases} g_y I \circ U(x, g) > u \\ 0 otherwise, \end{cases} \tag{6}$$

Where g_y indicates the instance label for the actual truth object $g$by $I \circ U$, defining a detector's quality. Cascade R-CNN for small object detection [33], [Author34(year)] makes it challenging to ask a single segment or regression to perform perfectly uniformly at all quality levels. The difficult regression task can be decomposed into more straightforward steps, forming a cascade regression problem. This formula acts as a cascade of specialized regresses.

$$f(x, \mathbf{b}) = f_\tau \circ f_\tau \circ f_\tau - 1 \circ .....f1(x, \mathbf{b}) \tag{7}$$

Where τ is the total number of cascade stages, every regression in the cascade is optimized b^t arriving at the corresponding scene instead of the initial distribution $\mathbf{b}^1$. This cascade enhances hypotheses progressively. In different regression to produce a bounding box of high $I \circ U$, we begin, for instance $(x_i, \mathbf{b}_i)$, cascade regression up-and-coming re-sample $(x_i', \mathbf{b'}_i)$ of great $I \circ U$. After varying levels of the regressed, we can find there is no overfitting. Positive examples are plentiful at all levels, and the detector of deep stages is optimized for a higher $I \circ U$ threshold.

Scenario2: Consider the concept of multi-granularity within ensemble neural networks, where a finite set of individual neural networks are employed to comprehend the same query. By utilizing various learning algorithms, improved predictive capabilities are achieved. The combined outputs of these distinct neural networks collectively determine the final outcome. Assume the input $x \in R^m$ joins distribution $p(x)$, if the output correlating to x is $d(x)$, the result correlating to the individual neural network $f_i(i = 1, 2, 3, ..., N)$ is f_i , here the weight correlating to $f_i(x)$, is ω_i , the production of Neural Networks ensemble $f_{ensemble}(x)$ correlating to x can be defined as follow:

$$f_{ensemble}(x) = \sum_{i=1}^{N} \omega_i f_i(x) \tag{8}$$

Generalization error of NNE:

$$\mathrm{E}_{ensemble} = \int P(x)(\int {}_{ensemble}(x) - D(x))^2 dx \tag{9}$$

Generalization error of $f_i(i = 1, 2, 3, ..., N)$:$\mathrm{E}_i = \int P(x)(f_i(x) - d(x))^2 dx$,weighted average of:

$$f_i(i = 1, 2, 3, .., N) = E_{average} = \sum_{i=1}^{N} \omega_i E_i \tag{10}$$

Diversity of:

$$\mathrm{f}_i(i = 1, 2, 3, .., N) = A_i = \int P(x) f_i(x) - f_{ensemble}(x))^2 dx \tag{11}$$

Diversity of NNE:

$$A_{ensemble} = \sum_{i=1}^{N} \omega_i A_i \tag{12}$$

There are some theories analysis defines as the computational formula:

$$E_{ensemble} = E_{average} - E_{ensemble} \tag{13}$$

As we see in Figure 4, Nonetheless, through the implementation of a branch feature selection technique aimed at recognizing the initial attribute granularity and optimal feature subsets across diverse datasets, it becomes feasible to reduce the computational expenses associated with high-dimensional data challenges. Figure 12 improves the generalization precision of NNE by not only considering the diversity between the networks. It uses r as an output error vector to measure how well each network generalizes. It can combine both also can improve the generalization ability of NNE4 right, one of multi-granularity neural ensemble network.

Scenario3: Let's consider feature learning for classification task; we refer to Figure 1, which is feature learning, especially on multi-granularity auto-encoding algorithms, essential for encoding and decoding networks. Also, we can consider the multi-granularity concept in terms of convolutional kernels. Whereby each type of kernel has specific features of interest, such as horizontal edge and vertical edge. By designing this network, we must ensure that the network extracts high-level features of diversity. As we see Figure 2 indicates, we chose the convolutional kernel of 1×1, 3×3, and 5×5 to develop a multi-granularity encoding model.

Figure 4 Illustrates the architecture of the deep multi-granularity encoding approach, wherein the initial step involves training a multi-granularity encoding model (MGE1). The outputs of its hidden layer then serve as inputs for the subsequent multi-granularity encoding (MGE2). In the second stage, a new MGE2 is developed, and further feature mapping is performed. Within our overarching framework, the encoder's neural network structure defines a sequence of encoders, while the decoder's neural network structure establishes a sequence of decoders. Through gradient descent optimization, the encoder and decoder networks are fine-tuned. Consequently, the entire architecture of the auto-encoders, encompassing both the encoder and decoder, induces a data bottleneck that ensures only the central information is reconstructed through the auto-encoders. This autoencoder setup effectively captures intermediate granularity-level visual representations from unlabeled remote sensing data. The task of learning features at this intermediate granularity level holds significant importance in accurately categorizing small-scale sensing images. An auto-encoder can take an input h and the first map is represented by $h \in {}^M$ for nonlinear mapping.

$$h = f(Wx + \beta) \tag{14}$$

Where W is a weight matrix to be estimated during training, β is a bias vector, and f is based on a nonlinear function, such as hyperbolic tangent function, and sigmoid function. The multi-granularity encoded feature represents h which is used to reconstruct the input x by a revised mapping, leading to the reconstructed information.

$$\gamma = f(W'h + \beta') \tag{15}$$

Where W' is always constrained to be the form $W' = W^T$, the exact weight is utilized for encoding the input and decoding the latent representation.

A. Remote Sensing Image Scene Classification

Various methods were examined to address land use image classification challenges in high-resolution overhead images, employing diverse techniques to analyze bag-of-visual-words (BoVW) approaches. In a manual feature-based strategy, distinct feature descriptors, such as the scale-invariant feature transform, are consistently employed (SIFT)[34], [35], [36] and Histogram of gradient (HoG). Differing from the conventional manual feature-based method, deep convolutional neural networks possess enhanced capabilities for feature representation. Currently,deep convolutional neural networks

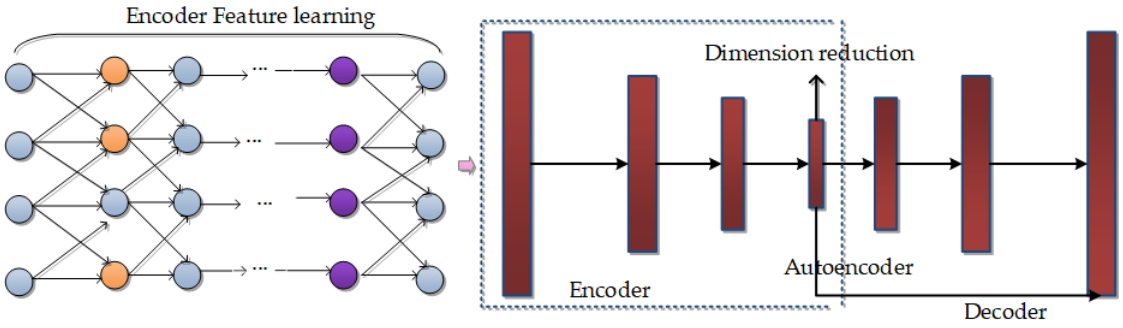

Fig. 1. Multi-Granularity encoding neural networks, described by Auto encoding

(DCNNs) have garnered noteworthy accomplishments in tackling classification difficulties within the realm of remote sensing[37]. There are [38] that use DCNNs to extract features of RS images and further explore its generalization latent to obtain great implementation. Nevertheless, certain methodologies integrate attention mechanisms into DCNNs, aiming to incorporate lesser-ranked features guided solely by global-level annotations [39], [40]. The second way of information is effectively used in RS scene classification duties [41], which is excellent. Feature encoders [42], in general, CNN's, the FOV determines segmentation accuracy. A constrained field of view (FOV) could lead to inaccurate identifications of objects, either generating false positives or missing out on objects (false negatives) due to insufficient contextual information. To enhance the FOV of Convolutional Neural Networks (CNNs), pooling and down-sampling layers are employed, but this leads to reduced output resolution and detail precision. Enhancing segmentation accuracy and refining proposed boundaries naturally involves broadening the field of view (FOV) without compromising resolution, while also enabling the handling of partial objects.

B. Multi-granularity Feature Extraction Methods

The fundamental goal of pattern recognition involves accurately assigning an input pattern to one of several output classes. Following the preprocessing phase, the object of character recognition extracts features. In the process of reducing dimensionality, the initial data rows are organized into more manageable clusters, and feature extraction is carried out. Drawing inspiration from the aforementioned techniques, we utilize multi-granularity feature extraction for handling remote sensing images. In scenarios where there's significant similarity between classes, the performance of DCNNs can drop significantly. To address this concern, a range of fine-grained feature extraction approaches are presented, tailored for multitasking and multi-attention object recognition. However, in many instances, only global annotations are available, making the identification of fine-grained attributes challenging due to the absence of semantic-level annotations. Consequently, multi-granularity feature extraction methods are introduced to enhance the region-based feature representation capability of DCNNs.

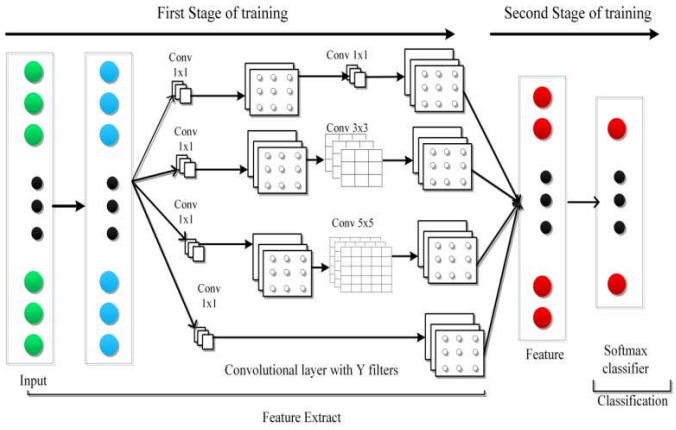

Fig. 2. Classification architecture for multi-kernel deep learning.

Fig.2 links the classifier to extract the final classification task. Softmax classifiers are always used for multi-classification functions of neural networks, where they can achieve a tremendous competitive performance eq.16, which K indicates the data has K classes. In the coding phase, the original input vector $x \in \Re^c$ via randomly adding Gaussian noisy obtain corrupt new input vector x, the enter the non-linear activation function through linear mapping. Assuming a training set

$$X = \{(x_1, y_1), (x_2, y_2), ..., (x_M, y_M)\},$$

, the overall cost function of MGE on data set X can be defined as eq. 17.

$$(x) = \frac{1}{\sum_{j=1}^{k} e_j^{\sum_{i=0}^{m} w_i x_i}} \begin{bmatrix} e_1^{\sum_{i=0}^{m} w_i x_i} \\ e_2^{\sum_{i=0}^{m} w_i x_i} \\ ..... \\ e_k^{\sum_{i=0}^{m} w_i x_i} \end{bmatrix} \tag{16}$$

The objective of Figure 16 is to achieve multiple representation mapping. For training these deep networks, it's necessary to employ a cost function that minimizes the reconstruction error and ensures the output images closely resemble the input images. By utilizing two distinct cost functions, remarkable achievements can be attained.

$$J(W, b) = \frac{1}{2M} \sum_{i=1}^{M} ||x^i - z^i||^2 + \frac{\lambda}{2} ||W||_2^2 \tag{17}$$

$$J(W,b) = -\frac{1}{M}\sum_{i=1}^{M} \tag{18}$$

The training process for the stacked depth MGE consists of two distinct phases: pre-training and fine-tuning. During the pre-training phase, multiple MGENNs are stacked and trained using unsupervised learning, while the fine-tuning phase involves supervised learning for training the soft-max classifier.

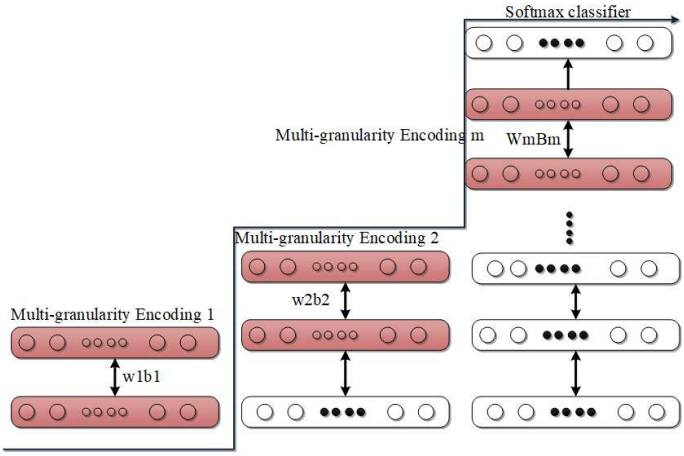

Fig. 3. Deep Multi-Granularity encoding neural network limited the nonlinear mapping capability of a multi-granularity network with only two hidden layers.

C. Feature Fusion Approach in Scene Classification

Many research studies employ feature fusion techniques to counteract the adverse impact of varying image resolutions and attain enhanced performance. Also, [54] offers a multi-scale convolutional neural network (CNN) structure consisting of a fixed-scale net and a varied-scale net to address the scale variation of objects in remote sensing data.[55] create an architecture with two branches that can incorporate both global-context and local-object characteristics.[56], [57] presents a method for fusing multi-layer features from pre-trained CNN models in order to classify RS scenes. This study focuses on a strategy for fusing features with varying granularity, localization, and region scales. By creating multi-subnets with alternative topologies, ensemble learning-based approaches provide an alternative method for extracting multi-granularity features. In [58] separate classification results are generated by multiple CNNs and are combined using occupation probability. Describes a new method of learning that uses deep sub-CNNs to learn targeted features.[59] uses a covariance pooling of CNN features to create an ensemble for high-resolution RS scene classification that is both effective and computationally inexpensive. We can assume the ensemble learning method in architecture to integrate multi-granularity and multilevel features from the above ensemble learning-based methods. In Figure 5.4, the left diagram illustrates the Multi-branch approach for constructing individual neural networks, while the right diagram depicts the Multi-branch technique for creating an ensemble of multi-granularity neural networks.

III. Multi-granularity Feature Encoding Network Method

Enhancing the precision of deep neural networks has been successfully demonstrated in various research studies, as discussed in the literature on multi-granularity feature learning. Building upon prior research, the approach involves amalgamating multiple CNN modules. The proposed method, known as Multi-granularity Feature Encoding Ensemble Network (MGFEEN), aims to enhance accuracy by leveraging multi-granularity features extracted by diverse networks. The methodology constructs distinct network levels, each contributing to the overall structure, as detailed earlier in Figure 5indicates that each set of networks in a given multilevel is connected to another level. The objective behind this concept is to disintegrate the parameter optimization process of networks into multiple granularities and levels, dividing it into steps. In this approach, each block within every network is trained and optimized separately, resulting in an enhanced accuracy performance.

Furthermore, in the dual-down design, the information from the second granularity level is extracted through the fully connected layer transfer. The second design on the right side showcases the enhanced two-level structure of MGFEEN, incorporating a transfer fully connected layer. For instance, M11 employs Inception-V3, while M12 is linked to ResNet-50. Additionally, M21 serves as an easily accessible convolutional neural network (CNN). The dataset used, known as the EACC-Dataset, is a distinctive Land Use Data Set comprising 2110 images sourced from Google Earth satellites. These images are divided into nine distinct object categories, each with RGB pixels measuring 256×256. The dataset encompasses 150 to 300 images per class, including agricultural, beach, buildings, commercial, desert, flood, forest, mountain, and river, as depicted in Figure 5.6, where images (1) through (9) exemplify the nine classes. As illustrated in Figure 5, there are i levels in which each set of networks in a certain multilevel connects to the next or preceding level. The objective of this strategy is to divide the optimization of network parameters into distinct phases involving various levels of granularity. In each phase, individual blocks within the networks are trained and optimized independently, thereby enhancing the overall accuracy performance. We proceed with the feature from multilevel $i-1$ to level i,there are two methods introduced:

- MGFE method: merging the output layer's values at granularity level $i-1$, which correspond to the class's probability value, and driving them to the granularity level i as can be seen in Figure 5 architecture;
- MGFEEN techniques generate an additional fully connected layer with neurons and convey features from the added fully connected layer to the granularity level, As demonstrated in Figure 5 structure.

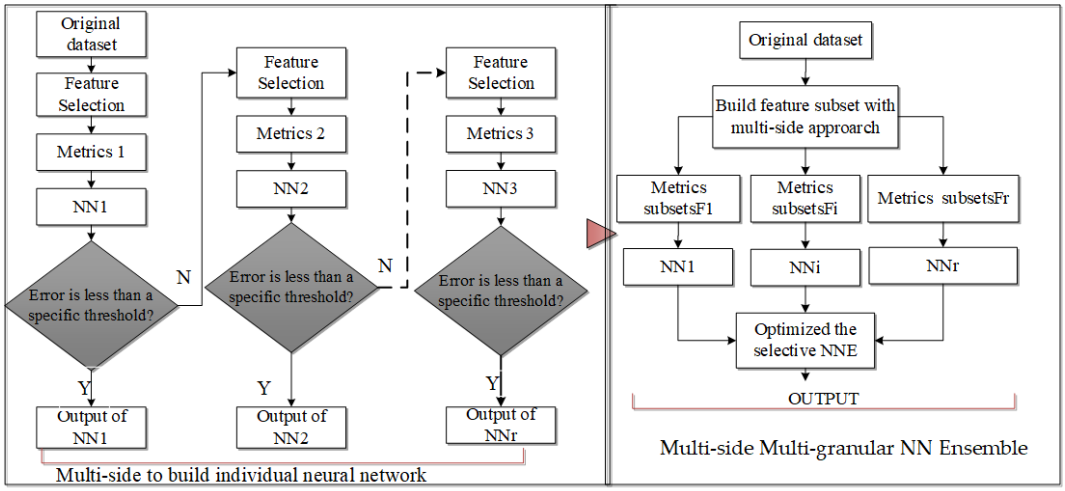

Fig. 4. Ensemble Neural Network (ENN).

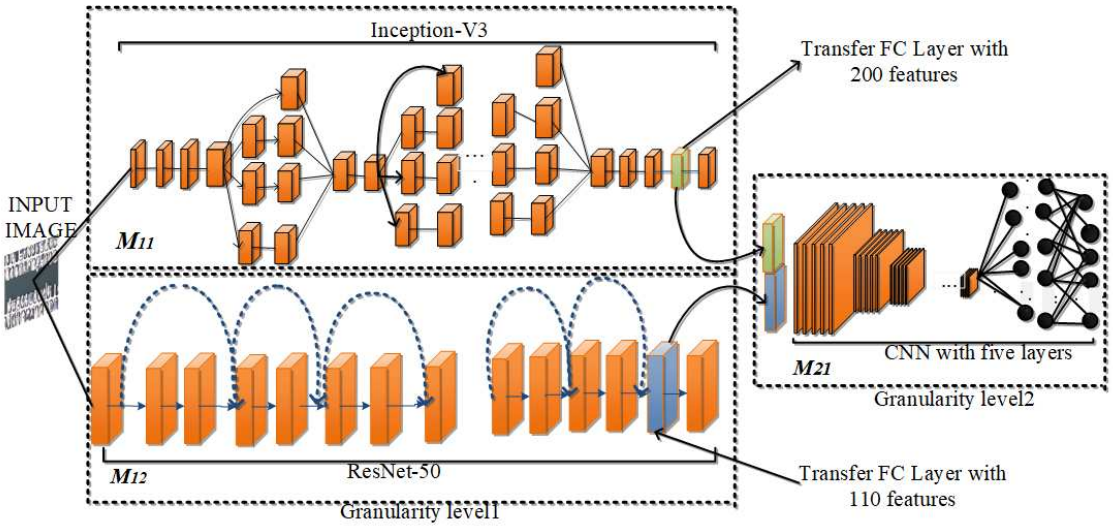

Fig. 5. MGFEEN method with an enhanced two-level structure with an additional fully connected layer. Inception-v3 as M11, Resnet-50 as M12 connected to the simple convolutional neural network as M21.

A. **Activation Function**

In the realm of multi-granularity and multilevel networks, the selection of diverse networks introduces variations and permits the utilization of distinct activation functions for extracting a range of features. To further enrich the arsenal of practical and effective activation functions beyond the commonly used rectified linear unit (ReLU), we introduce a modified activation function. Drawing inspiration from the architecture of neural networks, which draws parallels with the human nervous system, where signals are conveyed through minute electrochemical stimuli, our approach seeks to develop an activation function with a narrow range. This is intended to address the challenge of vanishing gradient descent and its associated issues. The following Equation12 evolved.

$$f(x) = \left\{ \begin{array}{l} 1, x \geq b \\ 0, x \leq a \\ \left(\frac{x-a}{b-a}\right)^n, a < x < b \end{array} \right\} \tag{19}$$

Here a, b and n are configurable parameters. This function's derivative is defined mathematically as following:

$$f(x) = \left\{ \begin{array}{l} 0, x \leq a \\ 0, x \geq b \\ n\left(\frac{x-a}{b-a}\right)^n, a < x < b \end{array} \right\} \tag{20}$$

In utilizing a convolutional neural network (CNN) for feature extraction, the suggested activation function, for configuring various parameters, can distribute in extracting a variety of features. We put the process in a separate experiment such as InceptionV3, DenseNet, and ResNet after modifying the ReLU activation layer to the suggested procedure. To gained results shows some modifications in the proposed function, see the following Equation14

$$f(x) = \left\{ \begin{array}{l} 0, cx \leq a \\ b\left(\frac{x-a}{b-a}\right)^n, a < cx < b \\ cx, cx \geq b \end{array} \right\} \tag{21}$$

There is a derivative in Equation22.

$$f(x) = \left\{ \begin{array}{l} 0, cx \leq a \\ bn\left(\frac{cx-a}{b-a}\right)^{n-1}, a < cx < b \\ c, cx \geq b \end{array} \right\} \quad (22)$$

The function expression in equation 22 quietness could not accomplish the initial goal of smaller activation function values similar to the human nervous system. Consequently, future researchers will conduct on this point.

B. Datasets

This study mainly evaluates our proposed approach to EACC scene classification dataset. The EACC Data set is a one-of-a-kind Land Use Data Set comprised of 2112 images obtained from the Google Earth satellite and classified into nine item classes utilizing 256*256 pixels in the RGB color space. The authors collected data from Easter Africa Community members in six countries (South Sudan, Rwanda, Uganda, Kenya, Burundi, and Tanzania). The images were obtained in Eastern Africa Community Participating Countries using the Google Earth Engine.Inc (EACC). Each of the following groups has between 150 and 300 images: forest, river, beach, buildings, commercial, desert, flood, mountain, and agricultural. Some samples from the dataset are depicted In Figure 6.

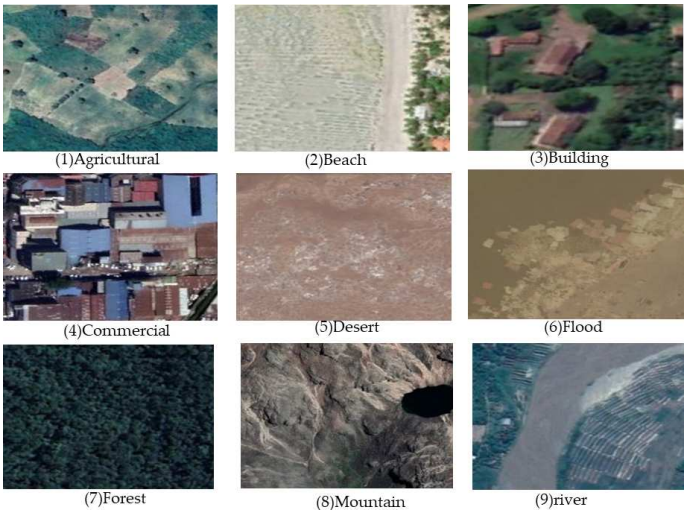

Fig. 6. Sample of the novel dataset collected from Easter Africa Community Countries (EACC)

C. **Implementation and Results**

All simulations were performed using the python 3.7, Keras, TensorFlow-GPU1.14.0 framework, and tested on a performance computer (Intel ® Core™ i5 CPU, 8GB RAM). We utilize ReNet50[67] and Inception-v3[68] as baseline models to perform significantly with the previous methods. We used these baseline models to construct MGFEEM with ten features of the output layer without the softmax activation function. Then, for each input image, ten features from each model

TABLE I
TEST ACCURACY OF DIFFERENT GRANULARITY LEVELS BASED ON CONVOLUTION AUTO-ENCODE (CONVAE) 20% TESTING FOR EACH DATASET. DIFFERENCES IN CLASSIFICATION ACCURACY WERE ANALYZED BETWEEN THE GRANULARITY MODULE WHERE LISTED IN FIGURE 2.

Method	UCM Dataset	SIRWHU Dataset	EACC Dataset
ConvAE(1X1)	95.81	94.16	95.90
ConvAE (3X3)	98.10	97.19	98.35
ConvAE(5X5)	98.30	98.06	98.90

TABLE II
SUMMARY OF DIFFERENT RESULTS OF THE ACTIVATION FUNCTION (RELU) AND THE ESTABLISHED ACTIVATION FUNCTION BY UTILIZING OTHER PARAMETERS IN THE EACC DATASET.

Parameters						
a	b	c	n	Act. fnx.	Variance	Average Acc.(%)
-	-	-	-	ReLU	**1.46**	**95.68**
-10	0.01	1	3	F1	**1.69**	**95.50**
-10	0.1	1	3	F2	1.89	94.30
-3	0.1	1	3	F3	2.17	93.98
-2	0.1	1	2	F4	2.65	94.83
-2	0.1	1	0.5	F5	3.13	93.16
-10	0.1	1	1/3	F6	4.18	93.32

are merged to generate a twenty-feature vector. This vector is then used as an input to an efficient module composed of five convolutional layers and one fully connected layer trained for 30 epochs. Combining these two modules with 98.06% and 98.02% accuracies in 98.50 % implements 0.01% over the most significant accuracy in the first granularity level.

The second network transfers the fully connected layer from the first granularity level for inception-v3 and the ResNet50. Within the same training details of baseline, the second granularity level is an easily CNN with 5CONV and 1FC layer trained using stochastic gradient descent (SGD) with ten epochs and an small learning rate of 0.01, see Figure 5 demonstrates the used structures. The proposed design indicated a performance accuracy improved from 98.06% using Inception-v3 and 98.02% of ResNet50 to 98.7%, which is meaningful. For instance Figure 7 depicts to compare the confusion matrix with different improvements between UCM-dataset and SIRI-WHU dataset, and table IV summarizes all the results.

In Table II the proposed activation can look like generalized ReLu as mentioned b=0 converted the function to the ReLU within six convolutional layers flowed with Bach normalization, max-pooling. In MGFEEN, during training, a percentage

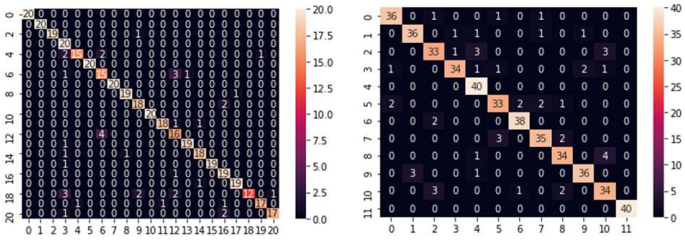

Fig. 7. Sample of a confusion matrix for the UCM dataset based on inceptionResNetV3. The right one is a Sample of the confusion matrix of the multi-granularity architecture of deep learning method1 using SIRI-WHU dataset.

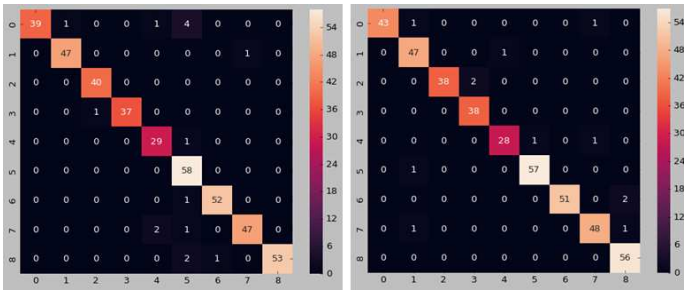

Fig. 8. Sample of a confusion matrix for EACC dataset using first granularity level. The second right is the confusion matrix of the multilevel structure of the multi-granularity feature encoding ensemble network (MGFEEN) using the EACC dataset.

TABLE III
SUMMARY OF RESULTS OF THREE APPROACHES AS WE INTERPRETED IN FIGURE 5

Model	**Granularity level1**	**Granularity level2**	**Overall accuracy**
InceptionResNetV2+DenseNet		89.75	
VGG16+DenseNet		96.91	
Inception-v3		98.06	
ResNet-50		98.02	
MGEENN	Inception-v3+ResNet-50	Five layers CNN	98.50
MGEENN approach2	Inception-V3+ResNet-50	Five layers CNN Additional FC	98.70

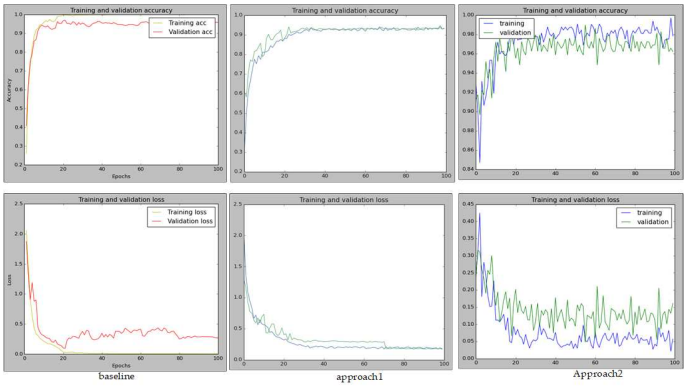

Fig. 9. Classification accuracy, for training efficiency of baseline, approach 1, and approach 2.

of features and neurons for testing is used. As well knows the issues of data training in remote sensing, the supervised deep learning model requires a massive amount of training samples. While in the RS classification, labeling the observed data to prepare training samples for each remote sensing classification is high time and cost-intensive.

MGFEEN's approach incorporated data augmentation techniques that we are developing to expand the quantity and performance of the training images, such as transfer learning and active learning. We focus on developing a high-precision deep learning model with a minimal sample size. Figure 4 illustrates the MGFEEN technique based on the Inception-v3 and ResNet-50 baseline models, approach one by utilizing the output layer's nine features without using soft-max. For each input image, the nine features from each network were combined into an 18-feature vector and fed into a basic network with five convolutional layers and one fully connected layer trained for 10 epochs. Once the MGFEEN results are combined with those from Inception-v3 and Resnet-50, we see that just one class has improved over the two previous models. The first level is composed of the inception-v3 and Resnet-50, as well as an additional completely connected layer. The second level is a simple CNN with five convolutional layers and one fully connected layer that was trained using SGD for 10 epochs at a learning rate of 0.01. We solve this constraint with the MGFEEN technique by retaining a greater number of features from level one and thereby retaining a greater number of elements from level one and providing exceptional performance. Additional progress can be achieved in the future by introducing an advanced customizable activation function. The latter can be investigated using MGFEEN by training each distinct network with a unique activation function. These may result in the extraction of a variety of features.

TABLE IV
RESULT OF THE DATASETS EXPERIMENTED WITH MGFEEN

Dataset	**Granularity level1 Acc.(%)**	**Granularity level2 Acc.(%)**	**Performance**
UCM	97.16/ 97.40	98.04	0.64
SIRWHU	95.30/ 95.40	97.8	2.4
EACC (9 Classes)	98.02 / 98.06	98.70	0.58 and 0.54

IV. Conclusion

In this paper, we present deep convolutional neural network (DCNN) assembling methods, and we propose the Multi-Granularity Feature Encoding Ensemble Network (MGFEEN). MGFEEN is based on establishing different networks at granularity levels, whereby each network is trained and optimized independently. This strategy optimizes the networks at granularity levels in order to optimize a larger network. Our proposed ensemble strategy aims to improve the performance accuracy of the outcomes as close as possible to the rectified linear unit (ReLU) activation function. MGFEEN was evaluated and tested using CNNs at two different granularity levels. ResNet-50 and Inception-v3 are used at the first granularity level, whereas a simple CNN is used at the second granularity level. Two fusion models were examined: the first model takes the output layers of the modules at granularity level 1 and maintains them at level 2. The second strategy is to construct a fully connected (FC) layer at granularity level 2.

The experimentation illustrated that the second method maintains high features and performs better from granularity levels 1 to 2. The experiments performed on the EACC dataset demonstrated that the accuracy increased from 98.06% and 98.2% using ResNet-50 and Inception-v3 to approximately 98.50%. The EACC dataset contains nine classes, and the improvement from granularity level one to granularity level two was 0.58% and 0.54%. This work will help remote sensing scientists improve larger remote sensing datasets, and further work will establish different advanced optimization techniques and activation functions. Additionally, we will experiment with MGFEEN by using different activation functions in the

independent modules of MGFEEN, which can help to extract additional features for large-scale remote sensing classification tasks.

CONFLICT OF INTEREST

This work has no Conflict of Interest

DATA AVAILABILITY

All data are available from the corresponding author upon reasonable request.

Acknowledgment

This research is supported by postdoc fellowship granted by the Institute of Computer Technologies and Information Security, Southern Federal University, project No P.D./22-01-KT. Also, Ahmed A. Abd El-Latif would like to thank Prince Sultan University for their support.

REFERENCES

[1] AParis, C. ; Weikmann, G. ;Bruzzone, L. *Monitoring of agricultural areas by using Sentinel 2 image time series and deep learning techniques*. *Image and Signal Processing for Remote Sens. XXVI* **2021**, vol. 11533, pp.18,doi: 10.1117/12.2574745.

[2] Ndayisaba, F. ; Nahayo, L. ;Guo, H. ; Bao, A. ; Kayiranga, A. ; Karamage, F.;Nyesheja, E.M.*Mapping and monitoring the Akagera Wetland in Rwanda*.*Sustain.*,**20217**;vol. 9, no. 2, pp. 1–13, doi: 10.3390/su9020174

[3] Wang, Qing ; Zhang, X. ; Chen, G. ; Dai, F. ; Gong, Y. ; Zhu, K. *Change detection based on Faster R-CNN for high-resolution remote sensing images*.*Remote sens. l.*,**2022**,vol. 9, no. 10, pp. 923–932, doi: 10.1080/2150704X.2018.1492172.

[4] Zhang, X. ; Fan, R. ; Ma, L. ; Liao, X. ; Chen, X. *Change detection in very high-resolution images based on ensemble CNNs*.*Int. J. of Remote Sens.*, **2020**,vol. 41, no. 12, pp. 4755–4777, doi: 10.1080/01431161.2020.1723818.

[5] LI R., LIUWJ ; YANG, L. *DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation*, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*,**2020**, vol. 11, n. 11, pp. 3954–3962, , doi: 10.1109/JSTARS.2018.2833382.

[6] Gong, X. ; Ju, X. ; Qian, K. ; Lu, T. ; Chen, Zh.*Remote Sensing Image Scene Classification along the High-speed Railway based on Convolutional Neural Network*,*J. Phys. Conf. Ser.***2021**, vol. 1684, n. 1, pp. 012112,doi: 10.1088/1742-6596/1684/1/012112.

[7] Putri, A.R.D. ; Sidiropoulos, P. ; Muller, J.P.*Anomaly detection performance comparison on anomaly-detection based change detection on martian image pairs*,*Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch* **2019**,vol. 42, n. 2/w13, pp.1437–1441,doi: 10.5194/isprs-archives-XLII-2-W13-1437-2021.

[8] Rabbi, J. ; Ray, N. ; Schubert, M. ; Chowdhury, S. ; Chao, D.*Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network*,*Remote Sensing*,**2021**, vol.12, n. 9, pp. 1432

[9] Voulodimos, A. ; Doulamis, N. ; D., Anastasios ; Protopapadakis, E.*Deep learning for computer vision: A brief review*,*Computational intelligence and neuroscience*, **2020**.

[10] Gollapudi,S., *"Deep Learning for Computer Vision,"Learn Comput. Vis. Using OpenCV***2019** , pp. 51–69, 2021, doi: 10.1007/978-1-4842-4261-2-3..

[11] Salman,H.; Grover,J.; Shankar,T.*"Hierarchical Reinforcement Learning for Sequencing Behaviors,"***2020** vol. 2733, pp. 2709–2733, doi: 10.1162/NECO.

[12] Sharma, S. ; Juneja, A. ; Sharma, N.*Using Deep Convolutional Neural Network in Computer Vision for Real-World Scene Classification*,*Proc. 8th Int. Adv. Comput. Conf. IACC 2018*,**2020**, pp.284–289,doi: 10.1109/IADCC.2020.8692121.

[13] Paoletti,M. E.; Haut,J. M.; Fernandez-Beltran R.;Plaza, J. ; Plaza,A. J.; Pla,F. *"Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," IEEE Trans. Geosci. Remote Sens.***2019**, vol. 57, no. 2, pp. 740–754, doi: 10.1109/TGRS.2020.2860125.

[14] Lai, Y-C. ; Huang, Z-Y. *Detection of a Moving UAV Based on Deep Learning-Based Distance Estimation*,*Remote Sensing*, **2021**, vol. 12, n. 18, pp. 3035

[15] Shafiq, M.A. ; Wang, Z. ; Amin, A. ; Hegazy, T. ; Deriche, M. ; AlRegib, G.*Detection of salt-dome boundary surfaces in migrated seismic volumes using gradient of textures*,*2019 SEG Annual Meeting*,**2019**, doi: 10.1190/segam2015-5927230.1

[16] Bernal, J. ; Kushibar, K. ; Asfaw, D.S. ; Valverde, S.; Oliver, A. ; Mart, R. ; Llad, X. *Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review*,*Artif. Intell. Med.*,**2019**, vol. 95, pp. 64–81, doi: 10.1016/j.artmed.2020.08.008.

[17] Rahman, S. ; Wang, L. ; Sun, C. ; Zhou, L.*Deep learning based HEp-2 image classification: A comprehensive review*,*Medical Image Anal.*,**2021**, pp.101764, doi: 10.1016/j.media.2021.101764.

[18] Hwang, J-I. ; Jung, H-S. *Automatic ship detection using the artificial neural network and support vector machine from X-band SAR satellite images*,*Remote Sensing*,**2020**, vol. 10, n. 11, pp. 1799.

[19] Miao, J. ; Wang, B. ; Wu, X. ; Zhang, L. ; Hu, B.; Zhang, J.Q. *Deep Feature Extraction Based on Siamese Network and Auto-Encoder for Hyperspectral Image Classification*,*Int. Geosci. Remote Sens.Symp.*,**2020**,pp.397–400,doi: 10.1109/IGARSS.2019.8899230

[20] Aslam, M. A. ; Salik, M. N. ; Chughtai, F.; Ali, N. ; Dar, S.H. ; Khalil, T. *Image classification based on mid-level feature fusion*,*15th Int. Conf. Emerg. Technol. ICET 2020*, **2019**, doi: 10.1109/ICET48972.2019.8994721.

[21] Petrovska,B.; Zdravevski,E.; Lameski,P.; Corizzo,R.; Štajduhar, I.; Lerga,J. *"Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification,"* Sensors (Switzerland)**2021**, vol. 20, no. 14, pp. 1–22, 2021, doi: 10.3390/s20143906.

[22] Aung, Su W. Y. ; Khaing, S. S. ; Aung, S. T.*Multi-label land cover indices classification of satellite images using deep learning*,*International Conf. Big Data Anal. and Deep Learning Applications*, **2020**,vol.744, pp.94–103

[23] Scott,G. J.; Marcum,R. A.; Davis,C. H.; Nivin,T. W. *"Fusion of Deep Convolutional Neural Networks for Land Cover Classification of High-Resolution Imagery," IEEE Geosci. Remote Sens. Lett.* **2020**, vol. 14, no. 9, pp. 1638–1642, doi: 10.1109/LGRS.2020.2722988.

[24] Krizhevsky,A. ; Sutskever,I. ; Geoffrey H.E., *"ImageNet Classification with Deep Convolutional Neural Networks," Adv. Neural Inf. Process. Syst.***2019**, pp. 1–9, doi: 10.1109/5.726791.

[25] Yao,Y. ; Hu,Q.; Yu,H. ; Grzymala-Busse,J. W. *"Rough Sets, fuzzy sets, data mining, and granular computing: 15th International Conference, RSFDGrC 2019 Tianjin, China, November 20–23, 2015 proceedings," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)***2019**, vol. 9437, doi: 10.1007/978-3-319-25783-9.

[26] Wang,D. ; Shen,Z.; Shao,J.; Zhang,W. Xue,X.; Zhang, Z. *"Multiple granularity descriptors for fine-grained categorization," Proc. IEEE Int. Conf. Comput. Vis.***2019**, vol. 2015 Inter, pp. 2399–2406, doi: 10.1109/ICCV.2019.276.

[27] Liu,H. ;Li, L.; Wu,C. ;Technology, I. *"Colar Image Segmentation Algorthims Based on Granular Computing Clustering,"***2019**, vol. 7, no. 1, pp. 155–168.

[28] Missen, M. M. S.*Combining granularity-based topic-dependent and topic-independent evidences for opinion detection*,*Université Paul Sabatier-Toulouse III*, **2018**

[29] Aplin,P.; Atkinson, P. M.; Curran,P. J. *"Fine Spatial Resolution Simulated Satellite Sensor Imagery for Land Cover Mapping in the United Kingdom,"* ,2022, vol. 4257, no. 98.

[30] Zhang,L.; Xu, F.;Fraundorfer, F. *"Deep Learning in Remote Sensing,"***december,2022**.

[31] Szigarski C; Jagdhuber T; Baur M, et al. *"Analysis of the radar vegetation index and potential improvements"*. *Remote Sensing*, **2022**, 10(11): 1776.

[32] Meher,S. K.; Pal,S. K. *"Rough-wavelet granular space and classification of multispectral remote sensing image," Appl. Soft Comput. J.***2019**, vol. 11, no. 8, pp. 5662–5673, doi: 10.1016/j.asoc.2019.03.027.

[33] Ren,Y. *"applied sciences Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN,"***2021**, doi: 10.3390/app8050813.

[Author34(year)] Ren, Y. ; Zhu, C. ; Xiao, S.Small object detection in optical remote sensing images via modified faster R-CNN,*Applied Sciences***2021**,vol 8,n. 5,pp 813.

[34] Yang,Y. ;Newsam, S. *"Bag-of-visual-words and spatial extensions for land-use classification," GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst.***2020**, pp. 270–279, doi: 10.1145/1869790.1869829.

[35] Hu, F. ; Xia, G-S. ; Yang, W. ; Zhang, L. *Recent advances and opportunities in scene classification of aerial images with deep*

models,*Int. Geosci. Remote Sens. Symp.* **2018**, pp. 4371–4374, doi: 10.1109/IGARSS.2020.8518336.

[36] Jiang, J. ; Wu, D. ; Jiang, Z.*A correlation-based bag of visual words for image classification*,*2020 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)* **2020**, pp. 891–894,doi: 10.1109/ITOEC.2020.8122482.

[37] Dong, Y. ; Jiao, W. ; Long, T. ; He, G. ; Gong, C. *An extension of phase correlation-based image registration to estimate similarity transform using multiple polar Fourier transform*,*Remote Sens.*,**2020** vol. 10, n. 11, pp. 1719,doi: 10.3390/rs10111719.

[38] Hu G. X.; Yang Z.; Hu L.; et al. *Small Object Detection with Multiscale Features*. *International Journal of Digital Multimedia Broadcasting,* **2021***:1-10.*

[39] Dong Y; Jiao W; Long T; et al. An extension of phase correlation-based image registration to estimate similarity transform using multiple polar Fourier transform*.* Remote Sensing, **2021**, 10(11): 1719.

[40] M. A. Ganaie; M. Hu; M. Tanveer*; and P. N. Suganthan*; *'Ensemble deep learning: A review'*, **2021**, [Online]. Available: http://arxiv.org/abs/2104.02395.

[41] Ganaie, M. ; Hu, M.*Ensemble deep learning: A review*,*arXiv preprint arXiv:2104.02395***2021**,Available: http://arxiv.org/abs/2104.02395.

[42] Vali, A. ; Comai, S. ; Matteucci, M.*Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review*,*Remote Sens.***2020**, vol.12, n. 15, pp. 2495, doi: 10.3390/RS12152495.

[43] Li, F. ; Feng, R. ; Han, W. ; Wang, L. *Ensemble model with cascade attention mechanism for high-resolution remote sensing image scene classification*,*Opt. Express***2021**, vol. 28, n. 15, pp. 22358–22387, doi: 10.1364/oe.395866

[44] Guo, Y. ; Ji, J. ; Shi, D. ; Ye, Q. ; Xie, H. *Multi-view feature learning for VHR remote sensing image classification.Multimedia Tools and Applications***2021**, pp.1–13,doi: 10.1007/s11042-020-08713-z.

[45] Zou, Z. ; Shi, T. ; Li, W. ; Zhang, Z. ; Shi, Z.*Do game data generalize well for remote sensing image segmentation?*.*Remote Sens.***2021**, vol. 12, n. 2, pp.275,doi: 10.3390/rs12020275

[46] Corrales, H. ; Llorca, D. F. ; Parra, I. ; Vigre, S. ; Quintanar, A. ; Lorenzo, J. ; Hernandez, N.*CNNs for fine-grained car model classification.I Lect. Notes Comput. Sci. including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics***2020**,vol. 12014 LNCS, pp. 104–112, 2021, doi: 10.1007/978-3-030-45096-0-13.

[47] Kattenborn, T. ; Eichel, J. ; Fassnacht, F. E.*Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery*.*Sci. Rep.***2020**, vol. 9, n. 1, pp. 1–9,doi: 10.1038/s41598-019-53797-9

[48] Srivastava,S.; Vargas Muñoz,J. E.; Lobry,S.; Tuia,D. *"Fine-grained landuse characterization using ground-based pictures: a deep learning solution based on globally available data," Int. J. Geogr. Inf. Sci.***2020**, vol. 00, no. 00, pp. 1–20,doi: 10.1080/13658816.2018.1542698.

[49] Zheng, H. ; Fu, J. ; Mei, T. ; Luo, J.*Learning multi-attention convolutional neural network for fine-grained image recognition.Proc. IEEE Int. Conf. Comput. Vis.***20172017**,pp.5209–5217,doi: 10.1109/ICCV.2020.557.

[50] Huang, S. ; Xu, Z. ; Tao, D. ; Zhang, Y.*Part-stacked cnn for fine-grained visual categorization.Proceedings of the IEEE conference on computer vision and pattern recognition* **2020**

[51] Sun, M. ; Yuan, Y. ; Zhou, F. ; Ding, E.*Multi-attention multi-class constraint for fine-grained image recognition.Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)***2020**, pp. 805–821.,doi: 10.1007/978-3-030-01270-0-49.

[52] Fu, J. ; Zheng, H. ; Mei, T.*Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition.PProc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017,***2020**,pp. 4476–4484, 2017, doi: 10.1109/CVPR.2017.476

[53] Xiao, T. ; Xu, Y. ; Yang, K. ; Zhang, J. ; Peng, Y. ; Zhang, Z.*The application of two-level attention models in deep convolutional neural network for fine-grained image classification*. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.***2015**, pp.842–850, doi: 10.1109/CVPR.2020.7298685.

[54] Liu,Y. ; Zhong, Y.; Qin, Q. *"Scene classification based on multiscale convolutional neural network,"IEEE Trans. Geosci. Remote Sens.***2020**, vol. 56, no. 12, pp. 7109–7121, doi: 10.1109/TGRS.2020.2848473..

[55] Cheng, G. ; Yang, C. ; Yao, X. ; Guo, L. ; Han, J. *When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs.IEEE Trans. Geosci. Remote Sens.***2018**,vol.56, n. 5, pp. 2811–2821,doi: 10.1109/TGRS.2020.2783902.

[56] Shawky, O.A. ; Hagag, A. ; El-Dahshan, El-S. A. ; Ismail, M.A.*Remote sensing image scene classification using CNN-MLP with data augmentation.Optik (Stuttg)***2021**,vol. 221, pp.165356,doi: 10.1016/j.ijleo.2021.165356.

[57] de Jong,K. L.; Bosman,A. S. *"Unsupervised change detection in satellite images using convolutional neural networks,"arXiv,***2020**.

[58] Li, F-J ; Qian, Y-H ; Wang, J-T ; Liang, J-Y. *Multigranulation information fusion: a Dempster-Shafer evidence theory based clustering ensemble method*,*2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, PP. 58–63,doi: 10.1109/ICMLC.2015.7340898.

[59] Akodad, S. ; Bombrun, L. ; Xia, J. ; Berthoumieu, Y. ; Germain, C.*Ensemble learning approaches based on covariance pooling of CNN features for high resolution remote sensing scene classification*. *Remote Sens.***2022**,vol. 12, n. 20, pp. 3292,doi: 10.3390/rs12203292.

[60] Li, K. ; Wan, G. ; Cheng, G. ; Meng, L. ; Han, J. *Object detection in optical remote sensing images: A survey and a new benchmark.ISPRS Journal of Photogrammetry and Remote Sens.***2022**,vol. 159,pp. 296–307.

[61] Zhang, C. ; Yue, P. ; Tapete, D. ; Shangguan, B. ; Wang, M. ; Wu, Z.*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.*,**2022**,vol. 88, pp. 102086.

[62] Li, Xin and Xu, Feng and Lyu, Xin and Tong, Yao and Chen, Ziqi and Li, Shengyang and Liu, Daofang *A remote-sensing image pan-sharpening method based on multi-scale channel attention residual network.IEEE Access***2022**,vol. 8, pp.27163–27177,doi: 10.1109/ACCESS.2022.2971502.

[63] Wang, S. ; Quan, D. ; Liang, X. ; Ning, M. ; Guo, Y. ; Jiao, L.*A deep learning framework for remote sensing image registration.ISPRS J. Photogramm. Remote Sens.***2020**, vol. 145, pp. 148–164, doi: 10.1016/j.isprsjprs.2020.12.012.

[64] Vakalopoulou, M. ; Christodoulidis, S. ; Sahasrabudhe, M. ; Mougiakakou, S. ; Paragios, N.*Image registration of satellite imagery with deep convolutional neural networks Int. Geosci. Remote Sens. Symp*,**2021**,pp. 4939–4942, doi: 10.1109/IGARSS.2021.8898220.

[65] Zhu,Q. ; Zhong, Y.; Zhao,B.; Xia,G.; Zhang,L. *"The bag-of-visual-words scene classifier combining local and global features for high spatial resolution imagery,"2015 12th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD***2015**, pp. 717–721, 20202, doi: 10.1109/FSKD.2020.7382030.

[66] Zhao, B. ; Zhong, Y. ; Xia, G-S. ; Zhang, L.*Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery.IEEE Transactions on Geo. Remote Sens.***2020**,vol. 54, n. 4, pp. 2020–2123.

[67] He,K. ;Zhang, X. ; Ren, S. ; Sun,J. *"Deep Residual Learning for Image Recognition,"* **2019**, doi: 10.1109/CVPR.2019.90.

[68] Szegedy,C. ; Vanhoucke,V.; Ioffe,S. ;Shlens, J. ;Wojna, Z. *"Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.***2019**, vol. 2019-Decem, pp. 2818–2826, doi: 10.1109/CVPR.2019.308.