Evolving Domain Granulation Based Graph Convolutional Neural Network with Co-Teaching for Remote Sensing Image Classification

D. Arun Kumar *†, S. Rahiman †, G. Anjitha†, and K. Joshna †
* IEEE Senior Member and GRSS Member
† Dept. of Electronics and Communication Engineering, K.S.R.M. College of Engineering, Kadapa, Andhra Pradesh, India - 516005

2024 IEEE India Geoscience and Remote Sensing Symposium (InGARSS) | 979-8-3503-9034-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/InGARSS61818.2024.10984159

***Abstract*—Domain adaptation (DA) is an important area of research in remote sensing image classification. Advancement in the remote sensing (RS) technology, produced images at different resolutions and from different domains within a class. The deep feature extractors like Resnet-50 are used to obtain features from the images. These features possesses uncertainty within. It is an important challenge for a classifier to deal with the uncertainty in the feature vectors. The uncertainty in the feature vectors is maximum when the images are obtained from multiple domains within a class. Motivated with this, we propose an evolving domain granulation (EDG) to deal with the uncertainty in the feature vectors from multiple domains with-in a class. Graph convolutional neural network (GCN) is known for its ability to process the feature vectors and implement DA in a non-regular grid structure. We introduce the concept of co-teaching with ensemble of classifiers in GCN to adapt the evolving domain granulated feature for un-supervised RS image classification. The model is named as evolving domain granulation based graph convolutional neural network with co-teaching (EDG-GCN). The superiority of proposed model over state-of-the-art models is tested with multispectral RS scene images and hyperspectral RS patch images.**

***Index Terms*—Evolving Domain Granulation, Graph Convolutional Neural Network, Multiple classifier based Co-Teaching, and Remote Sensing Image Classification**

I. Introduction

The recently developed deep learning models like convolution neural networks (CNNs) are used to classify patch/scene images. The CNN models like Resnet-50 [1] has deep layers to extract informative feature vector from the input image. The feature vector is processed by a classifier in the classification stage to label it. The feature vectors extracted from the patch/scene images have uncertainty. Data granulation is one of the methods to deal with the uncertainty in the data. Granulation is defined as grouping the objects based the properties like similarity and connectivity [2]. Pal and Mitra [3], suggested class-independent (CI) granulation method to deal with the uncertainty in the feature vectors. In CI, the data is granulated to three fuzzy granules, namely, *low*, *medium*, and *high*. CI granulation do not consider the class-belonging information of the granules. A better data granulation method, namely, class dependent (CD) granulation was suggested by Pal *et. al* [4]. In CD granulation, a feature in the feature vector is represented with its membership to c number of classes. CI and CD granulations are confined to overall information of the data and these two data granulation methods do not consider the domain-level granulation within a class. With this motivation, we introduce evolving domain granulation (EDG) method to deal with the uncertainty in the feature vectors at the domain level within a class. In EDG, a feature in the feature vector is represented by its membership to the domain granules with in a class.

The feature vector obtained by using reset-50 feature extractor is processed by a classifier and it is labeled in the classification stage of CNN. Different type of classifiers like neural networks, support vector machines, etc., are used in the classification stage of CNN to label the feature vectors. Among these classifiers, GNN is known for its better domain adaptability due to its non-grid based method of processing the data [5]. In the case of un-supervised domain adaptation, where, the class label of an image is not available, the concept of co-teaching is introduced to generate the pseudo label of an unknown feature vector, after which, based on the co-teaching, GCN is used to label the feature vector [6] and [7]. In [6] and [7], a single classifier, namely, multi-layer perceptron is used to generate pseudo labels of unsupervised data. A single classifier cannot consider the overall information to generate a better pseudo label. Motivated with this, we propose ensemble of classifiers based pseudo labeling in GCN for un-supervised domain adaptation. The novelty of the present work is three fold, i.e., (i) Evolving domain granulation to deal with the uncertainty in the data at the domain-level within a class, (ii) Ensemble of classifiers based co-teaching in GCN to generate pseudo labels for unsupervised DA, and (iii) Applicability of the proposed model to implement domain adaptation for patch/scence based RS images. Here, the proposed model is named as evolving domain granulation based graph convolutional neural network with co-teaching for remote sensing image classification.

II. Proposed Model for RS Image DA

The flow diagram of EDG-GCN is shown in Fig. 1. In the first step, a Resnet-50 feature extractor with the series of fifty convolutional and pooling layers is applied on input image to obtain the informative features. EDG is applied on the feature vector and each feature vector is represented with its membership to the domain granule within a class. These domain granulated features are passed through the ensemble of classifiers to generate pseudo label of the feature vector. Based on this pseudo label, the elements of affinity matrix in GCN are defined and furthermore, GCN is used to label the feature vector. The detailed architecture of EDG-GCN is shown in Fig. 2. In Fig. 2, a Resnet-50 feature extractor with three layers of

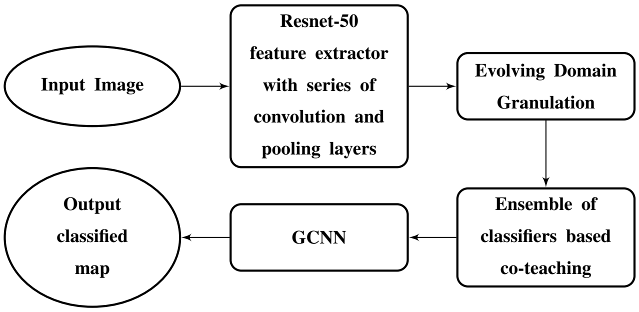

Fig. 1: Functional block diagram of EDG-GCN

convolutional and pooling layers (as an example) is applied on input image to obtain a feature vector of dimension 1×4. This feature vector is passed through the ensemble of classifiers to generate pseudo label. The concept of co-teaching is applied in GCN by defining the elements of affinity matrix based on the pseudo label. Here, GCN is trained based on the batch gradient descent method and furthermore, EDG feature vector is labeled.

A. Evolving Domain Granulation

The feature vector from a domain within a class is granulated using a trapezoidal membership function. There are two important functions implemented in the domain granulation, (i) Creation of a granule and (ii) Updation of a granule.

1) Creating a Domain Granule: A new domain granule is created when the feature vector is arrived from a new class. It is also created when the feature vector do not fit within the existing domain granules. This is shown in Fig. 3 (a). In Fig. 3, a pth granule of a class c is created for a feature value f_n along the feature axis. The lower bound and upper bound of a granule is computed as $f_n - \sigma$ and $f_n + \sigma$, where, σ is a constant (generally considered as the standard deviation of feature values for feature f_n). The membership value is maximum value (1) at the center of the granule i.e., at f_n. The membership value decreases on either sides of the center value.

2) Updating a Domain Granule: A granule is updated when a new feature value (f_n^{new}) lies within it. The domain granule is adjusted so that the membership of a feature value is maximum when it is between f_n^{new} and f_n. The membership value decreases on either sides of the f_n^{new} and f_n. This approach is implemented by considering all the feature vectors in the dataset. In EDC, a feature vector P with n number of features and c number of classes is represented by $n \times c$ number of domain granulated features as,

$$P = \begin{bmatrix} \max(\mu_1(f_1^1), \cdots, \mu_1(f_1^m)), \\ \max(\mu_2(f_1^1), \cdots, \mu_2(f_1^m)), \\ \cdots, \max(\mu_c(f_1^1), \cdots, \mu_c(f_1^m)), \\ \cdots, \max(\mu_c(f_n^1), \cdots, \mu_c(f_n^m)) \end{bmatrix}. \quad (1)$$

$\max(\mu_c(f_n^1), \cdots, \mu_c(f_n^m))$ is the maximum membership among m domain membership values of feature f_n in class c. $\mu_c(f_n^m)$ is the membership of a feature f_n to the m^{th} domain of class c.

B. Graph Convolutional Neural Network

The architecture of GCN is similar to a single hidden layer neural network (NN) like multi-layer perceptron (MLP). Pictorial representation of GCN architecture is shown in Fig. 4. In Fig. 4, GCN has an architecture of n:s:c, where, n is the number of granulated features in the informative feature vector, s is the number of nodes in the hidden layer, and c is the number of classes in the dataset. w_{ns} is the weight between n^{th} node of input layer and s^{th} node of hidden layer, and v_{sc} is the weight between s^{th} node in the hidden layer and c^{th} node in the output layer. Let, $X_{m \times n}$, represents m granulated feature vectors with n features in a batch. These feature vectors are mathematically represented by a graph $G = \{V, E, A\}$, where V, E, and A denotes nodes, edges, and adjacency of a graph, respectively [8]. A feature vector in X is considered as a node in a graph and its adjacency represents a edge weight in a graph. The elements of affinity matrix A_{ij} of m feature vectors is computed as, A_{ij}= 1, if x_i and x_j belong to a same class and A_{ij} is 0, otherwise. x_i and x_j are the ith and jth feature vectors in X. The diagonal degree matrix (D_{ii}) of A_{ij} is computed as, $D_{ii} = \sum_j A_{ij}$ The normalized adjacency matrix $\overline{A_{ij}}$ is given as, $\overline{A} = D^{-1/2}AD^{-1/2}$. The hidden layer output of GCN is computed as, $H = \overline{A}X$. The H matrix is feed-forwarded to the next layers of GCN and it is trained by using batch gradient descent algorithm.

C. Ensemble of Classifiers based Co-Teaching in GCN

In ensemble of classifiers based co-teaching in GCN, the unsupervised feature vector is passed through the ensemble of classifiers to generate pseudo class label to the feature vector. The pseudo label generated by the ensemble of classifiers is used to define the elements of affinity matrix (as explained in sec. II-B).

III. Results and Discussion

A. Datasets Used

The performance of EDG and its superiority over similar classification models is demonstrated with multispectral and hyperspectral RS image data sets. Four multispectral scene datasets, named Aerial Image (AID), Northwestern Polytechnical University (NWPU), University California Merced (UCM), and remote sensing scene classification (RSCNN7), and one hyperspectral dataset, named, Indian Pines (IP) is used in the present study. These data sets are described in Table I. The classes like residential, farm land, parking, forest, and river are common in AID, NWPU, UCM, and RSCNN7 datasets. These five classes were considered to implement DA. The scene image (with a single class in an image) from forest class of AID dataset is shown in Fig. 5 (a). The hyperspectral image of IP is shown in Fig. 5 (b). The hyperspectral images consist of small images (patches of size 3×3), each of those representing a small region/patch. The IP dataset is divided in to four sets, namely, Set-1, Set-2, Set-3, and Set-4 to implement DA.

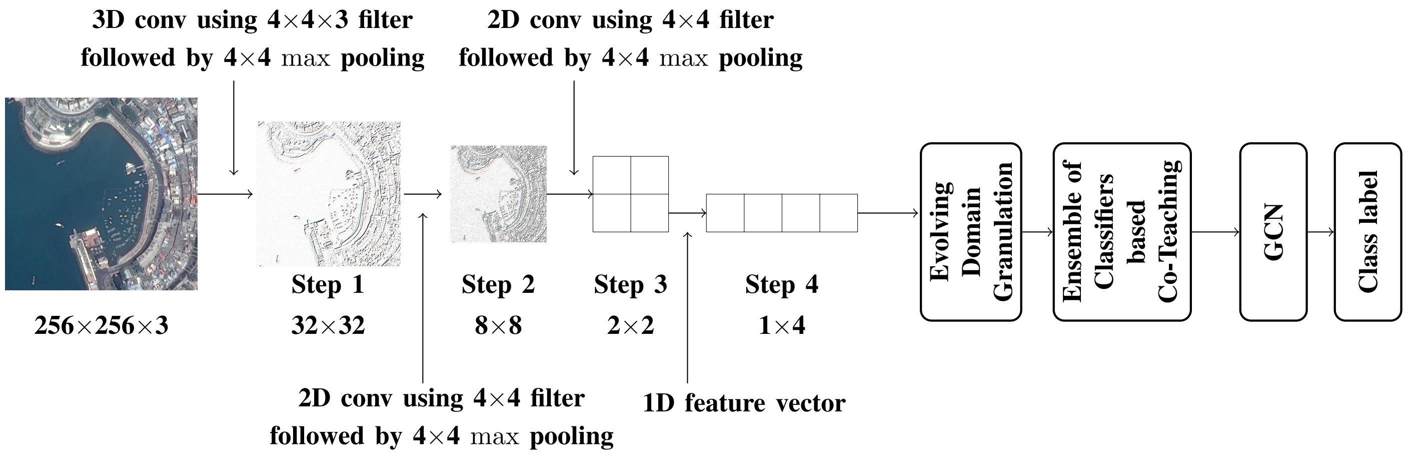

Fig. 2: Architecture of EDG-GCN

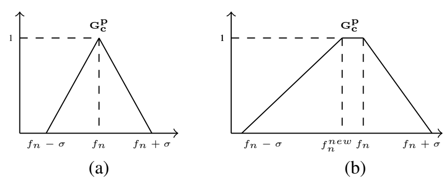

Fig. 3: (a) Creating a new domain granule and (b) Updating a domain granule. G_c^p is the p^{th} domain granule of class c

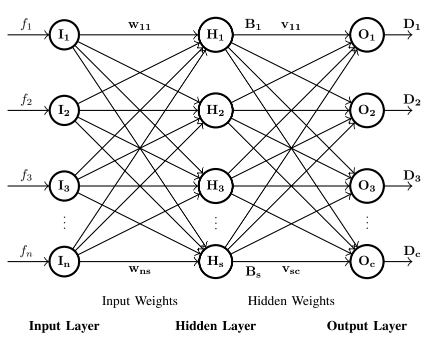

Fig. 4: Architecture of graph convolutional neural network

(a) (b)

Fig. 5: (a) RS scene image with single class (Forest - AID dataset), and (b) RS patch image with multiple classes (IP)

TABLE I: Description of the data sets

Type of classification	Name	Image size	Bands	Classes
Scene (Multispectral)	AID	600 × 600	3	15
	NWPU	256× 256	3	19
	RSCNN7	400 × 400	3	7
	UCM	256× 256	3	21
Patch (Hyperspectral)	IP	145 × 145	224	16

B. Model description

Five models were considered to implement the performance comparison of EDG-GNN. The description of the models is based on the type of feature extractor, type of data granulation, the type of classifier used for co-teaching, and the type of classifier in the classification stage. This is represented as, *Model : Feature Extractor + Type of Granulation + Type of co-teaching + Classifier stage*, *Model 1: Resnet-50+None+Single classifer+GCN (base model) [6], Model 2: Resnet-50+None+ensemble of classifiers+GCN, Model 3: Resnet-50+CI+ensemble of classifiers+GCN, Model 4: Resnet-50+CD+ensemble of classifiers+GCN, Model 5 (proposed): Resnet-50+EDG+ensemble of classifiers+GCN*. The models 1-5 have Resnet-50 as a feature extractor and GCN as a classifier in the classification stage. Model 1 is CNN with Resnet-50 feature extractor and un-granulated data is provided as an input. It has multi-layered perceptron (MLP) as a single classifier used in the co-teaching of GCN. Model 2 is CNN with Resnet-50 feature extractor and un-granulated data is provided as an input to it. It has ensemble of classifiers (the models like MLP, support vector machine (SVM), random forest (RF), and decision tree (DT) are used in the ensemble of classifiers). Model 3 is CNN with CI granulated input. Model 4 is CNN with CD granulated input and Model 5 is EDG-GCN (proposed model) with EDG based input data. The performance of these models is evaluated using the metri cs like overall accuracy (OA), overall accuracy standard deviation (OA_{STD}), kappa coefficient (K), kappa coefficient standard deviation (K_{STD}), F-score (F), computational time (T_c), and

TABLE II: Performances of the models in SSMT DA

Model	Single-Source Multi Target DA (AID, NWPU → RSCNN7, UCM)										
	Training				Performance Metrics						
	20%	40%	60%	80%	OA%	OA_{STD}	K	K_{STD}	F	$T_c(s)$	NI
M_1	78.88	79.74	80.93	81.86	80.35	3.02	0.74	0.19	0.71	5.02	80
M_2	79.46	80.94	81.89	82.45	81.18	2.67	0.76	0.16	0.76	5.32	80
M_3	80.86	81.67	82.45	83.78	82.19	2.02	0.79	0.14	0.78	5.17	80
M_4	81.98	82.87	83.78	84.96	83.39	1.98	0.84	0.11	0.81	5.03	80
M_5	82.83	83.73	84.59	85.83	84.24	0.63	0.86	0.08	0.83	4.28	80

number of training iterations (NI).

C. Performances of the models in SSMT (scene images)

An un-supervised single source-multi target (SSMT) DA is implemented with the four scene image data sets. One of these four datasets (AID) is used to prepare the source set and the remaining three of these datasets (NWPU, RSSCN7, UCM) are used to prepare the target set. The source (S) to target (T) DA is indicated as S→T. Here, the source (AID) and the target (NWPU, RSSCN7, UCM) DA is indicated as AID→NWPU, RSSCN7, UCM. The selection of target set among the three target datasets is based on the minimum mean distance of target data from the mean distance of source data. The training is implemented in a sequential order by considering one target set after the other. The performance of the models 1-5 is tested with four scene datasets in SSMT DA and the results are listed in Table II. Two facets of comparison is implemented in this study. In the first facet, comparison is made between model 1 and model 2 to demonstrate the improvement in the performance of GCN after using ensemble of classifiers in the co-teaching. In Table II, model 2 is 0.83% superior to model 1. The superiority of model 1 (base model [6]) over model 2 is due to the use of ensemble of classifiers in co-teaching of GCN which will consider the collective information to generate pseudo label. In the second facet, models 2-5 are compared to demonstrate the superiority of EDG over other data granulation methods. Among these models, model 5 has an highest OA with value 84.24%. It is 3.25%, 2.05%, and 0.85% better than model 2, model 3, and model 4, respectively. Model 5 is 3.89% better the model 1 (base model [6]). The superiority of model 5 over other models is justified in terms OA_{STD}, K, K_{STD}, and F. Also, we tested EDG-GCN with the state of the art models like UniDACNN[9], CGCNDA[7], MGCNDA[6], MSDA[10], and MLDA[11]. It is noticed that model 5 is superior to state-of-the-art models interms of the performance metrics. The performance of the models is tested with IP dataset in SSMT-DA. The dataset is divided in to four disjoint sets (set-1, set-2, set-3, and set-4), such that each set has image patches from all the classes. In this comparison, model 5 outperformed model 1-4 and state of the art models in terms of performance metrics. The results are depicted in Table III.

IV. Conclusion

A EDG-GCN model is proposed in the present study to implement SSMT DA using RS scene and patch image datasets. The concept of evolving domain granulation is used in the proposed model to effectively handle the uncertainty in the feature vectors obtained by using feature extractor (Resnet-50). An ensemble of classifiers based co-teaching in the proposed model could provide better pseudo labels to the un-supervised feature vectors. The EDG-GCN could implement SSMT DA better than the state of the art models because of these two characteristics i.e., EDG and ensemble of classifiers based co-teaching. It could produce an overall accuracy of 84.24% as tested with RS scene image dataset (SSMT-DA). The proposed model is 3.89% better than the base model ([6]). The superiority of the model over the state of the art models is similar to scene image data when tested with hyperspectral RS patch image dataset.

TABLE III: Performance of models with IP data set (set 1 → set 2, set 3, set 4)

Type	Data set	Model	OA%	OA_{STD}	F	$T_c(s)$	K	K_{STD}	NOI
Patch	IP	M_1	88.64	2.08	0.78	5.14	0.72	0.22	80
		MLDA[11]	90.14	1.78	0.81	4.96	0.75	0.11	80
		M_4	91.66	1.28	0.83	4.88	0.78	0.10	80
		M_5	92.74	0.72	0.85	4.87	0.84	0.08	1

V. Acknowledgement

Dr. D. Arun Kumar is thankful to the management of K.S.R.M College of Engineering for supporting the work under the seed fund project scheme (K.S.R.M/seed_project/2024-25/01). He is also thankful to Dr. B. S. Daya Sagar, Head and Professor, Indian Statistical Institute - Bangalore Center for his valuable inputs in the execution of the work.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[2] L. A. Zadeh, “Fuzzy sets,” *INFORM CONTROL*, vol. 8, pp. 338–353, 1965.

[3] S. K. Pal and S. Mitra, “Multilayer perceptron, fuzzy sets, and classification,” *IEEE Trans. Neural Networks*, vol. 3, pp. 683–697, 1992.

[4] S. K. Pal, S. K. Meher, and S. Dutta, “Class-dependent rough-fuzzy granular space, dispersion index and classification,” *PATTERN RECOGN*, vol. 45, pp. 2690–2707, 2012.

[5] Y. Xu, Y. Zhang, T. Yue, C. Yu, and H. Li, “Graph-based domain adaptation few-shot learning for hyperspectral image classification,” *Remote Sensing*, vol. 15, no. 4, p. 1125, 2023.

[6] S. Saha, S. Zhao, and X. X. Zhu, “Multi-target domain adaptation for remote sensing classification using graph neural network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[7] S. Saha, L. Mou, X. X. Zhu, F. Bovolo, and L. Bruzzone, “Semi-supervised change detection using graph convolutional network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 607–611, 2021.

[8] Z. Zhang, Y. Cai, W. Gong, X. Liu, and Z. Cai, “Graph convolutional extreme learning machine,” *IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2020.

[9] Q. Xu, Y. Shi, X. Yuan, and X. X. Zhu, “Universal domain adaptation for remote sensing image scene classification,” *IEEE Trans. Geosci. Remote Sens*, vol. 61, pp. 1–15, 2023.

[10] Y. Yin, Z. Yang, H. Hu, and X. Wu, “Universal multi-source domain adaptation for image classification,” *PATTERN RECOGN*, vol. 121, pp. 1–13, 2022.

[11] M. Cai, B. Xi, J. Li, S. Feng, Y. Li, Z. Li, and J. Chanussot, “Mind the gap: Multi-level unsupervised domain adaptation for cross-scene hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–6, 2024.