ACTA IMEKO
ISSN: 2221-870X
March 2022, Volume 11, Number 1, 1 - 8

Multilayer feature fusion using covariance for remote sensing scene classification

S. Thirumaladevi[1], K. Veera Swamy[2], M. Sailaja[3]

[1] *ECE Department, Jawaharlal Nehru Technological University, Kakinada - 533003, Andhra Pradesh, India*
[2] *ECE Department, Vasavi College of Engineering, Ibrahimbagh, Hyderabad - 500 031, Telangana, India*
[3] *ECE Department, Jawaharlal Nehru Technological University, Kakinada - 533003, Andhra Pradesh, India*

ABSTRACT
Remote sensing images are obtained by electromagnetic measurement from the terrain of interest. In high-resolution remote sensing imageries extraction measurement technology plays a vital role. The scene classification is one of the interesting and challenging problems due to the similarity of image structure and the available HRRS image datasets are all small. Training new Convolutional Neural Networks (CNN) using small datasets is prone to overfitting and poor attainability. To overcome this situation using the features produced by pre-trained convolutional nets and using those features to train an image classifier. To retrieve informative features from these images we use the existing Alex Net, VGG16, and VGG19 frameworks as a feature extractor. To increase classification performance further makes an innovative contribution fusion of multilayer features obtained by using covariance. First, to extract multilayer features, a pre-trained CNN model is used. The features are then stacked, downsampling is used to stack features of different spatial dimensions together and the covariance for the stacked features is calculated. Finally, the resulting covariance matrices are employed as features in a support vector machine classification. The results of the experiments, which were conducted on two difficult data sets, UC Merced and SIRI-WHU. The proposed Staked Covariance method consistently outperforms and achieves better classification performance. Achieves accuracy by an average of 6 % and 4 %, respectively, when compared to corresponding pre-trained CNN scene classification methods.

Section: RESEARCH PAPER

Keywords: Feature extraction; pre-trained convolutional neural networks; support vector machine; scene classification

Citation: S. Thirumaladevi, K. Veera Swamy, M. Sailaja, Multilayer feature fusion using covariance for remote sensing scene classification, Acta IMEKO, vol. 11, no. 1, article 33, March 2022, identifier: IMEKO-ACTA-11 (2022)-01-33

Section Editor: Md Zia Ur Rahman, Koneru Lakshmaiah Education Foundation, Guntur, India

Received December 25, 2021; **In final form** February 20, 2022; **Published** March 2022

Corresponding author: S. Thirumaladevi, e-mail: thirumaladeviece1@gmail.com

1. INTRODUCTION

Remote sensing scene categorization has gotten a lot of attention recently, and it may be utilized in a variety of practical applications like urban planning, defence, space applications, in which measurement technology plays a key role [1]. On the other hand, it is a difficult challenge, since scene images often have complicated spatial structures with great intra-class and slight inter-class variability. To solve this problem, numerous strategies for scene classification have been advised in recent years [2]. Recently, inspired by the tremendous achievement of convolutional neural networks (CNNs) relating to the computer vision field [3]. Deep neural networks have gained prominence in the remote sensing community due to their exceptional performance, particularly in scene classification and computer vision applications [4]. Developing a deep CNN model from scratch, on the other hand, frequently necessitates a large amount of training data, whereas commercially available scene image data sets of remote sensing scene image data sets are typically tiny. Deep CNN models have a high degree of generalization on a wide range of tasks because they are commonly trained on ImageNet [5], which contains millions of images (e.g., scene classification and object detection [6]). In this context, the idea of using off-the-shelf pretrained CNN models for example AlexNet [7], Visual Geometry Group (VGG) 16 [8], and VGG19 as feature extractors for scene categorization using remote sensing has gained attraction. The success is due to these models representing images using hierarchical architecture and can extract more representative Features. While these models can achieve categorization performance is excellent. Hu et al. [9] looked into two scenarios for using a pretrained CNN model

(VGG16). The final few fully connected layers are portrayed as final image attributes for scene classification in the first scenario. In the second case, the final convolutional layer's feature maps are encoded to represent the input image using a standard method of feature encoding, such as the improved Fisher kernel [10]. The Vector Support Machine (SVM) is used as the final classifier in both cases. To improve the efficacy of the proposed method, the features were extracted from multiple CNNs of the same image combined by Xue et al. [11] for classification. For feature fusion, Sun et al. [12] have used the gated bidirectional connection method. In [13], the image is represented by combining the last two fully connected (FC) layers of a CNN model.

Here we propose an innovative method, called the Stacked Covariance strategy to fuse features from different layers of a pre-trained CNN to classify remote sensing scenes. In the first phase, a pre-trained CNN model is used to extract multi-layered features and concatenate them. The covariance approach is used to aggregate the concatenated multiple feature vectors extracted from different layers. In contrast to traditional strategies, which only use first-order statistics to integrate feature vectors, the proposed strategy allows for the use of second-order statistics information. More representative features can thus be learned as a result. Then, the features are stacked, and covariance is calculated. Finally, for classification using an SVM classifier and improved the classification performance.

This is how the rest of the paper is organized. Section 2 explains the intended scene classification framework, novel aspects of our proposed technique. Section 3 contains the full experimental results for two data sets, Section 4 of this work concludes with some observations.

2. PROPOSED TECHNIQUE DESCRIPTION

The process of transforming the raw image into numerical features that can be processed while retaining the original information is referred to as feature extraction. With the upsurge of deep learning, the first layers of deep networks have largely replaced feature extraction, particularly for image data. Pre-trained networks with hierarchical architecture can extract a large number of features from an image, which is thought to transmit additional information that can be put to much better use to increase categorization accuracy. The Learned image characteristics are first retrieved from a pre-trained convolutional neural network and then used to train an image classifier. All pre-trained CNNs requisite fixed-size input images and specify the desired image size, as well as create augmented image data stores and use these data stores as input arguments to activations to automatically resize the training and test images before they are submitted to the network. Remove the pre-trained CNN's last FC layer FC8 and consider the rest as a fixed feature extractor. We feed an input image scene into the CNN and using the vector as a representation of global features of the input image, generate in advance a dimensional activation vector from the first or second FC layer. Finally, use the dimensional features to train a linear SVM classifier for scene classification. Figure 1 shows an illustration of this.

To improve the classification accuracy further proposed a modified pre-trained network design by combining information from several convolution layers. The shallower levels of the CNN model are more likely to represent low-level visual components (such as edges), whereas the deeper layers exemplify more abstract information in the images. Furthermore, in certain computer vision applications, combining different levels, from shallow to deep, can provide state-of-the-art performance, meaning that merging different layers of CNN can be very helpful.

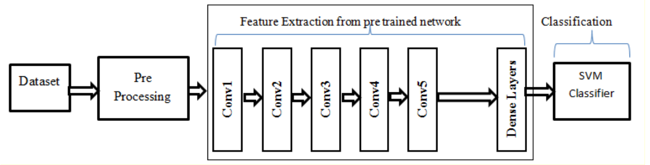


Figure 1. Classification using single-layered pre-trained CNN as a feature extractor.

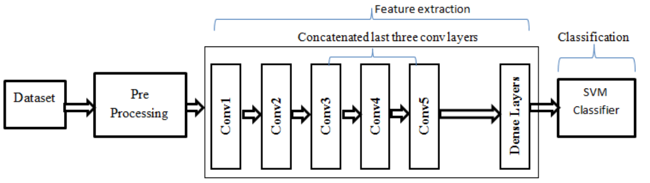


Figure 2. Classification using stacked multilayer pre-trained CNN as a feature extractor.

Our proposed approach uses a similar strategy to take advantage of the information held by multiple layers in this way. This is represented in Figure 2. Here Convolutional layers of the last three blocks of pretrained networks are adopted and concatenated the features extracted from these layers. Namely conv3, conv4, conv5 in case of AlexNet and "conv3-3", "conv4-3", and "conv5-3" in case of VGG16, VGG19. Different convolutional layers predominantly have distinctive spatial dimensions, they cannot be directly concatenated. To address this issue, downsampling with bilinear interpolation is used in conjunction with channel-wise average fusion. Obtained features that were reformed into a matrix along the channel dimension and aggregated using covariance. The proposed technique is described below.

CNN Model is a collection of functions in which each function f_n takes data samples X_n and a filter bank b_n as inputs and outputs X_{n+1} , where $n = 1, 2, \dots N$ and N is the number of layers represented as a number (1)

$$F(X) = f_N(\dots f_2(f_1(X; b_1); b_2) \ \dots \ b_N) \, . \tag{1}$$

The filter bank b_n for a pretrained CNN model was learned from a large data collection. The multiplayer characteristics are retrieved from an input image X as follows: $L_1 = f_1(X; b_1)$, $L_2 = f_2(X; b_2)$, and so on. As pretrained models, AlexNet, VGG16, and VGG19 are employed in this paper. The features produced from the convolutional layers of the last three blocks of pretrained networks are adopted and utilized. Different convolutional layers typically have different spatial dimensions; therefore, they can't be concatenated directly. Direct concatenation is not allowed when conv3 may have $L_1 \in {R_1}^{h \times w \times d_1}$, conv4 $L_2 \in {R_2}^{h \times w \times d_2}$, and conv5 has $L_3 \in {R_3}^{h \times w \times d_3}$. Downsampling with bilinear interpolation, as well as channel-wise average fusion, are used to solve this problem. Using downsampling with a d number of dimensions three convolutional layers that have been pre-processed have been obtained and channel-wise average fusion is performed then the stacked feature set is acquired as follows: $L = [L_1, L_2, L_3] \in {R_i}^{S \times S \times D}$, where $D = 7\ d$ and S is the predefined down-sampled spatial dimension. The covariance-based pooling can be written as [14]

$$P = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \mu)\,(Y_i - \mu)^{\mathrm{T}} \;\in R_i{}^{D\times D}\,, \tag{2}$$

where $[Y_1, Y_2, \dots Y_N] \in R_i{}^{D\times N}$ is the vectorization of L, $N = S^2$ and $\mu = (1/N)\sum_{i=1}^{N} Y_i \in R^D$, the covariance between the two separate features makes is represented by P, while the variance of each map is represented by diagonal entries. This method incorporates covariance (i.e., second-order statistics) to produce a more dense and discriminative exemplification. Second, the correlation between two distinct feature maps is represented by each entry in the covariance matrix. This is an easy approach to merge data from different feature maps that complement each other.

The suggested method varies from existing CNN-based algorithms that are pre-trained. Concatenating the CNN's unique convolutional features (from shallow to deep layers), feature maps from several layers are merged. As a result, the suggested technique performs much better in terms of categorization. Furthermore, because the covariance matrices do not lie in Euclidean space, they cannot be processed by the SVM. The covariance matrix, on the other hand, can be mapped into Euclidean space using the matrix logarithm operation [15].

$$\hat{P} = \mathrm{logm}(P) = U \log \sum U^{\mathrm{T}} \;\in R^{D\times D}\,, \tag{3}$$

where, $P = U \sum U^{\mathrm{T}}$ is the Eigen decomposition equation of P. The preceding operations are carried out both on the samples of training and testing. $\{V_i, S_i\}, i = 1, 2, \dots n$ and the testing sets are now taken into account where S_i represents the number of training samples and n represents the number of corresponding labels.

$\{V_i, S_i\}, i = 1, 2, \dots n$ is exploited to train an SVM model as

$$\begin{aligned} &\mathrm{Min}_{a,\zeta\,,b} \left\{1/2\;\|a\|_2 + P \sum \zeta_i\right\} \\ &S_i\big(\angle\varphi(V_i, a + b)\big) \geq 1 - \zeta_i \\ &\varepsilon_i > 0, V_i = 1, 2, \dots n \end{aligned} \tag{4}$$

where a and b are the parameters of a linear classifier (.) is the mapping function and ε_i are positive slack variables to assert with outliers in the training set.

With $\mathrm{k}\big(V_i, V_j\big) \;= V_i{}^{\mathrm{T}} V_j$

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} S_i\,\lambda_i\,k(v_i, v) + b\right). \tag{5}$$

3. EXPERIMENTAL RESULTS ANALYSIS DISCUSSION

3.1 Experimental Data Sets

We run tests on two tough Image data sets related to remote sensing scene images to see how well the suggested approach performs. 1) Land Use Data Set from UC Merced [16]: 2100 pictures are classified into 21 scene groups in the UC Merced Land Use (UC) [17] data set. Each class consists of 100 images in the RGB space with a size of 256 × 256 pixels. Each image has a one-foot pixel resolution. Figure 3. depicts sample images

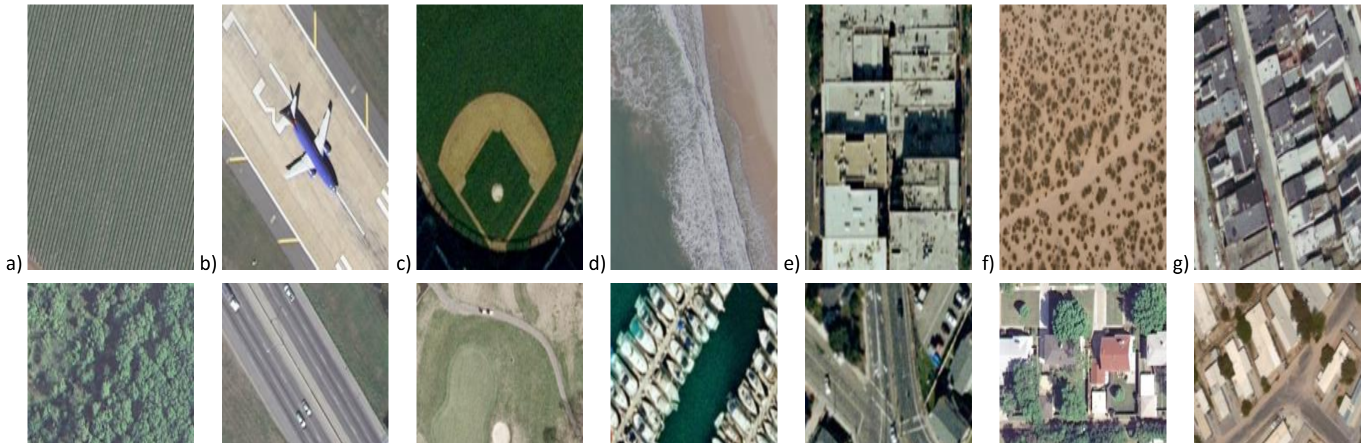


Figure 3. Land-use categories of 21 Example classes representation of UC Merced data set : a) agricultural13, b) airplane19, c) baseball diamond3, d) beach33, e) buildings21, f) chaparral13, g) denseresidential40, h) forest23, i) freeway23, j) golfcourse41, k) harbour31, l) intersection3, m) mediumresidential12, n) mobilehomepark12, o) overpass45, p) parkinglot32, q) river32, r) runway26, s) sparse residential64, t) storagetanks54, u) tenniscourt16.

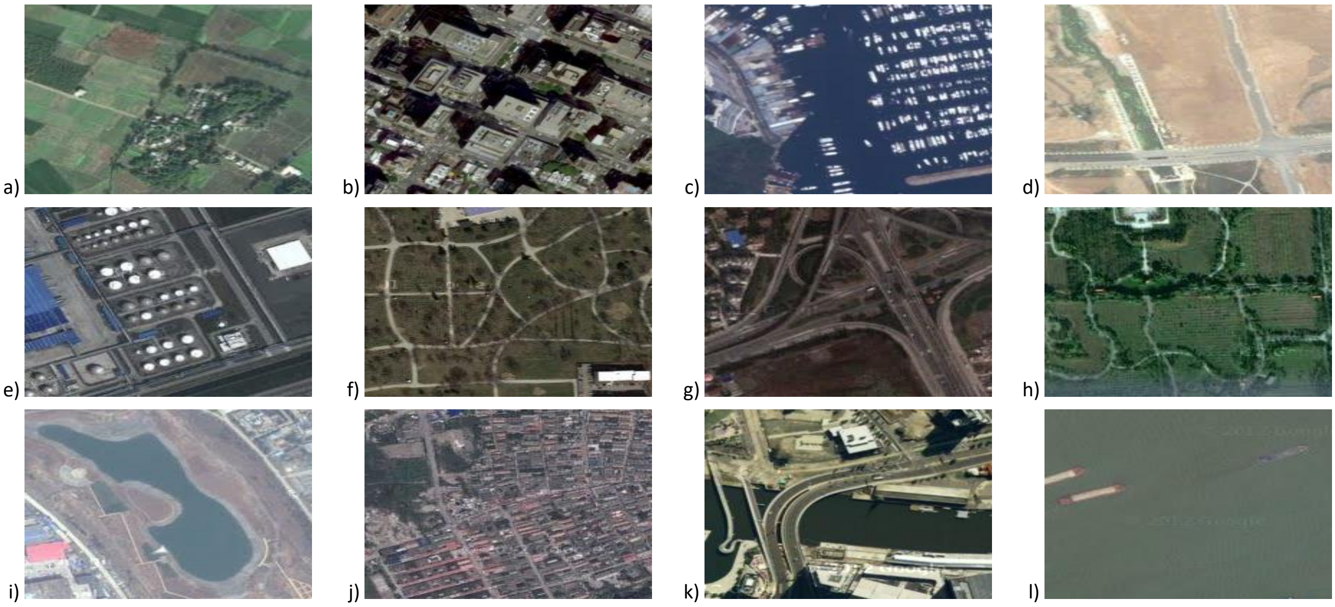


Figure 4. Example class representation of the SIRI-WHU dataset: a) agriculture1, b) commercial50, c) harbor64, d) idle_land76, e) industrial111, f) meadow120, g) overpass15, h) park29, i) pond37, j) residential1, k) river106, l) water97.

from each class, some categories (forest and sparse residential, for example) exhibit a significant level of interclass similarity, making the UC data set a difficult one to work with.

SIRI-WHU was obtained from Google Earth (Google Inc.) and covers urban regions in China, as well as SIRI-WHU [18]. There are 12 items in the data set. Each class has 200 photos, each cropped to 200 × 200 pixels and with a spatial resolution of 2 meters. In this study, 80 % of the training samples were chosen from the SIRI-WHU [19] Google data set, while the remaining amount of samples was kept for testing. The sample images of the SIRI-WHU data set are shown in Figure 4.

3.2 Experimental Setup

In our approach, multilayer features are extracted using three well-known CNN pretrained models: AlexNet [7], VGG-19 [8], and VGG-16 [8]. VGG-16 and VGG-19 three convolutional layers (e.g., "conv3–3," "conv4–3," and "conv5–3") , AlexNet's three convolutional layers (i.e., "conv3," "conv4", and "conv5") are used. For feature extraction, the scene images are preset to the size of the input layer such as 227 × 227 × 3 in the case of Alex Net and 224 × 224 × 3 forVGG-VG16 and 19. ImageNet is used to train both models. For illustration purposes, the UC data set and the SIRI-WHU data set are used, with 80 % training samples and 20 % testing samples selected. The most frequently used image classification assessment criteria are OA and confusion matrix, F1-score.

Confusion Matrix: This is a special matrix that is commonly used g to visualize the output. Each column in this matrix represents the estimated value, while each row signifies an authentic category. As a result, evaluating is relatively simple.

Overall Accuracy (OA): The number of appropriately categorized images Divide by the total number of images in the data set, regardless of which class they belong to.

F1 score: The *F1* score is a metric used to determine how accurate a test is. The harmonic mean of recall and precision is calculated using the test's precision and recall.

Based on the combination of real and anticipated categories, classification problem with several M, which comprises P positive instances and N negative instances, There are four types of cases: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The positive sample $P = TP + FN$ provides a positive sample that is expected to be positive, while TP represents a positive sample that is forecast to be negative. Similarly, TN denotes the number of negative cases that are identified as negative, while FP denotes the number of negative incidences that are predicted to be positive; thus, $N = TN + FP$ denotes the total number of negative samples.

The fraction of correct instances is the accuracy, and the calculation equation is

$$Accuracy = \frac{TP}{TP + FN + TN + FP}. \tag{6}$$

The fraction of actually positive instances in all cases projected to be positive elements is called precision. The formula is as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{7}$$

The recall calculation equation is the fraction of all positive samples that are projected to be positive.

$$Recall = \frac{TP}{TP + FN}. \tag{8}$$

The F1-score is a precision and recall evaluation indicator with a comprehensive calculation equation.

$$Recall = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \tag{9}$$

The confusion matrix, along with accuracy, is shown in Figure 5. The first column depicts a single fully connected layer is used as the final feature extractor for scene classification, whereas the second column depicts the proposed SC-based network classification. Experiments on the UC data set revealed that the OAs of pre-trained AlexNets is 79.76 %, VGG-19 is 81.19 %, and VGG-16 is 83.81 % while using SC and combining the last

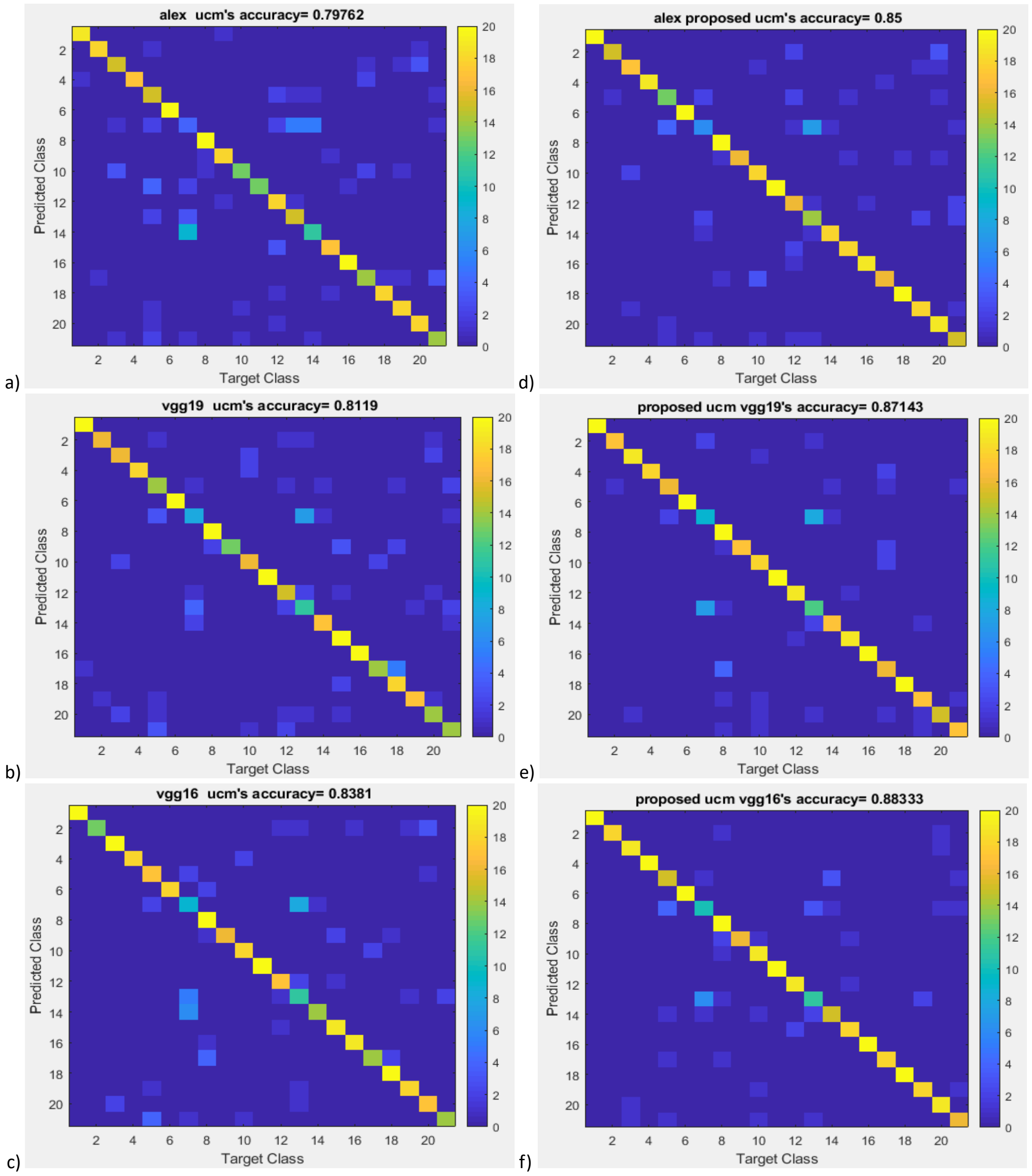


Figure 5. Confusion matrix of UC Merced Dataset using three pre-trained networks. First column corresponding to single-layered, a) Alex Net, b) VGG19, c) VGG16, second column corresponding to multi-layered fusion, d) SC-Alex Net, e) SC-VGG19, f) SC-VGG16.

Table 1. Comparison results for the two data sets UCM and SIRI-WHU.

	UCM Dataset with 80% training (Overall Accuracy %)		SIRI-WHU Dataset with 80% training (Overall Accuracy %)	
Network/Method	Pre-trained network FC7 as a feature extractor	Proposed SC network as a feature extractor	Pre-trained network FC7 as a feature extractor	Proposed SC network as a feature extractor
AlexNet	79.76	85	86.52	90
VGG-VD19	81.19	87.14	87.60	91.08
VGG-VD16	83.81	88.33	88.04	92.60

three Conv layers increases accuracy by 85 %, 87.14 %, and 88.33 %, respectively. The proposed method accomplishes perfect classification performance on the majority of classes, such as Agricultural, beach, chaparral, forest, harbour, parking lot, runway and improved classes are buildings, dense residential, baseball diamond, tennis court and on average 6 % accuracy is increased in the case of UCM.

Similar results can be obtained in the experiments conducted on the SIRI-WHU data set, confusion matrix for the fully connected layer is treated as a final feature extractor for scene classification, proposed SC-based classification are shown in Figure 6, where the pre-processed single layer of AlexNet is 86.52 %%, VGG-19 is 87.6 %, and VGG-16 is 88.04 % while using the proposed strategy increases the accuracy by 90 %,

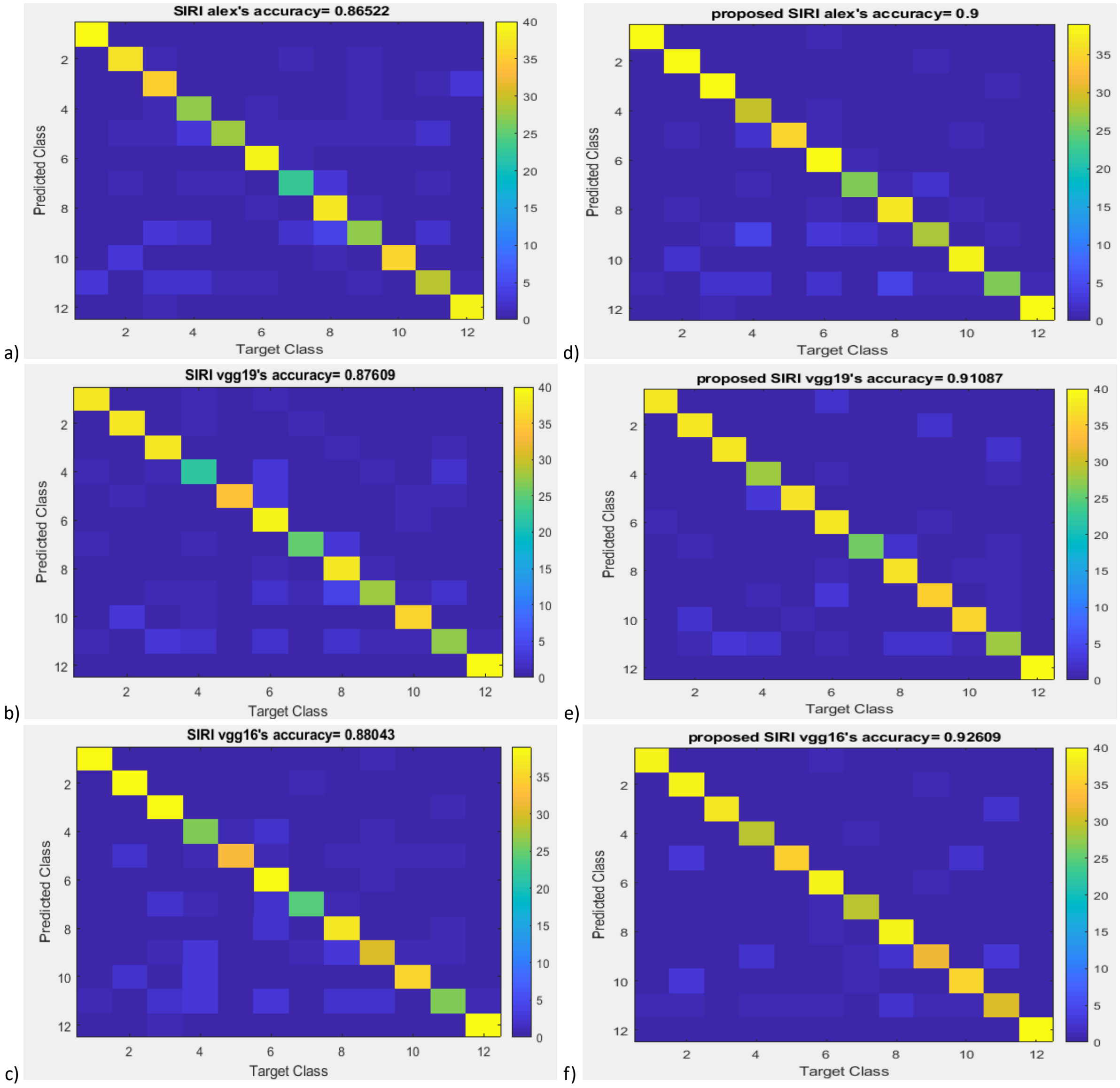


Figure 6. Confusion matrix of SIRI-WHU Dataset using three pre-trained networks. First column corresponding to single-layered, a) Alex Net, b) VGG19, c) VGG16, second column corresponding to multi-layered fusion, d) SC-Alex Net, e) SC-VGG19, f) SC-VGG16.

91.08 %, and 92.60 %, respectively. In most classes, the suggested technique achieves optimal classification performance like Agriculture, commercial, harbour, meadow, meadow and improved classes are Industrial, overpass, pond, overall 4 % accuracy is increased and comparison graphs are shown in Figure 7.

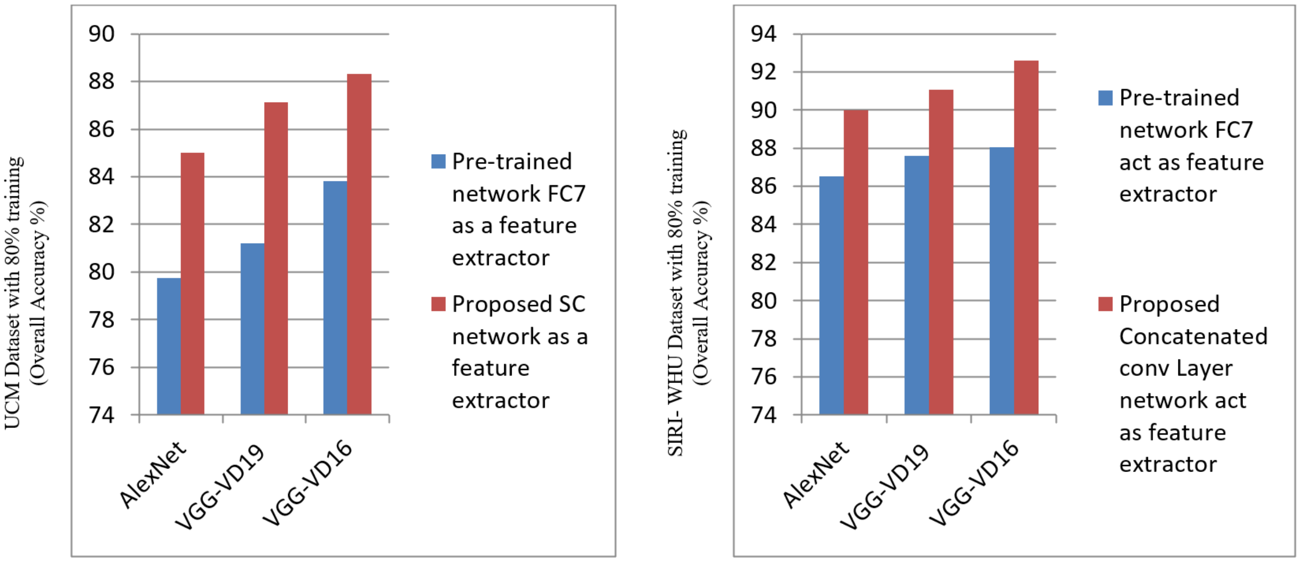
The related comparison results for the two data sets are shown in Table 1. The proposed scenario shows a clear improvement in OA when several Conv layers are combined.

As illustrated in Figure 8, F1 scores of improved classes of UCM data set. By using the proposed strategy, a considerable number of classes show noticeable improvement such as agricultural, beach, harbour, runway reached 100 % and dense residential-class improves approximately 40 %.

Likewise, Figure 9, shows the corresponding F1 scores of pre-trained single-layered and proposed networks on the SIRI-WHU data set. As can be witnessed the proposed strategy, exhibits obvious improvements in most of the classes For example, on the SIRI-WHU data set water reached 100% and harbour, idle_land, industrial, overpass, park classes reach above 90%.

4. CONCLUSION

In this research, we portray Stacked covariance, a new technique for fusing image features from multiple layers of a CNN for scene categorization using remote sensing data. Feature extraction is performed initially with a pre-trained CNN model, followed by feature fusion with Covariance in the presented SC-based classification framework. More dense features are recovered for classification because the proposed scenario takes into account second-order data. Each feature represents the covariance of two distinct feature maps and these features are applied to SVM for classification. Our extensive suggested SC method's effectiveness is validated by comparison with state-of-the-art methodologies using two publicly accessible remote sensing image scene categorization data sets. We recognize that utilizing the proposed SC technique, the accuracy attained for most classes shows obvious enhancements, indicating that this is a viable improvement strategy.



a) UCM Dataset

b) SIRI-WHU Dataset

Figure 7. Comparison of pre-trained networks and proposed SC framework-based networks of UC Merced (UCM), SIRI-WHU Datasets.

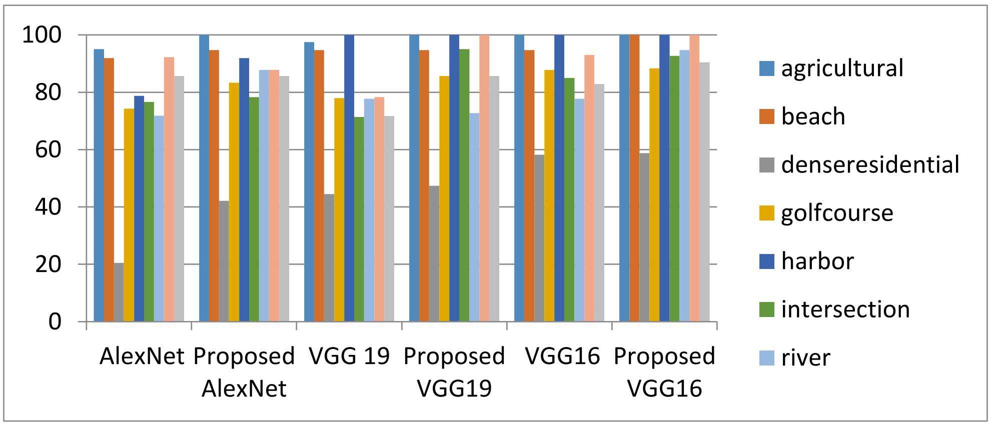


Figure 8. Comparison between F1 scores of UC Merced Data Set improved classes with pre-trained, proposed framework networks.

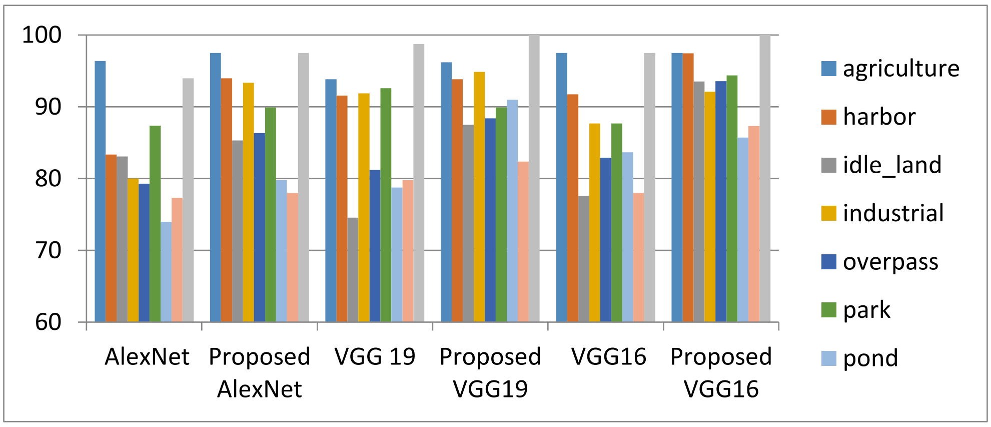


Figure 9. Comparison between F1 scores SIRI-WHU Data Set improved classes with pre-trained, proposed framework networks.

REFERENCES

[1] L. Fang, N. He, S. Li, P. Ghamisi, J. A. Benediktsson, Extinction profiles fusion for hyperspectral images classification IEEE Trans. Geosci. Remote Sens., vol. 56, no. 3, 2018, pp. 1803–1815. DOI: 10.1109/TGRS.2017.2768479

[2] X. Bian, C. Chen, L. Tian, Q. Du, Fusing local and global features for high-resolution scene classification, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 6, Jun. 2017, pp. 2889–2901. DOI: 10.1109/JSTARS.2017.2683799

[3] X. Lu, B. Wang, X. Zheng, X. Li, Exploring models and data for remote sensing image caption generation, IEEE Trans. Geosci. Remote Sens., vol. 56, no. 4, Apr. 2018, pp. 2183–2195. DOI: 10.1109/TGRS.2017.2776321

[4] B. M. Reddy, M. Zia Ur Rahman, Analysis of SAR images using new image classification methods, International Journal of Innovative Technology and Exploring Engineering, vol.8, no.8, 2019, pp. 760-764.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Miami, FL, USA, 20-25 June 2009, pp. 248–255.
DOI: 10.1109/CVPR.2009.5206848

[6] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in Proc. IEEE Int. Conf. Comput. Vis., Columbus, OH, USA, 23-28 June 2014, pp. 580–587.
DOI: 10.1109/CVPR.2014.81

[7] Y. Wang, C. Wang, L. Luo, Z. Zhou, Image classification based on transfer learning of convolutional neural network, Chinese Control Conference (CCC), Guangzhou, China, 27-30 July 2019, pp. 7506-7510.
DOI: 10.23919/ChiCC.2019.8865179

[8] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3-6 November 2015, pp. 1-13.
DOI: 10.1109/ACPR.2015.7486599

[9] S. Tammina, Transfer learning using VGG-16 with deep convolutional neural network for classifying images, International Journal of Scientific and Research Publications (IJSRP), vol. 9, 2019, no. 10, pp. 143-150.
DOI: 10.29322/IJSRP.9.10.2019.p9420

[10] S. Putluri, M. Z. Ur Rahman, S. Y. Fathima, Cloud-based adaptive exon prediction for DNA analysis, Healthcare Technology Letters, vol. 5, no. 1, 2018, pp. 25-30.
DOI: 10.1049/htl.2017.0032

[11] Y. Bi, B. Xue, M. Zhang, Genetic programming with image-related operators and a flexible program structure for feature learning in image classification, IEEE Transactions on Evolutionary Computation., vol. 25, no. 1, 2020, pp. 87–101.
DOI: 10.1109/TEVC.2020.3002229

[12] H. Sun, S. Li, X. Zheng, X. Lu, Remote sensing scene classification by gated bidirectional network, IEEE Trans. Geosci. Remote Sens., vol. 58, no. 1, pp. 82–96, 2019.
DOI: 10.1109/TGRS.2019.2931801

[13] G.Fiori, F.Fuiano, A.Scorza, J.Galo, S.Conforto, S.A Sciuto, A preliminary study on an image analysis based method for lowest detectable signal measurements in Pulsed Wave Doppler ultrasounds, Acta IMEKO, vol. 10, no.2, 2021, pp. 126-132.
DOI: 10.21014/acta_imeko.v10i2.1051

[14] L. Fang, N. He, S. Li, A. J. Plaza, J. Plaza, A new spatial– spectral feature extraction method for hyperspectral images using local covariance matrix representation, IEEE Trans. Geosci. Remote Sens., vol. 56, no. 6, Jun. 2018, pp. 3534–3546.
DOI: 10.1109/TGRS.2018.2801387

[15] V. Arsigny, P. Fillard, X. Pennec, N. Ayache, Geometric means in a novel vector space structure on symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl., vol. 29, no. , 20071, pp. 328–347.
DOI: 10.1137/050637996

[16] UC Merced Data Set. Online [Accessed December 2019] http://weegee.vision.ucmerced.edu/datasets/landuse.html

[17] I. M. E. Zaragoza, G. Caroti, A. Piemonte, The use of image and laser scanner survey archives for cultural heritage 3D modelling and change analysis, Acta IMEKO, vol. 10, no.1, 2021, pp. 114-121.
DOI: 10.21014/acta_imeko.v10i1.847

[18] SIRI-WHU Data Set. Online [Accessed August 2020] http://www.lmars.whu.edu.cn/prof_web/zhongyanfei/Num/Google.html

[19] Y. Liu, Y. Zhong, F. Fei, Q. Zhu, Q. Qin, Scene classification based on a deep random-scale stretched convolutional neural network, Remote Sensing, vol. 10, no. 3, Apr. 2018.
DOI: 10.3390/rs10030444