ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 11 Issue 4 April 2024

Decade Exegesis on Deep Learning Methods Adopted for Remote Sensing Image Classification and Retrieval

[1] Nisha Gupta, [2] Ajay Mittal, [3] Satvir Singh

[1] MRSPTU, Bathinda
[2] Aryabhatta Group of Institutes, Barnala
[3] IKGPTU, Jalandhar
Corresponding Author Email: [1] nishabs2019gmail.com, [2] mittalajay@gmail.com, [3] drsatvir.in@gmail.com

Abstract— ***There are main two approaches to extract features of images to implement content based retrieval system. First is the conventional machine learning methods another is deep learn- ing convolution neural network architectures. In this letter, the literature review about various feature extraction methods used for feature ex- traction of query image and the images stored in databases has been explored. Efforts are made for detailed survey of deep learning approaches for the purpose of extraction of most important salient features of images directly affecting the retrieval performance for classifying the remote sensing images and finally retrieving most rele- vant top images.***

Keywords: Convolution Neural Networks, Deep Learn- ing, Machine Learning.

I. INTRODUCTION

RSIR (Remote Sensing Image Retrieval) is retrieving similar images from remotely sensed image library/archives by visual features matching between query image and each sample of dataset images as presented in figure1. Remote sensing means monitoring the physical charac- teristics of an area by using camera at a distance. Images are captured from satellite or aircraft. Satellite or air- crafts cameras collect remotely sensed images. Databases related to remote sensing images are rapidly growing archives. Remote sensing image are very complex in na- ture having heterogeneous content representations. Re- mote sensing databases are the large scale libraries of abundant images having high dimensional features.

Feature extraction is the transformation of human perception into a numerical representation in form of a fea- ture vector manipulated by machines. This is the prime most and most crucial stage in choosing representative features of a remote sensing image retrieval system design. Major research issue with image retrieval systems is the semantic gap that is the difference between the high level concepts present in images and the low-level attributes used to characterize the image [1]. The main objective of the research is to draft techniques for transforming ab- stract ideas into features. The prominent image features can effectively characterize the contents of an image. It is most important to extract the candidate features crucial for describing the query image and each dataset image sample [2]. The retrieval performance in terms of storage needs, processing time, retrieval time, overall computa- tional time and high similarity index is directly affected by the extracted features. The dimensionality reduction procedure can effectively represent just significant image features as a comparative lower level feature vector. Re- searchers use a variety of feature descriptors to character- ize an image's visual content as a low dimensional feature vector.

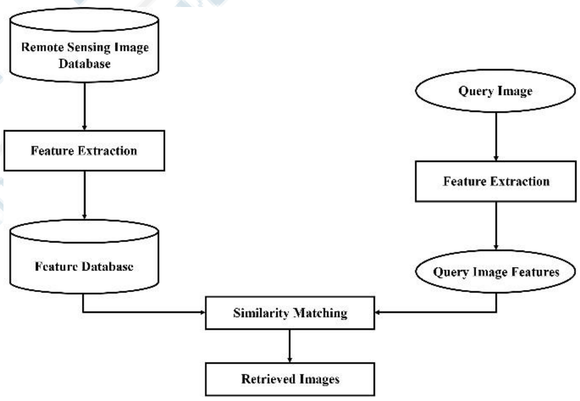

Fig. 1. General Framework for Remote Sensing Image Retrieval System

II. FEATURE EXTRACTION APPROACHES

A. Feature Extraction based on Global Feature Descriptors

The image features are divided into two categories that is global features and local features. Global features represent the whole image. Color, shape, texture and spatial information represent the whole image. Specific areas of a picture, such as borders, blobs and corners represent the local features [3].

B. Feature Extraction based on Local Feature Descriptors

The color feature is the basic feature used by re- searchers for image classification and retrieval. It per- forms well despite of image size and orientation. Tex- ture feature represents patterns of the image which is not based upon single intensity like color. Wavelet transform,Gabor filter,Markov random field GLCM(Gray Level Co-occurrence Matrix[4] and EHD (Edge Histogram Descriptor)[5] are the popularly used algorithms used for extracting texture features of the image by the re- searchers. Still computational complexity is the main con- cerning issue for texture features [6]. Shape is extracted on the basis of region or boundary of the image [7]. Shape ex- traction is done either within entire region or only sig- nificant parts of image. Fourier descriptor [8] and moment invariants [9] are popularly used shape extraction meth- ods to extract the shape features of an image. Shape de- scriptors are variant to translation and scale. Thus it is better to merge shape feature descriptor with other de- scriptors to excel accuracy. Invariant moments, consecu- tive boundary segments, aspect ratio, polygonal approx- imation, fourier descriptors, b-splines etc. are the popu- lar methods used to calculate shape descriptors. Deep Learning proved more efficient and powerful than tra- ditional methods. Deep learning architectures have mul- tiple processing layers that auto learns good important feature representation of images of massive images and obtains higher image classification accuracy than tradi- tional systems. Deep neural networks learn complex rep- resentations of data by discovering hierarchical patterns and features in the data. Deep Neural Network Architec- tures are demonstrated in figure2

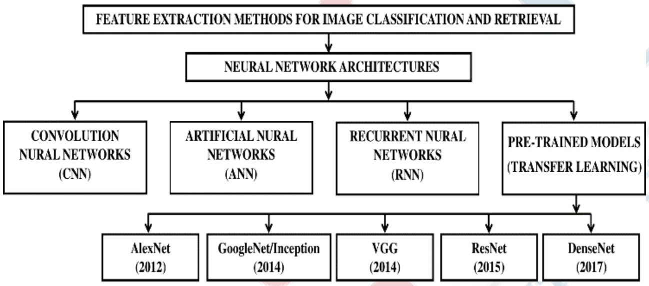

Fig. 2. Deep Learning based Feature extraction methods for Image Classification and Retrieval for Remote Sensing Images

C. Feature Extraction based on CNN (Convolution Neural Networks)

Convolution Neural Networks learn image features through its layered architecture. CNN's have the multiple layers. CNN's has fully connected, pooling, and convolu- tional layers. Filters are applied to input images using a convolutional layer to learn features. The first convolu- tion layers learn features like texture and edges. Complex features are learned by later layers. Pooling layer is re- sponsible for down sampling the incoming inputs. finally fully connected layer makes predictions about the input image's class or label. Last layers learn features like ob- jects. Last layers learn to connect higher features to indi- vidual classes. CNN's are invariant towards translation, scaling, and rotation [10].

D. Feature Extraction based on ANN (Artificial Neural Networks)

Artificial neural networks are linked through neurons and the links. ANN Three layers make up an ANN: input, hid- den, and output layers. There are n neurons in the input layer. Each neuron represents one independent variable in the network. There are as many neurons in the out- put layer as there are classes. There is a weight associ- ated with each neuronal connection. During each itera- tion of training process this weight is modified in arti- ficial neural network. An ANN must complete training and testing phases in order to function as a classifier [11]. **RNN (Recurrent Neural Networks)** are also the fam- ily of neural networks and work with sequential data in a feed forward fashion. RNN's use same weights for each element of the sequence and decreases the number of pa- rameters.

E. Feature Extraction based on Deep Learning Pre-Trained Models employ- ing Transfer Learning

A neural network trained on large dataset gains knowl- edge from this data and this acquired knowledge termed as weights of the network [12] Only the learned features in the form of weights can be extracted and then trans- ferred to any other neural network instead of training that neural network from the initial stage. Instead of building a model from scratch pre-trained models are trained on large dataset are used as a feature extractor by remov- ing the output layer and using the entire network as a fixed feature extractor by freezing the weights of initial layers while retraining only higher layers for new problem specific dataset. Correct weights for the network are iden- tified for the network by multiple forward and backward iterations. The weights and architecture acquired by pre-trained models previously trained on huge datasets may be used directly and apply the learned weights on our tar- get problem known as transfer learning. ImageNet dataset is a rich source of millions of labelled images across thousands of classes that enable the ImageNet dataset a valuable source of training deep learning pre-trained models[13].Knowledge is acquired by pre-trained model on ImageNet dataset that helps to acquire a rich set of learned features and weights helps in adapting the model to specific target task and enhances accuracy[14].Transfer learning can be employed between entirely different but relevant source domain and target domain samples. Pre- trained models are trained on source domain and then learning can produce much higher accuracy results on the target task [15]. Commonly used convolution pre-trained models used for feature extraction based upon trans- fer learning include AlexNet [16], GoogleNet [17] and VGG (Visual Geometry Group variants include VGG16 and VGG19) [18] and DenseNet [19]. These pre-trained models differ with respect to its layered structure and convolution approaches.

Decade exegesis on deep learning methods adopted for

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 11 Issue 4 April 2024

Remote Sensing Image Classification and Retrieval is presented in as below:

Thirumaladevi, S and Swamy, K Veera and Sailaja, M (2013) [20] Image classification is performed by using the transfer learning using pre-trained AlexNet and variants of Visual Geometry Group (VGG) networks VGG-16 and VGG-19. Features are retrieved from the pre-trained network. Fully connected layers are employed with Support Vector Machine classifier. Experiments are done using UCM and SIRI-WHU dataset. Author has achieved the accuracy of 93.57% on UCM and 91.34 on SIRI-WHU using AlexNet. Further accuracy achieved is 94.08% on UCM and 92.78% on SIRI-WHU using VGG-16 and 95% accuracy achieved on UCM and 93.4% accuracy achieved on SIRI-WHU using VGG-19 pre-trained model.

Szegedy et. al (2015) [17] presented Inception-v3 pre-trained model also named as GoogleNet well suited for image classification. The inception module employs a combination of different convolutional filters, including 1x1, 3x3, and 5x5 convolutions to capture spatial information effectively. The model also incorporates max pooling lay- ers to down sample the feature maps and reduces the spatial dimensions. The factorization technique used in Inception-V3 reduces the computational complexities of the network by replacing large convolutions with smaller convolutions. Factorization preserves the spatial information and reduces the computational cost. Auxiliary classifiers used in inception-V3 provide additional gradient flow during training and help in avoiding the vanishing gradient problem. Factorization plays a crucial role in re- ducing the number of parameters thereby enhancing the efficiency of the network without compromising its perfor-

mance. By substituting larger convolutions with smaller convolutions, the model can attain comparable or supe- rior results and proves computationally efficient.

Do, Thanh-Toan and Cheung, Ngai-Man (2017) [21] has proposed a fast embedding method that is the extension of several popular embedding methods such as VLAD (Vector of Locally Added Descriptors), TLCC (Local Co- ordinate Coding using Local Tangent), VLAT (Vector of Locally Aggregated Tensor). Embedded vector can be effi- ciently computed thus more efficiency in similarity match- ing with less retrieval time. FAemb method are evalu- ated with different local features descriptors SIFT (Scale Invariant Feature Transform) and CNN (Convolutional Neural Networks). The experimental results predict that the proposed fast embedding methods proves the better performance over the several standard public image retrieval benchmark datasets. CNN features are used on Holidays and SIFT features are used on Oxford5k dataset. Mean average precision on Holidays dataset is 74% and on Oxford dataset is 45.6%. FAemb can be applicable for small or medium sized datasets. It is not well suited for large scale datasets. This is the major concerning issue.

Li, Jun and Xu, Chang and Yang, Wankou and Sun, Changyin and Tao, Dacheng (2017) [22] has proposed the image re-ranking based method DMINTIR (Discrimina- tive Multi-view Interactive Image Re-ranking). It inte- grates multiple features describing image and user rel- evance feedback. Learned feature vector is obtained to assign ranks to target images in databases. Experiments are conducted on Oxford 5k and Paris 6k datasets. Au- thor has achieved the mean average precision of 85.34% on oxford 5k and 81.75% on Paris 6k dataset.

Huang, Gao and Liu, Zhuang and Van Der Maaten, Laurens and Weinberger, Kilian Q (2017) [19] introduced the DenseNet (Dense Convolutional Network) which links each and every layer to the other layers in a feed- forward fashion. L layers have connections with one be- tween in each layer and its successive layers in the tra- ditional convolutional networks. A network has direct connections(L(L+1))/2. Features-maps of each layer were taken as an input for the successive layers and feature- maps of previous layers are taken as an input for the present layer. DenseNets solves the problem of vanishing- gradient. Experiments are conducted on four benchmark datasets CIFAR-100(Canadian Institute For Advanced Research), SVHN, CIFAR-10 and Imagenet. Author has achieved the accuracy of 97.44% on AID, 99.50% on UC- Merced and 95.89% on Optimal and 94.98% on NWPU- RESIS45 dataset.

Dong, Shan and Zhuang, Yin and Yang, Zhanxin and Pang, Long and Chen, He and Long, Teng (2019) [23] has investigated pixel-level classification for classification of remote sensing images. The author has proposed ResNet-

101 deep learning pre-trained model. 2D dilation con-

volutions with various sample rates are applied to each scale feature layer to improve the multi-scale feature description. Following that, each intra-scale and inter-scale feature layer is successively processed by optimal chan- nel selection. Two datasets ISPRS and GID are used for experimentation. The author has achieved 77.74% pixel accuracy on GID and 86.67% pixel accuracy on ISPRS dataset.

Zhang, Jianming and Lu, Chaoquan and Li, Xudong and Kim, Hye-Jin and Wang, Jin (2019) [24] Convolu- tional Neural Networks (CNN) has two common issues. One is that these models causes over-fitting because they have several parameters and the other one is not deep sufficient to extract abstract information. For solv- ing of these two problems, author proposed DenseNet pre-trained model for remote sensing image classifica- tion. DenseNet uses less number of convolutional ker- nels to generate several reusable features. Dense connec- tions make the network deeper even more than 100 lay- ers. Data augmentation is used. Experiments are done on AID (Aerial Image Dataset) dataset, UCM dataset, NWPU-RESISC45 dataset, and Optimal-31 dataset. The author has achieved the accuracy of 98.67% (50% train- ing ratio), 99.50% (80% training ratio) on UCM dataset and 95.37% (20% training ratio), 97.19% (50% training ratio) on AID dataset and 95.41% (80% training ratio) on optimal-31 and 92.90 (10% training ratio), 94.95 (20% training ratio) on NWPU-RESISC45 dataset.

Liu, Yishu and Ding, Liwang and Chen, Conghui and Liu, Yingbin (2020) [25] has focused to address the issues of training CNNs needs significant labelled samples. The author proposed method of SBS-CNN (Similarity based Supervised Learning using Convolution Neural Network) based on similarity learning applying transfer learning to CNN training that transform similarity learning into deep ordinal classification with CNN experts pre-trained over large scale labelled everyday image sets. It jointly deter- mines image similarities and provides pseudo labels for classification. SBS-CNN has small network size, compact feature vectors and less retrieval time. Gradient descent is calculated at each stage to minimize error. Major re- search issues are that the Gradient descent is calculated at each stage to minimize error. The Negative values are given to ReLU (Rectified Linear Unit) activation function turns the values into zero which decreases the ability of model to fit or train from data properly. Author has used Everyday ImageSet , UC-Merced and PatterNet dataset for experimentation and best ANMR achieved is 0.2185 on benchmark dataset.

Li, Guoqing and Zhang, Meng and Li, Jiaojie and Lv, Feng and Tong, Guodong (2021) [26] proposed two CNNs architectures that are DenseDsc and Dense2net. These two CNNs are compactly connected and com- pact connectivity facilitates features to reuse in the net-

works. Efficient group convolution adopted by Dense2net and more efficient separable convolution using depth adopted by DenseDsc. Both of these techniques improved an efficiency of parameters. The proposed techniques are evaluated on CIFAR and ImageNet datasets. Au- thor has achieved the accuracy 74.2% on CIFAR-100 and 76.3% on ImageNet (top-1) using DenseDsc. 73.68% ac- curacy achieved on CiFAR-100 and 77% ImageNet using Dense2Net.

Sumbul, Gencer and Ravanbakhsh, Mahdyar and Demir, Begüm (2021) [27] proposed Selection of triplets of similar (positive) and dissimilar (negative) images to a reference image called an anchor. In triplet selection each raining is annotated by multiple class labels. Considering this issue author has proposed DAS-RHDIS (Diverse An- chor Selection-Relevant, hard, diverse positive and neg- ative image selection) a triplet sampling method using neural networks for multilabel remote sensing images. Ini- tially, a set of anchors diverse to each other in the em- bedding space is selected from the current mini batch. Secondly, different sets of positive and negative images are chosen for each anchor by evaluating the relevancy, hardness and diversity of the images among each other. Main advantage is the increase in learning speed since in- formative triplets allow fast convergence. Author has used two datasets UCMerced and IRS-BigEartNet for experi- mentation. The author has achieved the 56.8% accuracy, 65.3% precision, 70% recall and 67.5% F1-score on UC- Merced dataset. The author has achieved the 62.7% ac- curacy, 77.7% precision, 75.7% recall and 76.7% F1-score on IRS-BIgEarthNet dataset.

Sumbul, Gencer and Ravanbakhsh, Mahdyar and Demir, Begüm (2022) [28] The author has proposed a PLASTA-MTL (Plasticity-Stability Preserving and Met- ric Learning) Plasticity condition is associated with sen- sitivity to new information and stability condition is associated with protection from radical disruptions by new information of the learning process. This is achieved by defining two novel loss functions. PPL (Plasticity Preserving Loss Function) transforms the global image representation space to new information learned with each task. Another loss function SPL (Stability Preserv- ing Loss) is used to protect the global representation space. Experiments are done on two dataset: DLRSD and BigEarthNet-S2. Accuracy achieved on DLRSD (Dense Labelling Remote Sensing Dataset) dataset is 97.5% and on BigEarthNet-S2 is 97.7%.

Zhang, Zhiqi and Lu, Wen and Feng, Xiaoxiao and Cao, Jinshan and Xie, Guangqi (2022) [29] has proposed a fea- ture learning approach with unique distance metrics for remote sensing image classification and retrieval. For mea- suring interclass variance over intra class variance author has used loss function, balanced deep linear discriminant analysis. Author has used experiments on RED (Recipro-

cal Exponential Distance) for maintaining distance con- trast in high dimensionality. Proposed BDLDA can op- timize the Rayleigh Ritz quotient which measures inter- class variance over intra class variance. Author has used two convolution neural network MKANet Class Multi branch kernel and MobileNetV3. Experiments are done on two datasets RSSCN-7 and OPTIMAL-31. Author has achieved the precision of 0.8963 on RSSCN7 and 0.7447 on OPTIMAL-31 dataset using MKANet class classifica- tion network. Author has achieved the precision of 0.8573 on RSSCN7 dataset using MobileNetV3 classification net- work.

Tan, Pooi Shiang and Lim, Kian Ming and Tan, Cheah Heng and Lee, Chin Poo (2023) [30] proposed DenseNet-121 deep learning pre-trained using transfer learning tech- nique to extract significant features of the image. Com- putation resources are greatly reduced using DenseNet. Three benchmark datasets: soundscapes1, soundscapes2 and urbansound8k are used for experimentation. The pro- posed pre-trained model DenseNet-121 with multilayer perceptron outperforms existing works on soundscapes1, soundscapes2 and urbansound8k datasets with the F1- scores of 80.7%, 87.3% and 69.6% respectively. Deep Learning based feature extraction methods used for im- age classification and retrieval for Remote Sensing Images in past decade are summarized in table1

III. CONCLUSION

Recent image classification and retrieval research method- ologies highly concentrate upon the deep learning tech- niques as it has remarkable improvement in accuracy scores

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 11 Issue 4 April 2024

over traditional machine learning methods.

Table 1. Deep Learning based Feature extraction methods used for Image Classification and Retrieval for Remote Sensing Images in past decade

AUTHOR	TECHNIQUE	OBSERVATIONS		
		DATASET	EVALUATION METRICS	SCORE
S.Thirumaladevi [20] (2013)	AlexNet	UCM	Accuracy	93.57%
		SIRI-WHU	Accuracy	91.34%
	VGG16	UCM	Accuracy	94.08%
		SIRI-WHU	Accuracy	92.78%
	VGG19	UCM	Accuracy	95%
		SIRI-WHU	Accuracy	93.4%
C.Szegedy [17] (2015)	Inception-v3	ImageNet	Top-1 Error	17.2%
			Top-5 Error	3.58%
Do T. T. [21] (2017)	SIFT based Fast Embedding (FAemb)	Oxford 5k	MAP	45.6%
	CNN based Fast Embedding (FAemb)	Holidays	MAP	74%
J.Li [22] (2017)	DMINTIR	Oxford5k	Mean Average Precision	85.34%
	(Discriminative Multi-View Interactive Image Re-ranking)	Paris6k	Mean Average Precision	81.75%
G.Huang [19] (2017)	DenseNet	AID	Accuracy	97.44%
		UCM	Accuracy	99.50%
		Optimal	Accuracy	95.89%
		NWPU	Accuracy	94.98%
Dong [23] (2019)	ResNet-101	GID	Accuracy	77.74%
		ISPRS	Accuracy	86.67%
J.Zhang [24] (2019)	DenseNet	UCM	Accuracy (50% Training ratio)	98.67%
			Accuracy (80% Training ratio)	99.50%
		AID	Accuracy (20% Training ratio)	95.37%
			Accuracy (50% Training ratio)	97.19%
		Optimal-31	Accuracy (80% Training ratio)	95.41%
		NWPU-RESISC45	Accuracy (10% Training ratio)	92.90%
			Accuracy (20% Training ratio)	94.95%
Liu [25] (2020)	Similarity Based Supervised Learning Using Convolution Neural Network (SBS-CNN) applying Transfer Learning	PatterNet	ANMR	0.2185
G. Li, M.Zhang [26]	DenseDsc	CIFAR-10	Accuracy	94.05%

(2021)		CIFAR-100	Accuracy	74.24%
		ImageNet	Accuracy	76.3%
	Dense2Net	CIFAR-10	Accuracy	94.19%
		CIFAR-100	Accuracy	73.68%
		ImageNet	Accuracy (Top-5)	76.3%
			Accuracy (Top-1)	77.1%
G.Sumbul [27] (2021)	DAS-RHDIS (Diverse Anchor Selection-Relevant.and Diverse Positive Negative Image Selection)	UC-Merced	Accuracy	56.8%
			Precision	65.3%
			Recall	70%
			F1-Score	67.5%
		Big Earth Net	Accuracy	62.7%
			Precision	77.7%
			Recall	75.7%
			F1-Score	76.7%
G.Sumbul [28] (2022)	PLASTA-MTL (Plasticity-Stability Preserving Metric Learning),	Dense Labelling Remote Sensing Dataset (DLRSD)	Accuracy	97.5%
		BigEarthNet-S2	Accuracy	97.7%
Z.Zhang [29] (2022)	MKANet Class	RSSCN7	Precision	0.8963 (89.6%)
			mAP	0.9167 (91.6%)
	MobileNetV3	RSSCN7	Precision	0.8573 (85.7%)
			mAP	0.8586 (85.8%)
	MKANet Class	OPTIMAL-31	Precision	0.7447 (74.4%)
			mAP	0.7399 (73.9%)
			ANMR	0.0325 (32.5%)
P.S.Tan [30] (2023)	DenseNet-121	Soundscapes1	F1-score	80.70%
		Soundscapes2	F1-score	87.30%
		Urban-Sound8k	F1-score	69.60%

REFERENCES

[1] C. Bai, J.-n. Chen, L. Huang, K. Kpalma, and S. Chen, “Saliency-based multi-feature modeling for semantic im- age retrieval,” Journal of Visual Communication and Im- age Representation, vol. 50, pp. 199–204, 2018.

[2] D. Zhang, M. M. Islam, and G. Lu, “A review on auto-matic image annotation techniques,” Pattern Recognition, vol. 45, no. 1, pp. 346–362, 2012.

[3] M. Sivakumar, N. Kumar, and N. Karthikeyan, “An ef-ficient deep learning-based content-based image retrieval framework.,” Computer Systems Science & Engineering, vol. 43, no. 2, 2022.

[4] R. M. Haralick, “Statistical and structural approaches to texture,” Proceedings of the IEEE, vol. 67, no. 5, pp. 786– 804, 1979.

[5] C. S. Won, D. K. Park, and S.-J. Park, “Efficient use of mpeg-7 edge histogram descriptor,” ETRI journal, vol. 24, no. 1, pp. 23–30, 2002.

[6] A. Alzu’bi, A. Amira, and N. Ramzan, “Semantic content-based image retrieval: A comprehensive study,” Journal of Visual Communication and Image Representation, vol. 32, pp. 20–54, 2015.

[7] D. Tian, “Support vector machine for content-based im-age retrieval: A comprehensive overview.,” J. Inf. Hiding Multim. Signal Process., vol. 9, no. 6, pp. 1464–1478, 2018.

[8] D. Zhang and G. Lu, “Review of shape representation and description techniques,” Pattern recognition, vol. 37, no. 1, pp. 1–19, 2004.

[9] T. Suk and J. Flusser, “Affine moment invariants gener-ated by graph method,” Pattern Recognition, vol. 44, no. 9, pp.

2047–2056, 2011.

[10] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, et al., “Deep learning for computer vision: A brief review,” Computational intelligence and neuro- science, vol. 2018, 2018.

[11] R. Ashraf, M. Ahmed, S. Jabbar, S. Khalid, A. Ahmad, S. Din, and G. Jeon, “Content based image retrieval by us- ing color descriptor and discrete wavelet transform,” Jour- nal of medical systems, vol. 42, pp. 1–12, 2018.

[12] Y. Wang, R. Xiao, J. Qi, and C. Tao, “Cross-sensor remote-sensing images scene understanding based on transfer learning between heterogeneous networks,” IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1–5, 2021.

[13] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, “Deep learning earth observation classification using imagenet pretrained networks,” IEEE Geoscience and Remote Sens- ing Letters, vol. 13, no. 1, pp. 105–109, 2015.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Ima-genet classification with deep convolutional neural net- works,” Advances in neural information processing sys- tems, vol. 25, 2012.

[15] F. Zeng, S. Hu, and K. Xiao, “Deep hash for latent im-age retrieval,” Multimedia Tools and Applications, vol. 78, pp. 32419–32435, 2019.

[16] 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.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recogni- tion, pp. 1–9, 2015.

[18] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into con- volutional nets,” arXiv preprint arXiv:1405.3531, 2014.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Wein- berger, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.

[20] S. Thirumaladevi, K. V. Swamy, and M. Sailaja, “Remote sensing image scene classification by transfer learning to augment the accuracy,” Measurement: Sensors, vol. 25, p. 100645, 2023.

[21] T.-T. Do and N.-M. Cheung, “Embedding based on func- tion approximation for large scale image search,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 3, pp. 626–638, 2017.

[22] J. Li, C. Xu, W. Yang, C. Sun, and D. Tao, “Discriminative multi-view interactive image re-ranking,” IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3113–3127, 2017.

[23] S. Dong, Y. Zhuang, Z. Yang, L. Pang, H. Chen, and T. Long, “Land cover classification from vhr optical re- mote sensing images by feature ensemble deep learning network,” IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 8, pp. 1396–1400, 2019.

[24] J. Zhang, C. Lu, X. Li, H.-J. Kim, and J. Wang, “A full convolutional network based on densenet for remote sens- ing scene classification,” Mathematical Biosciences and Engineering, vol. 16, no. 5, pp. 3345–3367, 2019.

[25] Y. Liu, L. Ding, C. Chen, and Y. Liu, “Similarity-based unsupervised deep transfer learning for remote sensing image retrieval,” IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 11, pp. 7872–7889, 2020.

[26] G. Li, M. Zhang, J. Li, F. Lv, and G. Tong, “Efficient densely connected convolutional neural networks,” Pat- tern Recognition, vol. 109, p. 107610, 2021.

[27] G. Sumbul, M. Ravanbakhsh, and B. Demir, “Informa- tive and representative triplet selection for multilabel re- mote sensing image retrieval,” IEEE Transactions on Geo- science and Remote Sensing, vol. 60, pp. 1–11, 2021.

[28] G. Sumbul and B. Demir, “Plasticity-stability preserving multi-task learning for remote sensing image retrieval,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–16, 2022.

[29] Z. Zhang, W. Lu, X. Feng, J. Cao, and G. Xie, “A discriminative feature learning approach with distinguish- able distance metrics for remote sensing image classifi- cation and retrieval,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 889–901, 2022.

[30] P. S. Tan, K. M. Lim, C. H. Tan, and C. P. Lee, “Pre-trained densenet-121 with multilayer perceptron for acoustic event classification.,” IAENG International Jour- nal of Computer Science, vol. 50, no. 1, 2023.