IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

A Review Paper Object Recognition using Deep Learning

Lolakshi P K[1], Koushik Achar[2], Manikanta[3], Krupashree R[4], Laya R[5]
Faculty, Department of Information Science and Engineering
Student, Department of Information Science and Engineering
Alvas Institute of Engineering and Technology, Mijar, Moodubidre, India
koushikachar25@gmail.com, lolakshi@aiet.org.in, 4AL21IS023@gmail.com
4AL21IS024@gmail.com, 4AL21IS025@gmail.com

Abstract: *Deep learning has gained popularity amongacademics in the field of object identification due to its ability to overcome the limitations oftraditional techniques that rely on handcrafted features. Deep learning algorithms have advanced significantly in object identification during the past few years. This study presents modern and efficient deep learning frameworksfor object recognition. This paper presents the most current developments in deep neural network-based object identification algorithms.Benchmark datasets for performance evaluation are also explored. The article also highlights how the object recognition technique may be applied to certain item kinds. We finish with the advantages and disadvantages of present approaches and future scope in thisarea.*

Keywords: Deep learning

I. INTRODUCTION

Object recognition is a crucial task in computervision, as it involves the identification and classification of objects in images. The implementation of object recognition on machines is a labyrinthine task, making it necessary to design more potent and less complicated object recognition approaches. Asthe digital database of visual information grows, image analysis approaches are requiredto automatically get its semantic contexts. Good image feature descriptions are the backbone of a good object recognition system.In the past decade, the comprehensive study ofhigh resolution image classification has been carried out with handcrafted features from spatial and spectral domains. The gray-level co-occurrence matrix (GLCM) is used as texture- based descriptors to provide the spectral variation information required for efficient image classification. The extended morphological profiles proposed byBenediktsson et al. have also been used for spatial feature extraction The intra-class variation of the building database makes handcrafted features not an efficient solution.Therefore, handcrafted features are replaced by the feature extracted by sparse coding scheme proposed by Chenyadat. The sparse-constrained support vector machine (SVM) is another feature learning model presented by Tuiaetal. Deep features are more efficient and powerful than low-level features in scene classification, image classification, and face recognition. The mostly used object proposal approaches are based on super-pixel grouping. The typical supervised classification modelsare like a decision tree [23], random forest [24],and support vector machine (SVM]. A randomforest approach is based on the construction ofseveral decision trees during training and the integration of prediction of all the trees is usedfor classification. SVM uses finite training samples to tackle high dimensional data.

For image classification, Chen et al. proposed astacked auto encoder to predict the hierarchal feature of hyperspectral image in the spectral domain. A deep belief network (DBN) represents spectral based features forhyperspectral data classification. Mouetal. introduced recurrent neural network forclassification of hyperspectral images.

The CNN has the capability to automatically discover contextual 2-D spatial features for image classification. There are various supervised CNN-based models used for spectral-spatial classification of hyperspectral remote sensing images. Chen et al. proposed a supervised l2 regularized 3-D CNN-based feature extraction model used for classification purpose. Ghamisi et al. proposed self- improving CNN model. Zhao and Du et al. introduced a spectral, spatial

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

features-based classification framework. Various types of feature extractors have been used in the past based on shape, texture, and venation. Shape-based Elliptic Fourier and discriminant analysis was proposed to discriminate different plant types. Shape-based approaches were based on invariant moments and centroid-radii models. The merger of gray level co-occurrence matrix (GLCM) and LBP was proposed by Tang et al. to extract texture- based features.Age reckoning can be considered as regression or classification issue. Support Vector Regression (SVR), Canonical Correlation Analysis (CCA), and Classificial Nearest Neighbor (NN) and support vector machines (SVMs) are used as classification approaches.The semiautomated image intelligence processing (SAIP) system, a popular template base, has been criticized for its performance degradation in extended operating conditions (EOC). To address this issue, model-based moving and stationary target acquisition and recognition (MSTAR) systems have been developed using trainable classifiers like artificial neural networks (ANN), SVM, and Adaboost. Deep convolutional networks (ConvNet) have shown remarkable performance in object detection and recognition. This paper discusses two-step and single-step architecture object recognition techniques using deep learning, illustrating various object recognition applications,revealing different dataset types, andconducting comparative analysis.

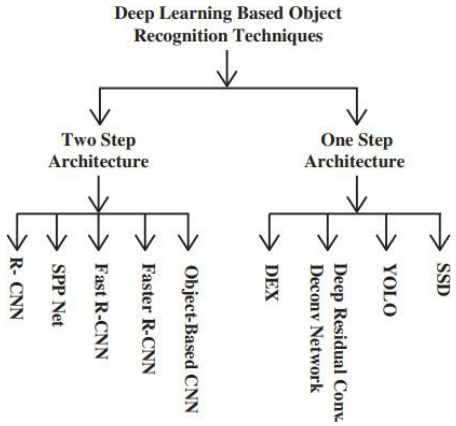

Fig. 1. Overview of deep learning based objectrecognition techniques.

II. TWO-STEP ARCHITECTURE

In 2014, Ross Girshick introduced R-CNN to improve candidate bounding box quality and extract high-level features using deep architecture. R-CNN has two stages: the Generation of Region Proposal stage and DeepFeature Extraction using CNN. R-CNN generates around 2K region suggestions by selective search, which is quick and efficient. Deep feature extraction using CNN retrieved deep features from clipped or distorted regions, resulting in robust 4096-dimensional features.

R-CNN adjusts the region suggestion to fit the specified input size, but this can reducerecognition accuracy. To address this issue, He et al. introduced SPPnet, a novel CNN architecture that reuses the 5th convolutional layer (conv 5) to convert random area suggestions to fixed-size feature vectors. Fast R-CNN pipelining uses hierarchical mini-batches and truncated singular value decomposition (SVD) to include the fc layer.

Faster R-CNN uses region proposal methods to forecast item locations in multiple detection networks. While Fast R-CNN and SPPnet are efficient detection networks, the calculation of region suggestions remains a challenge. The proposed RPN provides full-image convolutional features to the detection network, while RPN predicts object borders and objectness scores. Fast R-CNN serves as the detection network, and the Fast R-CNN and RPN are combined as a unitary network using a convolutional feature and 'attention' technique.

High-resolution photographs are crucial for disaster relief and urban planning due to their accuracy and speed of categorization and interpretation. Deep learning may effectively eliminate semantic gaps in complicated patterns in object-based CNNs, but deeplearning approaches do not record theboundaries between various things. To addressthis issue, combining deep feature learning withobject-based classification strategies is suggested. This strategy enhances the

Copyright to IJARSCT
www.ijarsct.co.in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

accuracyof high-resolution picture categorization.The approach consists of two steps: first, deep features are extracted using CNN, followed byobject categorization using these features.

III. SINGLE-STEP ARCHITECTURE

Szegedy et al. developed a DNN-based technique using basic bounding box interference to abstract objects with the assistance of a binary mask. Pinheiro et al.proposed a CNN model with two branches,while Erhan et al. suggested a regression-based multibox approach for region recommendations. Yoo et al. proposed AttentionNet, a CNN-based object identification tool.The Deep Expectation (DEX) method is introduced for estimating age without facial landmarks, utilizing the IMDB-WIKI databaseof face photos with age and gender labels. TheVGG-16 architecture is used as a CNN to predict an individual's age, consisting of 13convolutional layers with a 3x3 filter and 3 fully linked layers. The CNN is fine-tuned using the new dataset IMDB-WIKI.The Deep Residual Conv-Deconv Network proposes a unique neural network architecture for unsupervised spectral feature learning in hyperspectral images. The proposed network relies on an encoder-decoder mechanism, with the 3-D hyperspectral data encoded using a convolutional subnetwork and decoded using adeconvolutional network to replicate the original data. The convolution subnetwork consists of convolutional blocks, each containing a stack of layers and a 3x3 convolutional filter.The Fast YOLO paradigm predicts confidence across several categories and bounding boxes, dividing the picture into N × N grids and predicting objects centered in each cell. The network can process images at 45 FPS in real time, while the Fast YOLO version can handle155FPS.Liu et al. present a unique single shot multibox detector (SSD) technique to address spatial constraints on bounding box predictions. SSD uses anchor boxes with different aspect ratios and scales to discretize the output space of bounding boxes, as opposed to YOLO's fixed grid. The network combines predictions from many feature maps with varied resolutions to handlediverse object sizes.In summary, various techniques have been developed to improve the accuracy and efficiency of image classification.Convolutional blocks, deconvolutional networks, and the Fast YOLO paradigm are some of the most promising approaches to address these challenges.

IV. APPLICATIONS

Plant Identification

Botanists may quickly and readily identify unknown plant species using plantidentification systems, a computer vision area. Several research have explored the use of leaf data to predict plant species. This approach usesa convolutional neural network to extract valuable characteristics from leaves, followed by a deconvolutional network to evaluate the yield. This approach offers superior information on venation orders [83] compared to shape [33]. Leaf data is represented at manylevels (lower to higher) based on species class.This study contributes to creating a hybrid feature extraction model that improves the performance of plant categorization systems.

Age Estimation

Age is a crucial factor in shaping identity and social interactions. Age estimation depends on elements such as posture, facial wrinkles, vocabulary, and knowledge. Age estimate is utilized in a variety of applications, including intelligent human-machine interfaces, security, transportation, and medical. AI advancements improve the accuracy of age assessment using deep learning techniques.Deep learning algorithms [75] outperformstandard methods for estimating age in terms ofaccuracy and resilience.

Target Classification for SAR Images

The trainable classifier and feature extractor arethe two components of the SAR-ATR(synthetic aperture radar automated targetrecognition) method. The hand-drawn characteristics are often extracted and have an effect on the system's accuracy.Through automatic feature learning from massive amounts of data, deep convolutional networks produced the most advanced results in several computer vision and speech recognition assignments. Convolutional networks have significant overfitting issues when used for SAR-ATR. A-ConvNet, a novel all- convolutional network, is suggested as asolution to this problem. Rather of relying solely on completely linked layers, the A- ConvNet uses layers that are

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

weakly connected. When it came to classifying targets in the SAR image dataset, the A-ConvNet outperformed the standard ConvNet [84].

Face Recognition

Face recognition is the process of identifying a person by their face from a picture or database [85].The facial recognition problem is handled by machine learning techniques like deep neural networks due to the growing amount of the dataset.

With big datasets, the deep learning techniques perform noticeably better. Convolutional neural networks (CNNs) in particular achieve a very high recognition rate when it comes to facial recognition tasks.

V. DATASET

The study focuses on the performance of hyperspectral image classification approaches in three distinct cities: Beijing, Pavia, and Vaihingen. The datasets used include PASCALVOC 2007 and 2012 datasets, as well as the MSCOCO dataset. The Faster R-CNN approach isassessed using the PASCAL VOC 2007 dataset, while the plant identification approachis evaluated using the Malayakew leaf dataset and Flavia leaf datasets. The DEX work uses five distinct datasets for original and apparent age estimation, with the IMDB-WIKI being the largest data for age reckoning. The Indian Pines dataset is gathered over Indian pine sites in Northwestern India using an AVIRIS sensor, while the Pavia University dataset is acquired over the University of Pavia using an ROSIS sensor. The performance of hyperspectral image classification approaches is assessed using overall accuracy, average accuracy, and Kappa coefficient. The experiments are carriedout with A-ConvNet using the MSTAR criterion dataset under standard and extended operating conditions.

Researchers	Techniques	Database used	Result (% accuracy)
R. Girshick et al. [16]	R-CNN	PASCAL VOC, ILSIRC	79.8
X. Bai et al. [22]	SPPnet	PASCAL VOC 2007, ILSIRC	93.42
R. Girshick et al. [65]	Fast R-CNN	PASCAL VOC	89.3
S. Ren et al. [66]	Faster R-CNN	PASCAL VOC 2007, PASCAL VOC 2012	91.8
W. Zhao et al. [69]	Object based CNN	Beijing, Pavia, VaiHingen	99.04
Rasmus Rothe et al. [75]	DEX	IMDB-Wiki	96.6
L. Mou [76]	Deep residual conv-deconv	Pavia, Indian pines	87.39
J. Redmon quad et al. [78]	Yolo	PASCAL VOC 2012, COCO	90.6
W. Liu et al. [80]	SSD	PASCAL VOC 2012, COCO	83.2

Table1: Comparative analysis

VI. COMPARATIVE ANALYSIS

Table I provides an overview of some of the research conducted in the area of deep learning-based object recognition. The table lists the many methods that various researchers have employed. The researchers' published results are quite encouraging, but they were computedfor a specific kind of database. The key question is what will happen if similar techniques are applied to different databases. Itis therefore desirable to compare the many strategies that have been discussed in the literature.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 13, April 2025
Impact Factor: 7.67

VII. CONCLUSION

This paper discusses the development of deep neural network-based object recognition techniques, including Segnet and DEX. The two-step framework familiarizes architectures used for object recognition, while one-step frameworks like YoLO and SSD are reviewed. Benchmark datasets and application areas are discussed, offering promising future scope for object recognition. The paper suggests that future improvements in object recognition include focusing on contextual information to improve performance. Segnet can be designed to calculate uncertainty for prediction from deep segmentation networks, and the training dataset for DEX can be increased. More robust landmark detectors can lead to better face alignment. Future work could explore the capability of Deep Residual Conv-Deconv Network for Hyperspectral Image Classification using APs and estimation profiles to extract spatial information in a robust and adaptive way. Overall, this paper provides valuable guidance for future progress in deep learning-based object recognition

REFERENCES

[1]. Shokoufandeh, A., Keselman, Y., Demirci,M. F., Macrini, D. and Dickinson, S.J., 2012. Many to many feature matching in object recognition: A Review of three approaches.IET Computer Vision, 6(6), pp.500–51

[2]. Martin, L., Tuysuzoglu, A., Karl, W. C. and Ishwa, P., 2015. Learning based object identification and segmentation using dual energy CT images for security. IEEE Transaction on Image Processing, 24(11), pp.4069–4081.

[3]. Cheriyadat, A.M., 2014. Unsupervised feature learning for aerial scene classification. IEEE Transactions on Geoscience and RemoteSensing, 52(1), pp.439–451.

[4]. Ghamisi, Y., Chen and Zhu, X.X., 2016. A self-improving convolution neural network forthe classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 13(10), pp.1537– 1541.

[5]. Zhao, W. and Du, S., 2016. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing, 54(8), pp.4544–4554

[6]. Romero, A., Gatta, C. and Camps-Valls, G.,2016. Unsupervised deep feature extraction forremote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 54(3), pp.1349– 1362.

[7]. Tang, Z., Su, Y., Er, M. J., Qi, Zhang, F. L. and Zhou, J., 2015. A local binary pattern basedtexture descriptors for classification of tea leaves. Neurocomputing, 168, pp.1011–1023.

[8]. Larese, M.G., Namías, R., Craviotto, R.M., Arango, M.R., Gallo, C. and Granitto, P.M., 2014. Automatic classification of legumes using leaf vein image feature.pattern Recognition 47(1), pp.158–168

[9]. ang, X., Guo, R. and Kambhamettu, 2015.Deeply-Learned Feature for Age Estimation.IEEE Winter Conference on Applications of Computer Vision (WACV).

[10]. Rothe, R., Timofte, R. and Van Gool, 2016.Some Like It Hot-Visual Guidance for Preference Prediction. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)