Recognition and Detection of Multiple Objects from Images: A Review

Mini Rani, Er. Ankur Gupta

Abstract— **Due to the vast advancement in the field of computer vision technology, object detection and recognition systems gained significant interest of researchers. Although there are number of object recognition systems implemented in past researches but still there remains a constant demand for new better recognition systems. The task of object recognition is considered to be challenging and difficult task in the field of computer vision. Number of methodologies has been proposed in past that made it possible to recognize the objects form the images but to try new techniques and methods is a vast area of research in this field. This paper summarizes the main researches in the field of object recognition using various approaches and their results to help carry out further possible work.**

Index Terms— **Object Recognition, CNN, object detection, deep neural network..**

I. Introduction

Object recognition identifies the object form the images or videos. It is very easy for the human beings to identify the objects from the images and to differentiate them. However, this process is quite difficult for the machines because they don't have any capability to identify the objects until we train them to do so. It normally uses the various algorithms like machine learning algorithms to make the machines capable of identifying and differentiating the various objects. The object recognition is used in several fields for various purposes. For example, the technology like driverless cars uses the object recognition to identify the various objects present on the road like other cars, persons, traffic signs etc. [1]. Some other applications that uses object recognition are industrial inspection, robotics vision, and disease detection in bio imaging.

There are different problems to the recognition of entities. Usually, artifacts must be observed in different lighting and contrast conditions against cluttered, noisy backgrounds and other objects. Proper representation of features is a crucial step in an object recognition system as it increases efficiency by distinguishing the object in various lightings and scenarios from the context or other objects. The features of object recognition were classified into two classes-sparse and dense representations. Interest-point detectors are used to classify structures such as corners and blobs on the object of incomplete representations of features [1]. The technique of object recognition normally works by first detecting the objects from the image then recognizing it.

***Mini Rani**, M.Tech Scholar, Department of CSE, RIMT University, (e-mail: er.mini.verma678@gmail.com). Mandi Gobindgarh (Punjab), India.*

***Ankur Gupta**, Assistant Professor, Department of CSE, RIMT University, (e-mail: ankurgupta@rimt.ac.in). Mandi Gobindgarh (Punjab), India.*

A. *OBJECT DETECTION*

In many computer vision and pattern recognition applications, such as video identification, vehicle navigation, surveillance and autonomous routing, object detection and tracking play an important role. Object detection requires identifying objects in the frame of a video sequence and understanding patterns. In any tracking method, either in each frame or when the object first appears in the video, an object detection mechanism is required. The most common approach of object detection is to use data in a single frame.

Although some of the methods of object detection use the temporal information derived from analyzing a series of frames to reduce the number of false detections and increase the rate of accuracy. The following are few methods for object detection [17].

Frame differentiation: By calculating the difference between two successive images, the presence of moving objects in a frame is determined. Frame differentiation approach has a good adaptability for a variety of dynamic environments, but it also demonstrates errors in obtaining a complete description of the moving object responsible for the empty phenomenon, resulting in a very low level of accuracy in the detection of the moving object.

Optical Flow: This method involves calculating the optical flow field of the image and performing clustering processing according to the characteristics of the image's optical flow distribution. This method can get an object's complete movement information and is useful for detecting the moving object from the background with 85 percent accuracy, but this method has a few disadvantages including large quantities of calculations, sensitivity to noise, poor anti-noise performance, which makes it unsuitable for object detection and tracking in real time.

Point Detectors: In objects that have an expressive texture in their respective locations, point detectors are used to find some useful points. A useful point of interest is one that invariants lighting and camera perspective shifts. Many widely used detectors of points of interest include detector Moravecs, detector Harris, detector KLT, detector SIFT.

Background Subtraction: For each incoming frame in the video frames, a reliable form of object detection involves building a scene representation known as the background model and detecting deviations from the model. Any major change from the background model in an image region is noted as a moving object [17].

B. *FEATURE EXTRACTION*

It refers to the extraction of visual features that can provide robust and semantic representation. This features help to identify the various objects from any image. Various methodologies have been proposed in past to extract the features of the image like SIFT, Haar, GSCM etc. [9].

C. *APPLICATIONS OF OBJECT RECOGNITION*

a) Surveillance
b) Biometric recognition
c) Content-based image retrieval (CBIR)
d) Medical analysis
e) Industrial inspection
f) Recognition of optical character or document
g) Robotic
h) Contact with human computers
i) Smart vehicle systems

D. *DIFFICULTIES FACED IN OBJECT RECOGNITION*

Few of the difficulties faced in the process of object recognition are explained below:

a) **Positioning:** It is possible to change the position in the object image. When matching objects are used, these images must be treated equally by the program.
b) **Lighting:** In the course of the day, the lighting conditions can vary. The weather conditions in a picture can also influence the lighting. Indoor and outdoor photos may have different lighting conditions for the same item. These can affect the picture due to which the recognition can be incorrect.
c) **Rotation:** It is possible to rotate the image. The system needs to be able to cope with this difficulty.
d) **Obstruction:** The condition is referred to as occlusion when an object is not fully visible in an image. The object recognition system has to handle this type of problem.
e) **Mirroring:** The object recognition system must recognize the mirrored image of any object.

II. LITERATURE REVIEW

Yan, L. et al. (2017) [10] proposed a deep learning model for iterative object recognition. Based on user needs, the program can classify a number of objects. It will become ever more intelligent during user interaction and identify more classes of objects. By studying few examples, including five pictures of an object type, it has high performance. The recognition level is becoming higher and higher, and after studying the new samples, it can recognize ever more object categories. Through evaluating it on complex datasets, researchers have confirmed its efficient learning capability, high recognition levels, and wide range of applications. It also reported that our incremental learning method's object recognition level would be higher with user's interaction.

Tang, J., & Wen, G. (2016) [15] proposed a flexible object recognition approach through a classifier interface with multiple features. Through numerous characteristics involved with the classifier fusion, it shows the power of dealing with pose variance, occlusion and lighting adjustments. The results of the experiment indicate that the proposed system of object recognition can correctly and robustly identify the object in certain difficult situations.

Bai, Z. et al. (2015) [20] explored the use of ELM-LRF as a general framework for the identification of common objects. There are distinct merits for ELM-LRF compared to traditional methods: 1) non-specific task for not using task-specific information; 2) simple to use that it does not require pre-processing, such as design of suitable features, shape model construction or anything else; 3) highly efficient as only a small portion of the connection weights have to be calculated. In addition, unlike the newly emerging CNN, where link weights are iteratively tuned, most ELM-LRF weights are generated simply randomly and only the β output weights are deterministically determined. It significantly reduces compared to CNN: 1) Computational complexity; 2) an enormous set of training requirements. ELM-LRF's general model provides superior precision with exceptionally high speed in the experiments.

Wu, H. et al. (2017) [18] proposed one model for object recognition based on CNN refinement. Our method could achieve realistic recognition results in combination with optimized segmentation and CNN refinement model. More importantly, adequate experiments based on the database that conclude a large number of complicated pictures show the efficacy and robustness of our method. However, our method still has some drawbacks. On the one hand, for some complicated images, it is difficult to separate the target object from the background if there are obvious differences between foreground and background that could have a negative impact on the results of the recognition. More parameters, on the other hand, need to be adjusted for CNN refinement. We need to do further research in the future work to deal with the disadvantages.

Guo, Q. et al. (2016) [21] proposed the CNN-HMM hybrid to recognize the number of the street view building. Our method uses CNN, which has recently performed promisingly well in many problems with computer vision, and HMM, which is a classic sequence model. They are incorporated in a hybrid manner in this work. Using this model, training and recognition can be performed directly at word level, no labeling labor or sophisticated segmentation algorithms are required. On the SVHN dataset, we achieve promising results. Experiments show that the hybrid model will significantly increase the recognition quality by replacing GMM with CNN with HMM. And, by adopting a language model, this model can easily be extended to general scene text recognition. Yet, for example, HMM has a few drawbacks, missing background awareness, incorrect assumption of impedance, and low capacity for discrimination. Future work should be done by discriminatively learning HMM under the criterion of MMI or MCE or using Conditional Random Field (CRF) instead of HMM. On the other hand, when training CNN to integrate contextual information, neighboring frames should be considered. We used CNN's technology is fairly off-the-shelf. It is necessary to investigate better architecture.

Xu, N. et al. (2019) [16] recommended an efficient object detection strategy when using a deep reinforcement learning solution to help object recognition system dynamically change brightness. Experiments show the adaptive brightness change needs and the efficacy of the proposed effective object recognition method. Future research is specifically concerned with in-depth verification and further progress in both the imaging technique and the method of object recognition.

Cadoni, M. et al. (2019) [19] proposed a novel image-based approach for identifying objects, with basic characteristics extracted using iconic descriptive terms and distinguished by a consistency parameter used to optimize the model by prioritizing the most relevant points in more than one view. The template preserves some of the object's 3D knowledge by taking into account the continuity of the salient points, which proves to be more than adequate to identify objects of different nature, from small items to large buildings, with great variations in textures. The tests, carried out on a collected dataset, show the validity of the process in different resolutions on objects acquired through different systems. Recognition rates hit 91.17 percent at rank one, which compares favorably with the 85.75 percent recognition rate achieved through a competitive Bag-of-Words-based process. While the paradigm of recognition has been described as a probe image vs a model (generated from one to many images), the process naturally generalizes model vs model by simply building the model from a set of frames when more than one image is available as a probe, as well as video. These aspects will be the subject of future investigations, together with an improvement in the refinement of the model based on point persistence.

Ciocca, G. et al. (2018) [22] proposed an analysis of CNN-based food classification and retrieval functionality. To learn the features of the food domain, we use a residual network with 50 layers as a reference architecture. We argue that we need food domain representative food repository to have robust features for food related tasks. To check this, we find various food databases publicly available and categorize them according to the representativeness of their food-domain which we expressed through the maximum number of images, number of domain categories and number of category examples. We are also introducing a new food database, Food-475, which is a refinement of our Food-524 food database previously proposed. The features learned from the new database outperform those learned from other food datasets and the very large image database of ImageNet. Specifically, studies on the Food-475 and UNICT-FD1200 database test sets show that these apps show significant improvements in food identification and recognition tasks accuracy. It shows that the more representative the food domain database is and the more accurate it is to identify and retrieve features obtained from a CNN educated on that database.

Hao, W. et al. (2018) [23] presented an advanced image recognition model based on CNN in conjunction with ideas for bottom-up regions. Target area choice and weight enhancement were our paper's key contributions. Based on the above enhancements, appropriate tests using a greater number of images are also used to illustrate the reliability and robustness of our system. Further changes in the future, however, are still required. For example, if the target region is not selected accurately, due to ineffective selection, the error could be enhanced. In addition, the weight of the improvement is based on subjective evaluation, but there would be a potential for subjective evaluation.

Zhao, B. et al. (2018) [24] Analyzed weather recognition drawbacks as a single label classification method, and proposed a multi-label classification system for the weather recognition project. This allows one picture to belong to several categories of weather that can provide a much more detailed explanation of the climatic conditions. In particular, it is a CNN-RNN framework, where CNN is developed with a network-wise model to extract most associated visual feature, and a convolutionary LSTM is used to predict weather labels while preserving the visual feature's spatial information. In addition, we build two datasets to solve the problem of lacking information for the 405-weather recognition task. The experimental results have, in fact, confirmed the feasibility of the suggested solution. In the future work, we plan to introduce the function of weather recognition distribution prediction, which not only classifies the picture with multi-labels, but also predicts the abilities of various weather classes.

Year	Author(s)	Method Proposed	Results (Outcomes)
2017	Wu, H. et al. [18]	CNN refinement method based recognition	method could contribute to retrieve pictures efficiently
2019	Cadoni, M. et al. [19]	Object model is constructed and developed because of significant point resolution.	Recognition rates hit 91.17 percent at rank one, which compares favorably with the 85.75 percent recognition rate achieved through a competitive Bag-of-Words-based process.
2015	Bai, Z. et al. [20]	ELM-LRF was proposed as a general framework for object identification	Reduced complexity and required amount of training data as compared to CNN
2016	Guo, Q. et al. [21]	Presented an integrated model by combining Convolutional Neural Network (CNN) and Hidden Markov Model (HMM)	Hybrid model will significantly increase the recognition quality
2018	Ciocca, G. et al. [22]	use of CNN-based features for food retrieval and recognition	features learned from the new database outperform
2018	Hao, W. et al. [23]	Optimized CNN	Results show our method's superiority, especially compared to some other traditional methods.
2018	Zhao, B. et al. [24]	proposed a multi-label classification system for the weather recognition	Improved results of weather recognition
2016	Tang, J., & Wen, G. [15]	proposed a flexible object recognition approach through a classifier interface with multiple features	The results were found relevant as per the technique applied.
2017	Yan, L. et al. [10]	deep learning model for classification	Good learning capability of model along with higher learning rates.

2019	Xu, N. et al. [16]	Active method for object recognition by adjusting adaptive brightness.	Research observed that the performance of YOLOv3 is much better than other methods of detectors.

III. CONCLUSION

To conclude, it can be said that object recognition has vast scope in the field of research. Number of algorithms was developed and few are still in process by researchers that show the importance of recognition. Number of applications uses recognition method to complete their functioning. It includes biometric devices, face recognition, self-driving cars, drones and so on. This paper summarized some of the main researches based on CNN to recognize the object. The future work can include optimizing the CNN, to introduce new approach, or to combine previous approaches to show more significant results in this field. Due to number of applications of Object Recognition in various fields, it has vast scope of research for advancements. Number of algorithms and methodologies has been proposed in past researches that shows significant improvement but as research is a never ending process, there is always a scope for more improvements.

REFERENCES

[1] R. Bhuvaneswari and R. Subban, "Novel object detection and recogni- tion system based on points of interest selection and SVM classification," *Cognitive Systems Research*, vol. 52, pp. 985–994, 2018.

[2] M. Srinivas, Y. Y. Lin, and H. Y. Liao, "Learning deep and sparse feature representation for fine-grained object recognition," *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1458– 1463, 2017.

[3] A. Cheddad, H. Kusetogullari, and H. Grahn, "Object recognition using shape growth pattern," *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis*, pp. 47–52, 2017.

[4] A. Gleibman, 2013, arXiv preprint.

[5] W. Fang, Y. Ding, F. Zhang, and V. S. Sheng, "DOG: A new background removal for object recognition from images," *Neurocomputing*, vol. 361, pp. 85–91, 2019.

[6] S. Kumar, A. Balyan, and M. Chawla, "Object Detection and Recogni- tion in," *Images. IJDER*, vol. 5, no. 4, pp. 1029–1034, 2017.

[7] X. Wang, "Deep learning in object recognition, detection, and segmen- tation," *Foundations and Trends in Signal Processing*, vol. 8, no. 4, pp. 217–382, 2016.

[8] S. Brahimi, N. B. Aoun, and C. B. Amar, "Boosted Convolutional Neural Network for object recognition at large scale," *NeuroComputing*, vol. 330, pp. 337–354, 2019.

[9] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, 2019, learning systems.

[10] L. Yan, Y. Wang, T. Song, and Z. Yin, "An incremental intelligent object recognition system based on deep learning," *Chinese Automation Congress (CAC)*, pp. 7135–7138, 2017.

[11] S. Masuda, Y. Kaeri, Y. Manabe, and S. Kenji, "Scene Recognition Method by Bag of Objects Based on Object Detector," *Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 321–324, 2018.

[12] B. Jiang, X. Li, L. Yin, W. Yue, and S. Wang, "Object Recognition in Remote Sensing Images Using Combined Deep Features," *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 606–610, 2019.

[13] Y. W. Hsu, Y. S. Ciou, and J. W. Perng, "Object recognition system design in regions of interest based on AdaBoost algorithm," *20th International Conference on Information Fusion (Fusion) (pp*, pp. 1– 5, 2017.

[14] M. Nikhila and S. Rawat, "Recognition of salient object," *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*, pp. 283–286, 2016.

[15] J. Tang and G. Wen, "Object recognition via classifier interaction with multiple features," *8th International Conference on Intelligent Human- Machine Systems and Cybernetics (IHMSC)*, vol. 2, pp. 337–340, 2016.

[16] N. Xu, C. Huo, and C. Pan, "Adaptive Brightness Learning for Active Object Recognition," *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2162–2166, 2019.

[17] C. M. Sukanya, R. Gokul, and V. Paul, "A survey on object recognition methods," *International Journal of Science, Engineering and Computer Technology*, vol. 6, no. 1, pp. 48–48, 2016.

[18] H. Wu, R. Bie, J. Guo, X. Meng, and C. Zhang, "CNN refinement based object recognition through optimized segmentation," *Optik*, vol. 150, pp. 76–82, 2017.

[19] M. Cadoni, A. Lagorio, and E. Grosso, "Incremental models based on features persistence for object recognition," *Pattern Recognition Letters*, vol. 122, pp. 38–44, 2019.

[20] Z. Bai, L. L. C. Kasun, and G. B. Huang, "Generic object recognition with local receptive fields based extreme learning machine," *Procedia Computer Science*, vol. 53, pp. 391–399, 2015.

[21] Q. Guo, F. Wang, J. Lei, D. Tu, and G. Li, "Convolutional feature learning and Hybrid CNN-HMM for scene number recognition," *Neu- rocomputing*, vol. 184, pp. 78–90, 2016.

[22] G. Ciocca, P. Napoletano, and R. Schettini, "CNN-based features for retrieval and classification of food images," *Computer Vision and Image Understanding*, vol. 176, pp. 70–77, 2018.

[23] W. Hao, R. Bie, J. Guo, X. Meng, and S. Wang, "Optimized CNN based image recognition through target region selection," *Optik*, vol. 156, pp. 772–777, 2018.

[24] B. Zhao, X. Li, X. Lu, and Z. Wang, "A CNN-RNN architecture for multi-label weather recognition," *Neurocomputing*, vol. 322, pp. 47–57, 2018.