The Use of Machine Learning and Deep Learning for Object Recognition in Computer Vision

Darsh Jadhav
Computer Science
Bangor University
darshjadhav7@gmail.com

***Abstract*—Object recognition techniques aim to solve computer vision problems. Computer vision is the field in developing and implementing techniques to allow computers to understand information from images and videos. Object recognition is a evolving technique which adopts the use of machine learning and deep learning to solve problems. Machine learning is a multi-disciplinary scientific field which combines several other scientific fields in order to predict outcomes from training data. There are many applications of object recognition, ranging from facial recognition to autonomous vehicles. This paper aims to outline the general procedure that an object recognition system would go through, how object recognition requires machine learning and deep learning in order to work, the current applications of object recognition, and what the future entails for object recognition and computer vision.**

***Index Terms*—Machine Learning, Deep Learning, Object Recognition, Computer Vision**

I. Introduction

The scientific field of Computer Vision (CV) aims to develop and implement techniques that allow computers to process and understand information from images and videos. Machine learning is widely used in the world of computer vision as it has the ability to improve the perception that computers have of its surrounding stimuli, subsequently improves the transformation of the signals sensed to internal representations. This allows computer vision systems to bridge the gap between the internal representations perceived of the environment [1]. Object recognition (also referred to as object detection) is a widely applied computer vision technique used to identify which objects are present within digital images and videos [2]. Object recognition is the result of machine learning and deep learning algorithms being able to study images [3]. These algorithms gain a deeper understanding of these images, processing large amounts of information gained by collections of data, thus giving the ability for computer to be able to identify objects. These algorithms tend to classify objects seen in the images by extracting features, these features tend to be distinguishable from other objects. Object recognition is widely used in the field of computer vision, some examples include robotic vision, internet content censorship, and machinery inspection [4].

II. Literature Review

A. Machine Learning

Machine learning is a multi-disciplinary scientific field, combining the ideas from different fields such as statistics, neuroscience and biology, physics, and mathematics, in order for machines to learn from data [5]. The computers learn from their own actions to alter the predictions that they make in order to improve accuracy. There are three main types of machine learning algorithms, supervised, unsupervised, reinforcement. Supervised machine learning is training the algorithm with data which tends to be labelled. This means that the machine is learning and improving accuracy based on making mistakes. Unsupervised machine learning uses unlabelled data, allowing the model understand the data by itself. Unsupervised machine learning tends to be used when the intention is to find patterns in the data. Furthermore, a common use of unsupervised learning is clustering data, finding similarities between instances and grouping them together. Reinforcement machine learning can be explained as combination of supervised and unsupervised learning. When the algorithm is classifying objects, reinforcement learning informs the algorithm if the object was misclassified. The key detail is that the algorithm does not get told how to fix its mistake, this forces the algorithm to learn from its errors in order to improve the accuracy of the model.

B. Deep Learning

Deep learning is a massively evolving area of research within the field of machine learning, they utilise deep neural networks algorithms in order to learn data representations [6]. Deep learning is tends to be used as a solution to many problems within the broad field of computer vision [7]. Object recognition works by finding windows in which the object may be present in, deep learning can be applied to this computer vision problem. The history of deep learning can track back to the 1940s, where general learning problems were intended to be solved by simulating the human brain. Fast forwarding to the 1980s to 1990s where deep learning became more popular due to the introduction of the back-propagation algorithm [8]. Deep learning was declining as learning models were overfitting the data when training. Overfitting is where the learning model follows the data too closely, meaning that when new data is added, it fails to predict outcomes correctly. This

problem was due to the unavailability of large datasets, lack of competitiveness performance compared to other machine learning algorithms, and the limited computational power [8]. Deep learning managed to become more popular since around 2006 due to significant advancements in speech recognition [9]. Deep learning has been able to be tested to a large capacity as large datasets such as ImageNet has surfaced. An ImageNet application is object recognition, with the availability of large training data and deep learning neural networks, object recognition can be achieved [10].

C. Object Recognition

Objection recognition is known to be "one of the fundamental computer vision problems" [8], this is because of its ability to supply information of great value in order to understand images and videos. The typical model of object recognition tend to use three fundamental steps: Informative region selection, feature extraction, and classification. Informative region selection is the process in finding key parts of an image which contain relevant information about the object in question. This avoids the typical exhaustive approach of searching the whole image using multi-scale sliding windows as this can be computationally expensive and produce redundant slide windows [8]. Feature selection is the next step that is used in object recognition. Feature selection (also known as attribute selection) is a commonly used process in machine learning, evaluating the features of data which tend to be irrelevant or redundant. The aim of feature selection is to improve prediction accuracy of the machine learning model whilst finding the minimal amount of features that contribute to the concept [11]. Feature selection in object recognition is used to extract visual features from images, providing a semantic representation of the data. Classification is a large portion of object recognition. Classification is the attempt to classify the target objects with the classes (categories), the aim is to classify the objects to the highest possible accuracy.

Typical object recognition tends to fall into two sizable categories, instance recognition and class recognition [12]. Instance recognition is traditionally tends to be attained by matching two or three dimensional key points within the images [13]. The application of deep learning to instance The use of extracting lines, edges, and contours of the objects in the image tend to be compared with images that have been used in the training of machine learning algorithms. Class recognition (also referred to as category recognition) is somewhat of a difficult challenge. The problem requires the algorithm to correctly identify different objects of inconsistent dimensions and colours whilst being able to consistently avoid misclassifying objects and maintaining a high level of accuracy [14]. There are many applications of object recognition, some include: Optical character recognition, autonomous vehicles, and facial recognition.

III. Applications of Object Recognition

A. Autonomous Vehicles

Autonomous vehicles utilise object recognition, being able to differentiate whether an object is a pedestrian, vehicle, or a road traffic sign. Furthermore, these autonomous vehicles will have to navigate themselves, thus requiring them to be able to judge when to accelerate, decelerate, perform a manoeuvre, etc. Essentially the vehicle would have to be able to see for itself as it would be autonomous [15]. Autonomous vehicles use LiDAR sensors, a technology that uses laser beams around its surroundings, this forms a three dimensional point cloud that maps the surrounding environment. With these three dimensional point clouds, it can then go onto the object recognition stage. This will involve feature extraction and the use of deep learning convolutional neural networks (CNN) in order to classify these objects accurately [16].

B. Facial Detection and Recognition

Facial detection is largely used computer vision task, the ability of a computer to identify a face within an image. Facial detection can be seen as a specific type of object class detection [15]. Examples of facial detection systems are in mobile applications such as Facebook. Facial recognition is the use of biometric technology to recognise a face, specific to a person. This is an extra step over facial detection as it involves looking into further details of the face. Facial recognition is widely used in the smartphone industry, it is a common way to unlock one's device. Facial recognition is also being used in many other applications, some include: Banks, biometric surveillance, and airports. These tend to be in the interest of security as the system would be able to identify each individual present. The method in which facial recognition works is by comparing the target image with images in the database, this is done by iterating through each bit in the image [15].

C. Automated Surveillance

Object recognition is also widely used in automated surveillance, this involves the CCTV cameras tracking and recording based on the objects detected. This is due to the limitations in storage and memory size. As there is not an infinite amount of storage, the surveillance system will have to recognise certain objects, such as pedestrians or vehicles entering the premises. The automate surveillance systems tend to use both object recognition and detection [15].

IV. Discussion

With the improvement of computational power, increase in training data, and efficiency of algorithms, the applications of object recognition will increase. Rodrigo Verschae and Javier Ruiz-del-Solar outlined some problems that have not been solved which seem to be a possible direction for object recognition. They mentioned the problem of Open-World Learning and Active Vision. This problem is implementing a way to detect new classes and subclasses, whilst also being able to distinguish between the subclasses following the original new class that has been learned. As new objects are being

introduced, the existing object recognition systems will have to be updated. This will require continuous maintenance of the system. To bypass this, the proposal is to introduce deep learning models which will do this for us [17].

V. Conclusion

Object recognition is evolving, making use of various different machine learning algorithms and techniques in order to accurately classify objects. The use of object recognition in conjunction with machine learning and deep learning allows us to use this technique for many important applications, such as autonomous vehicles or surveillance systems. Creating the best possible machine learning model takes time, Pedro F. Felzenszwalb et al. used discriminatively trained part based models with object detection. In their experience, they mentioned that "maintaining performance seems to require gradual enrichment of the model". Furthermore, they talked about simpler models being able to outperform rich models in practice due to the rich models suffering from difficulties in training [18]. Machine learning models always have to be tailored to the problem at hand, there is not one model which works for every problem. Machine learning and deep learning plays an important role in object recognition, allows these systems to accurately classify objects within an image.

VI. References

[1] N. Sebe, I. Cohen, A. Garge and T. S. Huang, Machine Learning in Computer Vision, Springer, 2005.

[2] A. Andreopoulos and J. Tsotsos, "50 Years of Object Recognition: Directions Forward," Computer Vision and Image Understanding, vol. 117, p. 829–835, 2013.

[3] "Image Processing and Computer Vision: What is Object Recognition?," MathWorks, [Online]. Available: https://uk.mathworks.com/solutions/image-video-processing/object-recognition.html. [Accessed 15 March 2020].

[4] L. F. Robles, "Object recognition techniques in real applications," University of Groningen, 2016.

[5] S. Marsland, Machine Learning: An Algorithmic Perspective. Second Edition, CRC press, 2015.

[6] D. Gromann, L. E. Anke and T. Declerck, "Special Issue on Semantic Deep Learning," Semantic Web, vol. 10, no. 5, pp. 815-822, 2019.

[7] M. Nielsen, Neural Networks and Deep Learning, Determination press, 2015.

[8] Z.-Q. Zhao, P. Zheng, S.-t. Xu and X. Wu, "Object Detection with Deep Learning: A Review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232, 2019.

[9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," IEEE Signal Processing Magazine, vol. 29, no. 6, p. 82, 2012.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in IEEE conference on computer vision and pattern recognition, 2009.

[11] M. Dash and H. Liu, "Feature Selection for Classfication," Intelligent data analysis, vol. 1, no. 3, pp. 131-135, 1997.

[12] R. Szeliski, Computer vision: algorithms and applications, Springer Science & Business Media, 2010.

[13] D. Held, S. Thrun and S. Savarese, "Robust Single-View Instance Recognition," Stanford University, 2016.

[14] J. Winn and A. Criminisi, "Object Class Recognition at a Glance," Microsoft Research Ltd, 2006.

[15] A. Vahab, M. S. Naik, P. G. Raikar and P. S. R, "Applications of Object Detection System," International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 4, p. 4186, 2019.

[16] M. Aryal, "Object Detection, Classification, and Tracking for Autonomous Vehicle," 2018.

[17] R. Verschae and J. Ruiz-del-Solar, "Object Detection: Current and Future Directions," Frontiers in Robotics and AI , vol. 2, p. 29, 2015.

[18] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 9, pp. 1627-1629, 2009.