International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT2410230

Deep SORT Related Studies

[1]Abdul Majid, [1]Qinbo, [1]Saba Brahmani

[1]Department of Computer Science and Technology, Faculty of Information Science and Engineering, Ocean University of China

Corresponding Author : Abdul Majid, abdul.majid827@yahoo.com

ARTICLEINFO

Article History:

Accepted: 05 April 2024

Published: 18 April 2024

Publication Issue

Volume 10, Issue 2

March-April-2024

Page Number

358-363

ABSTRACT

Computer vision is the field of computer science in which computers are made capable to see and recognize like human being. Deep learning is using multiple layers for the purpose of understanding and recognizing various objects. Deep Simple Real Time Tracker is the area in which the objects are tracked in real time from multiple images and videos. Many researchers have contributed to the field and various algorithms have been proposed. The current study presents the deep SORT related studies in which the various algorithms have been presented for the sake of understanding and starting point for the researchers interested in computer vision and deep sorting. The single shot detection, feature extraction, have been explained along with the research conducted. Feature selection and extraction, matching recognition, object tracking through frames have been appended to the current study.

Keywords : Object Detection, Deep Learning, Object Tracking, Matching And Recognition, Simple Real Time Tracker.

I. INTRODUCTION

Deep sorting (Simple Real Time Tracker) in computer vision refers to the use of deep learning techniques [1], particularly deep neural networks [2], for sorting and tracking objects [1][4] in visual data [5], such as images or videos. The algorithms tend to track moving objects [4][6]. The goal is to assign unique identities to objects and maintain consistency in their tracking across frames or scenes [7]. This is particularly important in applications like object tracking in surveillance videos, autonomous vehicles, and augmented reality systems. There are various applications of deep SORT.

One of the applications is the object detection [8] in which process typically begins with object detection, where a deep learning model is employed to identify and locate objects [9] within an image or video frame [10]. Popular object detection models include Faster R-CNN, YOLO (You Only Look Once) [11][12], and SSD (Single Shot Multibox Detector). The SingleShot Detector is shown in Figure 1. Another application is the Feature Extraction in which objects are recognized

by extracting features from objects based on distinct features. Features are the distinctive characteristics that help differentiate one object from another. Convolutional Neural Networks (CNNs) are commonly used for feature extraction in computer vision tasks. The feature extraction produces the distinctive features which are then converted to high dimensional vectors resulting deep embedding. These embeddings are representations of the objects that capture their unique characteristics. A Siamese network or a triplet network is commonly used to learn and generate these embeddings.

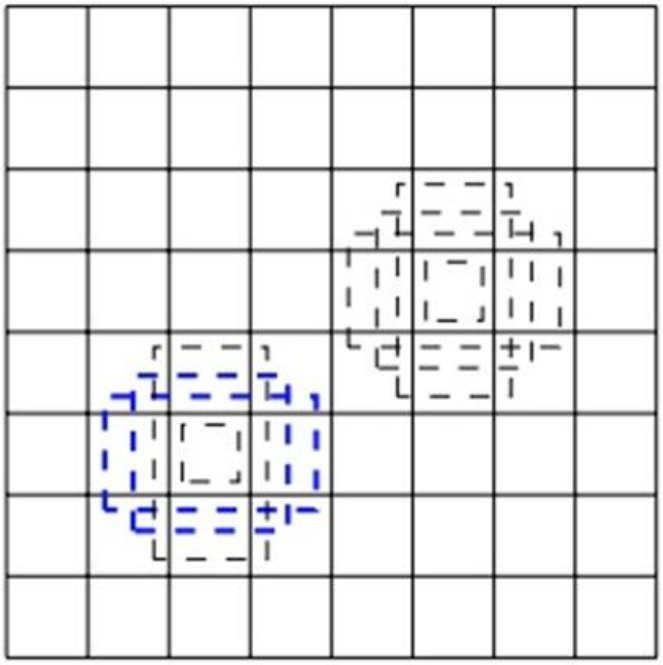


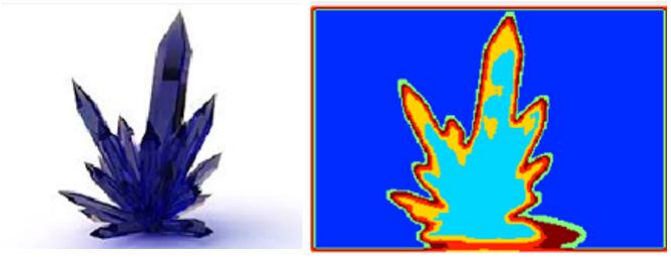
Figure 1 : Single Shot Detector (Redefined)

The next phase is the matching and sorting in which the embeddings are then used to match objects across frames or scenes. Deep SORT involves associating the embeddings of detected objects in the current frame with their counterparts in the previous frame. Various algorithms, such as the Hungarian algorithm, are often used to optimize this matching process and assign consistent identities to objects. The other applicable area is the object identity maintenance in which the videos are used to match the object identities. The video progresses and the object identities are maintained. Deep SORT continues to match and update object identities during the videos. This helps in maintaining a consistent tracking of objects, even when they might be partially occluded or undergo changes in appearance.

II. MATERIAL METHODOLOGY

1. Object Detection

Object detection is a fundamental component of deep sorting in computer vision [1]. In the context of deep sorting, the object detection step is responsible for identifying and locating objects within each frame of a video or series of images [4] [5]. This detection is a crucial prerequisite for subsequent steps, such as feature extraction, matching, and tracking. Here's how object detection fits into the deep sorting pipeline:



(a) object and detection through colour

(b) object and detection through edges

Figure 2. Object Detection

The very first step is the input data for the object detection. The image might be the single image or it may be the sequence of the multiple images in which number of objects may be available. The objects available in images will be tracked [8]. Deep Learning Models for Object Detection: State-of-the-art deep learning models for object detection are often employed in this step. Some popular models include

Faster R-CNN (Region-based Convolutional Neural Network), Utilizes region proposals and a region-based CNN for object detection. Another algorithm YOLO also called You Only Look Once which includes the probabilities-based prediction and classification in which the image is divided into multiple images and the grids are predicted. SSD (Single Shot Multibox Detector): Performs multi-scale feature extraction for detecting objects at different scales [12].

2. Feature Extraction

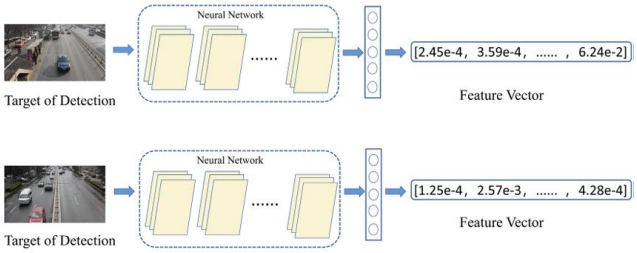


Figure.3 The process of encoding the features of the original input data [29]

Feature extraction in computer vision involves capturing relevant information from raw image data to create a compact and meaningful representation. These extracted features are used for various tasks such as image recognition, object detection, and image classification. The feature extraction has the step of preprocessing in which the resizing, and normalization is performed [14] [16]. Another is the data augmentation which has the variations in the data set and various transformations are applied such as zooming, rotation and others [15]. Various layers can be used for the Convolutional layers, pooling layers. The features are also called the discriminant characteristics which are first selected called feature selection and then these selected features are presented in the form of vector which is called feature extraction [14][15] [16].

3. Deep Embeddings

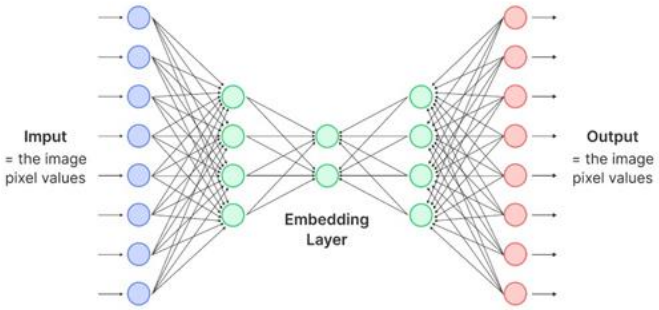


Figure.4 Embedding Layer along with input and output layer

Deep embeddings refer to the high-dimensional vector representations learned by deep neural networks, typically in the context of tasks like object recognition, image retrieval, or feature extraction. In deep learning, the term "embedding" often implies a transformation of input data into a space where meaningful relationships and patterns are more easily discernible. Deep embeddings are commonly employed in computer vision, natural language processing, and various other domains [17] [18]. Many of the applications are available for the embeddings including Image retrieval [19] [20] [21], face recognition [22], image retrieval [20] [21] and much more.

4. Matching and Sorting

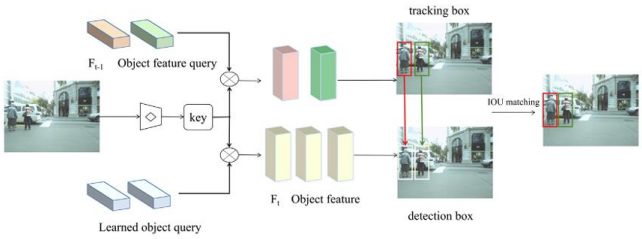


Figure 5: Online MOT algorithm [29] matching object process

Matching and sorting, in the context of computer vision and deep learning, typically refer to the process of associating objects across multiple frames in a video sequence and maintaining consistent identities for those objects. This is particularly important in tasks such as object tracking, where the goal is to follow the

movement of objects and assign unique identifiers to them over time [28].

5. Maintaining Object Identities

Maintaining object identities is a critical aspect of object tracking in computer vision applications. It involves assigning and preserving unique identifiers for objects over time, ensuring consistency as the objects move across frames in a video sequence. This process is crucial for understanding the trajectories and behaviours of objects in dynamic visual environments.

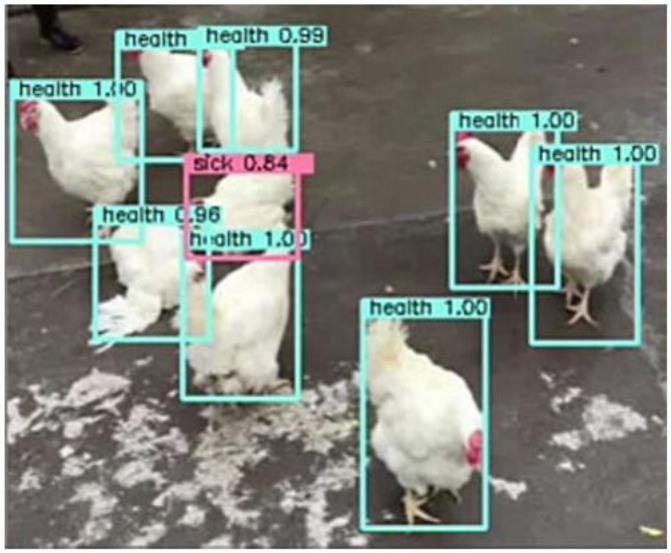


Figure 6 : Healthy and sick indoor broilers identification using Object tracking and maintenance (Adapted from [30])

The process of maintenance identification starts with a unique identifier which is to be assigned to each and every object especially found in the starting form. The unique identifier is regarded as the tag or the label given to the specified object. Then the next step is to match these object identifiers across each frame to check whether every frame is the same, changed or even vanished from the system. The process is based on the global descriptors [23]. With due course of the time, the frames are processed, and the objects are updated as per the properties. The changing properties are the velocity, size, position and many other properties. These properties show that the objects are maintained, and the objects are tracked continuously. In case of the objects are lost and removed or replaced from the system then the reidentification algorithms are applied so that the updating process can be continued, and every object is presented in the sequence [24]. The system is made possible with surety that the assigned unique numbers must remain unique during the process so that any confusion cannot arise during the process of maintenance and the consistency must be achieved. The other two algorithms can also be used during this maintenance including continuous monitoring during tracking and the optimization algorithms including Hungarian algorithm [25][26] and other optimization algorithms [27].

VI. CONCLUSION

Computer vision is the area which needs much attention even in the case the artificial intelligence has given many advancements, but the computer vision is still far from the reality and needs attention. Deep simple real time tracker is an advance field which uses deep technology and algorithms for tracking objects along with maintaining object identities of the objects during the video or the sequence of the images. Various activities and algorithms are proposed by the various researchers to make computer vision able to understand and track various objects in sequence of images or the videos. This study has presented the various activities of deep SORT including object detection, tracking, embedding, identification and maintaining identity across frames. The matching and classification is also the part of the study.

III. REFERENCES

[1]. Pereira, R., Carvalho, G., Garrote, L., & Nunes, U. J. (2022). Sort and deep-SORT based multi-object tracking for mobile robotics: Evaluation with new data association metrics. Applied Sciences, 12(3), 1319.

[2]. Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining

deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278.

[3]. Fort man, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," IEEE J. Ocean. Eng., vol. 8, no. 3, pp.173–184, 1983.

[4]. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," EURASIP J. Image Video Process, vol. 2008,2008.

[5]. Yu, W. Li, Q. Li, Y. Liu, X. Shi, and J. Yan, "Poi Multiple object tracking with high performance detection and appearance feature," in ECCV. Springer, 2016,pp. 36–42.

[6]. Ashqar, B.A., Abu-Naser, S.S., 2018. Image-based tomato leaves diseases detection using deep...

[7]. Akila et al.Detection and classification of plant leaf diseases by using deep learning algorithm.

[8]. Zhang, Y., Chen, Z., & Wei, B. (2020, December). A sport athlete object tracking based on deep sort and yolo V4 in case of camera movement. In 2020 IEEE 6th international conference on computer and communications (ICCC) (pp. 1312-1316). IEEE.

[9]. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in CVPR, 2008, pp. 1–8.

[10]. Milan, K. Schindler, and S. Roth, "Detection- and trajectory-level exclusion in multiple object tracking," in CVPR, 2013, pp. 3682–3689.

[11]. Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. Procedia Computer Science, 199, 1066-1073.

[12]. Huang, R., Pedoeem, J., & Chen, C. (2018, December). YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. In 2018 IEEE international conference on big data (big data) (pp. 2503-2510). IEEE.

[13]. Feng Liu, Guohui Li, Hong Yang, Application of multi-algorithm mixed feature extraction model in underwater acoustic signal, Ocean Engineering, Volume 296, 2024, 116959, ISSN 0029-8018,

[14]. Kunyan Li, Chen Kang, Deep feature extraction with tri-channel textual feature map for text classification, Pattern Recognition Letters, Volume 178, 2024, Pages 49-54, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2023.12.019.

[15]. Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020, July). Feature extraction methods: a review. In Journal of Physics: Conference Series (Vol. 1591, No. 1, p. 012028). IOP Publishing.

[16]. Salau, A. O., & Jain, S. (2019, March). Feature extraction: a survey of the types, techniques, applications. In 2019 international conference on signal processing and communication (ICSC) (pp. 158-164). IEEE.

[17]. Xu, H., Fu, H., Long, Y., Ang, K. S., Sethi, R., Chong, K., ... & Chen, J. (2024). Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Medicine, 16(1), 12.

[18]. Mallik, A., & Kumar, S. (2024). Word2Vec and LSTM based deep learning technique for context-free fake news detection. Multimedia Tools and Applications, 83(1), 919-940.

[19]. Levy, M., Ben-Ari, R., Darshan, N., & Lischinski, D. (2024). Chatting makes perfect: Chat-based image retrieval. Advances in Neural Information Processing Systems, 36.

[20]. Fang, S., Wu, G., Liu, Y., Feng, X., & Kong, Y. (2024). Dual enhanced semantic hashing for fast image retrieval. Multimedia Tools and Applications, 1-20.

[21]. Zhang, N., Liu, Y., Li, Z., Xiang, J., & Pan, R. (2024). Fabric image retrieval based on multi-modal feature fusion. Signal, Image and Video Processing, 1-11.

[22]. O'Neill, C. (2024). Disaster, facial recognition technology, and the problem of the corpse. New Media & Society, 26(3), 1333-1348.

[23]. Hardy, P., & Kim, H. (2024). LInKs" Lifting Independent Keypoints"-Partial Pose Lifting for Occlusion Handling With Improved Accuracy in 2D-3D Human Pose Estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3426-3435).

[24]. Hardy, P., & Kim, H. (2024). LInKs" Lifting Independent Keypoints"-Partial Pose Lifting for Occlusion Handling With Improved Accuracy in 2D-3D Human Pose Estimation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3426-3435).

[25]. Psalta, A., Tsironis, V., & Karantzalos, K. (2024). Transformer-based assignment decision network for multiple object tracking. Computer Vision and Image Understanding, 103957.

[26]. Huang, X., & Zhan, Y. (2024). Multi-object tracking with adaptive measurement noise and information fusion. Image and Vision Computing, 104964.

[27]. Marlin, S., & Jebaseelan, S. (2024). A comprehensive comparative study on intelligence based optimization algorithms used for maximum power tracking in grid-PV systems. Sustainable Computing: Informatics and Systems, 41, 100946.

[28]. Su, S., Han, S., Li, Y., Zhang, Z., Feng, C., Ding, C., & Miao, F. (2024). Collaborative multi-object tracking with conformal uncertainty propagation. IEEE Robotics and Automation Letters.

[29]. Xuan Wang, Zhaojie Sun, Abdellah Chehri, Gwanggil Jeon, Yongchao Song, Deep learning and multi-modal fusion for real-time multi-object tracking: Algorithms, challenges, datasets, and comparative study, Information Fusion, Volume 105, 2024,102247, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2024.102247.

[30]. Zhuang, X., Zhang, T., 2019. Detection of sick broilers by digital image processing and deep learning. Biosystems Engineering 179, 106–116. https://doi.org/10.1016/j. biosystemseng.2019.01.003.