I. J. Engineering and Manufacturing, **2024, 2, 34-45**
Published Online on April 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2024.02.03

Vehicle Object Tracking Based on Fusing of Deep learning and Re-Identification

Huynh Nhat Duy
Department of Computer Vision, University of Science, VNU-HCM
Email: 19C11003@hcmus.edu.vn
ORCID iD: https://orcid.org/0009-0001-7643-8629

Vo Hoai Viet*
Department of Computer Vision, University of Science, VNU-HCM
Email: vhviet @fit.hcmus.edu.vn
ORCID iD: https://orcid.org/0009-0002-7943-8621
*Corresponding Author

Received: 30 October, 2023; Revised: 19 December, 2023; Accepted: 19 February, 2024; Published: 08 April, 2024

Abstract: Object tracking is a popular problem for automatic surveillance systems as well as for the research community. The requirement of an object tracking problem is to predict the output including the object position at the current frame based on the input the position of the object at the previous frame. To present the comparison and experiment of some object tracking methods based on deep learning and suggestions for improvement between them in this paper, we had taken some important steps to conduct this research. First, we find out the studies related to deep learning-based object tracking models. Secondly, we examined image and video data sets for verification purposes. Thirdly, to evaluate the results obtained from existing models, we experimented with a little work related to object tracking based on deep learning networks. Fourth, based on the implemented object tracking models, we had proposed a combination of these methods. And finally, we summarize and give the evaluations for each object tracking model from the results obtained. The results show that object tracking based on Siammask model has the highest results TO score of 0.961356383 on VOT dataset and 0.969301864 on UAV123 dataset, but the possibility of errors is also high. Although the result of the combined method has few scores those are lower than the object tracking based on Siammask model, the combined method is more stable than the object tracking based on Siammask model when TME score of 16.29691993 on VOT dataset and 10.16578548 on UAV123 dataset. The Vehicle ReIdentification method results have scores that are not too overwhelming. However, the TME score is the highest with the TME score of 11.55716097 on the VOT dataset and 4.576163526 on the UAV123 dataset.

Index Terms: Vehicle Object Tracking, Surveillance Systems, Single-Object Tracking, Siammask, Vehicle-ReId

1. Introduction

In the field of computer vision, object tracking remains an important and challenging problem for researchers and automated surveillance systems alike. From traditional object tracking techniques to deep learning network-based object tracking techniques as well as benchmark datasets had updated and published in computer vision community, it has contributed to improving the applicability of the problem of practical systems in general (abnormal detection, human activity recognition, automated surveillance systems, ...) and of the field of computer vision in specific.

With the use of the traditional object tracking method, processing speed of model works fine in real time. However, the accuracy is near saturation, and it is difficult to achieve superior results using only traditional object tracking methods. Because traditional object tracking methods approach the problem based on low-level features, the details related to each part of the object do not contain too much information, making the accuracy of the model not high. Therefore, we have selected some deep learning network-based object tracking models with the expectation that all the metrics of these models will outperform traditional object tracking methods. To consider object tracking problems based on deep learning networks, we have experimented on the benchmark dataset with the specific object being the vehicle. By using a convolutional neural network, the temporal and spatial dependencies between the pixels were utilized, making the accuracy of the model achieve significantly superior results. However large computational volume is an issue to consider for a real-time system.

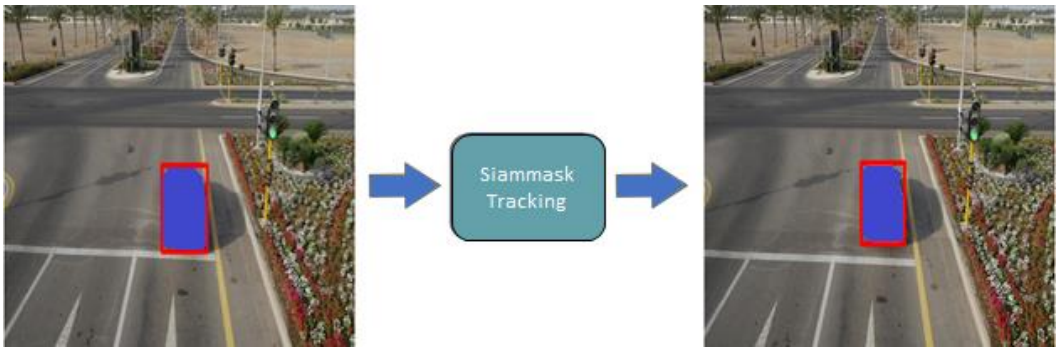


Fig. 1. Example for tracking object based on Siammask Model

There are a lot of studies and reviews related to object tracking problems based on deep learning network. However, the current deep learning network-based vehicle tracking reviews are still limited, we will review and build object tracking methods based on deep learning and Re-Identification methods (Vehicle Re-Id, Siammask), which is also the goal of this paper. We first research and survey deep learning-based object tracking models, the models selected includes two methods: i) Fast Online Object Tracking and Segmentation: A Unifying Approach [1]; ii) Multi-Domain Learning and Identity Mining for Vehicle Re-Identification [2]. Then, build the experiments and demonstrate the research models. From there evaluate the results and compare the experiments. The main contribution of this paper is a review of several deep learning-based object tracking models and their enhancements based on the combination of these methods on the vehicle-specific object to make a robust performance for tracking in vehicle domain. Specifically, fusion of Vehicle ReID and Siammask methods are applied to improve TME score on VOT and UAV123 dataset compared to previous approaches.

The remainder of this paper is organized into four sections; the literature review is presented in section 2. And next, we present some object tracking methods based on deep learning models as shown in section 3. In section 4 of the paper, we present datasets, experimental results, evaluation metrics and discussion. Finally, conclusions are explained in section 5.

2. Literature Review

In this section, we will cite some related reviews to make the point of our article. There is a lot of research as well as reviews [3, 4, 5, 6] related to object tracking problems. However, tracking specific object vehicle is still limited. And to our knowledge, there are no reviews related to deep learning network-based object tracking for vehicle objects. In [7], The authors experimented with some specific object tracking models such as Mean Shift 8, 9, 10] method, Kalman Filter [11, 12] method, Particle Filter [13, 14] method and hybrid methods between them for vehicle objects. In their study, they proved that their object tracking models need a more robust feature set to be able to represent the object's information and make it more accurate although the traditional object tracking methods are easy to implement and the computational volume is not large, detecting the link between two consecutive frames is a problem, if the feature set is not strong enough, the link between the current frame and the next frame will lead to unexpected results. Moreover, in their article, they also mention the use of deep learning networks to track vehicle objects in the future. Since the output of the deep learning model provides strong features including low-level features and high-level features, we will choose several deep learning models to experiment with in this article. By using the features that the deep learning network model provides, we expect that the ability to recognize the object to be tracked between two consecutive frames will be more effective than the traditional tracking methods. Therefore, we will experiment with some evaluation related to object tracking problems based on deep learning network in this article, so we can provide a more accurate view as well as future assessments and directions.

To carry out this review, we had surveyed several works to support experimentation, evaluation, and orientation such as Siammask [15], Vehicle Re-id and YOLO [16] models. The choice of data set is also important as it will affect our results. Therefore, the experimental process will be performed based on two data sets, namely VOT Challenge Dataset and UAV123 Dataset.

The VOT challenge [17] is an annual tracker benchmarking activity organized by the VOT initiative. All benchmark annotations were in accordance with the VOT2021 annotation process [17, 18] and were done manually with a precisely defined and repeatable way of comparing short-term trackers and long-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking, the bounding box annotation are done manually with resizing of the object for each frame sequence on the video. The VOT2021 challenge was composed of four sub-challenges focusing on different tracking domains: i) VOT-ST2021 challenge focused on short-term tracking in RGB; ii) VOT-RT2021 challenge focused on "real-time" short-term tracking in RGB; iii) VOT-LT2021 focused on long-term tracking, namely coping with target disappearance and reappearance; iv) VOT-RGBD2021 challenge focused on long-term tracking in RGB and depth imagery.

Fig. 2. Some images of the VOT Challenge Dataset

Different from popular data sets such as VOT [17], OTB5 [19], most of the data in UAV123 [20] collected includes the objects from an aerial view. The UAV123 [20] dataset contains a total of 123 video sequences and more than 110K frames making it the second-largest object tracking dataset after ALOV300++. All sequences are fully annotated with upright bounding boxes.

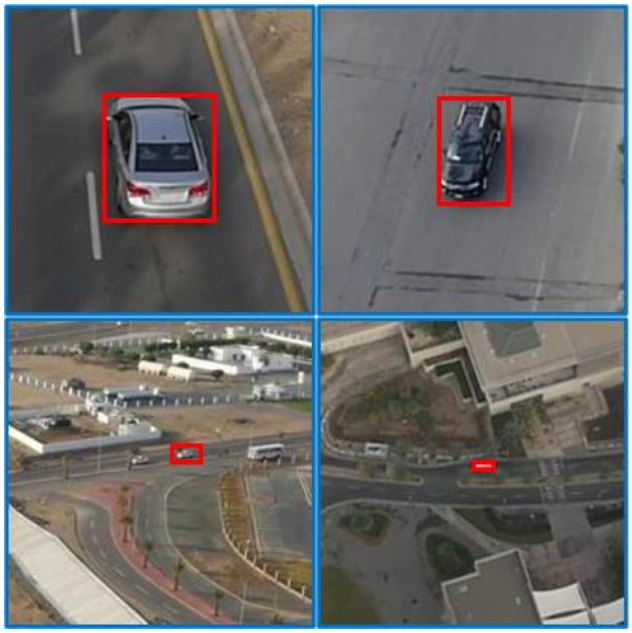


Fig. 3. Some images of the UAV123 Dataset

3. Methodology

In this section, we will present techniques related to deep learning networks as mentioned in section 2. And to ensure that the evaluation methods give the right results when the subject is lost for too long. re-detection of the object will be performed if the tracked object is lost after 20 frames during the construction of object tracking models.

3.1 Yolo

YOLO [16] is not only predicting labels for objects like classification problems, but it also determines the location of objects. Thus, YOLO [16] can detect many objects with different labels in an image instead of only classifying a single label for an image. YOLO [16] may not be the best algorithm, but it is the fastest of the class of object detection

models. It can achieve almost real-time speed without sacrificing accuracy compared to the top models. In this research, we need YOLO model to support Vehicle Re-Id problem.

3.2 Vehicle Re-Id

For building a connection between the frames containing the object on the same video, the model will find out the features representing the tracked object and build a feature vector to match the detected objects (to detect if it is the same object to be tracked or not). Since there are many topics related to Vehicle Re-Id problem [21, 22, 23], we will use a specific method, Multi-Domain Learning and Identity Mining for Vehicle Re-Identification (MDL-IMV Re-ID) [2] in this report. The Multi-Domain Learning and Identity Mining for Vehicle Re-Identification [2] method was developed based on the proposal Strong Baseline with Bag of Tricks [24] (BoT-BS) in Person ReID. The backbone of this model is based on the ResNet101 network [25] with the last average pooling layer removed. Furthermore, to be able to collect vehicle objects on the same frame, we used the YOLOv5 network model [16] for the process of detecting vehicle objects as well as determining the location of vehicles. those objects (algorithm flowchart was shown in Fig. 4).

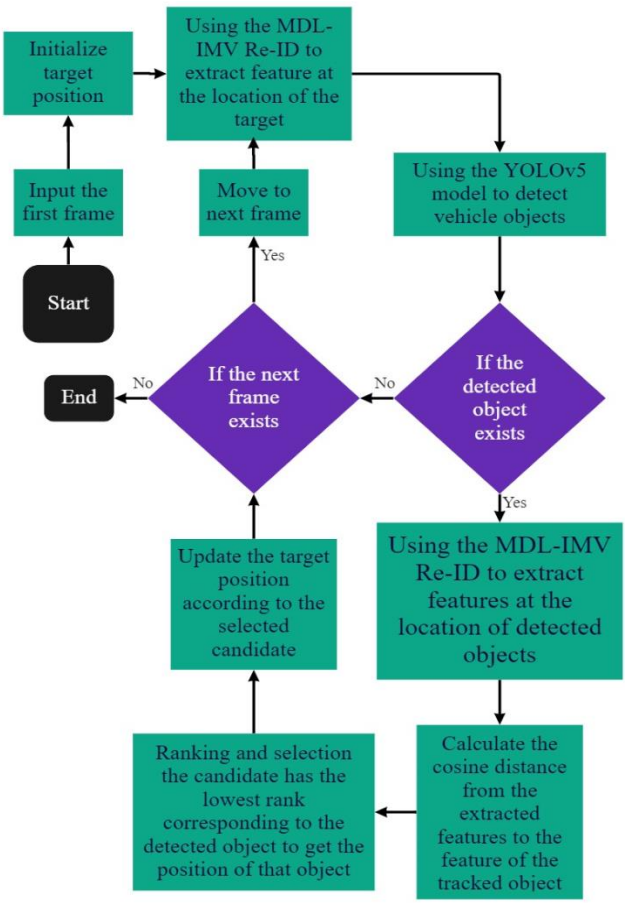


Fig. 4. Flow chart for tracking object base on Vehicle Re-Id

Algorithm Object tracking based on Vehilce-ReId

```
Input: first frame and initial position of the target.
Output: position of bounding box containing target at next frames.
Initialize object position y.
Use Vehicle Re-id to extract the feature corresponding to the target location y.
while all the frames of the video are not finishing executing do
      Use YOLOv5 to detect all vehicles in the current frame.
      if the vehicles exist do
            Feature extraction for each detected object.
            Calculate the cosine distance between the feature of Detected
            objects and the feature of the target y.
            Select candidate position z with the closest score to objective y.
            Update target position y.
```

```
            Extract the feature corresponding to the y target position.
      go to the next frame.
end.
```

3.3 Siammask

Siammask [1, 15] is an object segmentation and visual object tracking model, improved upon previous methods SiamFC [26, 27] (so-called twin neural network) and SiamRPN [28, 29] by adding a new branch to create a binary pixel mask. Siammask is a three-branch variant architecture (Figure 5) that uses ResNet-50 [30] as backbone, the Siammask is using the first 4 stages of ResNet, adjust layer and depth-cross-correlation resulting in a feature map of size 17×17 (create a multichannel response map).

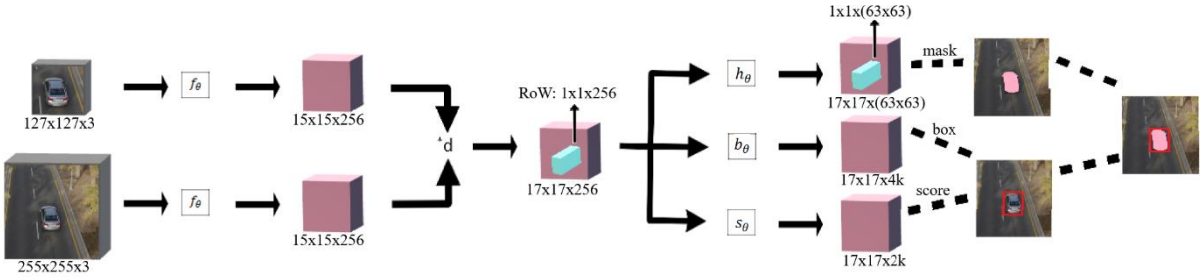


Fig. 5. Three-branch variant architecture of the Siammask model

Given an input image and a smaller sized image cropped from that input image, the Siammask model [15] tries to identify that cropped image in the next frame (shown in Fig. 6).

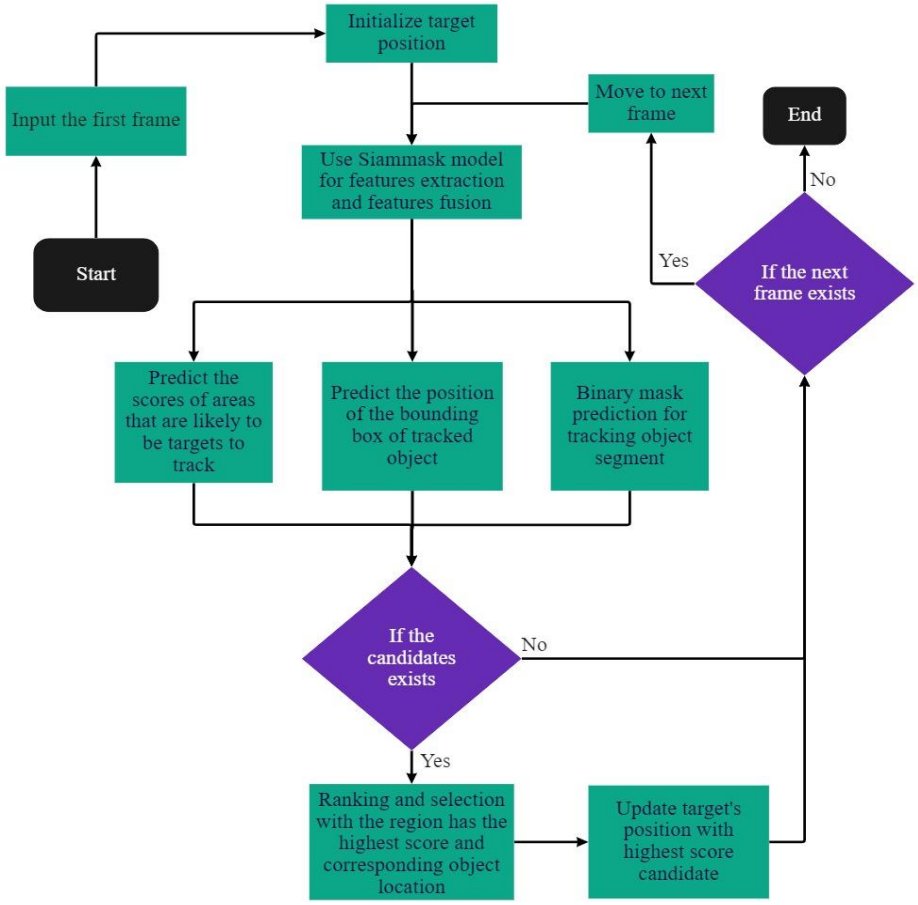


Fig. 6. Flow chart for tracking object base on Sammask model

Algorithm for Object tracking based on Siammask

```
Input: first frame and initial position of the target.
Output: position of bounding box containing target at next frames.
Mark the initial position of the target as template z at position of y.
while all frame of video sequence is complete do
      Uses CNN for feature extraction and feature fusion.
      Get prediction score of target area.
      Get prediction of bounding box of the target.
      Segment the target location.
      if the bounding box of the target is found do
            Update candidate z with highest score of target area.
            Update position of y.
      go to the next frame.
end.
```

3.4 Fusion of Vehicle Re-Id and Siammask

To minimize the problem of errors as well as increase the stability of the system when tracking objects by Siammask model while maintaining high accuracy, we propose a combination solution of Siammask and Vehicle Re-Id. The idea of the solution is as follows, the Siammask model is used to predict the object location and update the feature vector of the tracking target. If the tracked object is not found by Siammask model, the system will perform a feature match of the target by Multi-Domain Learning and Identity Mining for Vehicle Re-Identification [2] with the detected objects, get the best sample and update the location of the tracked object. The meaning of this proposal is to take advantage of the advantages of the two models above to improve the performance of the object tracking model. Furthermore, our expectation is that the weaknesses of the two models will be offset by each other when they are combined. The object tracking algorithm is based on the combination of Siammask and Vehicle Re-Id models is show in Fig. 7.

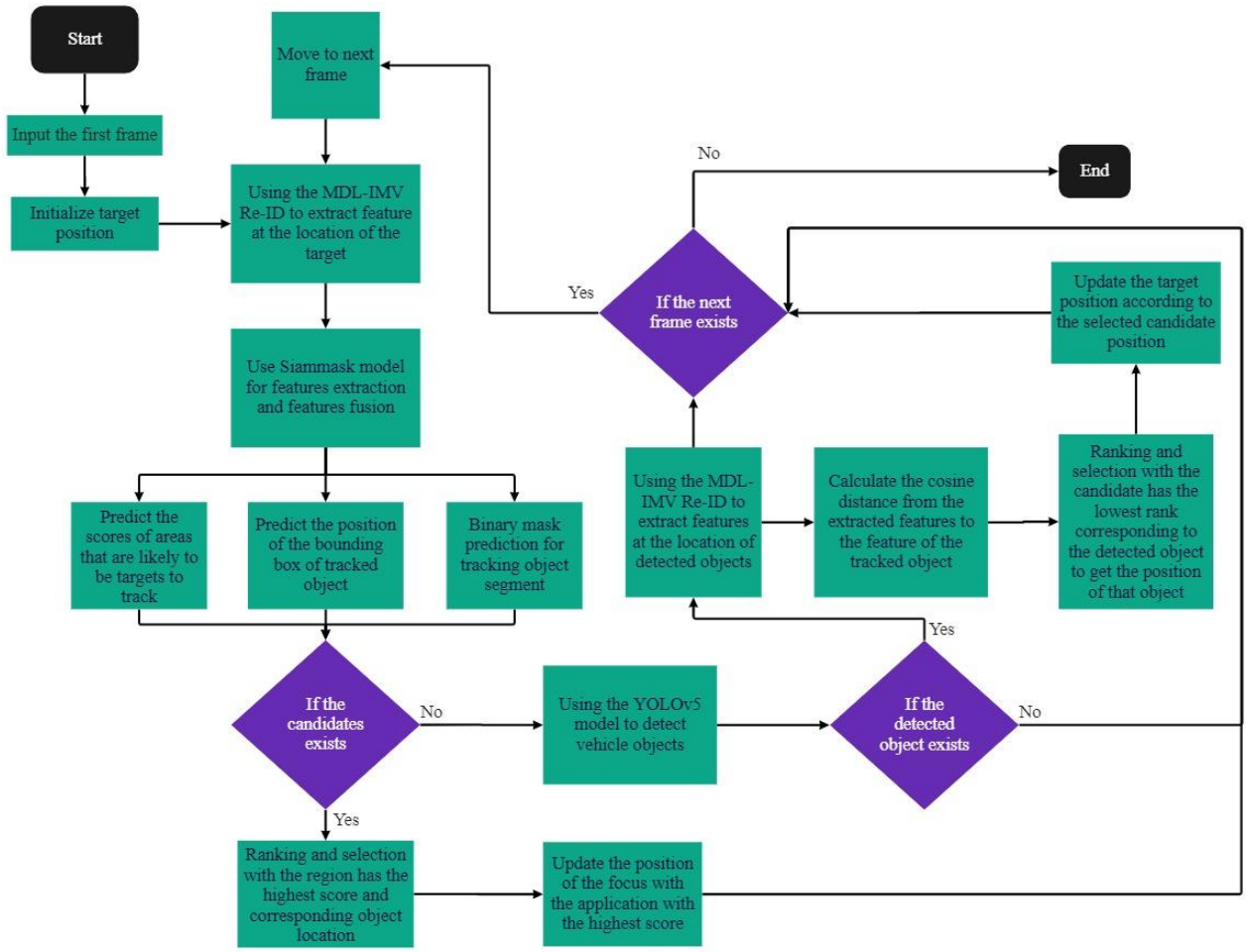


Fig. 7. Flow chart for tracking object base on combination of the Siammask and the MDL-IMV Re-ID

Algorithm for tracking based on fusing of Siamash, Yolo and Vehicle-ReId

```
Input: first frame and initial position of the target.
Output: position of bounding box containing target at next frames.
Mark the initial position of the target as template z at position of y.
Uses Vehicle Re-id to extract feature that corresponding to target position y.
while all frame of video sequence is complete do
      Uses CNN for feature extraction and feature fusion.
      Get prediction score of target area.
      Get prediction of bounding box of the target.
      Segment the target location.
      if bounding box of the target is found do
            Update candidate z with highest score of target area.
            Update position of y.
            Extract feature that corresponds to target position y.
      else
            Uses YOLOv5 to detect all vehicles at current frame.
            if the vehicles exist do
                  Extract feature for each detected object.
                  Calculate the cosine distance between detected object features
                  and the feature of target y.
                  Choose the candidate position z with the score closest to
                  target y.
                  Update position of y.
                  Extract feature that corresponds to target position y.
      go to the next frame.
end.
```

4. Experimental Results and Discussion

4.1 Dataset

In this section, we will present experimental results based on the process of study as well as building object tracking models. To be able to test and evaluate single object (vehicle) tracking models, we had tested results of these models on 2 datasets, VOT Challenge Dataset [17] and UAV123 [20] Dataset. The reason why we chose these 2 datasets is because in [7], we used these 2 datasets for the experimental process, so to be able to show the ability and performance for the object tracking model based on deep learning network on the same data set, we decided to choose 2 datasets VOT chalenge and UAV123 for the experimental process as well as evaluate the experimental results. These two datasets both have normal cases and hard cases such as the angle of the tracked object being changed a lot, occlusion, the camera moves out of the scene with the object tracked.

In addition, there are more cases the influence of object's headlight or the outside light that changes the color of the object of VOT dataset. And cases where the camera is far away from the object tracked or the object too small to detect and track of UAV123 [20] dataset. For each data set used to evaluate the results of several deep learning network-based object tracking models in this paper, we hope that the results will provide an overview of the outstanding scores as well as weaknesses of each model. On the other hand, the model evaluation that combines two object tracking models based on deep learning networks is not only to test and evaluate the results, but also to hope that the direction of combining models together will bring about positive results and better performance. In the future it is possible to combine object tracking model based on deep learning network with different methods.

a. UAV123 dataset with most cases of bird’s eye view

b. VOT Challenge dataset with most cases of closer view

Fig. 8. Pictures illustrating the videos of the dataset

4.2. Evaluation Metrics

Humans can recognize the location of the object to be tracked easily. However, it is a problem for an automated system to understand that event. On the other hand, to be able to assess whether a system is working well or not, we need to have corresponding measures in both space and time for that system.

Given an input video with n objects to track, k is the frame position of that video, i is the label containing the object to be tracked of the actual output and j is the label that the tracking system predicts. The determination of spatial overlap (TO) between ST and GT is defined:

$$A(GT_{ik}, ST_{jk}) = \frac{Area(GT_{ik} \cap ST_{jk})}{Area(GT_{ik} \cup ST_{jk})} \quad (1)$$

During the experiment, we evaluate the models based on the following criteria:

- Correct Detected Track (CDT), also known as True Positive (TP), which means that for the model to work properly, average sufficient spatial overlap (SSO) [7, 31] between GT (Ground-Truth Track) and ST (System Track) $\frac{Length(GT_i \cap ST_j)}{Length(GT_i)}$ must be more than a predefined track overlap threshold TR_{OV} and the temporal overlap (TO) $\frac{\sum_k^N A(GT_{ik}, ST_{jk})}{N}$ [7, 32], that means the number of intersections between ST and GT out of the total number of occurrences of GT must satisfy a given threshold T_{OV}.
- False Alarm Track (FAT) or False Positive (FP), a result is said to be FAT when the average SSO [7, 31] score between GT and ST is less than overlap threshold TR_{OV}, i.e, $\frac{Length(GT_i \cap ST_j)}{Length(ST_j)} \leq TR_{OV}$ and the number of intersections between ST and GT out of the total number of occurrences of ST (TO) is less than threshold T_{OV}.
- Track Detection Failure (TDF), A GT track is considered to have not been detected when the average SSO [7, 31] score between GT and ST is less than overlap threshold TR_{OV}, i.e, $\frac{Length(GT_i \cap ST_j)}{Length(GT_i)} \leq TR_{OV}$ and the number of intersections between ST and GT out of the total number of occurrences of GT (TO) is less than threshold T_{OV}.
- Track fragmentation [7] is a method used to check the continuous tracking level of an object in the tracking system. Latency [33] (time delay) of the system track (LT) is the time gap between the time that an object starts to be tracked by the system and the first appearance of the object.

$$\text{LT} = \text{start frame of } ST_j - \text{start frame of } GT_i \quad (2)$$

- Track matching error (TME) [7] is the average distance error between a system track and the GT track.

$$TME = \frac{\sum_{k=1}^N Dist(GTC_{ik}, STC_{jk})}{Length(GT_i \cap ST_j)} \quad (3)$$

Where $Dist(GTC_{ik}, STC_{jk})$ is the Euclidean distance between the centroids of GT and the system track. GTC and STC are respectively the coordinates of the center of the ground-truth track and system track.

- Track Completeness (TC) [7] is defined as the time span that the system track overlapped with GT track divided by the total time span of GT track, i.e, $TC = \frac{\sum_{k=1}^N O(GT_{ik}, ST_{jk})}{Number\ of\ GT_i}$, where:

$$O(GT_{ik}, ST_{jk}) = \begin{cases} 1 \; if \; A(GT_{ik} \, , \, ST_{jk}) > T_{OV} \\ 0 \; if \; A(GT_{ik} \, , \, ST_{jk}) \leq T_{OV} \end{cases} \quad (4)$$

4.3. Experimental Results

From the experimental results in table 1 and 2 as well as the comparison of analysis in table 3 of object tracking methods based on deep learning networks, we mention that object tracking model based on deep learning network has brought more positive results than traditional object tracking methods, making estimates for each model type based on the empirical process will also consolidate relevant literature and hypotheses. Moreover, the purpose of this discussion is to provide assessments, orientations as well as point out the strengths and weaknesses of some of the models that we have built.

Table 1. The results of the experiment for VOT challenge dataset

Method	The Temporal Overlap (for the case of TP)	The Temporal Overlap (for the case of FP)	Sufficient Spatial Overlap	Track Fragmentation	Track Matching Error	Track Completeness	Latency of The System Track
Vehicle Re-Id	0.475763172	0.45674837	0.358783198	98.95	**11.55716097**	0.464398604	0.238788084
Siammask	**0.961356383**	**0.93267016**	**0.658983168**	**14.9**	16.35113737	**0.892299137**	**0.085450688**
Fusion of Vehicle Re-Id and Siammask	0.959981677	0.931667845	0.655280043	28.65	16.29691993	0.887961502	0.125121895

Table 2. The results of the experiment for UAV123 dataset

Method	The Temporal Overlap (for the case of TP)	The Temporal Overlap (for the case of FP)	Sufficient Spatial Overlap	Track Fragmentation	Track Matching Error	Track Completeness	Latency of The System Track
Vehicle Re-Id	0.412672997	0.403401745	0.352496506	23.45454545	**4.576163526**	0.4043577	**0.230160798**
Siammask	**0.969301864**	**0.947525334**	0.716904096	**1.5**	10.68747544	0.917409012	0.276105723
Fusion of Vehicle Re-Id and Siammask	0.962222492	0.941348253	**0.720272309**	2.363636364	10.16578548	**0.917770435**	0.295623499

Table 3. Comparison of vehicle tracking model pros and cons

Method	**Pros**	**Cons**
Vehicle Re-Id	• The average distance between ST and GT is high.	• Gets complicated when there are many objects as vehicles. • Depends on the object detection model.
Siammask	• High precision • Fastest speed among VOS methods.	• Average distance between ST and GT is not high • Error occurs when motion is blur or not object tracked
Fusion of Vehicle Re-Id and Siammask	• Minimize the chance of error of Siammask method. • High precision.	• When the object loses track of the condition and there are many objects, resulting in slow processing. • Depends on the object detection model when the Siammask model loses track.

Based on the experimental results when building deep learning network models into the object tracking system, the accuracy has increased significantly, typically the Siammask model with CDT 0.961356383 on the VOT dataset and CDT 0.969301864 on the UAV123 dataset. However, the common point of object tracking methods based on deep learning networks usually include a massive number of weights (e.g., 100 MB for Siammask, 189.1 MB for Multi-Domain Learning and Identity Mining for Vehicle Re-Identification), which leads to a large amount of computation and easily slows down the program. The Vehicle Re-Id model is too dependent on the detection model (Table 3), if the object detection model detects wrong, the system can hardly track it. For the object tracking model based on the deep learning network to be put into practice, it is necessary to increase the computational power or combine several other object tracking models to reduce the computational weight of the model. And although the method of combining Siammask and Vehicle Re-Id has reduced the error of the Siammask model, but not much. Otherwise, with the precision of Fusion of Vehicle Re-Id and Siammask method is approximately the same as that of Siammask with CDT 0.959981677 on the VOT dataset and CDT 0.962222492 on the UAV123 dataset. Depending on dataset, Fusion of Vehicle Re-Id and Siammask method has higher indicators than Siammask such as SSO score of 0.720272309 on the UAV123 dataset and TC score of 0.917770435 on the UAV123 dataset while Siammask has SSO score of 0.716904096 on the UAV123 dataset and TC score is 0.917409012 on the UAV123 dataset. Moreover, the error of Fusion of Vehicle Re-Id and Siammask method is lower than that of Siammask as TME score is 16.29691993 on the VOT dataset and

TME score is 10.16578548 on the UAV123 dataset, while Siammask's TME score is 16,35113737 on the VOT dataset and 10.68747544 on the UAV123 dataset, respectively. Although the Fusion of Vehicle Re-Id and Siammask method does not significantly reduce the error compared to Siammask, with this result, it is quite acceptable and meets the requirements.

5. Conclusion

Generally, deep learning approach is a promising methodology for the state-of-the-art performance in many tasks in computer vision. In this research, we propose a fusing scheme of deep learning approaches for object detection, object tracking, and Re-Identification to create a robust methodology for object tracking in vehicle domain. Specifically, we use Siammask for object tracking, Yolo for object detection and Vehicle Re-Identification are fused into a unified framework. Experimentally, it is found that most of the scores of Siammask method are better than the other two methods and Vehicle Re-Id method has outstanding TME score. However, the combination method of Siammask and Vehicle Re-Id gives more stable results, moreover, the possibility of error is lower for the method of combining Siammask and Vehicle Re-Id than that of Siammask. On the other hand, when experimenting on the UAV123 dataset with deep learning network models, most of the indicators of the Siammask method are superior to the other two methods with the following TO measurement methods (in the case of TP) 0.969301864, TO (in case of FP) 0.947525334 and Track fragmentation 1.5. The Vehicle Re-Id method achieves the highest TME score 4.576163526. With the results of Siammask method based on Vehicle Re-Id showed the predominance of SSO indices 0.720272309401442 and TC 0.917770434662788. Siammask combined with Vehicle Re-Id has shown more stability than the other two methods and the low indicators are not too different from the Siammask method.

Object tracking is one of the growing research directions in the field of computer vision. However, to apply it in practice, it is necessary to study and solve many problems such as calculation ability and accuracy. In the future, incorporating traditional object tracking methods into deep learning network models can reduce computational costs as well as improve system speed.

Acknowledgment

This research is funded by the University of Science, VNU-HCM, Vietnam under grant number CNTT 2023-10.

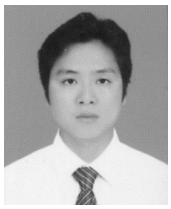
References

[1] Wang, Q., et al. Fast Online Object Tracking and Segmentation: A Unifying Approach. IEEE, 2020, pp. 1328–38.
[2] He, Shuting & Luo, Hao & Chen, Weihua & Zhang, Miao & Zhang, Yuqi & Wang, Fan & Li, Hao & Jiang, Wei. (2020). Multi-Domain Learning and Identity Mining for Vehicle Re-Identification. DOI:10.1109/CVPRW50498.2020.00299.
[3] Z. Soleimanitaleb, M. A. Keyvanrad and A. Jafari, "Object Tracking Methods: A Review," 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), 2019, pp. 282-288, doi: 10.1109/ICCKE48569.2019.8964761.
[4] K. R. Reddy, K. H. Priya and N. Neelima, "Object Detection and Tracking -- A Survey," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015, pp. 418-421, doi: 10.1109/CICN.2015.317.
[5] Held, David & Thrun, Sebastian & Savarese, Silvio. (2016). Learning to Track at 100 FPS with Deep Regression Networks.
[6] X. Hou, Y. Wang and L. Chau, "Vehicle Tracking Using Deep SORT with Low Confidence Track Filtering," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2019, pp. 1-6, doi: 10.1109/AVSS.2019.8909903.
[7] Vo Hoai Viet, Huynh Nhat Duy, " Object Tracking: An Experimental and Comprehensive Study on Vehicle Object in Video", International Journal of Image, Graphics and Signal Processing (IJIGSP), Vol.14, No.1, pp. 64-81, 2022.DOI: 10.5815/ijigsp.2022.01.06.
[8] Ravi Kumar Jatoth, Sampad Shubhra, Ejaz Ali,"Performance Comparison of Kalman Filter and Mean Shift Algorithm for Object Tracking", IJIEEB, vol.5, no.5, pp.17-24, 2013. DOI: 10.5815/ijieeb.2013.05.03 Reference.
[9] COMANICIU, D. AND MEER, P. 1999. Mean shift analysis and applications. In IEEE International Conference on Computer Vision (ICCV). Vol. 2. 1197–1203, doi: 10.1109/ICCV.1999.790416.
[10] H. Wang, X. Wang, L. Yu and F. Zhong, "Design of Mean Shift Tracking Algorithm Based on Target Position Prediction," 2019 IEEE International Conference on Mechatronics and Automation (ICMA), 2019, pp. 1114-1119, doi: 10.1109/ICMA.2019.8816295.
[11] Lim Chot Hun, Ong Lee Yeng, Lim Tien Sze and Koo Voon Chet (June 8th, 2016). Kalman Filtering and Its Real - Time Applications, Real-time Systems, Kuodi Jian, IntechOpen, DOI: 10.5772/62352.
[12] Feng Xiao; Mingyu Song; Xin Guo; Fengxiang Ge. Adaptive Kalman filtering for target tracking. 2016 IEEE/OES China Ocean Acoustics (COA), 2016, pp. 1-5, doi: 10.1109/COA.2016.7535797.
[13] Oğuzhan GültekİN, Bilge Günsel, "Robust object tracking by variable rate kernel particle filter", 2018 26th Signal Processing and Communications Applications Conference (SIU), 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404479.
[14] Marina A. Zanina, Vitalii A. Pavlov, Sergey V. Zavjalov, Sergey V. Volvenko, "TLD Object Tracking Algorithm Improved with Particle Filter". 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 2018, pp. 1-4, doi: 10.1109/TSP.2018.8441515.

[15] Z. Liang, C. Liang, Y. Zhang, H. Mu and G. Li, "Tracking of Moving Target Based on SiamMask for Video SAR System," 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), 2019, pp. 1-4, doi: 10.1109/ICSIDP47821.2019.9173432.
[16] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
[17] VOT Matej Kristan, Jiri Matas, Aleš Leonardis, Tomáš Vojíř, Roman Pflugfelder, Gustavo Fernández, Georg Nebehay, Fatih Porikli and Luka Čehovin, "A Novel Performance Evaluation Methodology for Single-Target Trackers", PAMI, vol. 38, no. 11, pp. 2137-2155, 1 Nov. 2016, doi: 10.1109/TPAMI.2016.2516982.
[18] Kristan, Matej & Matas, Jiri & Leonardis, Ales & Felsberg, Michael & Pflugfelder, Roman & Kamarainen, Joni-Kristian & Chang, Hyung & Danelljan, Martin & Čehovin Zajc, Luka & Lukežič, Alan & Drbohlav, Ondrej & Kapyla, Jani & Häger, Gustav & Yan, Song & Yang, Jinyu & Zhang, Zhongqun & Fernandez Dominguez, Gustavo & Abdelpakey, Mohamed & Bhat, Goutam & Zhu, Xue-Feng. (2021). The Ninth Visual Object Tracking VOT2021 Challenge Results. 2711-2738. 10.1109/ICCVW54120.2021.00305.
[19] Wu, Yi & Lim, Jongwoo & Yang, Ming-Hsuan. (2015). Object Tracking Benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence. 37. 1-1. 10.1109/TPAMI.2014.2388226.
[20] UAV123 Matthias Mueller, Neil Smith, and Bernard Ghanem, "A Benchmark and Simulator for UAV Tracking", ECCV, 2016.
[21] X. Liu, Y. Dong and Z. Deng, "Deep Highway Multi-Camera Vehicle Re-ID with Tracking Context," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 2090-2093, doi: 10.1109/ITNEC48623.2020.9085008.
[22] Z. Jamali, J. Deng, J. Cai, M. U. Aftab and K. Hussain, "Minimizing Vehicle Re-Identification Dataset Bias Using Effective Data Augmentation Method," 2019 15th International Conference on Semantics, Knowledge and Grids (SKG), 2019, pp. 127-130, doi: 10.1109/SKG49510.2019.00030.
[23] M. Wu, Y. Qian, C. Wang and M. Yang, "A Multi-Camera Vehicle Tracking System based on City-Scale Vehicle Re-ID and Spatial-Temporal Information," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 4072-4081, doi: 10.1109/CVPRW53098.2021.00460.
[24] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 0–0, 2019.
[25] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[26] Xu, Yinda & Wang, Zeyu & Li, Zuoxin & Yuan, Ye & Yu, Gang. (2020). SiamFC++: Towards Robust and Accurate Visual Tracking with Target Estimation Guidelines. Proceedings of the AAAI Conference on Artificial Intelligence. 34. 12549-12556. 10.1609/aaai.v34i07.6944.
[27] Li, Daqun & Yu, Yi & Chen, Xiaolin. (2019). Object tracking framework with Siamese network and re-detection mechanism. EURASIP Journal on Wireless Communications and Networking. 2019. 10.1186/s13638-019-1579-x.
[28] Li, Bo & Wu, Wei & Wang, Qiang & Zhang, Fangyi & Xing, Junliang & Yan, Junjie. (2019). SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks. 4277-4286. 10.1109/CVPR.2019.00441.
[29] B. Li, J. Yan, W. Wu, Z. Zhu and X. Hu, "High Performance Visual Tracking with Siamese Region Proposal Network," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8971-8980, doi: 10.1109/CVPR.2018.00935.
[30] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[31] L. M. Brown, A. W. Senior, Ying-li Tian, Jonathan Connell, Arun Hampapur, Chiao-Fe Shu, Hans Merkl, Max Lu, “Performance Evaluation of Surveillance Systems Under Varying Conditions”, IEEE Int'l Workshop on Performance Evaluation of Tracking and Surveillance, Colorado, Jan 2005.
[32] F. Bashir, F. Porikli. “Performance evaluation of object detection and tracking systems”, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), June 2006.
[33] Sven Ubik; Jiří Pospíšilík. Video Camera Latency Analysis and Measurement. IEEE Transactions on Circuits and Systems for Video Technology (Volume: 31, Issue: 1, Jan. 2021): 140 - 147. DOI: 10.1109/TCSVT.2020.2978057.

Authors’ Profiles

Huynh Nhat Duy graduated from the University of Science, VNU-HCMC, Vietnam in in 2015. Now, he is pursuing the master degree of Computer Science at University of Science, VNU-HCMC. His research interests include Image Processing and Computer Vision.



Vo Hoai Viet is a Lecturer and Senior Researcher at the University of Science, VNU-HCMC, Vietnam from 2012. He is currently working in Computer Vision at University of Science, VNU-HCMC, Vietnam. His research interests include Digital Image Processing, Programming Language, Computer Graphics, Computer vision, and Machine Learning.

How to cite this paper: Huynh Nhat Duy, Vo Hoai Viet, "Vehicle Object Tracking Based on Fusing of Deep learning and Re-Identification", International Journal of Engineering and Manufacturing (IJEM), Vol.14, No.2, pp. 34-45, 2024. DOI:10.5815/ijem.2024.02.03