Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism

Mohammed Thakir Mahmood 1, Saadaldeen Rashid Ahmed AHMED2, Mohammed Rashid Ahmed AHMED3

1 Computer Engineering, Karabuk University, Istanbul, Turkey
2 Computer Science, Tikrit University, Salahaldeen, I
3 Computer Engineering, Altinbas University, Istanbul, Turkey
E-mail: mohammed1991almashhadany@gmail.com

Abstract

Vehicle counting is an important process in the estimation of road traffic density to evaluate the traffic conditions in intelligent transportation systems. With increased use of cameras in urban centers and transportation systems, surveillance videos have become central sources of data. Vehicle detection is one of the essential uses of object detection in intelligent transport systems. Object detection aims at extracting certain vehicle-related information from videos and pictures containing vehicles. This form of information collection in intelligent systems is faced with low detection accuracy, inaccuracy in vehicle type detection, slow processing speeds. In this research, we propose a vehicle detection system from infrared images using YOLO (You Look Only Once) computational mechanism. The YOLO mechanism can apply different machine or deep learning algorithms for accurate vehicle type detection. In this study we propose an infrared based technique to combine with YOLO for vehicle detection in traffic. This method will be compared with a machine learning technique of K-means++ clustering algorithm, a deep learning mechanism of multitarget detection and infrared imagery using convolutional neural network

Keywords: Infrared, YOLO, intelligent transport system, spatial resolution, detection

1. Introduction

Infrared (IR) target tracking and detection is critical in video surveillance especially in transportation systems[1]. This infrared system has been utilized in military applications especially in IR imaging and guidance technology. This technology has attracted considerable attention due to its anti-interference ability, observability in all weather, high guidance precision and long detection distances[1]. Nonetheless, in contrast to the conventional visual image, the IR images have low spatial resolution, lack of textural information, and poor signal-to-noise (SNR) ratio. Additionally, tracking of vehicles in traffic which are moving fast using IR images may raise problems with target resolution and background motion[1]. In this study, we propose an observational framework for vehicle detection in traffic that utilizes infrared imaging and YOLO computational mechanism. This technique will be compared to other viable alternatives.
In order to overcome the problems associated with the current vehicle detection and tracking systems we will compare our proposed mechanism with kmean++ clustering algorithm that uses bounding boxes of varied sizes on the training dataset for detection of vehicles in traffic[2]. Secondly, a deep learning Multi-target detection approach that involves the YOLO mechanism under the Darknet framework[3]. Finally, an infrared imaging that uses convolutional Neural Network.

The primary contribution of this research is to develop the effect of an incredible algorithm of vehicle in IR target tracking. This research will provide important findings for development of IR-based vehicle detection and tracking systems that combine deep learning and YOLO computational technique. The second contribution of this study is to compare the various tracking algorithms to determine the most effective in vehicle detection in IR images.

2. RELATED WORK

In section we provide a review of the important algorithms used for object detection in intelligent transport system (ITS).

2.1 You only look once (YOLO)- Real-Time Object Detection

YOLO is a fast object detection algorithm that can be utilized in vehicle detection in an image. Although it has been utilized for a while it is not accurate for object detection as it is associated with loss of precision. It disadvantage is that it uses a single CNN network for the classification and localizing of an object employing the bounding boxes. Figure 1 below demonstrates the architecture of the YOLO technique.

![Figure 1: The architecture of the YOLO model [3]](image)

2.2 YOLOv2

YOLO technique guarantee real-time image processing with high accuracy but the method has a higher localization error with lower recall response. The YOLOv2 is an updated version of the YOLO technique. The YOLOv2 increases the accuracy and the recall response time as it incorporates new features listed below:

- A fully connected layer that are important and responsible for prediction of the boundary.
- Class prediction is accomplished at the boundary level rather than the cell level. The resulting elements will have four parameters of the boundary box.
- A pooling layer is removed to introduce a spatial output of the network to 13x13 from the initial 7x7
- Input image is varied from 418x418 to 416x416. The will result on an odd-numbered spatial dimensions which is important in case the picture is occupied by a large image in the center. S
- The last convolution layer in the image is replaced with three 3x3 convolutional layers that generate 1024 output channels

2.3 YOLO v3

This is an updated version of the YOLO that includes multi-label classification. The YOLOv3 produces non-exclusive output that has a score more than one. The YOLOv3 does not use softmax but rather an independent logistic classifier utilized to compute the likeness of the objects in the image. Furthermore, YOLOv3 employs a binary cross-entropy loss for each label rather than using the mean square error in the computation of the classification loss. Figure 2 demonstrates the neural architecture of YOLOv3.
2.4 Infrared Tracking using combined Tracking and Detection (CTAD)

This proposed mechanism is aimed at developing a tracking algorithm that can identify a vehicle fast and accurately using infrared image sequence[1]. The framework is created using a tracking system based on correlation filter and detector based on deep learning. This system is a combined tracking and detecting (CTAD)[4]. This collaboration leads to the algorithm having a robust discriminative power and high efficiency filter using deep learning. The technique was verified using experimentation for IR image sequences using a dataset for quantitative evaluation of the algorithm’s performance.

In this technique, a tracker-based on LCT will be used with a detector YOLOv3 utilized in the verification of the results as illustrated in figure 1 below. In this approach, the tracker T is the core of the algorithm that carries out the tracking of the target in most processes [13]. This tracker is responsible for the real-time requirements of the application. The tracker is made from the LCT algorithm which was selected based on various IR image sequences. This tracker address the problem related to long-term visual tracking in which the target vehicle in the image goes through heavy occlusion and substantial variation in appearance. This algorithm improves the tracking accuracy and reliability in complex environment.

The detector in this technique uses the deep learning technique of YOLOv3. This technique is important in verifying the tracking. The results in this verification system indicate that the system could detect targets contained in complex background. The detector varies the tracking results within specific frequency bands to reduce the need for heavy and complex computation. In this method the tracker operates independently. The CTAD technique can be summarized in the following pseudocode:

THE TRACKING THREAT IS INITIALIZED FOR THE TRACKER;

THE DETECTING THREAD FOR THE DETECTOR IS INITIALIZED;

RUN THE TRACKER;

IF INDEX OF THE FIGURE > Δn OR THE TRACKING FAILS THEN
RUN DETECTOR

VERIFY THE TARGET POSITION USING RESULTS OF DETECTOR;
ELSE
RUN TRACKER;
END

2.5 Kmean++ Clustering Algorithm (KCA)

In this clustering algorithm bounding boxes were used for training dataset, and six anchor boxes that have varied sizes are used for identification of objects in an image. The various scales of the vehicle may influence the vehicle detection model[2]. In this analysis, normalization was used to enhance the loss computation technique for the length and width of the bounding boxes. This technique enhances the feature extraction capability of the network[5]. The mechanism further used a multi-layer feature fusion and eliminated the repetitive convolution layer in the higher layers. The results depicted in this technique showed that the mean Average Precision (mAP) reached 94.78%. This model was also excellent in generalization of the test dataset.

This technique uses YOLOv2 for the vehicle detection model. The first stage entailed selection of anchor boxes [2]. This task was accomplished using the kmeans++ clustering algorithm for cluster analysis on the actual size of the vehicle and determine the bounding boxes from the BIT-vehicle training dataset [12]. This algorithm will then select the most suitable anchor boxes for the detection of the vehicles. To separate the boxes, the YOLOv2 distance function was applied instead of the Euclidean distance [17]. The IOU was adopted as the evaluation metric, as illustrated below:

\[ d(box, centroid) = 1 - IOU(box, centroid) \]

Through analysis of the clustering results, the value of k was finally set to 6. Thus, 6 anchor boxes of varied sizes would be used for positioning of the vehicle.

For vehicle detection, it is evident that the vehicle was approaching the surveillance camera during detection, thus the vehicle size may vary considerably. In training the YOLOv2, the various object size had a distinct impact on the entire model, introducing new large errors. Normalization calculations were used to calculate the width and the height of the bounding boxes as shown below.

\[
\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} o^{bj}_{ij} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} o^{bj}_{ij} \left[ \left( \frac{\omega_i - \hat{\omega}_i}{\omega_i} \right)^2 + \left( \frac{h_i - \hat{h}_i}{h_i} \right)^2 \right] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} n^{noob}_{ij} \left( C_i - \hat{C}_i \right)^2 + \lambda_{noob} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \left( \bar{C}_i - \hat{\bar{C}}_i \right)^2
\]

Where
\n\n- \( x_i \) and \( y_i \) are the center coordinates of the box of the \( i^{th} \) grid cell
- \( \omega_i \) and \( h_i \) are the width and height of the box of the \( i^{th} \) grid cell
- \( C_i \) is the confidence of the box of the \( i^{th} \) grid cell

\[
\sum_{i=0}^{S^2} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\]
\( \hat{x}_i, \hat{y}_i, \hat{\omega}_i, \hat{h}_i, \hat{C}_i \) and \( \hat{p}_i \) are the corresponding prediction of \( x_i, y_i, \omega_i, h_i, C_i \) and \( p_i \).

\( \lambda_{\text{coord}} \) is the weight of the coordinate loss.

\( \lambda_{\text{noobj}} \) is the weight of the bounding boxes.

\( B \) is the bounding boxes.

\( S^2 \) is the \( S \times S \) grid cells.

\( \Pi_{i}^{\text{obj}} \) indicate if the box is located in cell \( i \) or not.

\( \Pi_{j}^{\text{obj}} \) indicates that the \( j \)th box is responsible for prediction.

The element in this framework is the design of the network for vehicle detection. This is a multi-layer feature fusion because the variation in the vehicles has difference in color, contour, tire shape, and lamp shape [10]. The multilayer feature fusion strategy was used for reorganizing the local information.

### 2.5.1 Design of Network

In this KCA method, the process involves two significant steps: The multi-layer feature fusion, is the first step of identifying vehicles in traffic images. In in this section the difference between the vehicles is identified using contour, tire shape and lamp shape. The multi-layer feature fusion takes the general YOLOv2_vehicle model as illustrated in figure 4 below.

![Figure 4: The proposed network structure of the KCA model [2]](image)

The second step entails removing the repeated convolution layers in high layers. This step increases the amount of classes detected by this network. Additionally, this technique introduces the differences between the classes. This technique introduces a continuous and repeated 3x3x1024 convolution layers in the higher sections.

### 2.6 Deep Learning Multi-Target Detection (DLMTD)

This scheme is a multivehicle detection technique that uses YOLO under the Darknet framework. This approach enhances the YOLO-voc framework based on the variation of the target scene and traffic flow[3]. In this scheme the training model is based on the ImageNet where the parameters are tuned based on the training results and the vehicle features. This technique results in the creation of a YOLO-vocRV network that can be utilized for vehicle detection with a detection rate of 98.6% in free flow state, 96.3% in blocking flow state and 97.8% in synchronous flow state[3].

This technique also used YOLOv2 which can easily distinguish between the background and the target [11]. In this YOLOv2 the target location and the probability of the multi-target can be predicted in real-time [16].

The features in this technique were acquired using the CNN technique which eliminate the complex preprocessing that in need for images, the features that are obtained include: distortion variance, displacement invariance and scaling invariance [3]. The learning capabilities of the network make it easy for neurons to learn the weights for mapping planes, and while learning the network shares the weights to decrease the complexity of the technique.

Vehicle detection is completed using design concepts of the YOLOv2 that uses real-time detection and end-to-end training. The image will be divided into \( S \times S \) grid cell for learning feature [3]. When a vehicle falls in a cell then the said cell is responsible for detection of the vehicle. Additionally, the corresponding box is used for direct prediction for each target.
location in the feature map. The box regression will be utilized in the fine-tuning of the window and perform clustering statistics using the K-mean algorithm.

In this technique, the data set is fine-tuned based on classification. This fine tuning technique is utilized in the training of the vehicle dataset that will be used in the convolution neural network. Furthermore, the data is enhanced during the training phase using random scaling, exposure and saturation. Once diversity is established in the images the neural network is used to divide the image into various regions, which make it easy to predict the probabilities and borders and assign the bounding boxes based on the probabilities. The DLMTD model is illustrated below.

![Vehicle Detection Flowchart of the DLMTD Model](image1)

**2.7 Infrared using Convolutional Neural Network (CNN)**

This is a novel approach used to detect ground vehicle using aerial IR images that depend on the convolutional neural network [4]. The IR technique is evaluated using an IR dataset. The dataset used is publicly released [6]. The data was tested by building an end-to-end convolutional neural network which is constructed in this study. This technique was able to detect the stationary and moving vehicles in real urban traffic environment.

The technique proposed for object detection in infrared images is divided into three. The first technique entails manually segmenting the vehicles in the images using a labelling toolbox [4]. This labelling step are pivotal to the trainings. Secondly, the technique carries out sample region feature extraction in the convolution neural network. In this case data augmentation such as crops, sample expansion, exposure shifts, and rotation [10]. The training approach in this technique uses pre-trained classification network based on Image Net. Below is a flowchart of this technique.

![Proposed Vehicle Detection Technique from IR Images](image2)
3. COMPARISON

In section, we will compare the four mechanism of detecting vehicle in image taken from cameras from the intelligent transport system (ITS).

Table 1: Comparison of the detection algorithm

|                  | CTAD  | KCA      | DLMTD | CNN     |
|------------------|-------|----------|-------|---------|
| Tracking         | LCT   | Kmean++  | CNN   | CNN     |
| Detection        | YOLOv3| YOLOv2   | YOLOv2| Labeling|
| Error correction | Uses Regression models | Multi-Layer Feature Fusion | Intersection Over Union | Suppression |

Table 2: Comparison of the detection algorithms [1][2][3][4]

|                  | C TAD | K CA  | DL MTD | C NN  |
|------------------|-------|-------|--------|-------|
| Frames Per Second| 18.1  | 1     | 19.8   | 1     |
| Precision (%)    | 81.1  | 9.78  | 97.5   | 9.46  |

The techniques were compared based on the accuracy of detection (precision) and the speed of evaluation (FPS). Based on the literature review the DLMTD had the highest detection accuracy and the highest speed of evaluation[7]. Our technique had the most desired speed of 18.1 but had the second highest accuracy percentage. The combined tacking and detection (CTAD) performance can be improved by subjecting all the techniques to the infrared images that were used[9]. In the case of the IR images have low spatial resolution, lack of textural information, and poor signal-to-noise (SNR) ratio. These elements reduced the accuracy of the CTAD technique while the other mechanism used the normal images[8]. To have the second best speed illustrate the potential of this technique.

4. CONCLUSION

In this study, the survey compared some notable object detection techniques that can be applied on IR images to detect and track vehicle in ITS. The CTAD is a new technique that when couple with YOLO produce incredible results. The technique was evaluated in classical methods. The other techniques have been used to detect vehicle in normal images. The findings of this survey show that the is potential in the future to develop various techniques based on YOLO to detect vehicle on infrared images.

References

[1] Y. Hu, M. Xiao, K. Zhang, and X. Wang, “Aerial Infrared Target Tracking in Complex Background Based on Combined Tracking and Detecting,” *Math. Probl. Eng.*, vol. 2019, 2019, doi: 10.1155/2019/2419579.
[2] J. Sang *et al.*, “An improved YOLOv2 for vehicle detection,” *Sensors (Switzerland)*, vol. 18, no. 12, Dec. 2018, doi: 10.3390/s18124272.
[3] X. Li, Y. Liu, Z. Zhao, Y. Zhang, and L. He, “A deep learning approach of vehicle multitarget detection from traffic video,” *J. Adv. Transp.*, vol. 2018, 2018, doi: 10.1155/2018/7075814.
[4] X. Liu, T. Yang, and J. Li, “Real-time ground vehicle detection in aerial infrared imagery based on convolutional neural network,” *Electron.*, vol. 7, no. 6, Jun. 2018, doi: 10.3390/electronics7060078.
[5] J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, “An Operational System for Estimating Road Traffic Information from Aerial Images,” *Remote Sens.*, vol. 6, no. 11, pp. 11315–11341, Nov. 2014, doi: 10.3390/rs6111315.
[6] B. Maschinen, A. Investition, G. Beschaffungen, B. Ersatzbeschaffungen, and S. Mittelherkunft, “No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title.”

[7] L. Jiao et al., “A Survey of Deep Learning-based Object Detection.”

[8] C. S. Asha and A. V. Narasimhadhan, “Vehicle Counting for Traffic Management System using YOLO and Correlation Filter,” in 2018 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCCT 2018, Oct. 2018, doi: 10.1109/CONCCCT.2018.8482380.

[9] H. Song, H. Liang, H. Li, Z. Dai, and X. Yun, “Vision-based vehicle detection and counting system using deep learning in highway scenes,” Eur. Transp. Res. Rev., vol. 11, no. 1, pp. 1–16, Dec. 2019, doi: 10.1186/s12544-019-0390-4.

[10] S. Zhao, F. You, L. Shang, C. Han, and X. Wang, “Vehicle detection in aerial image based on deep learning,” p. 32006, 2019, doi: 10.1088/1742-6596/1302/3/032006.

[11] R. Laroca et al., “A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector.” Accessed: Apr. 25, 2020. [Online]. Available: https://web.inf.ufpr.br/vri/databases/ufpr-alpr/.

[12] C. J. Seo, “Vehicle Detection and Car Type Identification System using Deep Learning and Unmanned Aerial Vehicle,” Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 8, pp. 814–819, 2019.

[13] W. Liu et al., “G-RMI Object Detection,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.

[14] A. F. Zohra, S. Kamilia, A. Faycal, and S. Souad, “Detection And Classification Of Vehicles Using Deep Learning,” Int. J. Comput. Sci. Trends Technol., vol. 6, 2013, Accessed: Apr. 25, 2020. [Online]. Available: www.ijcstjournal.org.

[15] D. Gour and A. Kanskar, “Automated AI Based Road Traffic Accident Alert System: YOLO Algorithm,” Int. J. Sci. Technol. Res., vol. 8, p. 8, 2019, Accessed: Apr. 25, 2020. [Online]. Available: www.ijstr.org.

[16] Y. Jamtsho, P. Riyamongkol, and R. Waranusast, “Real-time Bhutanese license plate localization using YOLO,” ICT Express, Nov. 2019, doi: 10.1016/j.ictexc.2019.11.001.

[17] A. R. Caballo and C. J. Aliac, “YOLO-based Tricycle Detection from Traffic Video,” in Proceedings of the 2020 3rd International Conference on Image and Graphics Processing, 2020, pp. 12–16, doi: 10.1145/3383812.3383828.