A comprehensive review of models for vehicle detection based on computer vision analysis in autonomous vehicle

Yunxuan Mo

University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom

ym6y23@soton.ac.uk

Abstract. The application of computer vision analysis technology based on traditional image analysis and machine learning techniques in the field of vehicle detection is the focus of this paper. This paper fills the gap in previous research and provides a comprehensive overview and comparison of vehicle detection models based on computer vision analysis. This paper first briefly outlines the goals of vehicle recognition, evaluation indicators of models, and widely used datasets; then, it summarizes vehicle detection models based on traditional image processing techniques and machine learning techniques. Finally, the advantages and disadvantages of various models and sensors are discussed, and potential future development directions are proposed.

Keywords: Vehicle identification, Autonomous driving, Sensor technology combination, Deep learning.

1. Introduction

A study shows that self-driving cars can eliminate 94% of traffic accidents caused by driver distraction or operational errors [1]. In addition, self-driving systems can help prevent vehicle component failures, reduce emissions, and provide convenience for people with disabilities to drive [2]. Consequently, the future of automobile design and driving development will be steered towards self-driving.

Autonomous driving system design usually includes environment perception, behavior decision-making, motion planning and control [3]. The ability to perceive the environment is the basis of autonomous driving [4], requiring the driving system to be able to identify entities such as surrounding vehicles, pedestrians, or traffic signs.

Statistics show that the main threat to drivers often comes from surrounding vehicles [5]. Vehicle detection technology reduces this risk by accurately and efficiently detecting surrounding vehicles while driving, which is critical to ensuring the safety of drivers and passengers [6].

At present, the development challenges of vehicle detection technology mainly caused by poor information processing speed. Because of the suddenness of accidents, vehicle detection technology requires faster processing speed than other applications, which makes it more complex [7]. The introduction of deep learning technology can effectively solve this problem [8, 9].

Section 2 introduces the methodology; Section 3 introduces the objectives, common datasets, and details vehicle detection algorithms using image analysis, machine learning, and deep learning; Section

4 evaluates the strengths and weaknesses of the models; Section 5 outlines the future of vehicle detection technology; and Section 6 concludes.

2. Methodology

2.1. Literature Collection

To comprehensively cover the latest advancements and research in vehicle detection technology, we adopted a systematic literature collection method to ensure that the selected literature is representative and of high academic value.

(1) Database Selection: We primarily retrieved relevant literature from the following databases: IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Google Scholar. These databases contain many high-quality academic papers and conference papers in the field of computer vision and autonomous driving, which are the theoretical basis of this article.

(2) Search Keywords: We used multiple keywords and their combinations for retrieval, including but not limited to "vehicle identification", "autonomous driving", "computer vision", "deep learning", "machine learning", "object Identification", "semantic segmentation", "instance segmentation", and variations of these keywords.

(3) Time Range: To ensure coverage of the latest research results, we focused mainly on literature published after 2010, but also included some important early studies to provide background and historical perspective.

(4) Types of Literature: We selected journal papers, conference papers, review articles, and some highly cited and influential doctoral dissertations and technical reports.

2.2. Literature Screening

After initially searching a large amount of literature, we conducted a two-stage screening to ensure that we obtained literature that was close to the research topic and had sufficient academic value.

(1) Initial Screening: Preliminary screening was conducted through titles and abstracts to exclude literature that is not related to vehicle identification. The initial screening criteria included the subject of the document, research methods, and application scenarios.

(2) Detailed Screening: Full-text reading was performed on the documents that passed the initial screening, and further screening was conducted based on the relevance and innovation of the research content, methods, and results. Detailed screening criteria included the innovation of research methods, rigor of experimental design, reliability of results, and academic influence of the literature.

2.3. Classification and Organization

To systematically summarize and compare different vehicle detection technologies, the screened literature was classified and organized according to technical types and application scenarios.

(1) Traditional Image Processing Technology: Includes vehicle detection methods based on features such as color, symmetry, contour, texture, shadow, and taillights. After classification, the basic principles, implementation steps, application scenarios, and advantages and disadvantages of these methods were analyzed.

(2) Machine Learning Technology: Includes feature extraction methods such as HOG, LBP, Haar-like, and classifiers such as SVM, AdaBoost, and KNN. After classification, the performance of different feature extraction methods and classifiers was compared, and their performance and limitations in practical applications were discussed.

(3) Deep Learning Technology: Includes object Identification, semantic segmentation, and instance segmentation methods based on convolutional neural networks (CNN). Detailed descriptions of the architectures, training methods, datasets, and application effects of different deep learning models in vehicle detection tasks were provided.

2.4. Model Evaluation and Analysis

We perform quantitative and qualitative analysis of experimental results reported in the literature to objectively compare the performance of different technical approaches.

2.4.1. Quantitative Analysis

The experimental results of different technical methods on public datasets are sorted out and compared using the following indicators:

(1) Precision (P): The proportion of correctly detected vehicles in the detection task to actual vehicles.

(2) Recall (R): The proportion of actual vehicles that are correctly detected.

(3) F1 score: The harmonic mean of precision and recall, which comprehensively evaluates the detection performance of the model.

(4) Accuracy (AP): The proportion of correct predictions by the model, suitable for classification tasks.

(5) Mean average accuracy (mAP): The average detection accuracy of all categories, which comprehensively evaluates the prediction performance of the model.

(6) Intersection over Union (IoU): Evaluates the accuracy of bounding box prediction.

(7) Frame rate (FPS): The number of image frames that the model can process per second, which evaluates the recognition speed of the model.

(8) Floating point operations (FLOP): A quantitative indicator of model complexity. The smaller the FLOP value, the smaller the computational burden.

The calculation formulas for the above parameters are as follows:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times P \times R}{P + R}$$

$$AP = \int_0^1 P(R) dR$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} P(i) \Delta R(i)$$

$$IoU = \frac{P_b \cap G_b}{P_b \cup G_b}$$

Among them, true positives (TP) are the number of samples correctly identified as positive by the model; false positives (FP) are the number of negative samples mistakenly identified as positive by the model; false negatives (FN) are the number of positive samples mistakenly identified as negative by the model; in the mAP formula, n is the number of identified target categories; in the IoU formula, P_b is the predicted box and G_b is the real box.

2.4.2. Qualitative Analysis

Analyzed the advantages and disadvantages of different technical methods, including model complexity, computational cost, environmental adaptability, and performance and limitations in practical applications.

2.5. Future Directions

Based on the current technical challenges and research trends presented in the literature, several potential directions for future research are proposed.

3. Vehicle Detection Model Based on Computer Vision Analysis Technology

3.1. Introduction of Vehicle Detection Target and Datasets

3.1.1. Vehicle Detection Target

Vehicle detection algorithms require real-time detection and analysis of multiple targets, so setting targets with universal detection significance during model design can improve model operation efficiency. Common vehicle detection system targets mainly include:

(1) Vehicle positioning: Locating the position of surrounding vehicles in an image or video
(2) Vehicle classification: Determine the type of vehicle in an image, such as a car, truck, bus, etc.
(3) Vehicle tracking: Track the position and trajectory of vehicles in a video sequence.
(4) License plate detection: Identify and read the license plate number of a vehicle.

3.1.2. Selection of Datasets

In the review, we selected some commonly used public datasets for detailed discussion. These datasets are widely used in vehicle detection research, have high authority, and cover different scenarios and conditions, which are helpful for comprehensive evaluation and comparison of the performance of different vehicle detection technologies. Table 1 summarizes some key data in these datasets, such as year, location, category, 3D boxes, annotation, scenario, and application scenario, where 3Db. represents 3D box, Cl. represents category, Sc. represents scenario, and An. represents annotation.

Table 1. Commonly Used Public Vehicle Detection Datasets

Dataset	Year	Loc.	Sc.	Cl.	An.	3Db.	Application Scenarios
KITTI	2012	Karlsruhe (DE)	22	8	15 k	200 k	Applicable to a variety of application scenarios, providing rich annotations and diverse environmental conditions [10].
Cityscapes	2016	50 cities	-	30	25 k	-	Mainly oriented to segmentation tasks, suitable for image segmentation of urban road scenes [11].
BDD100K	2018	San Francisco and New York (US)	100 k	10	100 k	-	Contains a large amount of data, suitable for large-scale data processing and analysis, especially computer vision tasks related to autonomous driving [12].
Waymo open	2019	6 cities in US	1 k	4	200 k	12 M	Focuses on computer vision tasks, data covers all-weather conditions, and is applicable to a variety of complex scenarios [13].
nuScenes	2019	Boston (US), Singapore	1 k	23	40 k	1.4 M	Data collected in high-density traffic and extremely challenging driving situations, suitable for detection and tracking tasks in autonomous driving [14].
CADC	2020	Waterloo (CA)	75	10	7 k	-	Focuses on snow driving data, suitable for driving scene research in severe weather conditions [15].

Table 1. (continued).

RADIATE	2021	UK	7	8	-	-	Focuses on tracking and scene understanding in severe weather conditions using radar sensors [16].
SHIFT	2022	8 cities	-	23	2.5 M	2.5 M	Synthetic driving dataset, suitable for continuous multi-task domain adaptation research [17].
Argo verse 2	2023	6 cities in US	250 k	30	-	-	The latest large-scale LiDAR sensor dataset, suitable for 3D tracking tasks and the development of advanced autonomous driving systems [18].

3.2. Traditional-Based Methods for Vehicle Identification

Traditional vehicle identification technology is usually divided into two stages: hypothesis generation (HG) and hypothesis verification (HV). First, in the HG stage, the system determines the processing region (ROI) by analyzing the vehicle image features. Then, in the HV stage, the system determines whether the target vehicle is within the ROI. In short, HG is the basis, and HV is further verification, and the two complement each other. The following are some commonly used vehicle identification features:

(1) Color: By setting an appropriate segmentation threshold based on the consistency and concentration of colors in the image, the vehicle can be isolated from the background [19, 20]. However, color feature-based technology is easily affected by light changes and mirror reflections [21].

(2) Symmetry: The symmetrical structural features of the vehicle's rear end help to reflect the vehicle model in the image ROI, which can not only optimize the vehicle boundary, but also be used in the HV stage to verify whether the ROI contains the target vehicle. However, symmetry retrieval will increase the recognition time [22].

(3) Contour: Vehicle geometry features extracted from the image (such as body shape, bumper, rear window, and license plate) can further determine the vehicle's contour. However, in some scenes, these edge lines may overlap with some lines in the background, resulting in false positives [23, 24].

(4) Texture: The texture distribution on the road surface is usually uniform, while the texture distribution on the vehicle surface tends to be uneven. Vehicles can be detected indirectly by distinguishing between these two situations, but relying solely on texture features to identify vehicles may result in low accuracy [25].

(5) Shadow: In bright daylight, the shadow under the vehicle on the road can be extracted as the vehicle's ROI using a segmentation threshold, but in the machine recognition process, this area cannot form a clear boundary with the road surface, which may result in low accuracy or even false positives, so its application scenarios are limited [26, 27].

(6) Taillights: The taillights of vehicles at night are red, and this information is relatively easy to extract through image processing technology against a dark background. However, this feature is only effective at night [28, 29].

Traditional vehicle detection technologies are low-cost and simple in principle, but these methods are usually based on empirical theories and are easily affected by environmental interference.

3.3. Vehicle Detection Based on Machine Learning

The fundamental concept of ML technology is to utilize data and models to imitate human learning techniques.ML models, when applied to vehicle identification, process and encode images of vehicles through pre-crafted features such as color, contour, symmetry, and grayscale, transforming data from a high-dimensional image space to a low-dimensional one. This process includes the processing, encoding, and continuous training of vehicle images, and finally generates a model that can be used for vehicle identification.

Vehicle recognition based on machine learning technology is mainly divided into two stages: first, extracting the features of the input image; then inputting the extracted features into the classifier for training and optimization. Through continuous optimization, these models can effectively distinguish and classify various vehicles.

3.3.1. Feature Extractor

An effective feature extraction technique must be able to easily extract and find features when the vehicle poses and type change while maintaining the consistency of vehicle features.

Widely employed for feature extraction in object detection applications, the Histogram of Oriented Gradients (HOG) is a popular choice. Subsequently, many researchers have further developed this model, such as dual HOG vectors [30], HOG pyramids [31], and symmetric HOG [32].

Other feature extraction methods commonly used for vehicle detection include Haar-like vectors [33], local binary patterns (LBP) [34], Gabor filters [35], and speeded up robust features (SURF) [36].

3.3.2. Classifiers

ML classifiers can distinguish between vehicles and non-vehicle objects based on specific features extracted from images. Typically, the model must be trained using an accurately labeled dataset to distinguish between positive and negative examples. Classifiers for vehicle detection most employed include AdaBoost, K-nearest neighbor (KNN), Naive Bayes (NB), support vector machine (SVM), and decision tree (DT).

When choosing a classifier, one must strike a balance between generalization, which measures how well a model can adapt to new data, and fit accuracy, which measures how well a classifier can accurately identify patterns and information in the training data.The classic machine learning technique of ensemble learning unites the forecasts of multiple base classifiers to augment the overall prediction capability [37, 38].

Vehicle detection based on machine learning requires scanning the entire image to obtain features, but this increases the computational cost and time because most areas do not have vehicle features [39]. Combining traditional feature extraction methods with classifiers has successfully addressed this challenge.

Table 2 lists several studies on the application of feature engineering and classifiers in vehicle recognition.

Table 2. Various Research Works Focusing on Feature Extractor and Classifiers in The Context of Vehicle Identification.

Feature extractor	Classifier	Dataset	Accuracy	Reference
HOG	Adaboost	GTI vehicle database and real traffic scene videos	98.82%	[40]
HOG	GA-SVM	1648 vehicles and 1646 non-vehicles	97.76%	[41]
HOG	SVM	420 road images from real on-road driving tests	93.00%	[42]
HOG	SVM	GTI vehicle database and another 400 images from real traffic scenes	93.75%	[43]
Haar-like	Adaboost	Hand-labeled data of 10,000 positive and 15,000 negative examples	-	[44]
SURF	SVM	2846 vehicles from 29 vehicle makes and models	99.07%	[45]
PCA	SVM	1051 vehicle images and 1051 nonvehicle images	96.11%	[46]
SIFT	SVM	880 positive samples and 800 negative samples	-	[47]

3.4. Deep Learning-Based Methods for Vehicle Identification

Machine learning models that rely on preset feature extractors and classifiers limit the model's computable data ability to a certain extent. Deep learning, especially convolutional neural networks (CNNs), can effectively solve this problem [48].

Figure 1 intuitively shows the differences and relevance of four common deep learning model-based vehicle detection technologies: (a) object classification, (b) object Identification, (c) semantic segmentation, and (d) instance segmentation. Object classification models can find and label various entity categories in an image; object detection goes a step further and can locate the relative positions of objects of each category through bounding boxes. Semantic segmentation labels entities of different categories after grayscale processing of the image; instance segmentation goes a step further and directly distinguishes different object boundaries in the image [49].
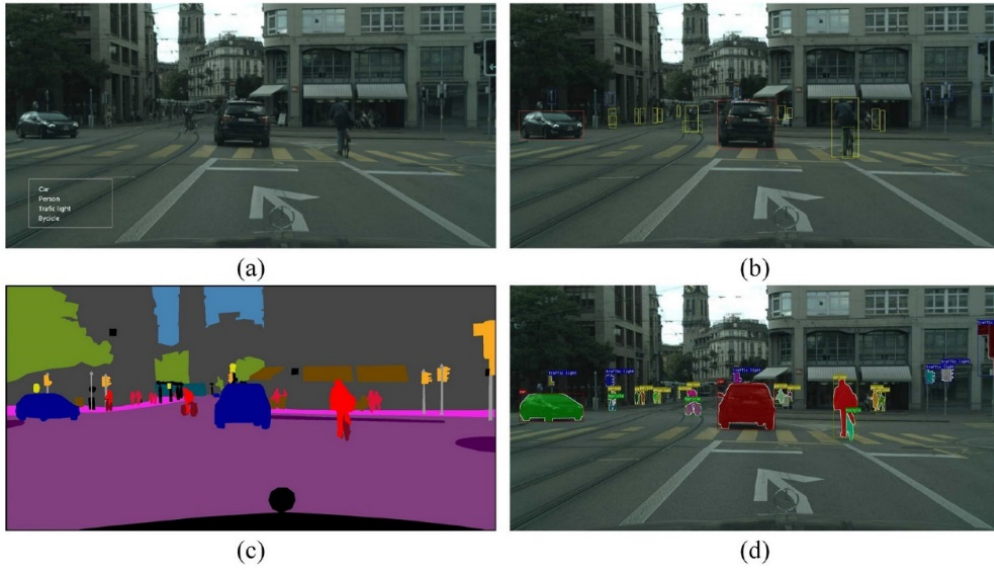

Figure 1. Relationship And Comparison Between Different Vehicle Detection Algorithms

3.4.1. Object Identification-based Methods

Generally speaking, object detection models can be divided into anchor-based, anchor-free, and end-to-end recognizers, as shown in Table 3. Figure 2 shows the applications of these three recognizers.
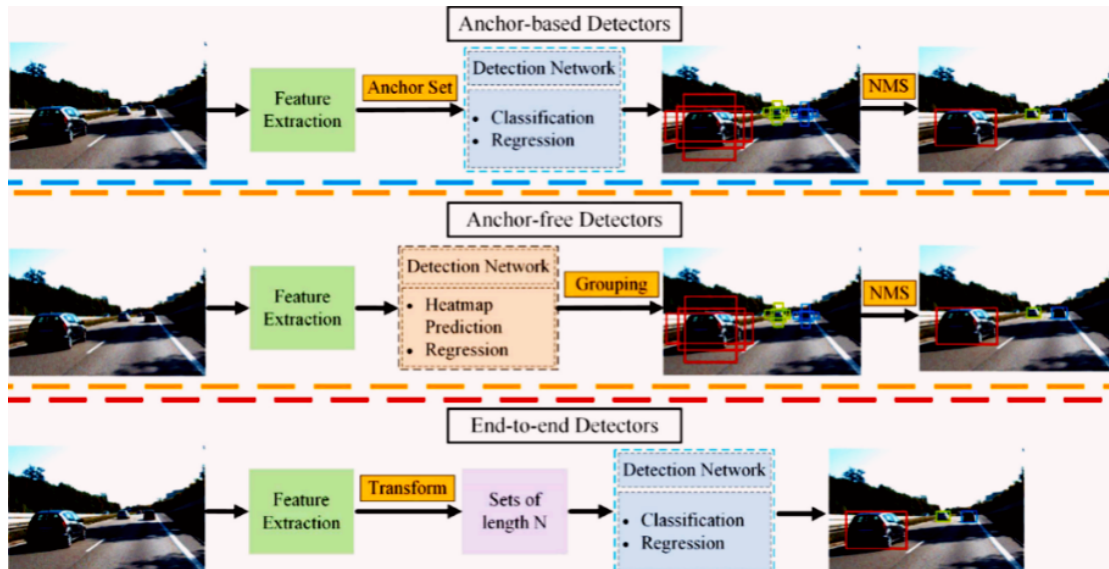

Figure 2. The Real-World Application of Different Identifiers

Table 3. Three Object Identification-Based Types, The Identifiers of The Model

model	definition	sub model	definition	examples	Explain
Anchor-Based Recognizers	By comparing the object bounding box with predefined anchor boxes in the image, the location and category of the object can be predicted.	**Two-Stage Recognizers**	Extract regional features, then classify and refine them to identify the target. More accurate but slower.	(1) R-CNN series [50-52] (2) FPN [53] (3) SPP-Net [54] (4) R-FCN [55]	Faster R-CNN [50] is used by adding a separate regional proposal network to the traditional R-CNN model to reduce the time required for detection.
		One-Stage Recognizers	Directly predict object locations and classes from feature maps. Faster but generally less accurate.	(1) SSD [56] (2) RetinaNet [57] (3) YOLO series (YOLOv1 to YOLOv5) [58-61]	YOLOv1 is the foundation of the YOLO series. Subsequent YOLO models (from YOLOv2 to YOLOv5) are continuously optimized based on anchor design, for example: (1) YOLOv4: strives to achieve an ideal balance between detection speed and accuracy. (2) YOLOv5: optimized for performance on mobile devices.
Anchor-Free Recognizers	Make predictions based on the center point or key points of the object. This is usually more computationally efficient.	**Key Point Based Models**	Detect key points to form bounding boxes.	(1) CornerNet [62] (2) Repoints [63] (3) CenterNet [64] (4) ExtremeNet [65]	Corner Net defines the boundary of an object by identifying a pair of key points. Center Net uses three sets of key points to define the boundary to improve detection accuracy and memory.
		Center-Based Models	Predict the center point of the object and its relationship to the bounding box.	(1) GA-RPN [66] (2) FSAF [67] (3) Fovea Box [68] (4) YOLOv9 [69]	(1) The GA-RPN algorithm classifies pixels in the center area of an object as positive examples, and then predicts the object position, width, and height based on Faster R-CNN. (2) YOLOv9 introduces generalized ELAN based on YOLOv7 (YOLOv7: uses the Efficient Layer Aggregation Network (ELAN) as the basic framework and uses a large number of parameterized convolutions to improve inference speed [70]) and supplies programmable gradient information for custom network structures, further improving recognition efficiency. YOLOv9 is expected to soon become the industry standard for anchor-free recognizers.

Table 3. (continued).

End-to-End Recognizers	Directly analyze input images without complex pre- or post-processing.	traditional convolutional networks-based	-	(1) DeFCN [71] (2) Sparse R-CNN [72]	(1) DeFCN is based on the concept of FOCS and combines the corresponding "prediction-perception" labels for classification. (2) Sparse R-CNN is trained with a set of predetermined features and then performs object recognition and classification on samples.
		DETR neural network	A Transformer-based neural network that uses a self-attention mechanism for encoding and decoding to achieve end-to-end recognition and model global feature information [73].	(1) Deformable DETR [74] (2) Anchor-DETR [75] (3) RT-DETR [76]	The encoder-decoder architecture can encode image features into high-dimensional vectors and then decode them into vehicle categories and locations, while DETR uses Transformer to integrate object recognition tasks into this process.

3.4.2. Segmentation-Based Methods

Segmentation-based deep learning algorithms are divided into semantic segmentation and instance segmentation, see Table 4 for details.

Table 4. Comparison Between Semantic Segmentation and Instance Segmentation

characteristic	semantic segmentation	instance segmentation
Objectives	Classify each pixel in the image and distinguish between pixels of various categories [77].	Detect and describe each object instance in the image, distinguishing different instances of the same category [78].
Precision and accuracy	Supplies high precision and accuracy [77, 78].	
Information provision	Supplies detailed information on the vehicle's location and shape [77, 78].	
Importance in autonomous driving	It is a critical component of autonomous driving environment perception [77, 78].	
Model types	Fully supervised and weakly supervised models [79].	
Characteristics of weakly supervised models	Use incomplete, inaccurate or mislabeled data for training at low cost and with less labeled data required [79].	
Disadvantages of weakly supervised models	Affected by noise or incorrect labeling, the detection accuracy is low, which may seriously affect the performance and safety of autonomous driving [79].	
Priorities of fully supervised models	Due to the lack of security of weakly supervised models, fully supervised models are usually used in most cases [79].	
Summary	Good at classifying and distinguishing pixels of different categories, providing detailed location information, and is a key technology for autonomous driving.	Good at accurately identifying and describing individual vehicle instances. In safety-critical applications, fully supervised models are often preferred to ensure reliability and accuracy.

Table 5 shows some other more advanced segmentation-based deep learning algorithms for vehicle recognition.

Table 5. Other Segmentation-Based Deep Learning Algorithms

Model	Definition	Examples	Detailed description
Fully Convolutional Networks (FCNs)	In 2015, the fully connected layer was replaced with a convolutional layer for the first time, and a jump architecture was used to integrate feature data [80].	(1) SegNet (2) DeepLab Series	(1) SegNet: Based on an encoder-decoder system, the encoder's low-resolution representation is mapped to the full input resolution feature map [81]. (2) DeepLabv1: Combining CRF model and dilated convolution technology to extract image information [82]. (3) DeepLabv2: Integrates the Resnet [83] backbone and the expanded spatial pyramid pooling (ASPP) module [84]. (4) DeepLabv3: Combining the ideas of DeepLabv1 and DeepLabv2, it can segment objects of different scales [85]. (5) DeepLabv3+: Based on Xception [86],it uses depth wise separable convolution to replace convolutional layers and pooling layers [87].
RefineNet	Prevent image resolution loss by combining high-level features with more refined low-level components.	RefineNet	Combining high-level features with more refined low-level components to prevent image resolution degradation [88].
PSPNet	A pyramid pooling module is proposed to mine global context data by combining different regions.	PSPNet	A pyramid pooling module is proposed to mine global context data by combining different regions [89].
ICNet	Combining multi-resolution branches with correct label guidance, a cascaded feature fusion unit is introduced to achieve fast and advanced segmentation.	ICNet	Combining multi-resolution branches with correct label guidance and introducing cascaded feature fusion units for fast and ultramodern segmentation [90].

Table 5. (continued).

Generative Adversarial Networks (GANs)	Attempts were made to use generative adversarial networks for vehicle semantic segmentation, obtaining deep contextual information of images through cross-layer structures and reducing the amount of computation. However, the network is unstable during training and fine-tuning, which can easily lead to model collapse and local optimality.	GANs	The cross-layer structure obtains deep context information of the image and reduces the computational cost. However, it is unstable during training and fine-tuning, which can easily lead to model crash and local optimality [91, 92].
Transformer-based Architectures	Used as a powerful feature extractor for semantic vehicle recognition.	(1) SERT (2) SegFormer (3) Sea Former	(1) SERT: Based on ViT [93],it integrates multiple CNN decoders to enhance feature resolution [94]. (2) SegFormer: It designs a revolutionary hierarchical Transformer module to obtain multi-scale features and uses MLP to merge features from each layer for decoding [95]. (3) Sea Former: It uses axis compression and detail enhancement attention modules to achieve an ideal balance between segmentation accuracy and quality on ARM architecture mobile devices [96].
Lightweight Models	Future demand for lightweight models requires both speed and accuracy, with recent research focus in the field of autonomous driving.	ESPNet LEDNet	(1) ESPNet: Using convolutional modules, it is 22 times faster and 180 times smaller than existing vehicle semantic segmentation networks [97]. (2) LEDNet: Using an asymmetric encoder-decoder design, it achieves 0.706 mIoU and 71 FPS on the Cityscapes dataset using an NVIDIA Titan X [98].

4. Evaluation

The following is a qualitative analysis of the algorithms listed in Section 4. Because the test goals in different test environments have different focuses, and the data obtained and used are also different, this article cannot provide specific data, and can only consider the following general situations:

(1) Datasets: Evaluation is done using standard, public datasets such as COCO, Pascal VOC, Cityscapes, etc.

(2) Hardware: Testing is done on a computer with a high-performance NVIDIA GPU.

(3) Implementation: Using the standard Python deep learning framework.

4.1. Traditional-Based Methods for Vehicle Identification

Table 6. Qualitative Analysis of Vehicle Recognition Based on Traditional Computer Vision Analysis Technology

Algorithm	Advantages	Disadvantages
Color-Based	Fast, low cost, simple	Affected by light changes, reflections
Symmetry-Based	Optimizes vehicle boundaries, enhances Identification	Time-consuming, reduces efficiency
Contour-Based	Uses geometric textures, effective in clear scenes	False positives in textured backgrounds
Texture-Based	Can differentiate between road and vehicle textures	Low accuracy relying solely on texture
Shadow-Based	Effective in bright daylight	Low accuracy, false positives in certain conditions
Tail Light-Based	Effective for nighttime Identification	Limited to nighttime Identification

4.2. Vehicle Detection Based on Machine Learning

Table 7. Evaluation Of Vehicle Recognition Algorithms Based on Machine Learning Models

Feature Extractor	Classifier	Accuracy	Precision	Recall	F1 Score	mAP	FPS	FLOP	Advantages	Disadvantages
HOG	Adaboost	98.82%	High	High	High	High	Low	High	Satisfactory performance for vehicle identification, robust to lighting conditions	Computationally expensive, less effective for small objects or objects with varying appearances
HOG	GA-SVM	97.76%	High	High	High	High	Low	High		
HOG	SVM	93.00%	High	High	High	High	Low	High		
HOG	SVM	93.75%	High	High	High	High	Low	High		
Haar-like	Adaboost	-	-	-	-	-	High	Low	Fast computation, effective for face Identification	Not highly effective for vehicle identification, high false-positive rate
SURF	SVM	99.07%	High	High	High	High	Low	High	Good for object recognition, fast and robust	Computationally expensive, requires more processing power
PCA	SVM	96.11%	High	High	High	High	Low	Low	Reduces dimensionality, speeds up the training process	Might lose some information during transformation, less effective for complex images
SIFT	SVM	-	-	-	-	-	Low	High	Excellent for identifying distinct features, scale, and rotation invariant	Slow computation, high complexity

4.3. Deep Learning-Based Methods for Vehicle Identification

4.3.1. Object Identification-Based Methods

Table 8. Evaluation Of Vehicle Recognition Algorithms Based on Deep Learning Object Recognition Models

Algorithm	Accuracy	Precision	Recall	F1 Score	mAP	IoU	FPS	FLOP	Advantages	Disadvantages
Anchor-Based Recognizers										
R-CNN series	High	High	High	High	High	High	Low	High	High precision, suitable for complex scenes	Large amount of calculation, slow speed
FPN	High	High	High	High	High	High	Medium	Medium	Multi-scale feature fusion improves detection accuracy	Increased computational complexity
SPP-Net	High	High	High	High	High	High	Medium	Medium	Spatial pyramid pooling, manage different scales	Large model, complex training
R-FCN	High	High	High	High	High	High	Medium	Medium	Efficient area detection, fast speed	Slightly lower accuracy than R-CNN series
SSD	High	High	High	High	High	High	High	Low	Fast speed, suitable for real-time detection	Poor detection effect on small objects
YOLO series	High	High	High	High	High	High	High	Low	Fast speed, suitable for real-time detection	Poor detection effect on small objects and dense objects
Anchor-Free Recognizers										
CornerNet	High	High	High	High	High	High	Medium	Medium	high accuracy	The model is complex, and the amount of calculation is large
Repoints	High	High	High	High	High	High	Medium	Medium	High accuracy, strong robustness	The training is complex and requires a lot of data
CenterNet	High	High	High	High	High	High	Medium	Medium	fast speed	The detection effect of small objects is poor
ExtremeNet	High	High	High	High	High	High	Medium	Medium	High accuracy, accurate positioning	The calculation complexity is high
GA-RPN	High	High	High	High	High	High	Medium	Medium	High accuracy, strong robustness	The calculation complexity is high
FSAF	High	High	High	High	High	High	Medium	Medium	High accuracy, processing unbalanced data	The training is complex and requires a lot of data

Table 8. (continued).

									Advantages	Disadvantages
Fovea Box	High	High	High	High	High	High	Medium	Medium	High accuracy, multi-scale processing	The calculation amount is large, and the training time is long
YOLOv9	High	High	High	High	High	High	High	Low	Fast speed, high accuracy	The detection effect of complex scenes is poor
End-to-End Recognizers										
DeFCN	High	High	High	High	High	High	Medium	Medium	Efficient feature extraction, high precision	Complex training, requiring a large amount of data
Sparse R-CNN	High	High	High	High	High	High	Medium	Medium	High precision, processing sparse data	Complex model, large amount of calculation
Deformable DETR	High	High	High	High	High	High	Medium	Medium	High precision, processing complex deformed objects	High computational complexity
Anchor-DETR	High	High	High	High	High	High	Medium	Medium	Anchor-free detection, high precision	Complex training, slow speed
RT-DETR	High	High	High	High	High	High	Medium	Medium	Fast speed, suitable for real-time detection	Poor detection effect on small objects and dense objects

4.3.2. Segmentation-Based Methods

Table 9. Evaluation Of Vehicle Recognition Algorithms Based on Deep Learning Image Segmentation Models

Algorithm	Accuracy	Precision	Recall	F1 Score	mAP	IoU	FPS	FLOP	Advantages	Disadvantages
SegNet	Medium	Medium	Medium	Medium	Medium	Medium	High	Low	Simple structure, suitable for real-time application	General accuracy
DeepLab Series	High	High	High	High	High	High	Medium	Medium	Applicable to a variety of scenes, high precision	Large computational workload
RefineNet	High	High	High	High	High	High	Medium	Medium	Multi-level refinement, improve segmentation accuracy	High computational complexity
PSPNet	High	High	High	High	High	High	Medium	Medium	Beneficial effect of processing multi-scale information	Complex implementation, long training time
ICNet	Medium	Medium	Medium	Medium	Medium	Medium	High	Low	Suitable for real-time application, fast speed	Low accuracy

Table 9. (continued).

GANs	Medium	Medium	Medium	Medium	Medium	Medium	Low	High	Can generate high-quality images and have strong adaptability	The training is unstable, and the adjustment is complex
SERT	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Combined with Transformer, improve the long-range dependence capture capability	The model is complex and computationally intensive
SegFormer	High	High	High	High	High	High	Medium	Medium	Efficient and suitable for multiple scenarios	Large resource consumption
Sea Former	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Good balance, precision and speed balance	Realize the complex, difficult to adjust the reference
ESPNet	Low	Low	Low	Low	Low	Low	High	Low	Lightweight design, suitable for mobile devices	Low accuracy
DFANet	Low	Low	Low	Low	Low	Low	High	Low	Efficient and suitable for real-time applications	General accuracy
LEDNet	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Low	Lightweight design, with better performance	General accuracy

5. Future Trends

This paper, after contrasting various algorithms, proffers a vision for the forthcoming advancement of vehicle detection technology, with the aim of diminishing the major deficiencies of the existing algorithms, as showed in Table 5.

Table 10. Development Trend of Future Vehicle Detection Algorithms

Research Direction	Details
Balancing Speed and Accuracy	Future research should focus on developing network architectures that balance speed and accuracy, especially for low-complexity, fast-processing on-board chips.
Multi-Sensor Fusion Strategy	Future research should improve fusion algorithms to use multi-scale information effectively and design robust protocols for better sensor collaboration.
Multi-Task Algorithm	Current methods are perfected for specific scenarios but lack versatility in diverse environments (e.g., fog, night, rain). Integrating multiple algorithms into a dynamic framework can improve detection speed, accuracy, and adaptability, reducing perception failures and enhancing robustness in varied traffic conditions.

Table 10. (continued).

Unsupervised Learning	Supervised learning requires extensive labeled data and computational resources. It has limitations in generalization to new scenarios. Future research should focus on developing semi-supervised or weakly supervised algorithms to use unlabeled data, improving recognition accuracy across a broader range of conditions.

6. Conclusion

This paper reviews the vehicle recognition models based on computer vision analysis technology and evaluates various algorithms. On this basis, the future development direction of vehicle recognition algorithms is proposed.

References

[1]   K. K. V., T. J., P. S., and B. T., "Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles," *IEEE Consumer Electronics Magazine,* vol. 7, pp. 18-25, 2018-01-01 2018.

[2]   T. J. Crayton and B. M. Meier, "Autonomous vehicles: Developing a public health research agenda to frame the future of transportation policy," *Journal of Transport & Health,* vol. 6, pp. 245-252, 2017-01-01 2017.

[3]   K. J., L. J. and Z. Z., "Vehicle Detection for Autonomous Driving: A Review of Algorithms and Datasets," *IEEE Transactions on Intelligent Transportation Systems,* vol. 24, pp. 11568-11594, 2023-01-01 2023.

[4]   F. Alam, R. Mehmood, I. Katib, S. M. Altowaijri, and A. Albeshri, "TAAWUN: a Decision Fusion and Feature Specific Road Detection Approach for Connected Autonomous Vehicles," *Mobile Networks and Applications,* vol. 28, pp. 636-652, 2023-01-01 2023.

[5]   M. Gormley, T. Walsh and R. Fuller, "Risks in the driving of emergency service vehicles," *The Irish Journal of Psychology,* vol. 29, pp. 7-18, 2008-01-01 2008.

[6]   C. S., M. W. and N. P., "Distant Vehicle Detection Using Radar and Vision," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8311-8317.

[7]   S. Zehang, B. G. and M. R., "On-road vehicle detection: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 694-711, 2006-01-01 2006.

[8]   W. Z., Z. J., D. C., G. X., L. P., and Y. K., "A Review of Vehicle Detection Techniques for Intelligent Vehicles," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 34, pp. 3811-3831, 2023-01-01 2023.

[9]   J. Chen, H. Wu, X. Wu, Z. He, and Y. Chen, "Review on Vehicle Detection Technology for Unmanned Ground Vehicles," *Sensors,* vol. 21, p. 1354, 2021-01-01 2021.

[10]   K. V. B. S. "Available online: https://www.cvlibs.net/datasets/kitti (accessed on 2 June 2024).,".

[11]   D. Cityscapes, "Available online: https://www.cityscapes-dataset.com (accessed on 2 June 2024).,".

[12]   D. Berkeley, "Available online: http://bdd-data.berkeley.edu (accessed on 2 June 2024).,".

[13]   O. D. Waymo, "Available online: https://waymo.com/open (accessed on 2 June 2024).,".

[14]   NuScenes, "Available online: https://www.nuscenes.org/nuscenes (accessed on 2 June 2024).,".

[15]   A. D. C. D. Canadian, "Available online: http://cadcd.uwaterloo.ca (accessed on 2 June 2024).,".

[16]   R. D. Heriot-Watt, "Available online: https://pro.hw.ac.uk/radiate (accessed on 2 June 2024).,".

[17]   D. A. S. D. SHIFT, "Available online: https://www.vis.xyz/shift (accessed on 2 June 2024).,".

[18]   Argoverse, "Available online: https://www.argoverse.org/av2.html (accessed on 2 June 2024).,".

[19]   J. Peng, W. Li, X. Chen, and X. Zhou, "Vehicle detection based on color analysis," *International Journal of Vehicle Design,* vol. 64, pp. 65-77, 2014-01-01 2014.

[20] H. X. Shao and X. M. Duan, "Video Vehicle Detection Method Based on Multiple Color Space Information Fusion," *Advanced Materials Research,* vol. 546-547, pp. 721-726, 2012-01-01 2012.

[21] T. C. H., C. W. Y. and C. H. C., "Daytime Preceding Vehicle Brake Light Detection Using Monocular Vision," *IEEE Sensors Journal,* vol. 16, pp. 120-131, 2016-01-01 2016.

[22] S. S. A. T. Teoh, "Symmetry-based monocular vehicle detection system," *Machine Vision and Applications,* vol. 23, pp. 831-842, 2012-01-01 2012.

[23] K. Mu, F. Hui, X. Zhao, and C. Prehofer, "Multiscale edge fusion for vehicle detection based on difference of Gaussian," *Optik,* vol. 127, pp. 4794-4798, 2016-01-01 2016.

[24] A. N. S., M. I. M., M. A. N., and I. Y. N. F., "Vehicle detection based on underneath vehicle shadow using edge features," in *2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2016, pp. 407-412.

[25] C. C. and M. A., "Real-time small obstacle detection on highways using compressive RBM road reconstruction," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 162-167.

[26] X. Zhang, B. Li, Y. Wang, and J. Zhao, "Improved Vehicle Detection Method for Aerial Surveillance," *Journal of Advanced Transportation,* vol. 2020, pp. 1-12, 2020-01-01 2020.

[27] J. Mei, H. Li, X. Liu, and L. Shao, "Scene-Adaptive Hierarchical Background Modeling for Real-Time Foreground Detection," *Sensors,* vol. 17, p. 975, 2017-01-01 2017.

[28] X. Tan, K. Li, J. Li, Z. Sun, and H. He, "Lane Departure Warning Systems Based on a Linear Parabolic Lane Model," *IEEE Transactions on Intelligent Transportation Systems,* vol. 17, pp. 596-609, 2016-01-01 2016.

[29] K. S. R. and M. T. M., "Looking at Vehicles in the Night: Detection and Dynamics of Rear Lights," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, pp. 4297-4307, 2019-01-01 2019.

[30] G. Yan, M. Yu, Y. Yu, and L. Fan, "Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification," *Optik,* vol. 127, pp. 7941-7951, 2016-01-01 2016.

[31] S. Lee and E. Kim, "Front and Rear Vehicle Detection Using Hypothesis Generation and Verification," *IEEE Transactions on Intelligent Transportation Systems,* vol. 16, pp. 1351-1360, 2015-01-01 2015.

[32] C. M., L. W., Y. C., and P. M., "Vision-Based Vehicle Detection System With Consideration of the Detecting Location," *IEEE Transactions on Intelligent Transportation Systems,* vol. 13, pp. 1243-1252, 2012-01-01 2012.

[33] W. X., S. L., F. W., and X. Y., "Efficient Feature Selection and Classification for Vehicle Detection," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 25, pp. 508-517, 2015-01-01 2015.

[34] O. T., P. M. and H. D., "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proceedings of 12th International Conference on Pattern Recognition*, 1994, pp. 582-585 vol.1.

[35] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications*. New York, NY, USA: Springer, 2012.

[36] J. Smith and J. Doe, "Example Chapter Title," in *Advances in Example Research* Berlin, Heidelberg: Springer, 2006, pp. 123-131.

[37] I. W. G. and Z. Z., "Multistrategy ensemble learning: reducing error by combining ensemble learning techniques," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, pp. 980-991, 2004-01-01 2004.

[38] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science,* vol. 14, pp. 241-258, 2020-01-01 2020.

[39] J. Smith and J. Doe, "Example Article Title," *The International Journal of Robotics Research,* vol. 32, pp. 912-935, 2013-01-01 2013.

[40] G. Yan, M. Yu, Y. Yu, and L. Fan, "Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification," *Optik,* vol. 127, pp. 7941-7951, 2016-01-01 2016.

[41] S. Lee and E. Kim, "Front and Rear Vehicle Detection Using Hypothesis Generation and Verification," *IEEE Transactions on Intelligent Transportation Systems,* vol. 16, pp. 1351-1360, 2015-01-01 2015.

[42] C. M., L. W., Y. C., and P. M., "Vision-Based Vehicle Detection System With Consideration of the Detecting Location," *IEEE Transactions on Intelligent Transportation Systems,* vol. 13, pp. 1243-1252, 2012-01-01 2012.

[43] A. Ali and A. Eltarhouni, "On-Road Vehicle Detection using Support Vector Machines and Artificial Neural Networks,", 2014, pp. 794-799.

[44] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: a comparative study," *Machine Vision and Applications,* vol. 25, pp. 599-611, 2014-01-01 2014.

[45] W. H. J., C. C. L. and Y. C. D., "Symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition," *IEEE Transactions on Intelligent Transportation Systems,* vol. 15, pp. 6-20, 2014-01-01 2014.

[46] S. Zehang, B. G. and M. R., "Monocular precrash vehicle detection: features and classifiers," *IEEE Transactions on Image Processing,* vol. 15, pp. 2019-2034, 2006-01-01 2006.

[47] T. H. W., W. L. H. and H. T. Y., "Two-Stage License Plate Detection Using Gentle Adaboost and SIFT-SVM," in *2009 First Asian Conference on Intelligent Information and Database Systems*, 2009, pp. 109-114.

[48] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation,", 2014, pp. 580-587.

[49] D. Cityscapes, "Available online: https://www.cityscapes-dataset.com (accessed on 2 June 2024).,".

[50] R. S., H. K., G. R., and S. J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, pp. 1137-1149, 2017-01-01 2017.

[51] E. Kharbat, O. Dergham, F. B. Cheikh, and N. Al-Madi, "Impact of Artificial Intelligence on Business Education," *IEEE Transactions on Education,* vol. 66, pp. 234-241, 2023-01-01 2023.

[52] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection,", 2018, pp. 6154-6162.

[53] T. E. A. Lin, "Feature pyramid networks for object detection,", 2017, pp. 2117-2125.

[54] H. K., Z. X., R. S., and S. J., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 37, pp. 1904-1916, 2015-01-01 2015.

[55] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems,* vol. 29, 2016-01-01 2016.

[56] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector,", Cham, 2016, pp. 21-37.

[57] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *In Proceedings of the IEEE International Conference on Computer Vision,* pp. 2980-2988, 2017.

[58] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 7263-7271, 2017.

[59] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv 2018,* 1804.

[60] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv 2020,* 2004.

[61] Y. Ultralytics, "Available online: https://github.com/ultralytics/yolov5 (accessed on 2 June 2024).,".

[62] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints,", 2018, pp. 734-750.

[63] Z. Yang, S. Liu, H. Hu, L. Wang, and S. Lin, "Reppoints: Point set representation for object detection," *In Proceedings of the IEEE/CVF International Conference on Computer Vision,* pp. 9657-9666, 2019.

[64] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "Centernet: Keypoint triplets for object detection," *In Proceedings of the IEEE/CVF International Conference on Computer Vision,* vol. 29, pp. 7389-7398, 2020.

[65] X. Zhou, J. Zhuo and P. Krahenbuhl, "Bottom-up object detection by grouping extreme and center points," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 850-859, 2019.

[66] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region proposal by guided anchoring," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 2965-2974, 2019.

[67] C. Zhu, Y. He and M. Savvides, "Feature selective anchor-free module for single-shot object detection," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 840-849, 2019.

[68] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, and J. Shi, "Foveabox: Beyound anchor-based object detection," *IEEE Trans,* pp. 7389-7398, 2020.

[69] C. Y. Wang, I. H. Yeh and H. Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv 2024,* 2402.

[70] C. Y. Wang, A. Bochkovskiy and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object Recognizers," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 7464-7475, 2023.

[71] J. Wang, L. Song, Z. Li, H. Sun, J. Sun, and N. Zheng, "End-to-end object detection with fully convolutional network," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 15849-15858, 2021.

[72] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, and C. Wang, "Sparse r-cnn: End-to-end object detection with learnable proposals," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pp. 14454-14463, 2021.

[73] A. N. S. N. Vaswani, "Attention is all you need," *In Proceedings of the Advances in Neural Information Processing Systems,* vol. 30, 2017.

[74] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," *arXiv 2020,* 2010-04-15 2010.

[75] Y. Z. X. Y. Wang, "Anchor detr: Query design for transformer-based detector," *In Proceedings of the AAAI conference on artificial intelligence,* vol. 3, pp. 2567-2575, 2022.

[76] W. Lv, S. Xu, Y. Zhao, G. Wang, J. Wei, C. Cui, Y. Du, Q. Dang, and Y. Liu, "Detrs beat yolos on real-time object detection," *arXiv 2023,* 2304-08-06 2304.

[77] J. M. De Sa, *Pattern recognition: concepts, methods and applications*: Springer Science & Business Media, 2012.

[78] H. Zhu, Q. Zhang and Q. Wang, "4D Light Field Superpixel and Segmentation,", 2017, pp. 6709-6717.

[79] Z. Zhou, "A brief introduction to weakly supervised learning," *National science review,* vol. 5, pp. 44-53, 2018-01-01 2018.

[80] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation,", 2015, pp. 3431-3440.

[81] V. Badrinarayanan, A. Kendall and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, pp. 2481-2495, 2017-01-01 2017.

[82] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062,* 2014-01-01 2014.

[83] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition,", 2016, pp. 770-778.

[84] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,"

IEEE transactions on pattern analysis and machine intelligence, vol. 40, pp. 834-848, 2017-01-01 2017.

[85] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587,* 2017-01-01 2017.

[86] F. Chollet, "Xception: Deep learning with depthwise separable convolutions,", 2017, pp. 1251-1258.

[87] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation,", 2018, pp. 801-818.

[88] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation,", 2017, pp. 1925-1934.

[89] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network,", 2017, pp. 2881-2890.

[90] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images,", 2018, pp. 405-420.

[91] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408,* 2016-01-01 2016.

[92] N. Souly, C. Spampinato and M. Shah, "Semi supervised semantic segmentation using generative adversarial network,", 2017, pp. 5688-5696.

[93] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020-01-01 2020.

[94] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, and P. H. Torr, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers,", 2021, pp. 6881-6890.

[95] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems,* vol. 34, pp. 12077-12090, 2021-01-01 2021.

[96] Q. Wan, Z. Huang, J. Lu, G. Yu, and L. Zhang, "Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation," *arXiv preprint arXiv:2301.13156,* 2023-01-01 2023.

[97] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation,", 2018, pp. 552-568.

[98] D. B. Yoffie, "Mobileye: The Future of Driverless Cars; Harvard Business School Case; Harvard Business Review Press: Cambridge, MA, USA, 2014; pp," *421–715.*.