**Literature Survey:**

* **Datasets relevant to application:**

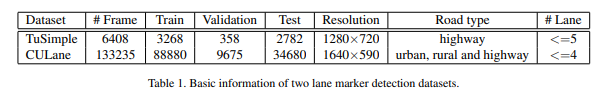
[**TuSimple Dataset:**](https://paperswithcode.com/dataset/tusimple)

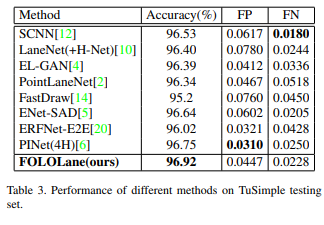
The TuSimple dataset consists of 6,408 road images on US highways. The resolution of the image is 1280×720. The dataset is composed of 3,626 for training, 358 for validation, and 2,782 for testing called the TuSimple test set of which the images are under different weather conditions.

Related Paper: [Focus on Local: Detecting Lane Marker from Bottom Up via Key Point](https://arxiv.org/abs/2105.13680v1)

Methodologies Used:

* Network for local geometry
  + Key point estimation
  + Local geometry construction
  + Network architecture
* Decoder for global geometry
  + Greedy decoder

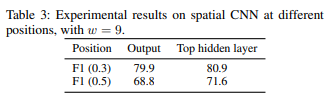
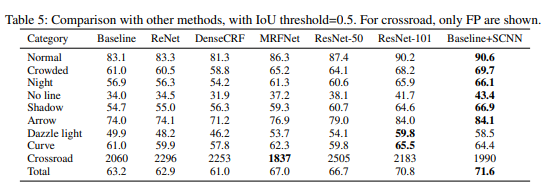




[**CULane Dataset**](https://xingangpan.github.io/projects/CULane.html)**:**

CULane is a large-scale challenging dataset for academic research on traffic lane detection. It is collected by cameras mounted on six different vehicles driven by different drivers in Beijing. More than 55 hours of videos were collected, and 133,235 frames were extracted. The dataset is divided into 88880 images for the training set, 9675 for the validation set, and 34680 for the test set. The test set is divided into normal and 8 challenging categories.

Related paper: [Spatial As Deep: Spatial CNN for Traffic Scene Understanding](https://arxiv.org/abs/1712.06080)

Experimental Results:

[**tvtLANE Dataset:**](https://drive.google.com/drive/folders/1MI5gMDspzuV44lfwzpK6PX0vKuOHUbb_?usp=sharing)

This dataset contains 19383 image sequences for lane detection, and 39460 frames of them are labeled. These images were divided into two parts, a training dataset contains 9548 labeled images and augmented by four times, and a test dataset has 1268 labeled images. The size of images in this dataset is 128\*256.

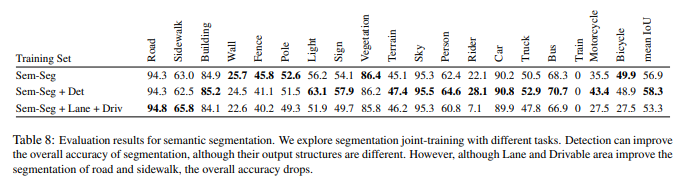
[**BDD100K**](https://bair.berkeley.edu/blog/2018/05/30/bdd/)**:**

BDD100K, the largest driving video dataset with 100K videos and 10 tasks to evaluate the exciting progress of image recognition algorithms on autonomous driving. The dataset possesses geographic, environmental, and weather diversity, which is useful for training models that are less likely to be surprised by new conditions. The dataset consists of 100,000 videos. Each video is about 40 seconds long, 720p, and 30 fps. The videos and their trajectories can be useful for imitation learning of driving policies, as in CVPR 2017 paper.

Related paper: [BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning](https://arxiv.org/abs/1805.04687)

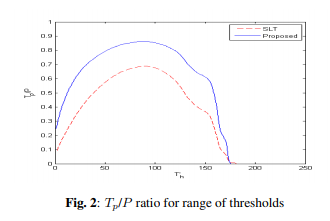
Methodologies Used:

* Image Tagging
* Object Detection
* Lane Marking
* Drivable Area
* Semantic Instance Segmentation
* Multiple Object Tracking
* Multiple Object Tracking and Segmentation
* Imitation Learning

Experimental Results:

* **Feature Extraction:**

[**Robust lane detection & tracking based on novel feature extraction and lane categorization**](https://ieeexplore.ieee.org/abstract/document/6855185)**(IEEE):**

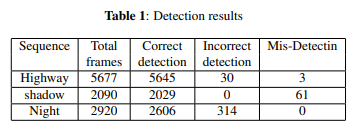
They introduce a robust lane detection and tracking algorithm to cope with complex scenarios and to decrease the effect of thresholds. For lane feature extraction, an extension to the symmetrical local threshold (SLT) is proposed to improve the feature map and obtain orientation information. Then, while creating a Hough accumulator, obtained orientation information is used to decrease computational complexity (≈ 60 times) and acquire a clearer accumulator. The left and right lanes are categorized by applying a mask on the Hough accumulator, which leads to low computational complexity and reduced sensitivity to thresholding.

Dataset Used: RoMa Datasets

Methodologies Used:

* Symmetrical Local Threshold
* Extension of Symmetrical Local Threshold
* Distance Transform
* Hough Transform
* Optimized Lane Categorization

Experimental Results:

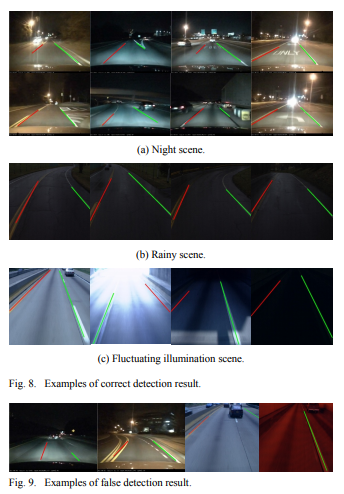


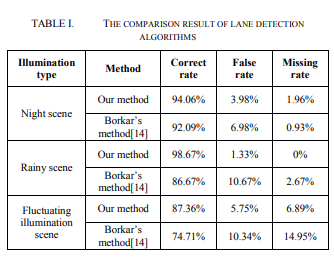
[**Lane detection algorithm based on local feature extraction**](https://ieeexplore.ieee.org/abstract/document/6775702)**(IEEE):**

An effective local feature extraction algorithm for lane detection is proposed in this paper. First, a lane region of interest (ROI) is determined by the location of the road surface that appeared in an image. Then, the light intensity and width of lane markings are taken as the local feature. A local threshold segmentation algorithm is utilized to extract lane-marking candidates followed by a morphological operation to obtain the accurate lane. An edge refining procedure is used to eliminate the interference and reduce computational cost. Finally, the lane marking is detected using Hough transform with some subsidiary conditions.

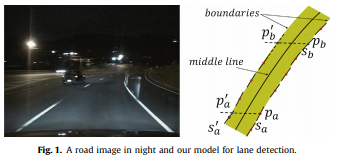
Methodologies Used:

* Region of Interest Selection
* Segmentation by the Median Local Threshold (MLT) Algorithm
* Interference Removed by Morphological Operation
* Edge Refining
* Lane Detection using Hough transform

Experimental Results:

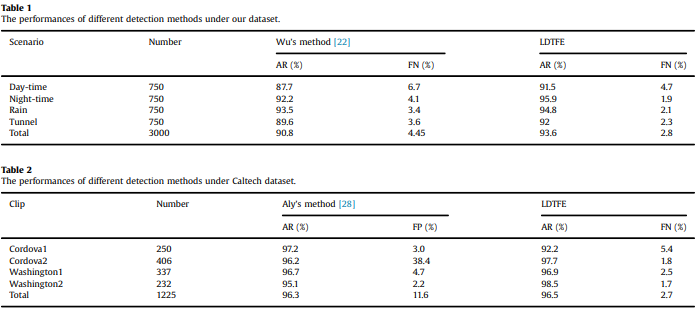


[**Robust Lane Detection using Two-stage Feature Extraction with Curve Fitting**](https://www.sciencedirect.com/science/article/abs/pii/S0031320315004690)**(ScienceDirect):**

They proposed a novel lane detection method. Their method regards lane boundary as a collection of small line segments. They proposed a modified Hough Transform to detect small line segments. Small line segments are clustered based on our proposed similarity measurement. Removing interferential clusters depends on the balance of small line segments.

Dataset Used: Their own custom dataset, Caltech Dataset

Methodologies Used:

* Two-stage feature extraction
  + Small line segment extraction
  + Small line segment clusters
* Identification of lanes
  + Finding the candidate clusters
  + Identifying the final lanes

Experimental Results:

* **Feature Tracking:**

[**Effective Lane Detection and Tracking Method Using Statistical Modeling of Color and Lane Edge-Orientation (IEEE):**](https://ieeexplore.ieee.org/abstract/document/5369883)

This paper proposes an effective lane detection and tracking method using statistical modeling of lane color and edge orientation in the image sequence. At first, we will address some problems of classifying a pixel into two classes (lane or background) and detecting one exact lane. Generally, the probability of a pixel classification error conditioned on the distinctive feature vector can be decreased by selecting more distinctive features. A proposed pixel classifier model (Bayes decision rule for minimizing the probability of error) uses two distinctive features, lane color, and edge orientation, for classifying a lane pixel from a background image.

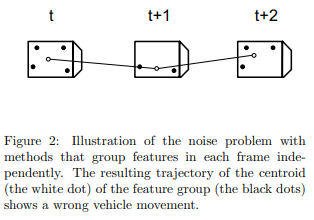
Methodologies Used:

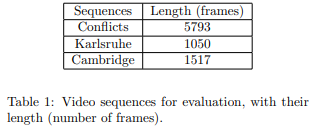
* Lane segmentation based on statistical Bayes decision rule
* Estimation of state-conditional densities for lane color and edge-orientation features
* Postprocessing
* Adaptive learning of estimated PDF

[**A feature-based tracking algorithm for vehicles in intersections (IEEE):**](https://ieeexplore.ieee.org/abstract/document/1640414)

Intelligent Transportation Systems need methods to automatically monitor road traffic and especially track vehicles. Most research has concentrated on highways. Traffic in intersections is more variable, with multiple entrances and exit regions. This paper describes an extension to intersections of the feature-tracking algorithm. Vehicle features are rarely tracked from their entrance in the field of view to their exit. Our algorithm can accommodate the problem caused by the disruption of feature tracks. It is evaluated on video sequences recorded on four different intersections.

Methodologies Used:

* Grouping features

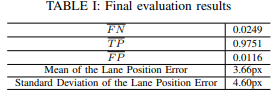
Experimental Results:

[**A novel multi-lane detection and tracking system (IEEE):**](https://ieeexplore.ieee.org/abstract/document/6232168)

In this paper, a novel spline-based multi-lane detection and tracking system is proposed. Reliable lane detection and tracking is an important component of lane departure warning systems, lane-keeping support systems, or lane change assistance systems. The major novelty of the proposed approach is the usage of the so-called Catmull-Rom spline in combination with the extended Kalman filter tracking. The new spline-based model enables accurate and flexible modeling of the lane markings. At the same time, the application of the extended Kalman filter contributes significantly to the system's robustness and stability.

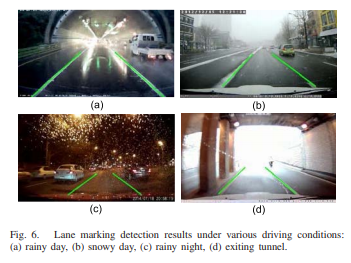
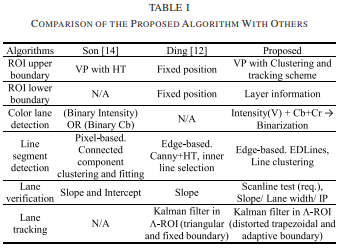
Methodologies Used:

* Using Catmull-Rom Splines for Lane Marking Representation
* State Vector
* Motion Model
* Measurement model
* Initialization of the Tracker
* Lane Merging Detection
* Clustering of Lane Markings

Experimental Results:

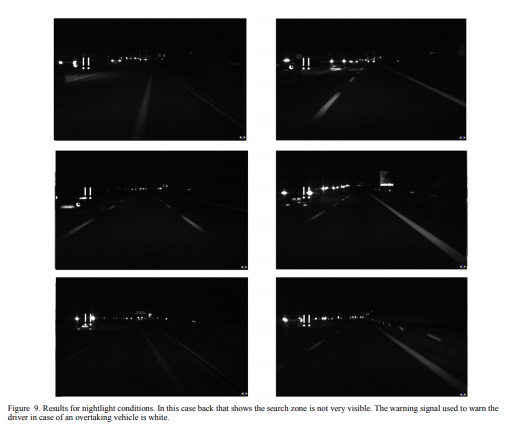
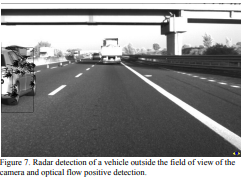
[**Robust Lane Detection and Tracking for Real-Time Applications (IEEE):**](https://ieeexplore.ieee.org/abstract/document/8303759)

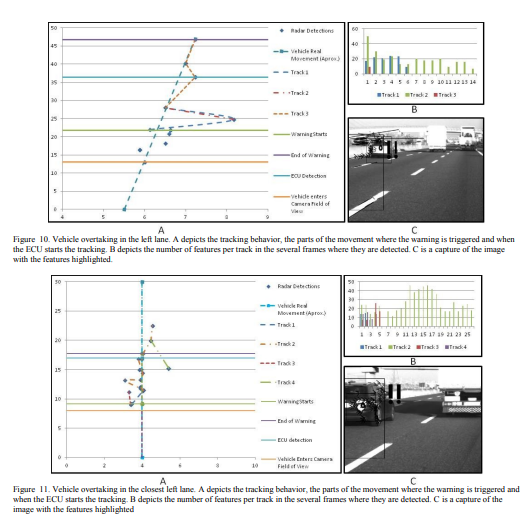
An effective lane-detection algorithm is a fundamental component of an advanced driver assistant system, as it provides important information that supports driving safety. The challenges faced by the lane detection and tracking algorithm include the lack of clarity of lane markings, poor visibility due to bad weather, illumination and light reflection, shadows, and dense road-based instructions. In this paper, a robust and real-time vision-based lane detection algorithm with an efficient region of interest is proposed to reduce the high noise level and the calculation time.

Experimental Results:

* **Optical Flow:**

[**Data fusion for overtaking vehicle detection based on radar and optical flow (IEEE):**](https://ieeexplore.ieee.org/abstract/document/6232199)

In this paper an approach to a real application is presented, able to fulfill the requirements of such demanding applications. Most of the commercial sensors available nowadays are usually designed to detect front vehicles but cannot detect overtaking vehicles. The work presented here combines the information provided by two sensors, a Stop&Go radar, and a camera. Fusion is done by using the unprocessed information from the radar, and computer vision-based on optical flow. The basic capabilities of the commercial systems are upgraded giving the possibility to improve the front vehicles detection system, by detecting overtaking vehicles with a high positive rate.



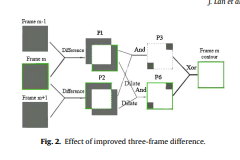
‘

[**The use of optical flow for road navigation (IEEE):**](https://ieeexplore.ieee.org/abstract/document/660838)

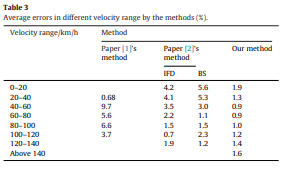
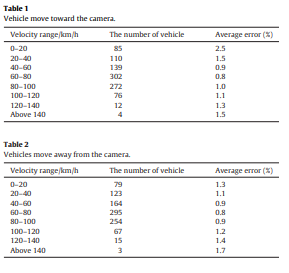
This paper describes procedures for obtaining a reliable and dense optical flow from image sequences taken by a TV camera mounted on a car moving in usual outdoor scenarios. By using correlation-based techniques and by correcting the optical flows for shocks and vibrations, useful sequences of optical flows can be obtained. When the car is moving along a flat road and the optical axis of the TV camera is parallel to the ground, the motion field is expected to be almost quadratic and have a specific structure.

Inferences: Egomotion parameters can be computed with simpler procedures when clear landmarks, such as road markers, are present in the scene. The major requirement for the recovery of a reliable optical flow is the presence of enough texture in the images. Well defined structures and high contrasts are not necessary. As a consequence, the optical flow can be computed in a variety of different scenarios and this versatility appears to be the major advantage of the proposed approach.

[**Vehicle speed measurement based on gray constraint optical flow algorithm (ScienceDirect):**](https://www.sciencedirect.com/science/article/abs/pii/S0030402613009005)

****This paper presents a novel vehicle speed measurement method, which contains the improved three-frame difference algorithm and the proposed gray constraint optical flow algorithm. By the improved three-frame difference algorithm, the contour of moving vehicles can be detected exactly. Through the proposed gray constraint optical flow algorithm, the vehicle contour's optical flow value, which is the speed (pixels/s) of the vehicle in the image, can be computed accurately. Then, the velocity (km/h) of the vehicles is calculated by the optical flow value of the vehicle's contour and the corresponding ratio of the image pixels to the width of the road. The method can yield a better optical flow field by reducing the influence of changing lighting and shadow.

|  |  |
| --- | --- |
| Basic Optical Flow Equation: | Gray Constraint Equation: |
|  |  |

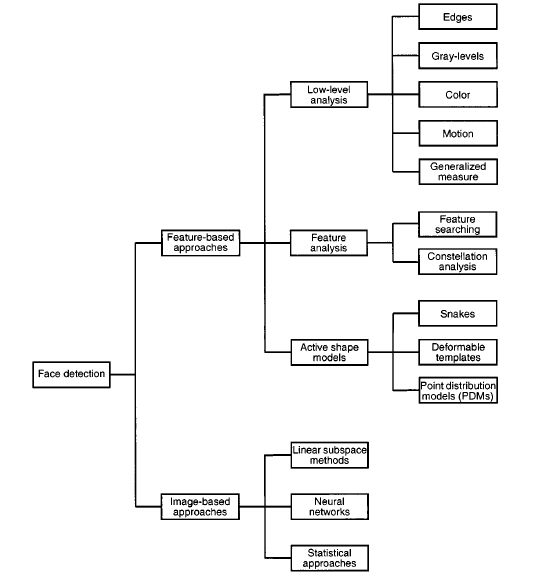
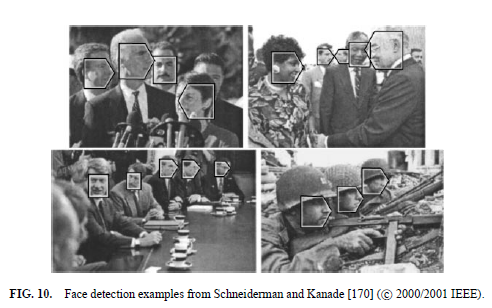
****Experimental Results:

* **Face Detection:**

[**Face Detection: A Survey (ScienceDirect):**](https://www.sciencedirect.com/science/article/abs/pii/S107731420190921X)

In this paper, they present a comprehensive and critical survey of face detection algorithms. Face detection is a necessary first step in face recognition systems, to localize and extract the face region from the background. It also has several applications in areas such as content-based image retrieval, video coding, video conferencing, crowd surveillance, and intelligent human-computer interfaces.

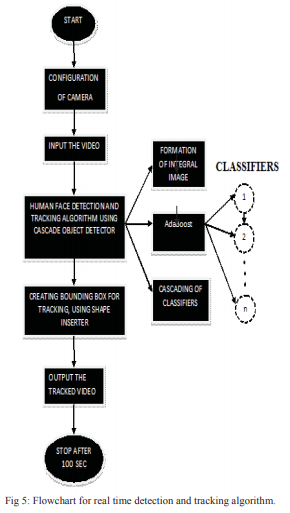
Approach

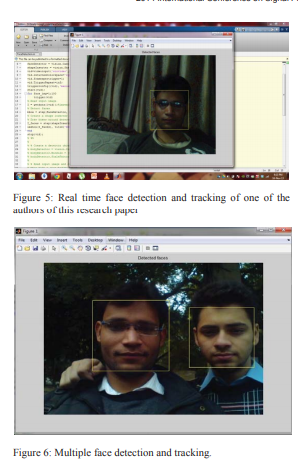
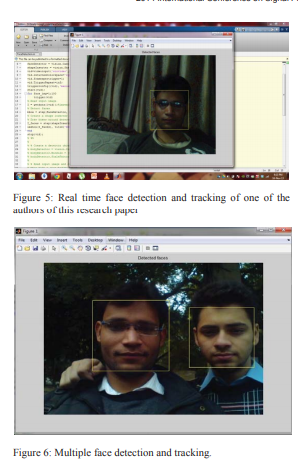
Experimental Results:

[**Real-time human face detection and tracking (IEEE):**](https://ieeexplore.ieee.org/abstract/document/6777046)

This paper describes the technique for real-time human face detection and tracking using a modified version of the algorithm suggested by Paul Viola and Michael Jones. The paper starts with the introduction to human face detection and tracking, followed by the apprehension of the Viola-Jones algorithm and then discussing the implementation in real video applications. Viola jones's algorithm was based on object detection by extracting some specific features from the image.

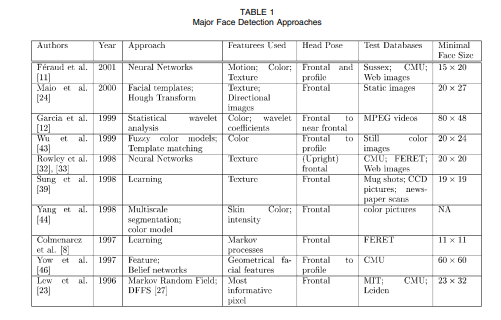
Approach:

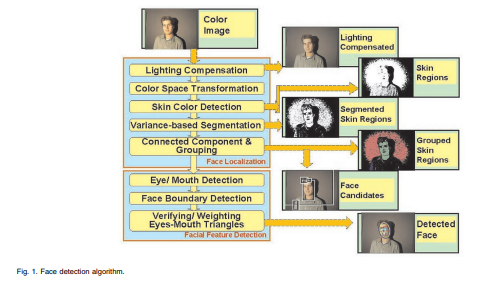


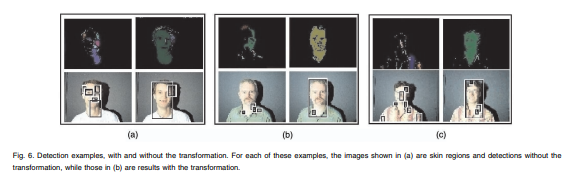
Experimental Results:

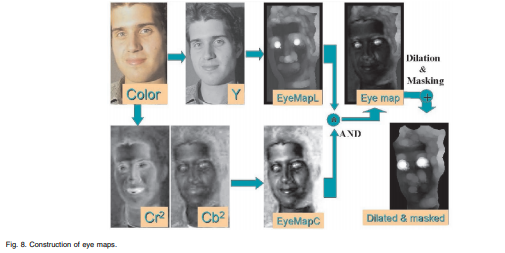
[**Face detection in color images(IEEE):**](https://ieeexplore.ieee.org/abstract/document/1000242)

Human face detection plays an important role in applications such as video surveillance, human-computer interface, face recognition, and face image database management. They propose a face detection algorithm for color images in the presence of varying lighting conditions as well as complex backgrounds.

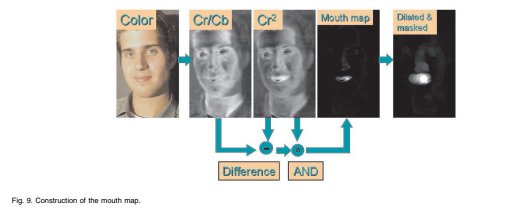




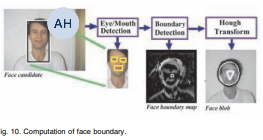




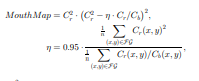
Localization of facial features



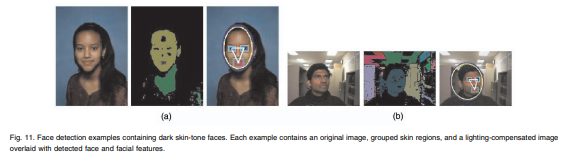
Eye Map

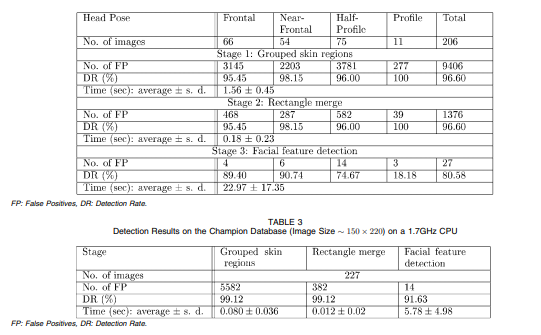
Formulae

Mouth Map

Formulae

Face Boundary and face score



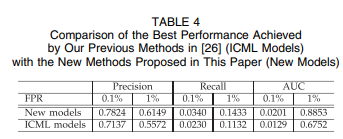


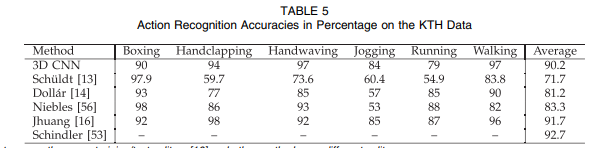
* **Action Recognition:**

[**3D Convolutional Neural Networks for Human Action Recognition (IEEE):**](https://ieeexplore.ieee.org/abstract/document/6165309)

In this paper, they develop a novel 3D CNN model for action recognition. This model extracts features from both the spatial and the temporal dimensions by performing 3D convolutions, thereby capturing the motion information encoded in multiple adjacent frames. The developed model generates multiple channels of information from the input frames, and the final feature representation combines information from all channels.

Dataset Used: TRECVID Dataset, KTH Dataset

Experimental Results:



[**Pose Based Action Recognition of Vulnerable Road Users Using Recurrent Neural Networks (IEEE):**](https://ieeexplore.ieee.org/document/9308462)

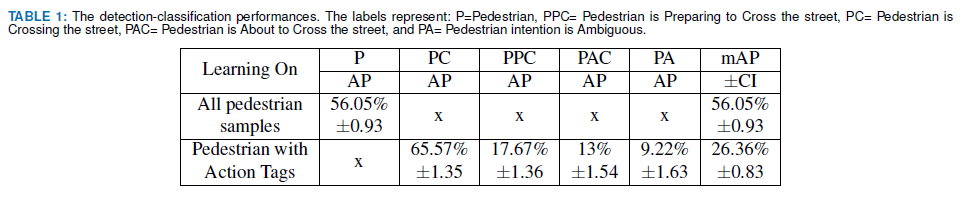
Their proposed approach can classify basic movements after different and especially short observation periods. The classification will then be successively improved in case of a longer observation. This allows countermeasures, such as emergency braking, to be initiated early if necessary. The benefits of using 3D poses are evaluated by comparison with a method based solely on the head trajectory. We also investigate the effects of different observation periods.

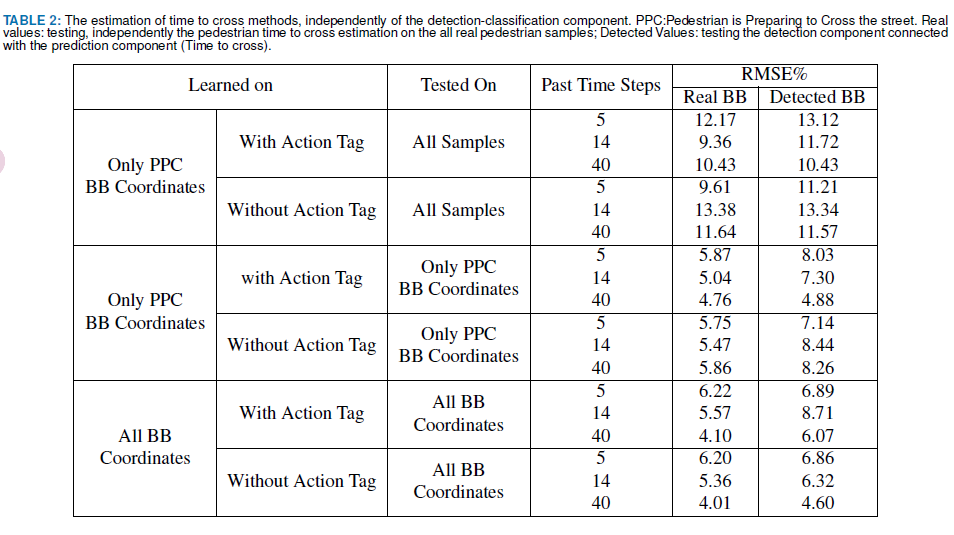
[**Multi-Task Deep Learning for Pedestrian Detection, Action Recognition and Time to Cross Prediction (IEEE):**](https://ieeexplore.ieee.org/document/8854084)

They propose 1) a pedestrian detection and action recognition component-based, on RetinaNet; 2) an estimation of the time to cross the street for multiple pedestrians using a recurrent neural network. For each pedestrian, the recurrent network estimates the pedestrian's action intention to predict the time to cross the street. Based on their experiments on the JAAD dataset, and show that integrating multiple pedestrian action tags for the detection part when a merge with a recurrent neural network (LSTM) allows a significant performance improvement.

Methodologies Used:

* Train all pedestrian Bounding Boxes (BB) samples with the RetinaNet for pedestrian detection purposes
* Split the pedestrian Joint Attention for Autonomous Driving (JAAD) data set into four classes for pedestrian action functionality
* Train a Long Short-Term Memory (LSTM) model using only BB coordinates in order to estimate the time to cross of each pedestrian.

Experimental Results:



[**Survey of pedestrian action recognition techniques for autonomous driving (IEEE):**](https://ieeexplore.ieee.org/abstract/document/8954864)

In this survey, they present a detailed description of the architecture for pedestrian action recognition in autonomous driving and compare the existing mainstream pedestrian action recognition techniques. They also introduce several commonly used datasets used in pedestrian motion recognition. Finally, they present several suggestions for future research directions.

Methodologies used:

Pedestrian detection in autonomous driving:

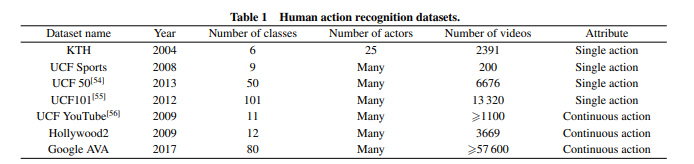
* Traditional algorithms
  + HOG feature and SVM algorithm
  + DPM algorithm
* Deep learning algorithms
  + Two-stage method
  + One-stage algorithm

Action recognition:

* Traditional method
  + Optical flow characteristics
  + Gradient characteristics
  + Dense Trajectories (DT) algorithm
* Deep learning method based on two-stream network

Datasets used: KTH, the UCF series, Hollywood2, and Google AVA

Experimental Results:

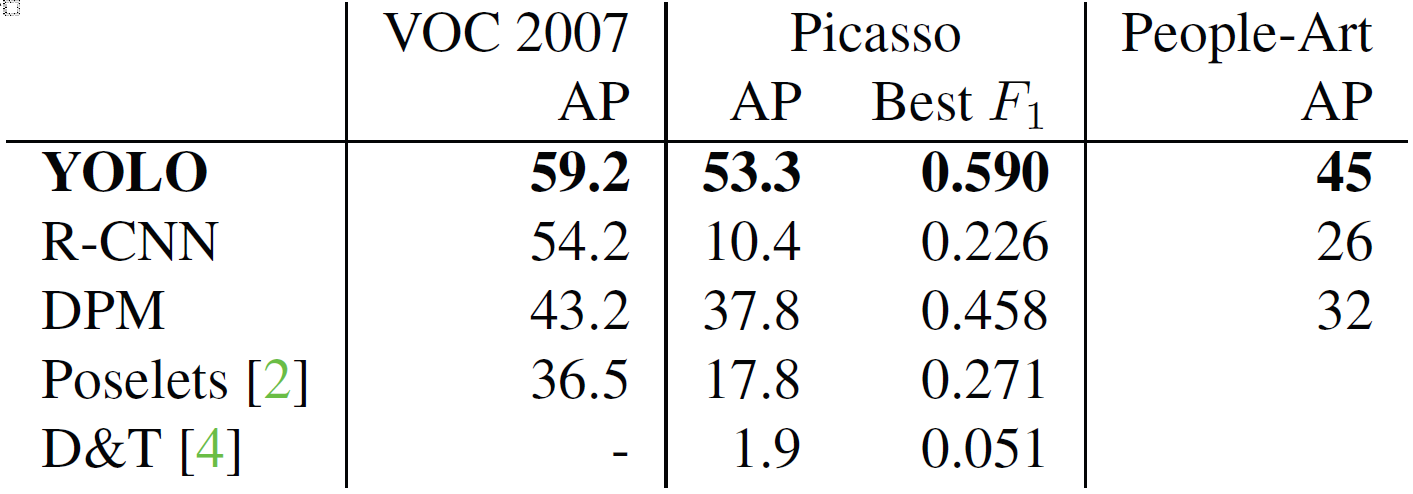


* **Object Recognition:**

[**Practical object recognition in autonomous driving and beyond (IEEE):**](https://ieeexplore.ieee.org/abstract/document/6301978)

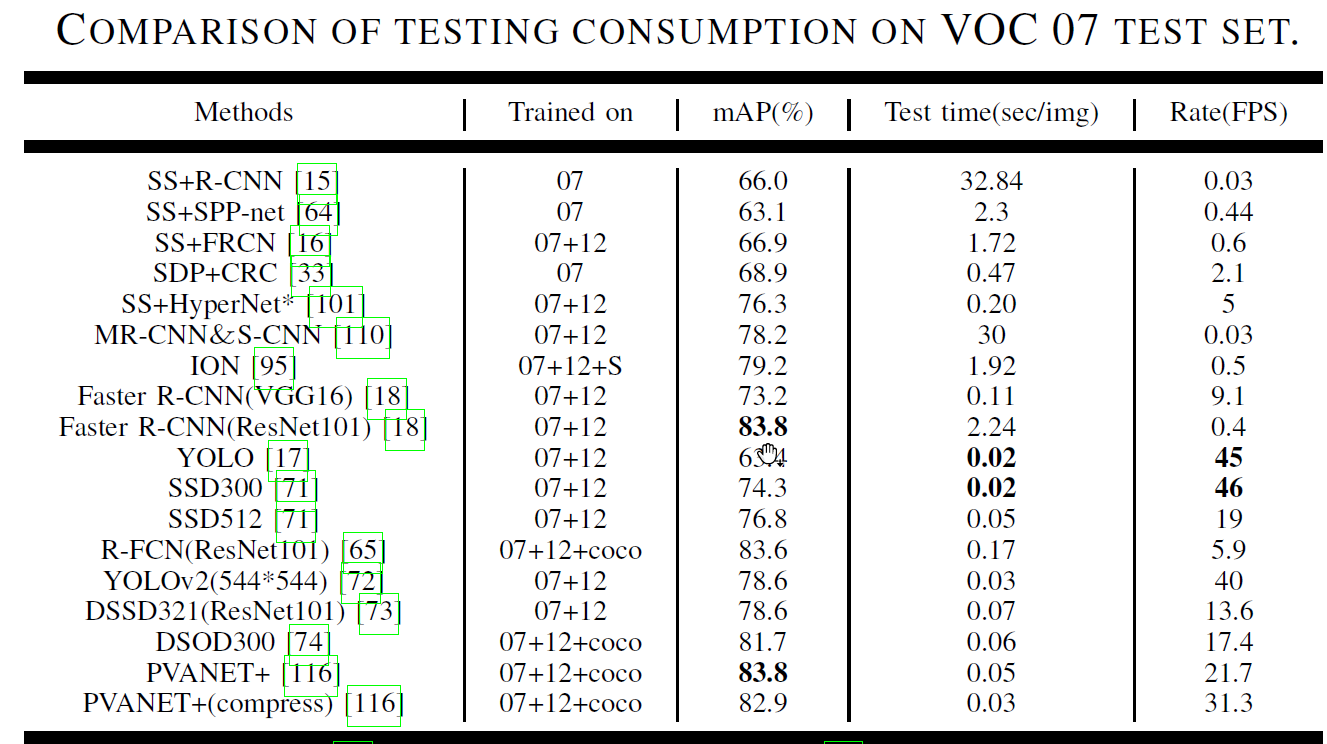
This paper is meant as an overview of the recent object recognition work done on Stanford's autonomous vehicle and the primary challenges along this particular path. The eventual goal is to provide practical object recognition systems that will enable new robotic applications such as autonomous taxis that recognize hailing pedestrians, personal robots that can learn about specific objects in your home, and automated farming equipment that is trained on-site to recognize the plants and materials that it must interact with.

[**You Only Look Once: Unified, Real-Time Object Detection (IEEE):**](https://ieeexplore.ieee.org/document/7780460)

YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, they frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. Their unified architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives in the background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork.

[**Object Detection With Deep Learning: A Review (IEEE):**](https://ieeexplore.ieee.org/document/8627998)

Their review begins with a brief introduction to the history of deep learning and its representative tool, namely, the convolutional neural network. Then, we focus on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance further. As distinct specific detection tasks exhibit different characteristics, we also briefly survey several specific tasks, including salient object detection, face detection, and pedestrian detection. Experimental analyses are also provided to compare various methods and draw some meaningful conclusions. Finally, several promising directions and tasks are provided to serve as guidelines for future work in both object detection and relevant neural network-based learning systems.



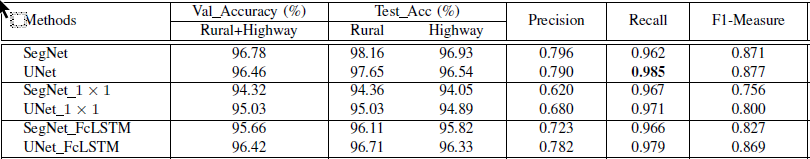
* **Deep Learning Architecture:**

[**Deep learning: Architectures, algorithms, applications (IEEE):**](https://ieeexplore.ieee.org/document/7477319)

This article consists of a collection of slides from the author's conference presentation. Some of the topics covered include Machine learning 101: Neural nets, backdrop, RNNs; Applications; Structured prediction; Unsupervised learning; "Neural Programs"; Architecture exploration; Towards hardware-friendlier DL; and Software.

[**Robust Lane Detection From Continuous Driving Scenes Using Deep Neural Networks (IEEE):**](https://ieeexplore.ieee.org/document/8883072)

They investigate lane detection by using multiple frames of a continuous driving scene and propose a hybrid deep architecture by combining the convolutional neural network (CNN) and the recurrent neural network (RNN). Specifically, information of each frame is abstracted by a CNN block, and the CNN features of multiple continuous frames, holding the property of time-series, are then fed into the RNN block for feature learning and lane prediction. Extensive experiments on two large-scale datasets demonstrate that the proposed method outperforms the competing methods in lane detection, especially in handling difficult situations.



[**L-UNet: An LSTM Network for Remote Sensing Image Change Detection (IEEE):**](https://ieeexplore.ieee.org/document/9301184)

Since ConvLSTM shares, similar spatial characteristics with the convolutional layer, L-UNet, which substitutes partial convolution layers of UNet-to-Conv-LSTM and Atrous L-UNet (AL-UNet), which further using Atrous structure to multiscale spatial information is proposed. Experiments on two data sets are conducted and the proposed methods show the advantages both in quantity and quality when compared with some other methods.