

UNIVERSITY OF GLASGOW



RESEARCH METHODS & TECHNIQUES

GROUP COURSEWORK

---

# Computer Vision and Deep Learning for Pedestrian Detection

---

Oliver Alan Stafurik	2444536S
Simona Holubkova	2483856H
Filip Vyrosto	2479054V
Jakub Jelinek	2478625J

December 1, 2022

## Literature Review

Object detection is a fundamental topic in computer vision and deep learning. It serves as a foundation for the detailed development of several research challenges. Pedestrian detection is a specific application of the object detection problem and in 2014, it was formally advanced to the deep learning level. Detection algorithms based on deep learning are often implemented by constructing regional recommendation boxes and then further predictions are made on these suggestion boxes [1], sometimes called boundary boxes.

Intelligent transportation systems have been created in recent years to assist in the reduction of the number of traffic in metropolitan areas, and the rate of accidents, injuries, and deaths caused by them. According to He et al. [1], these systems must be capable of allowing a human population to travel freely without the chance of an accident, while also lowering fuel consumption and environmental pollution. He et al. [1] suggest a number of difficulties and solutions for pedestrian recognition, including recognising pedestrians in exceptionally small images by adding geographical location specification and pedestrian attributes into Caltech and Inria datasets.

Caltech and Inria datasets are also utilised by Brunetti et al. [2] to compute the histogram of gradient feature extraction for pedestrian identification, where each generated histogram is treated as an image representation and a cascade of classifiers is used to distinguish each sub-region. Brunetti et al. [2] demonstrate that 3D motion capture systems are required for applications where a person's motion must be collected with high levels of precision, while the 2D collection may result in an excessive loss of information.

The four main pillars in pedestrian identification are feature extraction, deformation management, occlusion handling, and classification. Wanli Ouyang and Xiaogang Wang [3] suggest they can be jointly learned to maximise their capabilities through collaboration. Deep neural network (DNN) fusion architecture for fast and robust pedestrian detection is another opportunity for enhancement. The proposed network fusion architecture allows for the parallel processing of multiple networks for speed [4]. Yifu Zhang et al. [5] invented BYTE - an association method with a simple tracker ByteTrack scoring higher than any other previously designed tracker on MOT across any FPS range to improve multi-object tracking (MOT). ByteTrack primarily enhances the identification of objects with a low confidence score, which highly improves the detection of pedestrians in real-world situations where they are often at least slightly occluded. A combination of all these methods can give us a great base for pedestrian detection that can be used for deployment in further applications.

Pedestrians are the most important and vulnerable moving objects on streets and roadways, where thousands of people congregate daily. It is critical not only to identify the pedestrian but also to calculate his future motion path. According to Dominguez-Sanchez et al. [6], no work has been done utilising deep learning approaches to classify pedestrians based on their motion direction. Dominguez-Sanchez et al. [6] used their dataset to test, train, and fine-tune AlexNet, and then produced a unique input-filtered picture based on the post-processing of static video frames. The Advanced Driver Assistance System might be the next step towards future advancements. Once the pedestrian's motion path has been analysed and predicted, it is conceivable to apply it to the disciplines of intelligent driving [7].

One of the biggest problems with intelligent driving lies in the sheer computing resource limitations as a moving vehicle can provide a very limited amount of power to run the hardware. This can be mitigated by advancement in the hardware itself, but also by techniques reducing the performance requirements of the deep learning models themselves. Luo et al. [8] proposed an enhanced model of a very lightweight network G-YOLOX by reducing as many parameters as possible. Building up on YOLOX, a one-shot anchor-free detector, it uses a ghost model of an existing neural network

to synthetically optimise it and reduce its overall size. This network can be then run on mobile systems such as moving vehicles with slightly worse performance but achieve incomparable speed and resource requirements.

Detecting pedestrians from a moving vehicle is a difficult challenge with many factors influencing its reliability. One of the major complications is weather and lighting conditions, which training images occur rarely in common datasets. Li et al. [9] propose 3 new models which outperform existing solutions in hazy weather and poor lighting conditions scenarios. As the analysis by Kohli et al. [10] of the 2018 Uber self-driving vehicle crash which resulted in a pedestrian death suggests, solving such problems could save lives in the future. They applied various image enhancement techniques to the video from the frontal camera and then benchmarked the result by feeding it into common models in order to decide whether the crash was preventable. The best-performing model was able to detect the pedestrian no earlier than 0.86 s before the crash which is too late given the high speed of the vehicle. The proposed models by Li et al. [9] published more than a year after the study could provide us with new results that could potentially change the conclusion of the study in favour of the pedestrian, however, this has not yet been tested.

In conclusion, deep learning applications of pedestrian detection have become an important field of study in recent years due to the ever-growing use of Intelligent Transport Systems and self-driving cars. Many companies and institutions are trying to tackle challenges which arise such as low detection rates, optimization and resource issues, direction detection and poor light and weather conditions. Improving these methods further could help make roads a safer place by preventing collisions such as the Uber 2018 self-driving car crash.

## Paper Summaries

### Applications of deep learning techniques for pedestrian detection in smart environments: a comprehensive study

#### Reference

Fen He et al. “Applications of deep learning techniques for pedestrian detection in smart environments: a comprehensive study”. In: *Journal of advanced transportation* 2021 (2021)

#### Short Summary

Intelligent Transport System (ITS) aims to reduce the volume of traffic in metropolitan areas, reduce the risk of accidents, reduce fuel consumption, etc. ITS is used for automatic road management and real-time operations in events such as natural disasters and unnatural disasters. Further application of ITS such as management, electronic toll collection management, public transportation management, passenger communication management, and traffic flow management. Many technologies are used within ITS such as IoT, Neural Networks, Image Processing, Data Mining etc.. 5 main subsystems compose the ITS; ATIS, ATMS, AVHS, APTS and CVO. The main concern of the systems is to enable the ever-growing human population to move freely and safely within ITS. Thus, the need for pedestrian detection (PD) arises, and with it, its challenges.

#### Why did I read this paper?

Provides a general overview of ITS and its components. Also gives insight into different problems within this field concerning PD. Also published relatively recently (Oct. 2021)

#### Personal view of the paper

Gives the reader a nice overview of the entire problem of ITS as well as an introduction and explanation of the difference between Machine Learning and Neural Networks. The formulation of challenges and proposed solutions is very clear and insightful. Gives hints where more research with ITS and PD should take place.

#### What problem does this paper address?

No specific problem rather summarizes all problems within ITS concerning pedestrian detection (such as the need for fast detection, lightweight systems, etc...). Points in the direction where current research takes place.

#### Is it an important problem?

Yes, pedestrian detection is at the heart of ITS as pedestrians are the majority of the traffic that needs to be monitored and guided within ITS. Pedestrians/Cyclists also form the most vulnerable part of traffic.

#### What is the significance of the result and its solution?

In general, according to the studies conducted in this paper, the most important challenges in identifying pedestrians on the street using the proposed technologies and methods, especially deep learning techniques, can be expressed in Table 1 and some of the solutions that seem useful in solving these challenges are suggested in the table.

**What are the claimed novel contributions of the paper?**

This paper does not include any unique contributions.

**What previous work is the basis for this research?**

Authors reference many papers from the area of transportation, some problems are formulated in other papers however, no paper groups challenges and solutions in this way thus this is not fully based on any other paper.

**What methodology has been used?**

No methodology has been used per se, in the paper authors rather identify problem solution pairs that seem to be most important within ITS and PD.

**Does the methodology seem appropriate for this problem?**

Yes, the methodology seems appropriate, given the nature of this paper. The ITS is outlined in length and hints at its limitations.

**What conclusions are drawn from the results?**

Table 1 shows a summary of challenges and proposed solutions with respect to pedestrian/cyclist detection.

**Are they valid?**

Yes, given other papers within this field trying to tackle many of the problems mentioned in this paper.

**What did I learn?**

The complexity of the ITS, the need for PD and a clear distinction between Machine Learning and Neural Networks.

**What (if anything) would I have done differently?**

More illustrations where possible so as to better understand described challenges. Other than that, not much.

# Computer vision and deep learning techniques for pedestrian detection and tracking: A survey

## Reference

Antonio Brunetti et al. “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey”. In: *Neurocomputing* 300 (2018), pp. 17–33

## Short Summary

The automotive industry is ever-growing, therefore it is eminently important to improve techniques for vision-based pedestrian tracking. Every aspect of pedestrian detection using computer vision comes down to two main systems - the acquisition system (camera resolution, field of view, technology) and the environmental conditions (light conditions, weather conditions, etc.). Nowadays, most systems use 2D acquisition systems for pedestrian detection connected to ML techniques, however, in applications where the motion of a person should be acquired with high levels of accuracy, e.g. pedestrian protection systems, 3D systems are needed in order to preserve information that 2D tracking could not take into consideration. This paper analyses how Deep Learning methodologies work and emphasize that more tests on benchmark data sets need to be done in the future.

## Why did I read this paper?

This paper is generally cited by others (250+ times) and offers a wide range of citation indexes (200+ references). The paper was published on 26 July 2018, therefore still has an academic impact in its related field.

## Personal view of the paper

The Paper was mostly well-written, but some parts of it were hard to understand for someone unskilled in this field, however, everything was well summarised at the end of each section.

## What problem does this paper address?

This paper introduces the distinctions between 2D and 3D vision systems, as well as indoor and outdoor systems. The authors briefly explore Deep Learning approaches, such as Convolutional Neural Networks utilised in pedestrian recognition and tracking, at the end of the study. The problem that finishes the study is an investigation of Machine Learning algorithms on several benchmark datasets.

## Is it an important problem?

There is a rising interest in autonomous automobiles, and many conglomerates spend the majority of their funds on building better, more appropriate autonomous systems. In 2015, over 5000 pedestrians were killed and over 130,000 were wounded in car-related collisions in the United States alone. As a result, one of the most significant study fields in computer vision is people recognition and tracking.

## What is the significance of the result and its solution?

The findings emphasise the significance of testing pedestrian detection algorithms on various datasets in order to assess the robustness of the computed groups of features used as input to classifiers.

### **What are the claimed novel contributions of the paper?**

This paper establishes a new, powerful approach to designing a novel application for pedestrian detection. It uses deep architectures to extract features that are later used as input to the simple learner for pedestrian classification.

### **What previous work is the basis for this research?**

Given that the study is a survey, it makes extensive use of references. There are 206 references, the earliest of which is from 1973. However, there is a strong emphasis on Histogram Oriented Gradient features, which employ test sets from Caltech and Inria; almost 50 references are included in this area.

### **What methodology has been used?**

The paper contains no experiments. However, the process adopted is known as assessment by demonstration. The authors give examples in the paper to show previously discovered features from other research, such as HOG features extraction for pedestrian identification or a comparison for Log-Average Miss Rate evaluation considering dataset performances.

### **Does the methodology seem appropriate for this problem?**

The technique appears to be adequate, however, an alternative model - modelling and simulation - may have been utilised as well. Because there are so many references in the study, it may have been able to examine all complicated systems and construct a model that simulates whether the new approach to designing a modern pedestrian detection application is superior to the old ones described in the research. This would result in empirical evidence proving the assertion. Even if this is an excellent concept, this study gives a lot of background knowledge that may be expanded on in the future.

### **What conclusions are drawn from the results?**

The reviewed studies emphasise the need of investigating how recent techniques for pedestrian identification perform, as well as compare them to features-based approaches using benchmark datasets. Although reported studies demonstrate promising results in autonomous pedestrian identification, further designs must be created and validated.

### **Are they valid?**

Yes. The findings on benchmark datasets are carefully documented in the study. It also allows some room for future works, which the authors believe are vital to work on.

### **What did I learn?**

Before reading this paper, I was unfamiliar with all of the strategies utilised in pedestrian detection. The article was useful in explaining the many approaches used for pedestrian tracking, such as the distinctions between 2D and 3D tracking, inside and outdoor tracking, and how Machine learning is utilised for pedestrian identification.

### **What (if anything) would I have done differently?**

Because I wasn't familiar with everything in the paper, several sections, such as the HOG feature comparison, were a little confusing at first. Those obscure sections of the paper were quite difficult to understand, and I had to go over them several times. Furthermore, as I said in the technique part, I would want to see a model developed to compare the results.

# Joint deep learning for pedestrian detection

## Reference

Wanli Ouyang and Xiaogang Wang. “Joint deep learning for pedestrian detection”. In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 2056–2063

## Short Summary

The previous solution for Pedestrian Detection did not explore interactions between extracted features and instead learned them either individually or sequentially. This paper proposes a new approach to joint learning with respect to visual components (deformation and features). 84x28 images are gathered, with each image having 3 channels. A new way of generating detection maps is proposed (variability in convolution size, 3 levels). Next, proposed an extra Deformation Layer. Multi-stage training strategy is used for classification. Results were tested on Caltech and ETH data sets. Input channel design and joint learning design outperform contemporary state-of-the-art CNN by 9%

## Why did I read this paper?

One of the most cited papers on Pedestrian detection ( $\pm 790$ ). The proposed groundbreaking idea with respect to extracted features and explored areas which were poorly understood at the time. Served as a basis for many other papers following its release.

## Personal view of the paper

The paper is well-written and explains complex topics in an understandable way. The authors have a clear vision of what they want to achieve with this research. Good visual explanation is used where appropriate to aid in understanding.

## What problem does this paper address?

High miss rate for pedestrian detection of contemporary neural networks. Poor understanding of how extracted features “work together”.

## Is it an important problem?

The problem is important as its improvement increases the accuracy of detection (pedestrians in this case). Helps with safety in Intelligent Transport Systems. Also explores poorly understood areas of pedestrian detection with respect to feature extraction and interaction.

## What is the significance of the result and its solution?

This paper shows that more research is needed in this area. Moreover, it shows how promising this could be when with some not-too-complex changes, the detection rate was improved drastically.

## What are the claimed novel contributions of the paper?

Proposes a novel idea for improvement of the learning model. Jointly learning extracted features instead of sequentially/individually resulting in a new neural network. Understanding of the poorly explored section of pedestrian detection (feature extraction and interaction).

## What previous work is the basis for this research?

Some parts of the proposed solution are based on existing neural networks, however much of it is a novel and groundbreaking creation of the authors.



**What methodology has been used?**

Artefact creation. Create a new neural network with new processing of extracted image features (new deformation layer, new approach to gathering detection maps, etc...).

**Does the methodology seem appropriate for this problem?**

Yes as it creates a novel solution to a problem and then tests it on Caltech and ETH data-sets

**What conclusions are drawn from the results?**

9% average miss rate improvement on the Caltech data set. Learning features jointly does improve performance. Indicates the interaction between features should be explored more.

**Are they valid?**

Yes. The authors outline how the tests are executed and the performance of their solution is compared to other state-of-the-art ones. Results are clearly presented.

**What did I learn?**

How extracting features work via different classifications such as occlusion, deformation models, etc. . . . However, many topics are still beyond my understanding for me to fully grasp them.

**What (if anything) would I have done differently?**

Performance on Caltech is well documented but I would welcome more insight into the ETH data set and performance there.

# Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection

## Reference

Xianzhi Du et al. "Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection". In: *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, pp. 953–961

## Short Summary

The authors proposed a DNN architecture that is faster while keeping the accuracy high with three steps. Firstly, a single-shot detector generates all possibilities of candidates for pedestrian's bounding box with coordinates and confidence score, with multiple false positives that are later rejected. Then, instead of using hard classification that shows a significant decrease in accuracy, a soft rejection is deployed which re-evaluates the confidence score based on output from multiple classifiers that run in parallel. After, candidates are met with semantic segmentation to improve the detection, but this comes at a cost of time. Still, with this approach, they outscored all the current state-of-the-art models in both accuracy and speed. The main innovation is the implementation of the SSD and the soft rejection as these two factors are solving the issue that has been the elephant in the room for other approaches - neglecting not popular, distant candidates as pedestrians.

## Why did I read this paper?

This is an exciting and recently published article which enhanced traditional classification models with a generator for all possible objects. It was written by an interesting group of people hence it seemed promising at the first sight. It also fits well within the topic I am researching and gives a good overview of the current state of the art.

## Personal view of the paper

This paper shows enormous potential. Through the research that I have done, it is one of the first of its type that is reasonable and proposes huge improvement, especially in terms of accuracy. Regarding speed, this topic is quite controversial as all that matters is the current state of the hardware we are able to use during everyday life.

## What problem does this paper address?

Object detection problems, in particular, the speed and the accuracy in pedestrian detection as the previous solutions addressed only just one of the issues properly.

## Is it an important problem?

As more autonomous systems are being introduced to our everyday lives, it is necessary to keep improving these technologies to be able to fully rely on them without constant supervision.

## What is the significance of the result and its solution?

This model outperformed popular solutions that are currently used. Plus, the introduction of the soft rejection is novel around these architectures, so it seems like a clever improvement. If others would try to implement it, we could see that this makes up the extra step that F-DNN made and hence, the older, still well-performing models, would outperform the one we are reading about in this article.

### **What are the claimed novel contributions of the paper?**

The authors proposed that merging multiple approaches could give better results without affecting speed. They also applied soft rejection for candidates, using a classification network to consider any possible candidate and then give assign them relevancy, which is being re-evaluated throughout the process. Semantic segmentation also did not seem to feature in many other solutions, so this reinforcement of pedestrian detection accuracy is a good step.

### **What previous work is the basis for this research?**

The inventions in this paper as how the method was used seem ground-breaking in this area. The work takes away from the research that has been done on particular methods that were fused in this paper and from one on the current state of the problem.

### **What methodology has been used?**

Artefact creation. Authors took up on already existing neural network (NN) which was enhanced by fusion of other NNs and introduction of soft rejection of candidates.

### **Does the methodology seem appropriate for this problem?**

Yes, the authors wanted to provide improved solution to the problem by merging different state-of-the-art approaches.

### **What conclusions are drawn from the results?**

The authors have achieved better performance in this topic than the previous works that were done.

### **Are they valid?**

The speed and accuracy of proposed solutions were compared with the current performance leading models, Tables 2-5 in the paper show this comparison and prove the conclusion valid.

### **What did I learn?**

Even though it might not seem like a good idea to merge multiple models as it would slow down the processing time, the authors managed to lightweight the processing and still have good performance time, compared to similar solutions.

### **What (if anything) would I have done differently?**

I would not add anything, I believe that the extensive research performed by the authors covered multiple approaches and they decided what was the most accurate and efficient way to achieve better results.

# Bytetrack: Multi-object tracking by associating every detection box

## Reference

Yifu Zhang et al. “Bytetrack: Multi-object tracking by associating every detection box”. In: *European Conference on Computer Vision*. Springer. 2022, pp. 1–21

## Short Summary

The authors propose a new data association method for multi-object tracking (MOT) called BYTE. This method does not ignore objects with low detection scores but tries to detect occluded objects in them. The association runs in two steps. First, it matches the high score values and then it performs a second matching on unmatched tracklets.

They also propose a highly accurate, fast and simple tracker ByteTrack based on BYTE. This tracker performs well in MOT benchmarks such as MOT17.

## Why did I read this paper?

The paper describes multi-object tracking and shows its use case in pedestrian detection.

## Personal view of the paper

This paper is very well written, easy to read and describes an interesting approach to the problem. During the reading, I was thinking that this approach just makes sense and I was wondering why hasn't everybody used it before. It just feels natural.

## What problem does this paper address?

The paper tries to detect partially occluded objects which produce low confidence detection scores in classical detection approaches.

## Is it an important problem?

Yes, this has many real-life applications, especially in pedestrian detection where it is fairly uncommon to have a completely clear view of people.

## What is the significance of the result and its solution?

The proposed tracker had state-of-the-art performance in both HiEve and BDD100K detection benchmarks at the date of publication.

## What are the claimed novel contributions of the paper?

The paper proposes a simple, effective and generic association method which does not ignore the low-confidence objects and uses a two-step matching mechanism to track them as well as the high-score ones. It also proposes a tracker using the method and detector YOLOX to form a high-performance tracker.

## What previous work is the basis for this research?

The paper builds upon general object detection in MOT (multi-object tracking) and datasets such as MOT17 and data association methods such as SORT, Re-ID or DeepSORT.

**What methodology has been used?**

Artefact creation - The authors propose a new method for the association of objects which they named BYTE. It operates in two steps. The first one associates only the high-scoring objects and the second one uses motion similarity on occluded objects. The authors claim that the innovation lies in the junction area of detection and association.

**Does the methodology seem appropriate for this problem?**

Yes, as mentioned above, I found this technique very natural and it just makes sense.

**What conclusions are drawn from the results?**

The authors claim that BYTE can be easily applied to almost any existing tracker to improve their association of low-scoring objects which they support by data from testing. They also prove their performance by implementing a simple tracker themselves.

**Are they valid?**

The paper supports the claims with experimental results and scores from general benchmarks.

**What did I learn?**

There is a new technique that improves the detection rate of low-scoring objects which highly improves overall detection accuracy and has a relatively low false-positive rate.

**What (if anything) would I have done differently?**

I would try to optimise some of the state-of-the-art trackers and extend them to use BYTE and use as a proof-of-concept instead of implementing just a simple tracker.

# Pedestrian movement direction recognition using convolutional neural networks

## Reference

Alex Dominguez-Sanchez, Miguel Cazorla, and Sergio Orts-Escolano. “Pedestrian movement direction recognition using convolutional neural networks”. In: *IEEE transactions on intelligent transportation systems* 18.12 (2017), pp. 3540–3548

## Short Summary

This paper is focused on pedestrian recognition problem addressed together with popular CNNs like AlexNet, GoogLeNet and ResNet, Histograms of Oriented Gradients (HOG) and other pixel-based techniques. Using this approach, authors of this article were not only able to estimate a pedestrian’s path, but also to roughly estimate pedestrian intentions towards specific situations such as crossing at intersections. This can lead to the future development intelligent transportation systems and improvements in ADAS to avoid any possible risks.

## Why did I read this paper?

I wanted to look for papers with a different perspective on the challenge of pedestrian detection and this paper satisfied my wonders. Here, we are already working with images, or better say frames, containing a pedestrian in movement. The task was to determine if the pedestrian is moving to the left, to the right or going forward. During tackling this problem, readers are given a quick rundown of CNN, as well as the proposal of 3 different instances of CNN. It is related to the topic of this literature review, as the problem we are researching is complex, with this specific paper already mentioning future work like the implementation of the Advanced Driver-Assistance System (ADAS) or improved data augmentation.

## Personal view of the paper

This is a high quality paper, which addresses an important challenge of computer vision in pedestrian detection. Not only clear explanation of the complex solution is present, but also a brief introduction to CNN. It also provided good illustrations to help reader visualise the work that is being proposed.

## What problem does this paper address?

After we’re able to detect a pedestrian, we need to focus on recognition of pedestrian’s trajectory. We have to look where the pedestrian will be going to correctly asses the risk of the situation and alert systems involved in the traffic. This paper looks exactly into how we can decide on what direction the pedestrian is headed to. Compared to one previous approach, the authors decided to focus on 3 movements: going left, going right and going forward instead of other motions like crossing, bending in, stopping etc.

## Is it an important problem?

Pedestrians on the streets are almost never in still position. Hence identifying where they are heading is necessary to create a whole picture of the environment and use this to create a good automated solutions on the roads.

## What is the significance of the result and its solution?

Not much research has been done on this topic, hence the significance combined with the great results of this paper is very high. Especially these days, when some vehicles already contain ADAS

and are used in many countries, being notified in time to stop the vehicle to avoid any possible clash with a pedestrian (or another object) is crucial.

### **What are the claimed novel contributions of the paper?**

Modifications to the current state-of-the-art networks that take as an input frames with the flow estimation around the detected pedestrian and also new dataset as the existing ones were not sufficient for this purpose. Authors also claim that deep learning has not been applied to categorise pedestrians based on their direction of movement in any work prior to theirs.

### **What previous work is the basis for this research?**

Regarding previous work and articles, mostly general CNN and optimization research is used as the base of this paper. Authors provided a new data-set created by themselves, used already known models but fine-tuned them for their purposes using various mechanisms.

### **What methodology has been used?**

Artefact creation. Multiple known techniques are reused and fused into one solution to provide the best outcome.

### **Does the methodology seem appropriate for this problem?**

Yes, because it develops a solution to a further problem that is identified during pedestrian detection. Authors have performed an experiment using a computer-vision-based approach to set a baseline, tested multiple CNN architectures and evaluated their performance on a dedicated data-set. They fine-tuned the proposed CNN to improve the accuracy of the network and then tried the suggested approach on the state-of-the-art network architectures to get the best results.

### **What conclusions are drawn from the results?**

Performance and accuracy of different neural networks used. According to the tables containing the results of the validation tests.

### **Are they valid?**

Data provided by the authors show that the conclusion that their approach is a good way to tackle this challenge.

### **What did I learn?**

Architecture of CNN, how the HOGs can be used and how I can calculate the motion direction with the use of consecutive steps. One thing that I realized reading the article is that the color of the frame is not necessary.

### **What (if anything) would I have done differently?**

This seems to be a very complex task, combining multiple approaches and I am still at the beginning of learning deep learning, hence I am unable to provide further suggestions or modifications from my side.

# A review of intelligent driving pedestrian detection based on deep learning

## Reference

Di Tian et al. “A review of intelligent driving pedestrian detection based on deep learning”. In: *Computational intelligence and neuroscience* 2021 (2021)

## Short Summary

The review summarizes several articles on the topic of pedestrian detection based on deep learning and its use case in intelligent driving. It describes the current state where despite pedestrian detection technology making great progress, it is still far from perfect and nowhere near human perception.

It evaluates the pedestrian detection algorithm based on deep learning, its common use datasets and metrics used to evaluate it, describes the main limitations of its performance and tries to predict the future of this field.

The authors predict that pedestrians of different sizes and occlusion issues are going to be solved relatively soon. The networks are going to become more lightweight to be able to run on mobile terminals such as smart cars. Future detection technology should also try to understand the relationships between the objects in the scene and not just single pedestrian objects. The last prediction the authors made was that the combination of image-based solutions with other sensors such as lidars is going to play a significant role in future of intelligent driving.

## Why did I read this paper?

This paper is a very nice introduction to the topic of deep learning-based pedestrian detection in intelligent and autonomous driving which is one of the most researched use cases of this kind of detection. The paper summarizes the current state and also proposes the authors’ view of future development.

## Personal view of the paper

I liked the summary that this paper offers. The authors did not research a very interesting topic of infrared images that can be combined with vision-based detection to improve accuracy in difficult conditions, but they acknowledge it and gave a reason for it (lack of datasets). I quite liked the summary of their view of future development which helped me find research points in later publications.

## What problem does this paper address?

This review gives the reader an insight into the problem of autonomous driving and particularly pedestrian detection. This is a problem that saves lives and could minimise the number of deaths on roads in future.

## Is it an important problem?

“road accidents would be the fifth leading cause of death by 2030” [10] with almost 8 000 pedestrians deceased in incidents in 2020 just in the United Kingdom according to NHS. This is a significant number of human lives lost that could be prevented.



### **What is the significance of the result and its solution?**

The researchers collected a set of influential resources to tackle this problem and propose a list of research areas for the future. It also acknowledges the limitations of computing resources that might be reduced with technological advancements soon.

### **What are the claimed novel contributions of the paper?**

The paper does not claim to introduce any novelties, however, it aims to analyse deep learning-based algorithms proposed in recent years, introduce common datasets and explain why performance is the key limitation of pedestrian detection in moving vehicles and explains possible future development of the field.

### **What previous work is the basis for this research?**

Given that the paper is a review, it is based on 127 other resource articles from the most prestigious research facilities and universities.

### **What methodology has been used?**

The paper describes various detection techniques, types of neural networks, their evaluation and datasets used for testing. It is not a systematic review.

### **Does the methodology seem appropriate for this problem?**

I would say that this is an appropriate method how to do a literature review.

### **What conclusions are drawn from the results?**

The paper finishes by listing 4 branches of possible research development for the future.

### **Are they valid?**

The future will tell.

### **What did I learn?**

As somebody with very limited previous knowledge of the topic, this review allowed me to get a decent insight into the techniques involved in deep learning and its use case in pedestrian detection.

### **What (if anything) would I have done differently?**

I think that the review could elaborate a bit more on the motivation behind the research topic.

# G-YOLOX: A Lightweight Network for Detecting Vehicle Types

## Reference

Qiang Luo et al. "G-YOLOX: A Lightweight Network for Detecting Vehicle Types". In: *Journal of Sensors* 2022 (2022)

## Short Summary

The authors of this article present an enhanced model of a lightweight network for different vehicle detection. Initially, they explained the problem of the popular CNN used for object detection and a background to the specific issue they addressed. Then they briefly explained the previous works of other lightweight structures that were a base for building the G-YOLOX network structure. Its predecessors are multiple versions of MobileNet, an efficient and low computational model suitable for mobile and embedded vision applications. This network structure also introduces Mosaic, a data augmentation technique to improve the performance of a detector. The next one is EfficientNet, where they balanced the depth, width, and resolution of the neural network to the network performance and YOLO, a one-stage object detection network. Then, the Gaussian YOLOv3 structure's algorithm enables real-time operation and better detection by redesigning the loss function and setting the boundary boxes with Gaussian parameter. In G-YOLOX, authors replaced the coupled head, usually used in YOLO structures, with decoupled head, which changes structure to two parts (shared layers and decoupled layers) and improves memory storage as well as detection. G-YOLOX is also anchor-free, meaning it first locates predefined key points and then generates bounding boxes of possible outputs. Using an anchor-free mechanism significantly reduces the number of design parameters that need heuristic tuning. It also uses a ghost module to generate feature maps with fewer parameters and reduced the computational costs of the network. Using Ghost-4 module produces only a quarter of the parameters of a regular CNN. Results were tested on a new dataset built by authors and showed almost 70% decrease in parameters needed with a 40% decrease in the size of the weight file. The viability of the original concept of a one-stage anchor-free detector has been proven.

## Why did I read this paper?

Not only a new structure is proposed here, but it gave me a nice overview of other models previously used and gave a good background to this topic. Although the article does not specifically mention pedestrian detection, I believe this approach can be leveraged in future attempts to build an efficient detection model that would be easily deployed on the road. It is necessary to exploit the path of lightweight models if we want to move towards autonomous driving. Current models that perform very well require a lot of CPU and this cannot be utilized by current vehicles, especially if society wants to move towards more eco-friendly options like electric cars. If we had to satisfy current requirements to run a traditional CNN for pedestrian detection, it would increase the demand in visiting charging stations more often, having them located closer to each other than needed and overcrowded in time, using more electricity and we know that the current global source of electricity does not come from renewable resources.

## Personal view of the paper

I was a bit sceptical when I first went through this paper as the first thing, I remembered was that the detection success was decreased. Reading the article later, I found out that the authors propose a great trade-off between speed and complexity and detection accuracy and that this network structure can be deployed on mobile devices as well. I would personally welcome the article to be a little bit longer to explain more in-depth the mechanisms behind the improvements in speed and changes in detection.

## **What problem does this paper address?**

The practicality of network structures used for detection and their usage in embedded systems.

## **Is it an important problem?**

Yes, as long as we as a society aim to use as much AI on the road as possible, we need to address this issue as well. Using more metal is not a permanent solution considering environmental aspects as well as resources. From the technical perspective, almost no solution is viable if it is very hard to deploy and maintain in everyday life.

## **What is the significance of the result and its solution?**

As mentioned in the overview and shown in the results of the article, we can observe 67% decrease in parameters compared to the latest version of a similar lightweight structure, which also leads to a decrease by 40% of weight file size.

## **What are the claimed novel contributions of the paper?**

Authors use the latest model YOLOX to extract feature map information and a ghost model to decrease the number of parameters and make it more lightweight. They also switch to decoupled heads that showed an increase in performance. Another novelty is the new rich dataset, containing different vehicles in different conditions during the daytime. With this dataset, the proposed model can run on a device with 12GB of VRAM.

## **What previous work is the basis for this research?**

The base of the G-YOLOX is built on YOLOX network structure proposed in the article YOLOX: exceeding YOLO series in 2021 by Zheng Ge, Songtao Liu, Feng Wang, Zeming Li and Jian Sun.

## **What methodology has been used?**

Artefact creation. Authors took the previous work mentioned in section above and implemented their ideas to improve the performance of the network.

## **Does the methodology seem appropriate for this problem?**

Yes, this article attempts to create/improve an existing solution to a known problem. The evaluation was performed on a dataset designed especially for this topic, new designs were developed and tested in comparison to other modern solutions.

## **What conclusions are drawn from the results?**

The proposed work suggests a working network structure that needs fewer parameters and drastically improves the speed and size of stored files with just a small decrease in detection. “The viability of the original concept of a one-stage anchor-free detector has been proven.”

## **Are they valid?**

According to produced tables with results, the authors’ conclusions are correct.

## **What did I learn?**

What is a ghost model (generator of ghost feature maps using linear transformations) and how this can improve the speed of CNN and what other lightweight network structures exist (MobileNet, EfficientNet, YOLO).

**What (if anything) would I have done differently?**

Based on my current knowledge, I can't think of anything that would improve the proposed structure in the article.

# Deep learning approaches on pedestrian detection in hazy weather

## Reference

Guofa Li, Yifan Yang, and Xingda Qu. “Deep learning approaches on pedestrian detection in hazy weather”. In: *IEEE Transactions on Industrial Electronics* 67.10 (2019), pp. 8889–8899

## Short Summary

Poor light and weather conditions increase the risk of accidents on roads. Many driving assistants and self-driving cars use conventional networks. However, many networks work well only in clear light and weather conditions. This paper proposes a new improved network to tackle the mentioned deficit. The authors developed 3 pedestrian detection models based on You Only Look Once (YOLO). Also, created a new dataset with images in hazy weather and poor light conditions for testing and further research (HazyPerson + INRIA). Modified YOLO with depth-wise and point-wise convolution, use of priori boxes. The 3 models are, Simple-Yolo, VggPriotiBoxes-Yolo, MNPrioriBoxes-Yolo. For validation 1195 pedestrian images in hazy weather have been collected of which 1052 were for training, and 143 were for testing. Augmentation skills are used to expand the data set (creates HazyPerson data-set). Results show that the detection performances of the methods based on the convolutional neural network are much better than the traditional HOG+SVM and Haar+Adaboost methods. MNPrioriBoxes-Yolo gives the best results.

## Why did I read this paper?

Addresses one of the known issues with pedestrian detection, namely the large accuracy decrease in hazy weather. Also well-cited paper ( $\pm 107$ ). Additionally, I was surprised by how relatively recent the paper is.

## Personal view of the paper

The paper addresses a huge problem in pedestrian detection as most accidents happen in hazy weather. Improving in this direction could hugely contribute to safer roads (self-driving cars, Intelligent Transport Systems, etc...). The paper has also great potential to further expand and improve upon its findings.

## What problem does this paper address?

Poor light and weather conditions increase the risk of accidents on roads. Many driving assistants and self-driving cars use conventional networks. However, many networks work well only in clear light and weather conditions.

## Is it an important problem?

Hazy weather increases the traffic accident rate, thus improved detection for Intelligent Transport Systems in such conditions is needed. Also, clear weather can be rare in some parts of the world.

## What is the significance of the result and its solution?

The proposed models clearly outperformed contemporary, state-of-the-art solutions. This is the first major step in this direction and its deployment could largely improve the safety of the roads. Moreover, as a by-product of this research, the newly enhanced data set was created with images of pedestrians in hazy weather (HazyPerson + INRIA).

### **What are the claimed novel contributions of the paper?**

The authors propose 3 new CNN models which largely improve the detection rates in poor visual conditions. The proposed solution is also lightweight. Moreover, a new data set including images in hazy weather is created giving options to other researchers to test their solutions in all light/weather conditions.

### **What previous work is the basis for this research?**

Some of the research was based on already existing Yolo CNN and the idea of Priori Boxes. However, many of the improvements made were completely new and innovative.

### **What methodology has been used?**

Artefact creation. 3 new CNN were created all of which built on existing ideas improving them. This was then tested on a newly created data set which was integrated with already well-established data sets.

### **Does the methodology seem appropriate for this problem?**

Yes as it tries to create a solution to a known problem then tests it and compares it with other existing solutions.

### **What conclusions are drawn from the results?**

Results show that proposed models based on the usage of the convolutional NN outperform existing solutions using HOG+SVM and Haar+Adaboost methods. MNPrioriBoxes-Yolo gives the best results. However, priori boxes solutions are very dependent on the size of the boxes which are set by the architect.

### **Are they valid?**

Given the performance of improved data sets which include more images in hazy weather, the proposed solutions perform better than other contemporary ones.

### **What did I learn?**

Yolo which is a rather lightweight solution can be effectively used to detect pedestrians in hazy weather. First time hearing and learning about Priori-Boxes which improve even more the performance of Yolo. Also learned how fast Yolo can be compared to other NN given my limited knowledge of other papers and their results.

### **What (if anything) would I have done differently?**

Nothing as much of this is beyond my knowledge.

# Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash

## Reference

Puneet Kohli and Anjali Chadha. “Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash”. In: *Future of Information and Communication Conference*. Springer. 2019, pp. 261–279

## Short Summary

The authors investigated a very interesting question. Whether the Uber self-driving car crash could have been avoided. In 2018, an Uber self-driving car hit a pedestrian who unfortunately passed away. This paper investigates why the system did not detect the pedestrian soon enough to stop and avoid the accident.

The authors try to enhance the video feed by various methods such as histogram equalization, Canny edge detection or Gamma correction. They also tried to use some advanced techniques such as the Multi-exposure fusion framework, however, these took up to 1 second per frame (with a 24 fps video feed) which is not feasible in a scenario with such importance with human lives in question.

The authors also investigated 4 detection algorithms: RetinaNet, HOG+SVM, SSD and YOLO. Most of which were able to detect the pedestrian slightly less than a second before the crash. The processing time, however, was not even close to real-time with YOLO as the fastest algorithm processing just below 9 frames per second.

The authors did not answer the question in the end directly, but from the results they showed, they did not manage to prove that the crash could have been avoided.

## Why did I read this paper?

The topic of this paper fits very well our literature review. It is about a real-life problem involving pedestrian detection in conditions that even people generally struggle to detect them. It poses a question in the abstract that seemed interesting and tried to answer it later on.

## Personal view of the paper

The paper describes very well the various methods for image enhancements and detection algorithms. It then applies them to a very interesting problem. It shows that pedestrian detection in the low light condition is a bit more difficult than it seems when using just vision-based detection. Reading this paper allowed me to make an image of the performance and complexity of the algorithms and the possible outputs of the enhancement techniques. I think that it is a very well-written paper and even though the result they conclude with would not save the pedestrian, the paper has significant value.

## What problem does this paper address?

It investigates whether the 2018 Uber car crash was preventable.

## Is it an important problem?

The driver of the vehicle was indicted with negligent homicide in 2020. The trial was planned for 2021. However, it was delayed, because the investigation is complex, and the process is ongoing. If this paper could help solve the problem, it could have helped solve the investigation.

**What is the significance of the result and its solution?**

The paper did not answer the posed question in the end. The result shows that performance is the main limiter in the scenario. Given that the detection has to happen in a moving vehicle with limited power, this is an important result.

**What are the claimed novel contributions of the paper?**

The paper suggests a modification to the SVM adaptive thresholding technique by Tian et al.

**What previous work is the basis for this research?**

The paper builds on the various techniques for both image enhancement and object detection. All of which are referenced as previous work.

**What methodology has been used?**

Hypothesis testing

**Does the methodology seem appropriate for this problem?**

I did not like the idea of them testing the techniques only on the Uber crash video and the Caltech dataset. However, the choice of techniques was appropriate.

**What conclusions are drawn from the results?**

The proposed detection techniques detect the pedestrian at the same time as the state-of-the-art proprietary algorithm from Intel. They are however still too slow for a real-time use case. The experiments showed that the image enhancement techniques do not improve the detection rate or the performance at all.

**Are they valid?**

Given my best knowledge and the data they provided, I assume that they are correct.

**What did I learn?**

Real-time pedestrian detection in low-light conditions is hard, especially at high speeds. Most algorithms are far too demanding so they cannot be used for real-time detection.

**What (if anything) would I have done differently?**

I would test the algorithms on more datasets.



## References

- [1] Fen He et al. “Applications of deep learning techniques for pedestrian detection in smart environments: a comprehensive study”. In: *Journal of advanced transportation* 2021 (2021).
- [2] Antonio Brunetti et al. “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey”. In: *Neurocomputing* 300 (2018), pp. 17–33.
- [3] Wanli Ouyang and Xiaogang Wang. “Joint deep learning for pedestrian detection”. In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 2056–2063.
- [4] Xianzhi Du et al. “Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection”. In: *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, pp. 953–961.
- [5] Yifu Zhang et al. “Bytetrack: Multi-object tracking by associating every detection box”. In: *European Conference on Computer Vision*. Springer. 2022, pp. 1–21.
- [6] Alex Dominguez-Sanchez, Miguel Cazorla, and Sergio Orts-Escolano. “Pedestrian movement direction recognition using convolutional neural networks”. In: *IEEE transactions on intelligent transportation systems* 18.12 (2017), pp. 3540–3548.
- [7] Di Tian et al. “A review of intelligent driving pedestrian detection based on deep learning”. In: *Computational intelligence and neuroscience* 2021 (2021).
- [8] Qiang Luo et al. “G-YOLOX: A Lightweight Network for Detecting Vehicle Types”. In: *Journal of Sensors* 2022 (2022).
- [9] Guofa Li, Yifan Yang, and Xingda Qu. “Deep learning approaches on pedestrian detection in hazy weather”. In: *IEEE Transactions on Industrial Electronics* 67.10 (2019), pp. 8889–8899.
- [10] Puneet Kohli and Anjali Chadha. “Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash”. In: *Future of Information and Communication Conference*. Springer. 2019, pp. 261–279.