

Article

Augmentation of Severe Weather Impact to Far-Infrared Sensor Images to Improve Pedestrian Detection System

Paulius Tumas ¹, Artūras Serackis ^{1,*} and Adam Nowosielski ²

¹ Department of Electronic Systems, Vilnius Gediminas Technical University, 10223 Vilnius, Lithuania; vilniustech@vilniustech.lt

² Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, 71-210 Szczecin, Poland; anowosielski@wi.zut.edu.pl

* Correspondence: arturas.serackis@vilniustech.lt

Abstract: Pedestrian detection is an essential task for computer vision and the automotive industry. Complex systems like advanced driver-assistance systems are based on far-infrared data sensors, used to detect pedestrians at nighttime, fog, rain, and direct sun situations. The robust pedestrian detector should work in severe weather conditions. However, only a few datasets include some examples of far-infrared images with distortions caused by atmospheric precipitation and dirt covering sensor optics. This paper proposes the deep learning-based data augmentation technique to enrich far-infrared images collected in good weather conditions by distortions, similar to those caused by bad weather. The six most accurate and fast detectors (TinyV3, TinyL3, You Only Look Once (YOLO)v3, YOLOv4, ResNet50, and ResNext50), performing faster than 15 FPS, were trained on 207,001 annotations and tested on 156,345 annotations, not used for training. The proposed data augmentation technique showed up to a 9.38 mean Average Precision (mAP) increase of pedestrian detection with a maximum of 87.02 mAP (YOLOv4). Proposed in this paper detectors' Head modifications based on a confidence heat-map gave an additional boost of precision for all six detectors. The most accurate current detector, based on YOLOv4, reached up to 87.20 mAP during our experimental tests.

Keywords: FIR pedestrian detection; image noise; data augmentation; bad weather; confidence heat-map; ADAS; YOLO; ResNet50; ResNext50; DnCNN



Citation: Tumas, P.; Serackis, A.; Nowosielski, A. Augmentation of Severe Weather Impact to Far-Infrared Sensor Images to Improve Pedestrian Detection System.

Electronics **2021**, *10*, 934.
<https://doi.org/10.3390/electronics10080934>

Academic Editors: Dong Seog Han, Kalyana C. Veluvolu and Takeo Fujii

Received: 25 March 2021

Accepted: 9 April 2021

Published: 14 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Each year, over 80,000 lives are tragically lost on roads, according to the World Health Organization (WHO) yearly report [1]. European regional status report on road safety shows that over 221 people are killed on roads every day in the European region, and thousands more are injured or disabled, with long-lasting effects. According to the research, 30% of killed road users are pedestrians and cyclists. The main reasons for fatalities are rapid urbanization and motorization, poor safety standards and infrastructure, lack of strong enforcement, drivers being distracted or under the influence of drugs or alcohol, a failure to wear seat belts or helmets, and lack of access to timely post-crash care. Speeding is another critical element causing lack of time to avoid the accident, and early-stage detection of collision could drastically minimize the chance of accident [2–4]. Lastly, severe weather conditions like rain, snow, fog are visibility affecting factors causing drivers to adapt to the conditions. However, the study of Das et al. [5] showed that fog or smoke is 3.24 times more likely to result in a severe injury and is 1.53 times more likely to cause a multiple-vehicle crash. A similar study, prepared by Sun et al. [6] analyzed rain influence for the driver and, depending on road type, the risk to have an accident increase to 2.61 times.

1.1. Existing Solutions for Driver Assistance

Autonomous vehicles may reach full automation (Level-5) only when they can perform all driving functions under all conditions. It includes the ability to operate autonomously in adverse weather conditions and at night. The far-infrared (FIR) thermal sensors can deliver the sensing capabilities that Level-5 autonomy demands. It is expected, that night vision system revenue growth worldwide between 2015 and 2025 will reach around 18% [7]. A number of related patents are registered by AdaSky Ltd. [8–11], Autoliv Development and AB Flir Systems Inc. [12,13], to name a few. FIR based solutions are already available for Level-1 and Level-2 vehicle automation levels. Usually, these are pedestrian detection systems, combining visual spectrum, and FIR camera. Automakers like Audi, BMW, and Daimler install Autoliv designed FLIR Pathfinder nighttime driving assistance. Such a system is based on an FIR spectrum FLIR camera with a resolution of 324×256 and a frame rate of 30 Hz [14]. There are not many details available about accuracy, performance, and type of detector used, but in one publication [15] Autoliv engineers refer to a Cascade classifier. AdaSky uses convolutional recurrent neural networks for image pre-processing [11] and additional detector for pedestrian, vehicle, and animal detection.

1.2. Infrared Imaging

Infrared imaging is an important topic in computer vision since radiation of these electromagnetic waves could be detected by the sensors all day long. The most highly available and cheapest cameras on the market are near-infrared (NIR) cameras. It is a visual spectrum camera without an infrared spectrum filter. Such camera for pedestrian detection systems could provide from 44 to 73 m visibility during the night [16]. However, NIR vision's primary drawbacks include their susceptibility to glare, blooming, and streaking from active light sources such as oncoming traffic, traffic lights, streetlights, and reflective objects such as road signs. Also, NIR illuminators may cause glare to other drivers using the same type of system and may cause damage to eyes at short distances (<1 m) [17].

A short-wave infrared (SWIR) spectrum image sensors are rarely used in pedestrian detection applications. Bertozzi et al. [18] tried to use SWIR cameras in poor visibility conditions. Research has demonstrated that reduced visibility phenomena as haze and fog feature quite different behaviors in the SWIR spectrum but still of no practical utility for automotive applications. Whereas better visibility through haze may be achieved by employing SWIR sensors, but it is a negligible benefit for pedestrian detection, hazing a long-distance phenomenon, no improvements can be obtained in foggy conditions. Similar results observed by [19] where authors shown that detection rates obtained are no better than reported in the revised state-of-the-art works.

Mid-wave infrared (MWIR) is popular bandwidth used in military applications such as guided missile technology. The 3000–5000 nm portion of MWIR band is the atmospheric window in which the homing heads of passive IR 'heat-seeking' missiles are designed to work, homing on to the IR signature of the target aircraft, typically the jet engine exhaust plume [20]. There are not many published research works related to MWIR use for pedestrian detection use, probably because of specific military use. However, Nguyen et al. [21] tried to set up a dual detector based on MWIR and long-wave infrared (LWIR) spectrum. The work provided a very robust pedestrian tracking. However, none of the tested sequences provided ambient conditions such as fog, smoke, haze, and precipitation, all of which are expected to impact the two bands differently.

LWIR cameras can provide a better fit for person detection, especially in complex outdoor scenarios with masking background texture or lack of illumination according to [22]. In general, the human appearance in LWIR images is not homogeneously bright due to clothes and other effects. Instead, there are smooth gray-value transitions inside the human blob and, in case of a merge also to surrounding bright background regions. Also, according to Beeck et al. [23], LWIR images provide visible pedestrians in severe weather conditions (e.g., fog, heavy rain).

The far-infrared (FIR) cameras are mainly used for pedestrian detection because far-infrared rays are less susceptible to moisture than rays of other wavelength bands [24]. Unlike NIR or visible spectrum cameras, FIR cameras are susceptible to disturbing light, such as oncoming headlights.

1.3. Contribution

The main contributions of this paper can be summarized as follows:

1. A new approach of severe weather feature extraction and augmentation approach by using the Denoising Convolutional Neural Network (DnCNN).
2. Results of pedestrian detection using state-of-the-art detectors testing in a wide variety of situations by using two biggest FIR domain datasets.
3. A set of trained models for pedestrian detection, which shows the current state-of-the-art performance and are ready for real-time applications.

2. State-of-the-Art

Modern object detectors proposed in literature may be categorized into two categories. In Figure 1 we have represented a schematic diagram of one-stage and two-stage detectors. The shared part between the two categories is the backbone and the feature map layer parts. At this stage an object classification is usually implemented by following structures like Cross Stage Partial Networks (CSPNet) [25], Darknet [26], ResNet [27], ResNeXt [28], and VGG [29]. The Feature map layer is a bridge between the Backbone and Head where different layers are interconnected and composed of several paths. Typically, at this stage, researchers include different Feature Pyramid Networks [26,30–32] and Path Aggregation Network [33] (PANet). The Head is the part where actual detection is taking place.

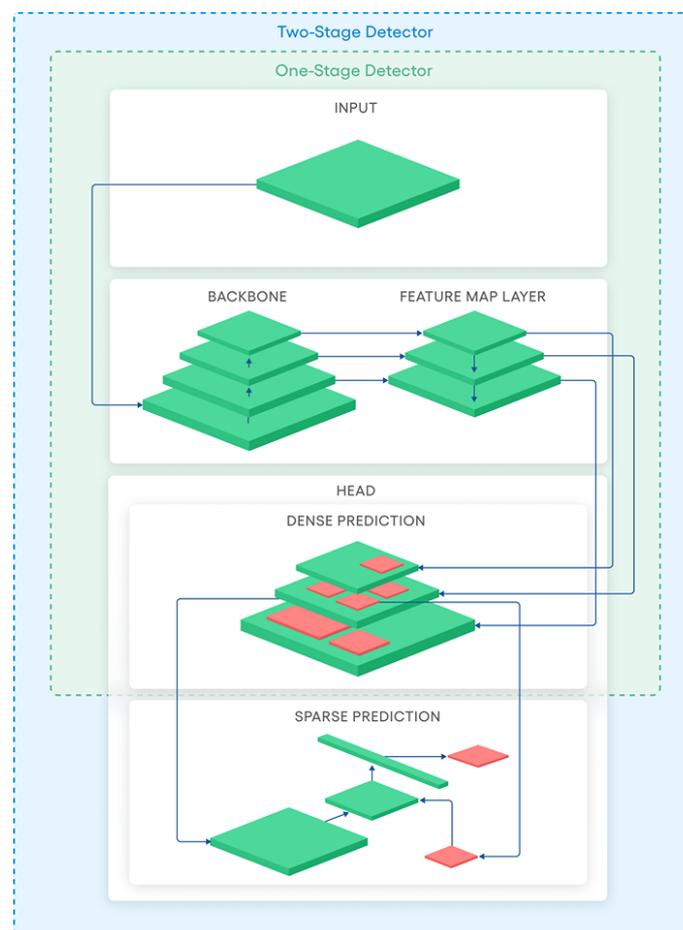


Figure 1. A schematic diagram for comparison of one-stage and two-stage object detectors.

The state-of-the-art of one-stage object detectors includes several versions of You Only Look Once (YOLO) [26,34–36], SSD [37], and RetinaNet [38]. The state-of-the-art two-stage detectors are R-CNN [39], Faster R-CNN [40], and R-FCN [41].

Two criteria are usually used to evaluate object detectors' performance: speed measured in frames per second (FPS) and accuracy evaluated by mean Average Precision (mAP). It is a metric typically used for PASCAL challenges [42] where Average Precision (AP) for one object class is calculated having an IoU threshold of 0.5 and the mAP is calculated by averaging AP over all object classes. Typically, two-stage detectors are very slow because of a complex feature extraction stage and complicated structure. In addition, two-stage detector are able to process images reaching only from 0.1 to 5 FPS [39–41,43,44]. However, such detectors has comparatively excellent accuracy of up to 69 mAP [45,46] on many popular benchmarking datasets (Caltech [47], KITTI [48], ImageNet [49], PASCAL VOC [50], and MS COCO [51]). On the other hand, the single-stage detectors are trying to trade accuracy in order to gain speed. YOLO and the minimized version called TINY YOLO were one of the first detectors able to achieve real-time performance with up to 43 mAP, outperforming SSD by 14 mAP [36].

One of the challenges for both detector categories' backbone and feature map layers is the multi-scale object detection. To address this problem, researchers usually stack different size layers on each other to extract features at different scales and join them later in a feature map layer. For example, the YOLOv2 backbone has 19 layers, and YOLOv3 has 53 layers. Such modification has affected the detector's accuracy on the MS COCO dataset by giving a 9.8 mAP increase. However, the speed has decreased from 40 to 20 FPS [52].

For pedestrian detection, researchers try to combine visual spectrum with FIR domain data [53–56] and apply YOLOv1, YOLOv2, YOLOv3, Faster R-CNN, R-FCN, or slightly modified versions where accuracy is ranging from 66 to 79 mAP. An actual FIR domain accuracy was tested by thermal camera manufacturer FLIR and SSD detector and reached 79.4 mAP [57]. A modified version of SSD was also tested [58] and reached from 87.68 to 97.5 mAP and YOLOv2 ranged from 58.5 to 80.5 mAP.

3. Materials and Methods

Numerous research showed that a key to successful machine learning-based image object detection is a rich dataset with diverse labeled examples used for model training. However, it is not easy to collect a dataset with many examples covering various situations captured by the FIR image sensor. The first reason—severe weather conditions prevent data collection due to rain and dirt which covers the sensor. The second reason—an image from the sensor is captured after pre-processing, where the charge collected in the FIR sensor matrix is equalized over the range to form an image. This pre-processing causes different images achieved at various environmental temperatures, and the same objects might look differently on a hot summer day and on cold winter day.

The annotation process is time-consuming and requires manual and repetitive work, which usually introduces errors. Finally, data diversification does not cover all situations. For these reasons, application-specific data augmentation could be an advantageous technique to generate additional unique samples. Data augmentation makes the dataset full of different samples, more balanced, transferring annotation, and filling the dataset gaps.

3.1. Pedestrian Datasets

In order to have an accurate pedestrian detector, a rich of features dataset is needed. Features could be described as situations where the image was captured for example during severe rain. Also, it is important to have various pedestrians poses like walking people in the group, crossing the street, or riding an bicycle.

There are up to ten datasets dedicated for pedestrian detection application in the FIR domain: CVC-09 [59], CVC-14 [24], FLIR-ADAS [60], KAIST [53], KMU [55], LSFIR [61], OTCBVS [62], RISWIR [63], Terravic Motion IR [64], SCUT [65], and ZUT-FIR-ADAS (ZUT) [66]. The biggest thermal dataset is SCUT, containing up to 211k frames, captured

with 384×288 resolution sensor and having more than 448k pedestrian annotations collected in Guangzhou, China. The dataset has a strictly predefined labeling protocol and has six classes (walk person, squat person, ride person, people, person, and combined annotation person/people). The second biggest dataset is the ZUT dataset. It contains widest variety of fine-grained annotated images (110k frames and 122k annotations), which were captured in the four biggest European Union countries with a 320×240 resolution sensor during severe weather conditions. The dataset also provides a synchronized Controller Area Network (CAN bus) data, including driving speed, brake pedal status, and outside temperature for further driver assistance system development.

However, the existing dataset does not cover all possible situations and data diversification could reveal possible deficiencies of detectors or rather improve existing performance.

3.2. Fusion of the Two Biggest Datasets

To train a mathematical model of the detector, we decided to merge ZUT [66] and SCUT [65] datasets into one. From ZUT, we took an 8bit+low pass version frames, containing only “Pedestrian”, “Occluded”, “Cyclists”, “Motorcyclist”, and “Scooterist” classes and merged them into one class. Such aggregation of the classes was needed because the SCUT FIR dataset includes only pedestrian-related classes (walk person, ride person, etc.).

The preparation of the SCUT dataset was performed in three stages. Firstly, we iterated through all SCUT training dataset and removed frames containing annotations of people groups and people annotations similar to the square shape. This step was needed because ZUT annotation methodology differs from SCUT annotation methodology. Secondly, we merged all classes into a single People class. Finally, we scaled down all the images to resolution of 640×480 pixels since SCUT dataset images have interpolated resolution of 720×576 pixels.

The summary of the number of video frames and a total number of annotations in selected datasets is given in Table 1. From ZUT dataset we extracted 69,455 frames (88,624 annotations) for training and 40,103 frames (33,808 annotations) for testing. From the SCUT dataset, we took 78,942 frames (118,377 annotations) for training and 76,381 frames (122,537 annotations) for testing. To show the spatial distribution of image annotations among these datasets, We have generated a heat map (see Figure 2) representing annotation location and size in image. From this heat map, it is visible, that most annotations are located on right side of the road (red and dark red color)—this is because China and Europe (where these FIR images were collected) are left-side drive countries. Also, we illustrate in Figure 2 that the spatial distribution of annotations remains similar in the train and test subsets.

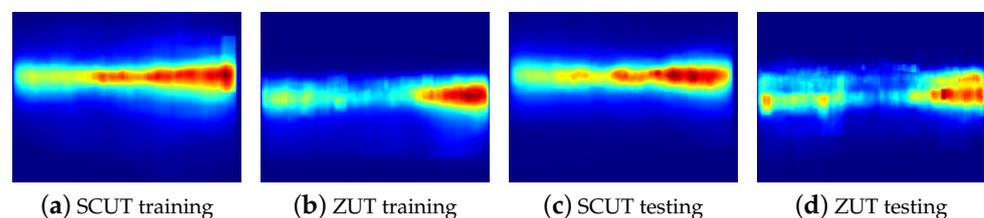


Figure 2. Annotation distribution per dataset.

Table 1. Dataset used for the training sample distribution.

Dataset	Training Frames	Traininig Annotations	Testing Frames	Testing Annotations
ZUT	69,455	88,624	40,103	33,808
SCUT	78,942	118,377	76,381	122,537
Total	148,397	207,001	116,484	156,345

3.3. Selection of the Detector Architecture

Six well-known convolutional neural network architectures were selected to investigate the most accurate and real-time ready pedestrian detector: a conventional TinyV3 [26], a TinyV3 with additional head (TinyL3 [67]), YOLOv3, YOLOv4, ResNet50 [27,68] and Cross Stage Partial Network (CSPNet) [25] applied on ResNeXt50 [28,69]. Also, we made the following modifications to the neural network configuration:

1. the input was set to 640×480 ;
2. annotations were auto-rotated by 5 degrees;
3. contrast and brightness left unchanged;
4. filter sizes were recalculated for single class use;
5. anchor ratios were recalculated using the k-means algorithm.

For selecting the best detector, we trained each detector until 300,000 steps by saving training weights every 1k step. After the training, we measured mAP at IoU = 50 for each saved step and the best result was used for further evaluation. The same methodology was used after augmentation, but the training steps were increased to 500,000 steps.

For evaluation of the detector performance we decided to measure Frames Per Second (FPS), mAP, average mAP, Average IoU, Re-call, Precision, and F1-score metrics. Also, we calculated True Positive (TP), False Positive (FP), and False Negative (FN) rates, using 50% of the IoU threshold. The FPS measurements were performed on Intel i7-8750H eighth-generation processor and NVIDIA RTX2070 8 GB graphics card. Training and validation were performed on Darknet Deep Neural Network (DNN) framework [70].

3.4. Dataset Augmentation

The manual collection of additional images in severe weather conditions is complicated and time-consuming process, especially the annotating phase. We decided to investigate a synthesis option for FIR images.

By investigating the spatial distribution of annotations on the left side is less intensive on a heat map (see Figure 2), we flipped images and their annotations horizontally. After mirroring the dataset images, the next step was to enrich a training dataset with severe weather samples.

A typical way to generate a wider variety of samples would be by using a Generative adversarial network (GAN). However, we have investigated an alternative approach. Since we are aiming to generate a severe weather features into an image that is visually close to noise and contains linear predefined features, we tried to invert the functionality of denoising DNN. DnCNN [71], DRU-net [72], FDnCNN [73], and DDP [74] are well known DNN-based denoisers, providing excellent Gaussian denoising, super-resolution transformation capabilities. For this reason we took DnCNN architecture, given in Figure 3, and performed a modification of training input function $y = x + v$ to $y = x$, where x is an input image and v is a random Gaussian noise.

To train a neural network dedicated to adding severe weather-related FIR image distortions, we took naturally distorted FIR images from the ZUT dataset. These natural images contain sequences of heavy rain, drizzle, and fog. We trained the slightly modified DnCNN until the loss function stopped converging. We applied DnCNN to generate distorted samples from images in the combined dataset, including the flipped ones.

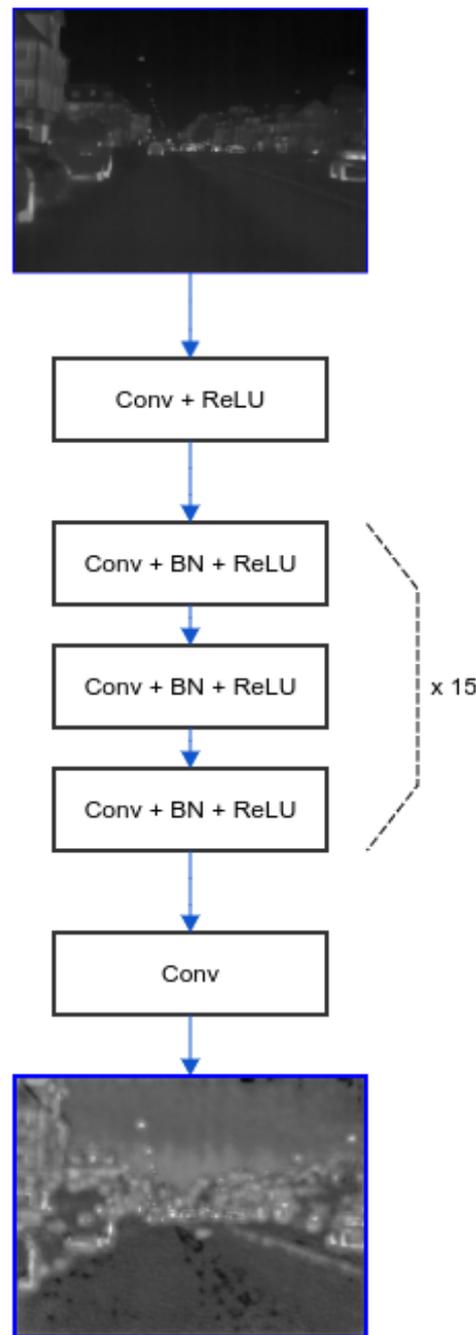


Figure 3. A structure diagram of the Denoising Convolutional Neural Network (DnCNN)-based image denoiser.

The generated samples are visually similar to original situations where rain and dirt distort captured image. In Figure 4 we have provided samples used for the training: input images (Figure 4c,d), real images captures in heavy rain with wet and dirty sensor (Figure 4a,b) and the output, modified by the DnCNN (Figure 4e,f). From the sample, shown in Figure 4c, we see that the pedestrian is cold and there is not much thermal radiation visible, but the corresponding sample in Figure 4e, generated by DnCNN, shows that features of a cold pedestrian are kept. Similar behavior is noticed with warm pedestrians: Figure 4d shows input image and Figure 4f shows the generated output.

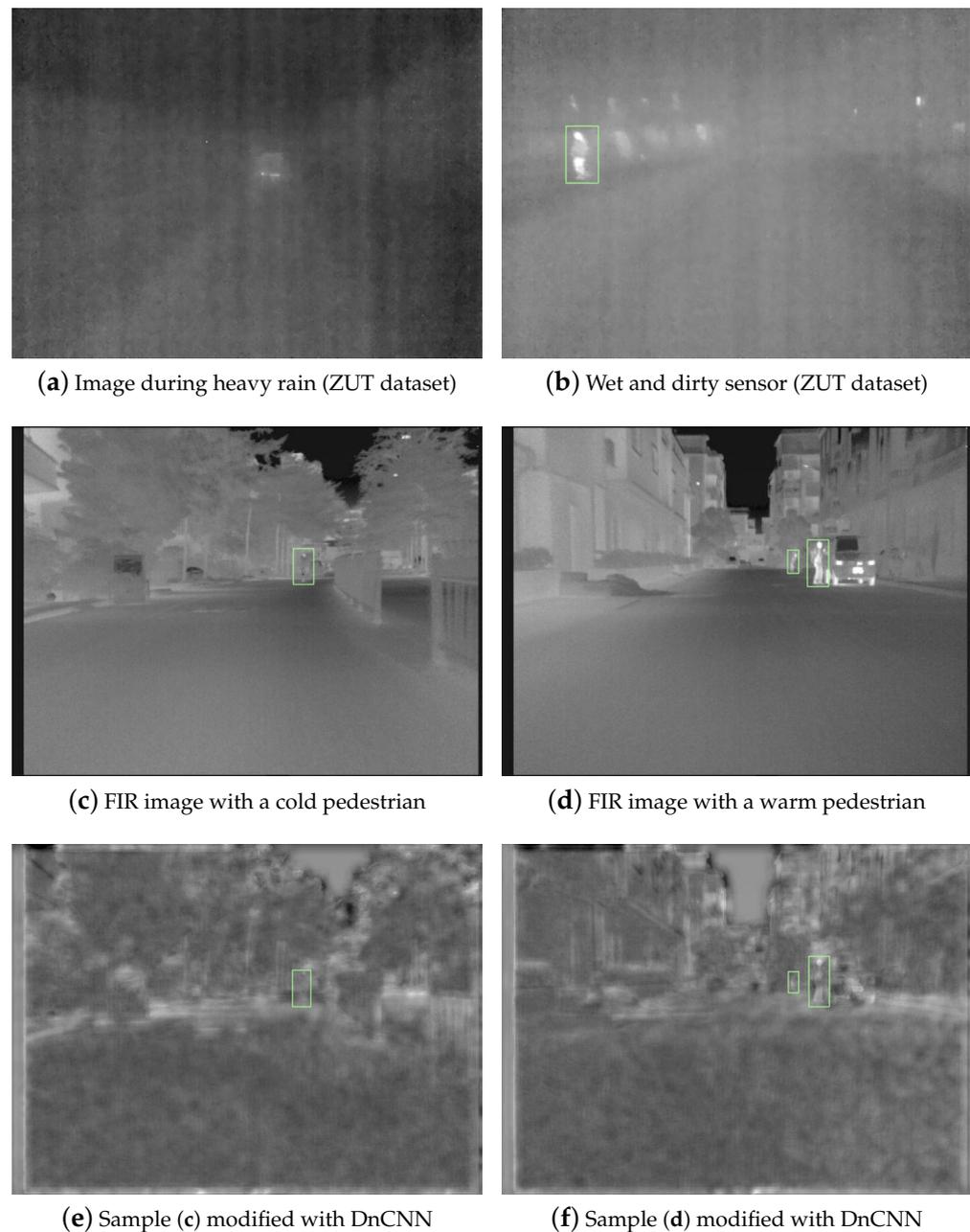


Figure 4. Comparison of captured and samples, augmented with DnCNN.

4. Results

In Tables 2 and 3 we have presented initial results on a fused dataset. The most accurate detector, tested on initial dataset was YOLOv4. It reached 86.05 mAP and achieved 15.97 FPS on average. The second highest precision was received by ResNet50, with reached 81.00 mAP and achieved 19.82 FPS on average. Detector structures with a minimized Backbone, such as TinyL3, outperformed YOLOv3 (the predecessor of YOLOv4) by 0.34 mAP and reached 43.1 FPS on average. The fastest detector was based on TinyV3 architecture. It worked at 55.57 FPS on average. However, it was the least accurate and showed 73.25 mAP. ResNext50 did not outperform the other detectors in any metrics. Looking from the training overview, the fastest was ResNext50 where only 27,000 training steps were needed to get decent accuracy. The longest time for the training took YOLOv3, where 209,000 steps were needed.

Table 2. Comparison of investigated detectors' efficiency using different scores.

Detector	FPS	mAP	Aver. IoU	Recall	Precision	F1-Score	Step
TinyV3	55.57	73.25	66.57%	0.63	0.89	0.74	172,000
TinyL3	43.1	80.14	62.52%	0.71	0.84	0.77	31,000
YOLOv3	17.88	80.48	68.95%	0.70	0.91	0.78	209,000
YOLOv4	15.97	86.05	59.28%	0.84	0.78	0.81	170,000
ResNet50	19.82	81.00	62.34%	0.75	0.82	0.78	194,000
ResNext50	17.70	77.07	65.57%	0.65	0.86	0.74	27,000

Table 3. Comparison of recognition rates for investigated detectors.

Detector	FPS	mAP	TP	FP	FN
TinyV3	55.57	73.25	86,009	11,109	49,604
TinyL3	43.1	80.14	96,265	18,311	39,348
YOLOv3	17.88	80.48	95,201	8978	40,412
YOLOv4	15.97	86.05	113,778	32,318	21,835
ResNet50	19.82	81.00	101,059	21,578	34,554
ResNext50	17.70	77.07	87,850	13,755	47,763

In the Table 4 we present detectors' performance with the augmented dataset. The best performance was observed by YOLOv4, reaching an accuracy of 87.02 mAP. The next most accurate detector is ResNext50, which reached 86.45 mAP and gained the highest accuracy boost with the augmented dataset (9.38 mAP). YOLOv3 is the third most accurate detector (83.87 mAP), which outperformed ResNet50 by 1.15 mAP and TinyL3 by 1.89 mAP. The ResNet50 is the fourth most accurate detector (82.72 mAP), which outperformed TinyL3 by 0.74 mAP. The TinyL3 detector showed 81.98 mAP and is more accurate than TinyV3 by 3.73 mAP. Finally, the training process revealed that ResNext50 was one of the first detectors to achieve results the fastest. YOLOv3 took the longest time to train, where 416,000 steps were needed.

Table 4. Retrained results.

Detector	mAP	TP	FP	FN	Aver. IoU	Recall	Precision	F1-Score	Step
TinyV3	78.22	98,158	19,810	37,455	61.98%	0.72	0.83	0.77	308,000
TinyL3	81.98	98,603	14,445	37,010	65.84%	0.73	0.87	0.79	307,000
YOLOv3	83.87	101,814	13,676	33,799	67.50%	0.78	0.88	0.81	416,000
YOLOv4	87.02	114,212	32,051	21,401	58.95%	0.84	0.78	0.81	300,000
ResNet50	82.72	103,827	20,451	31,786	63.89%	0.77	0.84	0.80	344,000
ResNext50	86.45	112,292	31,797	23,321	59.37%	0.83	0.78	0.80	54,000

Furthermore, a spatial detector confidence distribution of FP and FN, visible in Figure 5, revealed locations where the detector is failing to identify pedestrians or rather makes a false positive detection. This information could be a powerful way to dynamically adjust the detector's confidence threshold based on the image object location. For this reason, we slightly modified the detector's head to adjust the confidence threshold based on combined FP and FN heatmaps visible in Figure 6, which we call confidence heat-map. A decision to accept or reject detection is made by taking a detected pedestrian rectangle center coordinates and compare probability in the confidence heat-map. If detected object probability is lower than heat-map value—we skip detection, but if it is bigger or equal—we accept detection. As the Table 5 shows we additionally gained from 0.18 to 1.02 mAP where the most significant boost was observed by YOLOv3 detector.

Table 5. Retrained results with confidence heat-map.

Detector	mAP
TinyV3	78.74
TinyL3	82.73
YOLOv3	84.91
YOLOv4	87.20
ResNet50	83.44
ResNext50	86.67

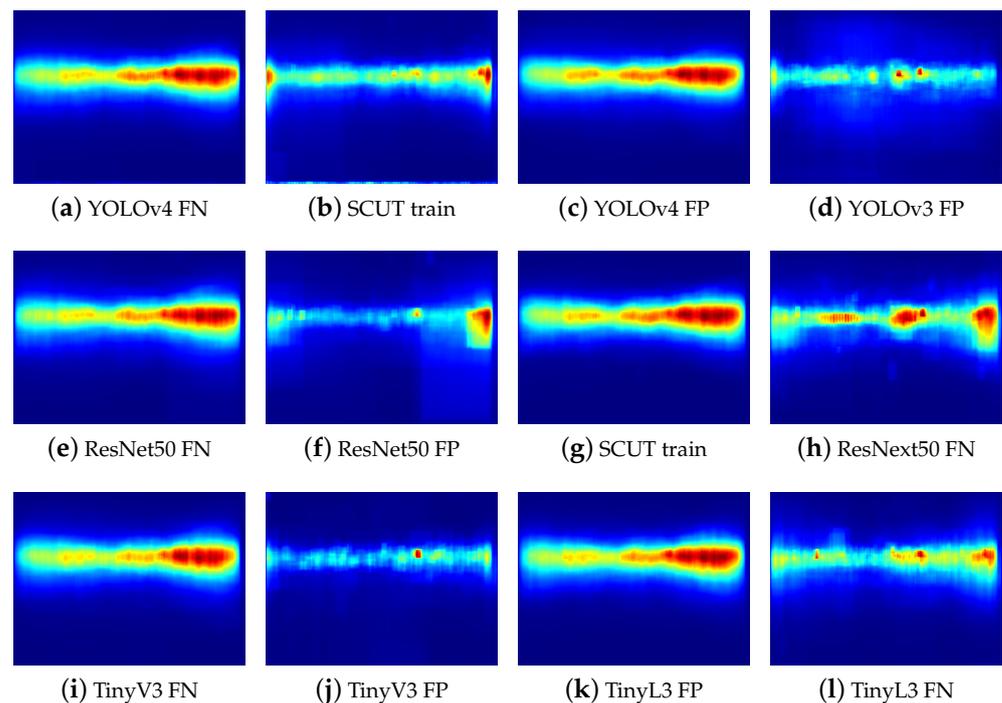


Figure 5. Spatial detector confidence distribution of False Positive (FP) and False Negative (FN).

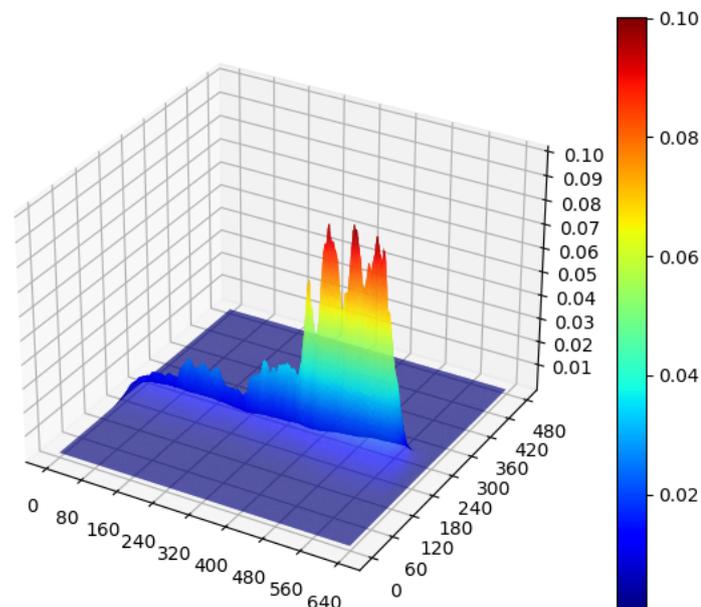


Figure 6. A confidence heatmap.

In addition, we were interested to see how well detectors perform in severe weather conditions before augmentation and after. In Figure 7 we present how well detectors recognize pedestrians and cyclist during the rain. As shown, YOLOv4, ResNet50 recognize all pedestrians with a probability larger than 50%. However, ResNext50, YOLOv3, TinyV3, and TinyL3 miss one or two pedestrians. Figure 8 shows the results after augmentation and YOLOv4, TinyV3, ResNet50, and ResNext50 recognize all pedestrians where YOLOv3 and TinyL3 miss only one pedestrian. Also after augmentation we see increased probabilities for pedestrians, which identifies more accurate detectors.

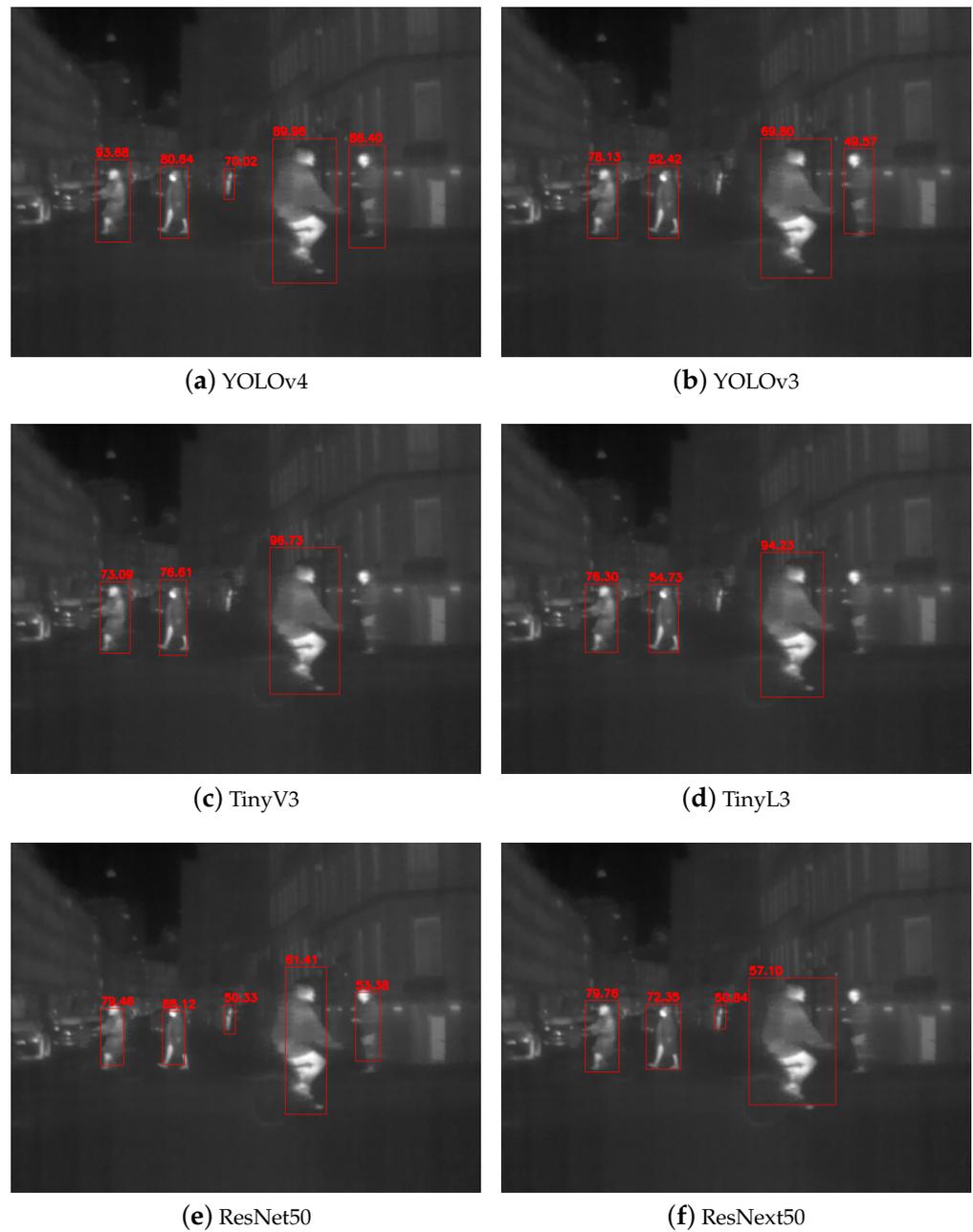


Figure 7. Comparison of detector performance before augmentation. Frame taken from ZUT dataset during the rain.

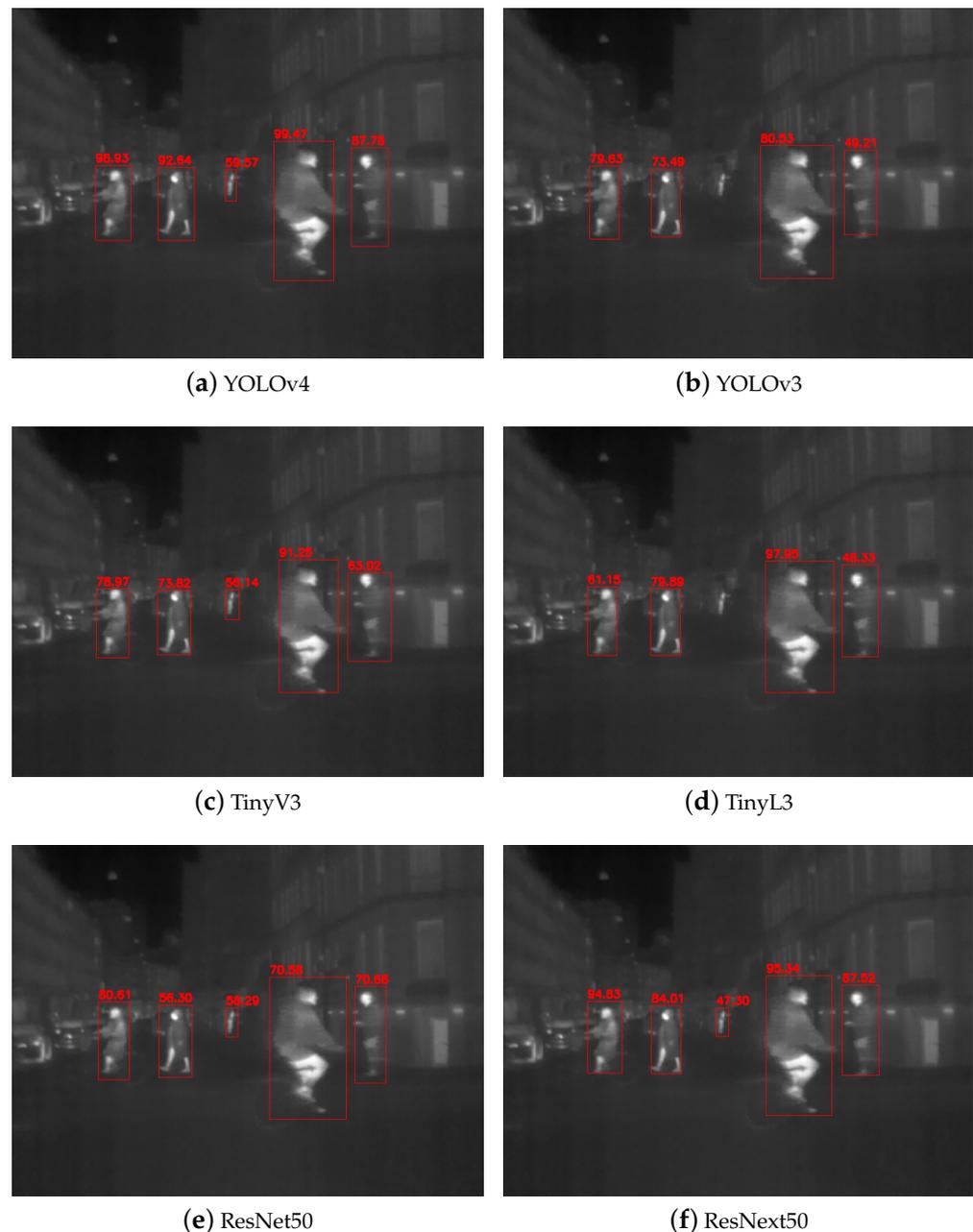


Figure 8. Comparison of detector performance after augmentation. Frame taken from ZUT dataset during the rain.

5. Discussion

The main idea behind this paper was to propose an efficient data augmentation solution for simulating bad weather conditions that impact the image received from the FIR sensor. Fortunately, our expectations that DnCNN will perform augmentations with inverted functionality and cause experimental investigation results confirmed more efficient detector training. The performed research also showed that DnCNN could help solve histogram equalization tasks, brightness and contrast auto-tune, and contour segmentation, essential for pedestrian detection in the FIR domain.

The analysis of annotations' heat-map in our fused FIR dataset with an exceptional amount of labeled frames and pedestrians gave us an idea to add a nonlinear decision threshold for the detector. This comparatively simple upgrade of detectors increased mAP for all tested state-of-the-art detectors (ready for real-time applications) without noticeable processing speed degradation.

The confidence heat-map could be extended for advanced driver-assistance system (ADAS) application usage in situations when severe weather conditions occur. For example, a dedicated confidence heat-map could be designed for rain, fog, frost, and snow conditions and applied to minimize false detections. In this way, a single neural network could benefit in multiple situations.

6. Conclusions

In the paper, we proposed a new field of DnCNN application. We demonstrated a novel approach of severe weather feature extraction and augmentation using DnCNN.

In our investigations, we merged the two biggest FIR domain datasets, ZUT and SCUT, into one. We analyzed the heat-maps taken from these datasets and employed them to identify the properties of annotations. Heat-maps, estimated from the combined dataset, revealed the detectors' locations with errors. We have successfully applied this information for detector accuracy investigations, allowing us to modify a detector Head to minimize a false detection. Finally, accuracy increased by 9.38 mAP.

The DnCNN application for efficient severe weather feature extraction and existing sample transformation proved its value in pedestrian detection using state-of-the-art detectors, evaluated under wide variety of situations. The detectors' training on merged and augmented SCUT and ZUT datasets demonstrated high potential. A wider variety of annotations resulted in a more robust detector capable of working in a broader range of weather conditions and situations, providing up to 87.2 mAP, ready for real-time applications. From the results presented, it can be concluded that data synthesis could contribute to other distortions generation, reflecting different weather conditions.

Author Contributions: Conceptualization and methodology, all authors; validation, P.T.; analysis, A.S. and A.N.; writing—original draft preparation, P.T. and A.S.; project administration, A.N.; supervision, A.N.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look Once
GAN	Generative Adversarial Network
FPS	Frames per Second
DnCNN	Denosing Convolutional Neural Network
FIR	Far Infrared
NIR	Near Infrared
MWIR	Mid-Wave Infrared
SWIR	Long-Wave Infrared
ADAS	Advanced driver-assistance systems
WHO	World Health Organization
mAP	mean Average Precision
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

References

1. European Regional Status Report on Road Safety. 2019. Available online: <https://www.euro.who.int/en/publications/abstracts/european-regional-status-report-on-road-safety-2019/> (accessed on 3 January 2021).

2. Khan, M.A.; Khan, S.F. IoT based framework for Vehicle Over-speed detection. In Proceedings of the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 4–6 April 2018; IEEE: New York, NY, USA, 2018; pp. 1–4. [\[CrossRef\]](#)
3. Kumar, T.; Kushwaha, D.S. An efficient approach for detection and speed estimation of moving vehicles. *Procedia Comput. Sci.* **2016**, *89*, 726–731. [\[CrossRef\]](#)
4. Breen, J.M.; Næss, P.A.; Hansen, T.B.; Gaarder, C.; Stray-Pedersen, A. Serious motor vehicle collisions involving young drivers on Norwegian roads 2013–2016: Speeding and driver-related errors are the main challenge. *Traffic Inj. Prev.* **2020**, *21*, 382–388. [\[CrossRef\]](#)
5. Das, S.; Brimley, B.K.; Lindheimer, T.E.; Zupancich, M. Association of reduced visibility with crash outcomes. *IATSS Res.* **2018**, *42*, 143–151. [\[CrossRef\]](#)
6. Sun, X.; Hu, H.; Habib, E.; Magri, D. Quantifying crash risk under inclement weather with radar rainfall data and matched-pair method. *J. Transp. Saf. Secur.* **2011**, *3*, 1–14. [\[CrossRef\]](#)
7. Suisse, C. ADAS System Revenue Growth Worldwide between 2015 and 2025, by Sub Component [Graph]. 2016. Available online: <https://www.statista.com/statistics/678634/adas-system-revenue-growth-by-sub-component/> (accessed on 3 January 2021).
8. Kuperman, V.; Brandwine, N. Compact Infrared Camera for Automotive Safety and Driving Systems. U.S. Patent 10,175,112, 14 December 2017.
9. Ulman, S.; Amotz, K.; Sterin, M.; Livny, O.; Itamar, A. Method and System for Generating Multiple Synchronized Thermal Video Streams for Automotive Safety and Driving Systems. U.S. Patent App. 15/961,125, 2019.
10. Kuybeda, O.; Madar, E. Techniques for Correcting Fixed Pattern Noise in Shutterless FIR Cameras. U.S. Patent 10,511,793, 17 December 2019.
11. Kuybeda, O. Scene-Based Nonuniformity Correction Using a Convolutional Recurrent Neural Network. U.S. Patent App. 16/275,577, 2019.
12. Terre, B.; Kostrzewa, J.; Källhammer, J.E.; Höglund, T. Infrared Camera Systems and Methods. U.S. Patent 7,340,162, 2008.
13. Terre, B.; Kostrzewa, J.; Källhammer, J.E.; Höglund, T. Infrared Camera Systems and Methods Using Environmental Information. U.S. Patent 8,077,995, 13 December 2011.
14. FLIR Path Finder Kit. Available online: http://www.safetyvision.com/sites/safetyvision.com/files/FLIR_PathFindIRII_User_Guide_1.pdf (accessed on 11 June 2019).
15. Forslund, D.; Bjärkefur, J. Night vision animal detection. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 737–742. [\[CrossRef\]](#)
16. Tsimhoni, O.; Bärgrman, J.; Flannagan, M.J. Pedestrian detection with near and far infrared night vision enhancement. *Leukos* **2007**, *4*, 113–128. [\[CrossRef\]](#)
17. Yagi, S.; Kobayashi, S.; Inoue, T.; Hori, T.; Michiba, N.; Okui, K. *The Development of Infrared Projector*; Technical Report; SAE Technical Paper: Warrendale, PA, USA, 2003. [\[CrossRef\]](#)
18. Bertozzi, M.; Fedriga, R.I.; Miron, A.; Reverchon, J.L. Pedestrian detection in poor visibility conditions: Would SWIR help? In Proceedings of the International Conference on Image Analysis and Processing, Naples, Italy, 9–13 September 2013; pp. 229–238.
19. Miron, A.D.; Benshair, A.; Fedriga, R.I.; Broggi, A. SWIR images evaluation for pedestrian detection in clear visibility conditions. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, 6–9 October 2013; pp. 354–359.
20. Negied, N.K.; Hemayed, E.E.; Fayek, M.B. Pedestrians’ detection in thermal bands—Critical survey. *J. Electr. Syst. Inf. Technol.* **2015**, *2*, 141–148. [\[CrossRef\]](#)
21. Nguyen, C.T.; Havlicek, J.P.; Fan, G.; Caulfield, J.T.; Pattichis, M.S. Robust dual-band MWIR/LWIR infrared target tracking. In Proceedings of the 2014 48th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2–5 November 2014; pp. 78–83.
22. Teutsch, M.; Muller, T.; Huber, M.; Beyerer, J. Low resolution person detection with a moving thermal infrared camera by hot spot classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 209–216.
23. Van Beeck, K.; Van Engeland, K.; Vennekens, J.; Goedemé, T. Abnormal behavior detection in LWIR surveillance of railway platforms. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.
24. González, A.; Fang, Z.; Socarras, Y.; Serrat, J.; Vázquez, D.; Xu, J.; López, A.M. Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors* **2016**, *16*, 820. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391. [\[CrossRef\]](#)
26. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.

29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
31. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
32. Liu, S.; Huang, D.; Wang, Y. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
33. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
34. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
35. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
36. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
37. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37. [[CrossRef](#)]
38. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
39. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [[CrossRef](#)]
40. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
41. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *arXiv* **2016**, arXiv:1605.06409.
42. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. Available online: <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html> (accessed on 1 April 2021).
43. Hoang Ngan Le, T.; Zheng, Y.; Zhu, C.; Luu, K.; Savvides, M. Multiple scale faster-rcnn approach to driver’s cell-phone usage and hands on steering wheel detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 46–53. [[CrossRef](#)]
44. Kaarmukilan, S.; Poddar, S.; Thomas, K.A. FPGA based Deep Learning Models for Object Detection and Recognition Comparison of Object Detection Comparison of object detection models using FPGA. In Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 11–13 March 2020; pp. 471–474.
45. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A survey of deep learning-based object detection. *IEEE Access* **2019**, *7*, 128837–128868. [[CrossRef](#)]
46. Soviany, P.; Ionescu, R.T. Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction. *arXiv* **2018**, arXiv:1803.08707.
47. Dollár, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 743–761. [[CrossRef](#)]
48. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The kitti vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361. [[CrossRef](#)]
49. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *arXiv* **2014**, arXiv:1409.0575.
50. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
51. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2014**, arXiv:1405.0312.
52. YOLOv2 Comparison versus YOLOv3. Available online: https://pyimagesearch.com/wp-content/uploads/2020/01/tiny_yolo_comparison.png (accessed on 3 January 2021).
53. Choi, Y.; Kim, N.; Hwang, S.; Park, K.; Yoon, J.S.; An, K.; Kweon, I.S. KAIST multi-spectral day/night data set for autonomous and assisted driving. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 934–948. [[CrossRef](#)]
54. Shopovska, I.; Jovanov, L.; Philips, W. Deep Visible and Thermal Image Fusion for Enhanced Pedestrian Visibility. *Sensors* **2019**, *19*, 3727. [[CrossRef](#)] [[PubMed](#)]
55. Jegham, I.; Khalifa, A.B. Pedestrian detection in poor weather conditions using moving camera. In Proceedings of the 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia, 30 October–3 November 2017; pp. 358–362. [[CrossRef](#)]
56. Takumi, K.; Watanabe, K.; Ha, Q.; Tejero-De-Pablos, A.; Ushiku, Y.; Harada, T. Multispectral object detection for autonomous vehicles. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, Mountain View, CA, USA, 23–27 October 2017; pp. 35–43. [[CrossRef](#)]

57. FREE FLIR Thermal Dataset for Algorithm Training. Available online: <https://www.flir.com/oem/adas/adas-dataset-form/> (accessed on 15 February 2021).
58. Chen, Y.; Shin, H. Pedestrian Detection at Night in Infrared Images Using an Attention-Guided Encoder-Decoder Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 809. [[CrossRef](#)]
59. Socarras, Y.; Ramos, S.; Vázquez, D.; López, A.; Gevers, T. Adapting Pedestrian Detection from Synthetic to Far Infrared Images. In Proceedings of the ICCV Workshop on Visual Domain Adaptation and Dataset Bias (VisDA), Sydney, Australia, 7 December 2013.
60. FLIR Thermal Sensing for ADAS. Available online: <https://www.flir.com/oem/adas/> (accessed on 11 June 2019).
61. Khellal, A.; Ma, H.; Fei, Q. Pedestrian classification and detection in far infrared images. In Proceedings of the International Conference on Intelligent Robotics and Applications, Portsmouth, UK, 24–27 August 2015; pp. 511–522.
62. Davis, J.W.; Keck, M.A. A two-stage template approach to person detection in thermal imagery. In Proceedings of the 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1, Breckenridge, CO, USA, 5–7 January 2005. [[CrossRef](#)]
63. Miron, A.D. Multi-Modal, Multi-Domain Pedestrian Detection and Classification: Proposals and Explorations in Visible over StereoVision, FIR and SWIR. Ph.D. Thesis, INSA de Rouen, Université Babes-Bolyai, Cluj-Napoca, Romania, 2014.
64. Terravic Motion IR Database. Available online: <http://vcipl-okstate.org/pbvs/bench/> (accessed on 11 June 2019).
65. Xu, Z.; Zhuang, J.; Liu, Q.; Zhou, J.; Peng, S. Benchmarking a large-scale FIR dataset for on-road pedestrian detection. *Infrared Phys. Technol.* **2019**, *96*, 199–208. [[CrossRef](#)]
66. Tumas, P.; Nowosielski, A.; Serackis, A. Pedestrian Detection in Severe Weather Conditions. *IEEE Access* **2020**, *8*, 62775–62784. [[CrossRef](#)]
67. Tiny YOLOv3 Configuration with 3 Layers. Available online: https://github.com/AlexeyAB/darknet/blob/master/cfg/yolov3-tiny_3l.cfg (accessed on 3 May 2020).
68. ResNet50 Configuration. Available online: <https://raw.githubusercontent.com/WongKinYiu/CrossStagePartialNetworks/master/in%20progress/csresnet50-panet-spp.cfg> (accessed on 3 May 2020).
69. ResNext50 Configuration. Available online: <https://raw.githubusercontent.com/AlexeyAB/darknet/master/cfg/csresnext50-panet-spp.cfg> (accessed on 3 May 2020).
70. Darknet YOLO Implementation. Available online: <https://github.com/AlexeyAB/darknet/commit/dcf3ea30f195e0ca1210d580cac8b91b6beaf3f7> (accessed on 3 January 2020).
71. Xu, Q.; Zhang, C.; Zhang, L. Denoising Convolutional Neural Network. In Proceedings of the 2015 IEEE International Conference on Information and Automation, Lijiang, China, 8–10 August 2015; pp. 1184–1187. [[CrossRef](#)]
72. Jafari, M.; Auer, D.; Francis, S.; Garibaldi, J.; Chen, X. DRU-net: An Efficient Deep Convolutional Neural Network for Medical Image Segmentation. *arXiv* **2020**, arXiv:2004.13453.
73. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [[CrossRef](#)]
74. Zhang, K.; Li, Y.; Zuo, W.; Zhang, L.; Gool, L.V.; Timofte, R. Plug-and-Play Image Restoration with Deep Denoiser Prior. *arXiv* **2020**, arXiv:2008.13751.