Performance Evaluation of CNN Based Pedestrian and Cyclist Detectors On Degraded Images

Tomáš Zemčík *zemcikt@feec.vutbr.cz*
Faculty of Electrical Engineering and Communications
Dept. of Automation and Instrumentation
Brno University of Technology, Brno, Czech Republic

Lukáš Kratochvíla *kratochvila@feec.vutbr.cz*
Faculty of Electrical Engineering and Communications
Dept. of Automation and Instrumentation
Brno University of Technology, Brno, Czech Republic

Šimon Bilík *bilik@feec.vutbr.cz*
Faculty of Electrical Engineering and Communications
Dept. of Automation and Instrumentation
Brno University of Technology, Brno, Czech Republic

Ondřej Boštík *bostik@feec.vutbr.cz*
Faculty of Electrical Engineering and Communications
Dept. of Automation and Instrumentation
Brno University of Technology, Brno, Czech Republic

Pavel Zemčík *zemcik@fit.vutbr.cz*
Faculty of Information Technology
Dept. of Computer Graphics and Multimedia
Brno University of Technology, Brno, Czech Republic

Karel Horák *horak@feec.vutbr.cz*
Faculty of Electrical Engineering and Communications
Dept. of Automation and Instrumentation
Brno University of Technology, Brno, Czech Republic

Abstract

This paper evaluates the effects of input image degradation on performance of image object detectors. The purpose of the evaluation is to determine usability of the detectors trained on original images in adverse conditions. SSD and Faster R-CNN based pedestrian and cyclist detector performance with images degraded with motion blur, out-of-focus blur, and JPEG compression artefacts, most commonly occurring in mobile or static traffic systems. An experiment was designed to assess the effect of degradations on detection precision and cross class confusion. The paper describes the two datasets created for this evaluation, evaluation of a number of detectors on increasingly more degraded images, comparison of their performance, and assessment of their tolerance to different types of image degradation as well as a discussion of the results.

Keywords: Object Detection, Image Degradation, Pedestrian Detection, Cyclist Detection, SSD, Faster R-CNN.

1. INTRODUCTION

Image object detection is one of the keys to solving technical problems in machine vision-based traffic application. CNN object detection architectures currently seem to be the most promising ones towards robust and reliable image object detection in very difficult traffic environments. With the latest advances in this field, the pressures of regulatory bodies to implement them into traffic safety improving applications is higher than ever, as demonstrated by EU's "Vision Zero" plan to have zero road accident deaths by 2050 [1]. Pedestrians and cyclists represent the most vulnerable road users making up for approximately 26% of all road accident-related deaths in 2018 according to WHO [2]. In the EU this number is even higher, at 32%, so it is no surprise that detecting them is a central issue in Autonomous Vehicles (AV), Advanced Driver Assistance Systems (ADAS), Intelligent Traffic Systems (ITS), and Intelligent Vehicle Systems (IVS). With the shift towards more automated transportation a lot of research in this area is being done, such as [3 - 5].

From the semantic point of view, both pedestrians and cyclists are very similar, as both classes represent humans in different poses with or without a bicycle. Therefore, their appearance in images is very similar. This similarity means that common features exist and can aid detector training. On the other hand, while similarities exist across classes, also an enormous variety occurs within individual instances of each class making distinguishing between pedestrians and cyclists, especially in certain orientations, a challenging task [6].

Despite the aforementioned similarity in appearance, however, for a higher-level AV or ADAS system, they represent dramatically different objects that have very different behavioural patterns in traffic and as such need to be dealt with very differently in the trajectory planning algorithms, thus it is needed to have a robust and accurate detector.

2. OBJECT DETECTION IN TRAFFIC APPLICATIONS

Since the publication of pedestrian targeted Viola Jones detector in 2003 [7], which is by many regarded the first pedestrian-in-image detector [8], major advances in the field have been made. In consequent years, the attention has shifted from hand-selected Haar-like features and boosted classifiers through HOG features using SVM classifiers [9, 10] to deformable part models [11]. In recent years, however, the most attention is concentrated on Deep Learning models based around CNNs. The DL approach finally allows for the feature extraction to be trained along with the classifiers [8].

Up until the rise of CNNs, detection methods of pedestrians and cyclists (and other classes) have been largely kept separate requiring multiple runs of different algorithms on a single image to detect various classes of road users [12]. With the capabilities of modern CNN meta-architectures, no reason not to implement multiclass capability into today's detectors exists anymore and with deployed AV and ADAS, this is a de-facto standard [12].

In practical deployment, ideal conditions rarely apply and many applications need to perform well even with poor input image quality. This paper evaluates performance of state-of-the-art object detection architectures on varying quality datasets. We do not intend to create a detector to compete with the state-of-the-art detectors.

Effects of image degradation on various CNN classification architectures were previously studied [13, 14]. With results showing that while all tested architectures were affected, not all are affected by the same way, with Roy at al. [13] concluding deeper networks are affected more. Pei et al. [14] concluded that while algorithms exist for removing the tested degradations from image (motion blur and haze), the improvement in CNN classification performance was insignificant. The effect of image degradation on object detection networks has not been that extensively studied.

Single Shot multibox Detector (SSD) and Faster Region-based Convolutional Neural Network (R-CNN) have been selected to be the candidates based on performance as measured by Tilgner [15].

2.1 R-CNN Object Detector Family
Girshick et al. have presented their Region-based Convolutional Neural Network (R-CNN) in 2014 [16]. The R-CNN workflow consists of 3 steps, where in the first one, an input image is divided into 2000 regions-of-interest (RoI). Consequently, all 2000 subimages are run through a CNN feature extraction and finally classified via SVM.

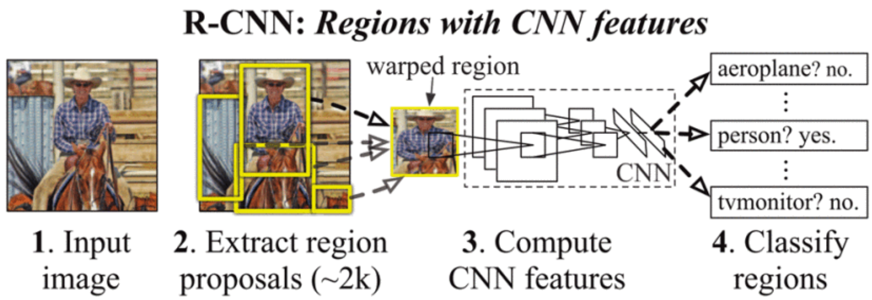

FIGURE 1: R-CNN Workflow [16].

R-CNN outperformed most of its contemporaries in terms of accuracy but running the CNN on all RoIs separately is obviously rather computationally expensive (tens of seconds per image), so primarily to address this drawback, Girshick introduced Fast R-CNN [17] in 2015.

Fast R-CNN improves upon the original method by only running the CNN once on each input image generating a convolutional feature map, by adding bounding box regression into the network, and then by resizing the selected region of interest in the feature map to be classified using a fully connected neural network. These improvements managed to improve computational performance by an order of magnitude without any significant loss in detection accuracy [17].

In 2016 Ren et al. published another major improvement in the R-CNN object detector family - the Faster R-CNN [18]. Similarly to Fast R-CNN, the convolutional feature map is only computed once per image but unlike in the above method, the region proposals are handled by a custom Region Proposal Network (RPN). These changes add the benefit of the method being implementable in GPUs and such acceleration improves the computational performance by yet another order of magnitude into the frame rate range of units of frames per second (FPS), making it a viable option for real time detection.

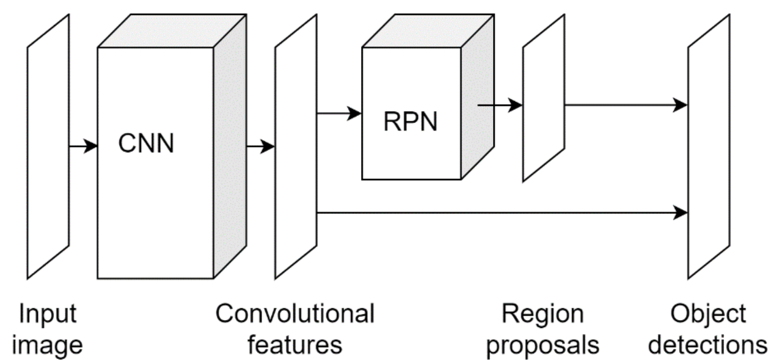

FIGURE 2: Faster R-CNN Workflow.

2.2 SSD Object Detector
Single Shot Multibox Detector (SSD) was published by Liu et al. in 2016 as a competitive real-time object detection method and at that time, it outclassed its contemporaries including Faster R-CNN both in terms of mAP and framerate on VOC2007 [19].

SSD uses a pre-trained image classification CNN base as a feature extractor to which further progressively smaller convolutional layers are attached. This progressive decrease in feature map size means that objects of very different scales can be detected [19]. As opposed to preceding object detection approaches, single shot detectors like YOLO [20] and SSD [19] do away with a trained region proposal approach and generally apply a variant of a grid based RoI generator, in SSD, a set of default bounding boxes is generated for each feature map. These default boxes are of pre-selected aspect ratios in order to accommodate different object classes [19].

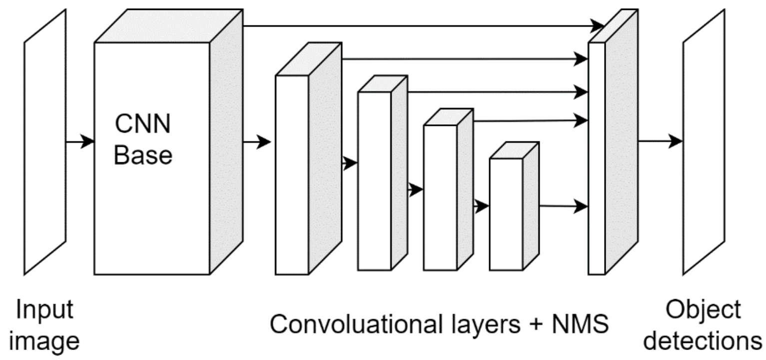

FIGURE 3: SSD Workflow (NMS - Non Maximum Suppression).

From all the default boxes classification, scores for all classes are extracted and Non Maximum Suppression is used to select only the strongest of the closely located detections.

SSD, when first published, outperformed Faster R-CNN on Pascal VOC and COCO datasets both in mAP and framerate. SSD300 performed with mAP 74.3% compared to 73.2% of Faster R-CNN, at 46 FPS compared to 7 FPS [19].

3. IMAGE DEGRADATIONS

In traffic images, different image degradations appear in many cases, such as JPEG compression artifacts, motion blur, and out-of-focus blur, that have been selected as the representative ones.

JPEG compression is still one of the most widely used compression algorithms. Based on 8x8 block DCT, the compression leaves rather prominent artifacts in images compressed at higher compression rates [21]. These artifacts are mostly ringing (caused by missing higher frequencies as more of the DCT is truncated) and block-to-block discontinuities. Although more modern and better compression standards have since been introduced, JPEG is still used across the industry especially in static traffic monitoring systems where lowering data sizes is important. Methods exist for JPEG artifact removal [22], but they require additional computational costs.

Motion blur is a blur caused by objects in the imaging moving too fast relatively to the exposure time. Motion blur affects mobile and static traffic systems similarly, while in static installations it is the objects to be detected that can move, in mobile systems it can be both motion of the camera and motion of the scene at the same time causing the blur. Similarly to the JPEG artifact removal, motion blur removal has been studied and removal methods are available [23].

Out-of-focus blur is the final type of degradation discussed. It is caused by objects in the scene being outside of the camera's depth-of-field. This type of degradation affects static traffic cameras the most. This type of blur can be readily removed as well [24].

As stated above, all selected degradations are to an extent removable and their avoidance or removal is advisable. Should they still occur, their removal might be computationally costly and time consuming and that is why investigating their effect on object detection is important.

4. DATASETS & EXPERIMENTS

To assess performance of the selected architectures, an experiment was devised where pedestrian and cyclist detectors trained on undegraded data applied on newly compiled datasets with varying magnitude of degradation.

Detectors were evaluated in terms of their average precision for each class. Average precision being used as defined by the Pascal VOC challenge [25, 26], and in terms of their cross-class confusion (cases where a pedestrian is detected to be a cyclist and vice versa).

4.1 Object Detectors

Two object detection models were trained of the selected architectures. The models trained by the authors were trained to detect both pedestrians and cyclists, while those selected for comparison only detect pedestrians.

A Faster R-CNN based detector has been selected for evaluation. The particular variant selected uses Resnet101 [27] as a feature extractor. The detector was trained on a combination of the left colour image of the stereo pair in the KITTI 2012 stereo dataset [28] and authors' own pedestrian dataset. Data augmentation using horizontal flip and limited skew were used to further improve the training. The training has been done through Tensorflow Object Detection API [29]. We will further refer to this detector as *FRCNN Zemcik*. For comparison, two more faster R-CNN detectors were used, *FRCNN Tilgner* [15] and *FRCNN Kitti zoo* [29], both using Resnet101 as a feature extractor.

Representing SSD architecture an SSDLite model *SSD Zemcik* has been trained on the same data. The model utilizes MobileNetV2 as the CNN backbone [30]. The same architecture trained by Tilgner will be used for comparison as *SSD Tilgner* [15].

4.2 Datasets

Testing datasets were collected to assess quality of all measured detectors. The datasets were designed to be representative of a multitude of traffic situation and individual images were not selected to purposely help or hinder detection. Two datasets have been compiled for this purpose:

Small Pedestrian and Cyclist Dataset - SPCD - consists of 200 images gathered using a state-of-the-art mid-range dashcam. The image data has been gathered in urban and suburban conditions in Brno, Czech Republic. This dataset only contains daytime images and all images contain at least one instance of a pedestrian or a cyclist.

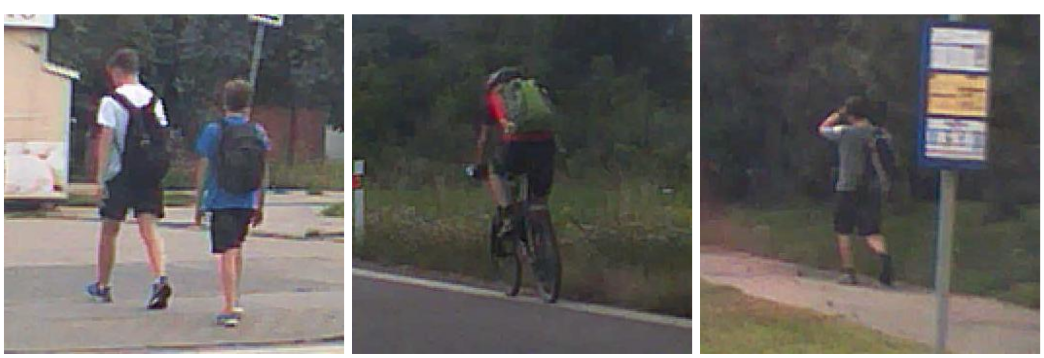

FIGURE 4: Selected Images from SPCD.

The Low Quality Traffic Dataset – LQTD - consists of over 400 images taken using a purposely selected low resolution low image quality dashcam with further image degradations, such as item reflections on the windscreen or more dirty windscreen achieved by placing the camera out of the wiper arch. Similarly to the SPCD, it was gathered in daytime and all weather. Majority of the images contain an instance of a pedestrian, a cyclist or both.

FIGURE 5: Selected Images from LQTD.

All three types of degradations were applied on all images of both datasets consecutively in varying severity. JPEG compression was applied with quality graduated from 100% to 10% to achieve images with increasing severity of JPEG artifact degradation.

FIGURE 6: JPEG compressed image with quality of 100%, 50%, and 10% respectively.

Motion blur was applied using a horizontal linear averaging filter from 2 to 20 pixels in width.

FIGURE 7: Motion blurred images with kernels of 0px, 10px and 20px.

The out-of-focus blur was approximated by application of a linear averaging filter with circular aperture simulating a Point Spread Function of an out-of-focus lens. The circle radius was selected from 1 to 10 pixels to correspond to the width of the motion blur kernel size.

FIGURE 8: Out-of-focus blurred images with kernel radius of 0px, 5px and 10px.

JPEG (quality)	**100%**	**90%**	**80%**	**70%**	**60%**	**50%**	**40%**	**30%**	**20%**	**10%**
SPCD	1.000	0.987	0.980	0.973	0.961	0.960	0.950	0.926	0.870	0.811
LQTD	1.000	0.975	0.962	0.952	0.941	0.933	0.923	0.906	0.878	0.811
Motion (kernel)	**2px**	**4px**	**6px**	**8px**	**10px**	**12px**	**14px**	**16px**	**18px**	**20px**
SPCD	0.973	0.957	0.934	0.904	0.875	0.851	0.830	0.812	0.797	0.784
LQTD	0.966	0.940	0.908	0.880	0.857	0.837	0.821	0.808	0.797	0.787
OoF (kernel rad)	**1px**	**2px**	**3px**	**4px**	**5px**	**6px**	**7px**	**8px**	**9px**	**10px**
SPCD	0.973	0.936	0.882	0.830	0.767	0.732	0.706	0.681	0.653	0.650
LQTD	0.957	0.902	0.834	0.787	0.737	0.708	0.686	0.664	0.648	0.635

TABLE 1: Mean Structural Similarity Index Measure for all degradations on both datasets (SPCD and LQTD) with respect to JPEG compression quality, motion blur kernel width and out-of-focus blur kernel radius.

The severity of the degradation was quantified using Structural Similarity Index Measure (SSIM) [31]. Table 1 gives the mean SSIM for both SPCD and LQTD datasets with each degradation severity.

5. RESULTS

The results of the experiments described above are presented in the following charts.

Figure 9 gives the class average precision as per the Pascal VOC metric [25, 26] for each classifier using both datasets. Pedestrian detectors are shown in solid line, while cyclist detectors are in dashed line. In all cases, the severity of the degradation increases from right to left.

Figure 10 gives the cross-class confusion for each of the classifiers. The confusion was quantified using the same Pascal VOC average precision metric, only the pedestrian and cyclist ground truth labels were switched. The solid line represents cyclists in the image that were erroneously detected and classified as pedestrians, while dashed line represents pedestrians erroneously labelled as cyclists.

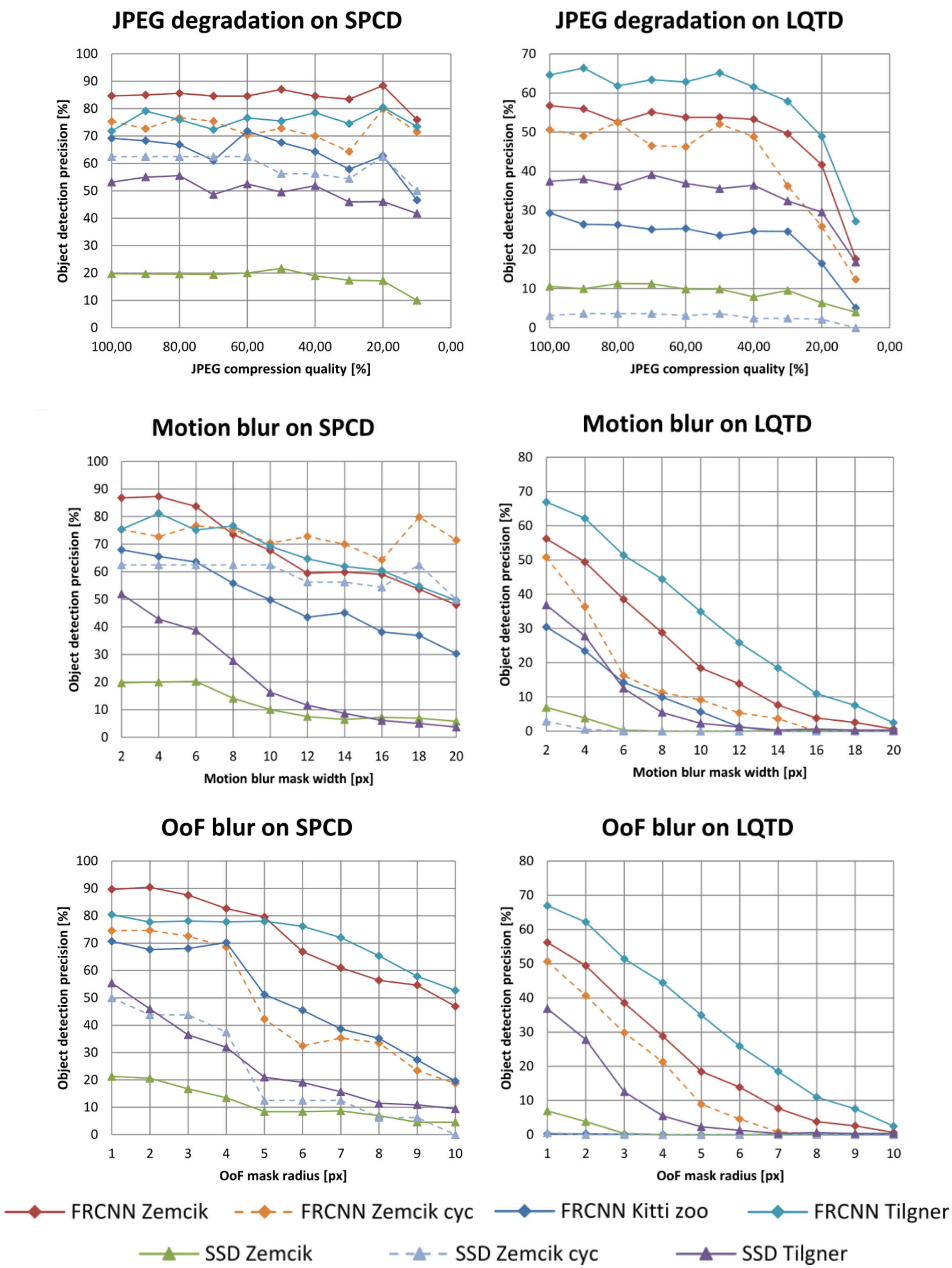

FIGURE 9: Average precision of all classifiers on both datasets.

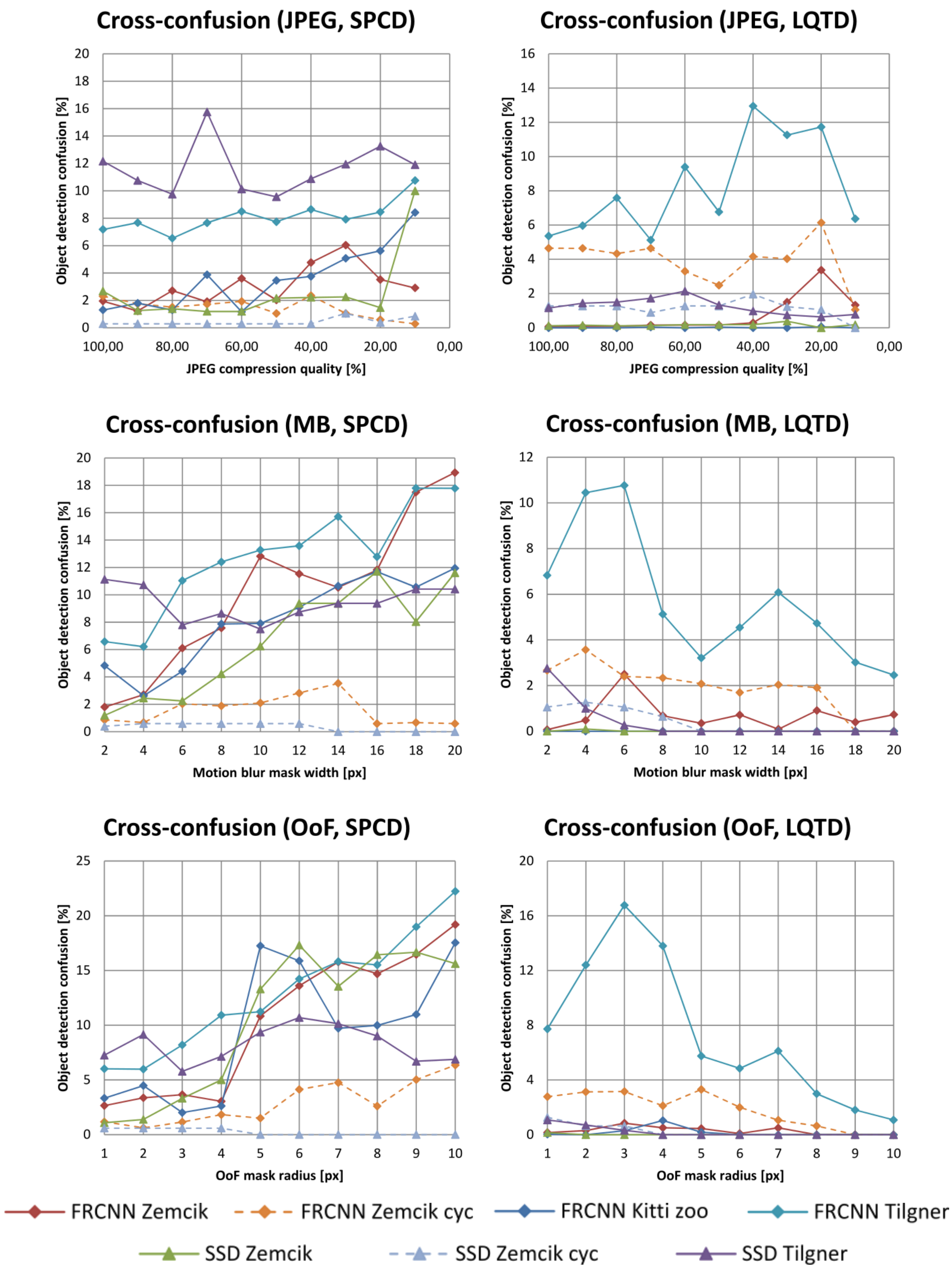

FIGURE 10: Cross-confusion of all classifiers on both datasets.

From figure 9 it is apparent that in all cases additional degradation reduces overall performance of all object detectors. It is interesting to note that while pedestrian detection precision decreases comparably for motion blur and out-of-focus blur in both datasets respectively, in the SPCD dataset the cyclist detection precision is much more affected by out-of-focus blur than motion blur.

Figure 10 shows that in the LQTD dataset there is no apparent trend for cross-class confusion to increase with increasing severity of the image degradation. On the SPCD dataset however, it is apparent that cross-class confusion in case of cyclists erroneously labelled as pedestrians increases significantly (by as much as 15% increase in Pascal VOC defined average precision for inverted labels) with increasing kernel sizes of both motion blur and the out-of-focus blur.

As mentioned above the cross-class confusion is not as critical as the detection itself, but it still has a bearing on how a higher-level path planning system sees the detected object, it is therefore desirable to keep the confusions to a minimum.

Knowing the extent of data degradation robustness the architectures have, it is now possible to design the ADAS systems so that the more performance hindering degradations are eliminated in the design (especially in the image acquisition and pre-processing stages). This also makes it possible to save resources trying to eliminate certain degradations that after all are not critical to the performance of the system (such as high quality JPEG compression).

6. CONCLUSIONS

In this paper, effects of image degradation were evaluated on selected object detection CNN architectures trained for detecting pedestrians and cyclists in typical traffic imagery. Not surprisingly, the performance of both SSD and Faster R-CNN based detectors dropped with increasing degradation severity.

The measurements show that artifacts of JPEG compression do not degrade the performance significantly with high quality - low compressions and only take effect with the rarely used high compressions. The effect was also much more pronounced with the lower resolution dataset because of the fixed block size used in JPEG compression.

The effects of motion blur were again more severe on the lower resolution LQTD dataset. On the higher resolution SPCD dataset, a significant growth of pedestrian and cyclist confusions occurred in both directions with the increase in the magnitude of the motion blur. The effects of out-of-focus blur were very similar to the effect of the motion blur.

Better understanding of the effects of image degradation on CNN based object detectors will help create more robust and efficient ADAS systems and will help push us one step closer towards fully automated transportation systems.

In future work, we would like to further investigate the effects of image degradation on traffic object detection, and we would also like to investigate how different approaches to the degradation effect mitigation (data augmentation with degradations pre-training and image degradation removal before detection) compare, and we would like to implement the selected degradation effect mitigation mitigating solutions in a practical system.

7. REFERENCES

[1] European Commission (2019). EU Road Safety Policy Framework 2021-2030 - Next steps towards "Vision Zero". Luxembourg: Office for Official Publications of the European Communities.

[2] World Health Organization (2018). Global Status Report on Road Safety 2018 [Online]. Available:https://www.who.int/publications/i/item/9789241565684#:~:text=The%20Global%20status%20report%20on,people%20aged%205%2D29%20years. [Nov 15, 2020]

[3] V. Viswanathan & R. Hussein (2017, Apr.). Applications of Image Processing and Real-Time embedded Systems in Autonomous Cars: A Short Review, *International Journal of Image Processing (IJIP),* [Online] *Volume (11) : Issue (2)*. Available: https://www.cscjournals.org/manuscript/Journals/IJIP/Volume11/Issue2/IJIP-1114.pdf. [Jan 23, 2021]

[4] S. Kim, I. Wisanggeni, R. Ros & R. Hussein (2020, Jan.). Detecting Fatigue Driving Through PERCLOS: A Review, *International Journal of Image Processing (IJIP),* [Online] *Volume (14) : Issue (1) : 2020*. Available: https://www.cscjournals.org/manuscript/Journals/IJIP/Volume14/Issue1/IJIP-1194.pdf. [Jan 23, 2021]

[5] Z. Cheng, L. Jeng & K. Li (2018, Nov.) Behavioral Classification of Drivers for Driving Efficiency Related ADAS Using Artificial Neural Network, *2018 IEEE International Conference on Advanced Manufacturing (ICAM),* [Online] *pp. 173-176*. Available: https://ieeexplore.ieee.org/document/8614836. [Nov 15, 2020]

[6] O. Fard & S. Hamidreza. Efficient multi-class object detection with a hierarchy of classes. Aubière, 2015, pp. 36-40.

[7] Viola, Jones, & Snow (2003, Oct.). Detecting pedestrians using patterns of motion and appearance. In *Proceedings Ninth IEEE International Conference on Computer Vision* [Online]. (pp. 734-741 vol.2). Available: https://ieeexplore.ieee.org/document/1238422 [Nov 15, 2020]

[8] R. Benenson, M. Omran, J. Hosang, & B. Schiele. (2014, Nov.). Ten Years of Pedestrian Detection, What Have We Learned?. [Online]. Available: https://arxiv.org/abs/1411.4304 [Nov 13, 2020]

[9] N. Dalal, & B. Triggs (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 886-893 vol. 1).

[10] Wojek C., Schiele B. (2008) A Performance Evaluation of Single and Multi-feature People Detection. In: Rigoll G. (eds) Pattern Recognition. *DAGM 2008. Lecture Notes in Computer Science, vol 5096. Springer, Berlin, Heidelberg.* [Online]. Available: https://doi.org/10.1007/978-3-540-69321-5_9 [Nov 23, 2020]

[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, & D. Ramanan (2010). Object Detection with Discriminatively Trained Part-Based Models *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(9), 1627-1645.

[12] W. Tian, & M. Lauer (2015, Sep.). Fast Cyclist Detection by Cascaded Detector and Geometric Constraint. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems* [Online]. (pp. 1286-1291). Available: https://ieeexplore.ieee.org/document/7313303 [Dec 05, 2020]

[13] Roy, P., Ghosh, S., Bhattacharya, S., & Pal, U. (2018, Jul.). Effects of Degradations on Deep Neural Network Architectures. [Online]. Available: https://arxiv.org/abs/1807.10108. [Dec 01, 2020]

[14] Y. Pei, Y. Huang, Q. Zou, X. Zhang, & S. Wang (2019, Nov.). Effects of Image Degradation and Degradation Removal to CNN-based Image Classification, *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, [Online]. *1-1.* Available: https://ieeexplore.ieee.org/document/8889765. [Nov 18, 2020]

[15] M. Tilgner. “Pedestrians Detection in Traffic Environment by Machine Learning” Master’s thesis, Brno University of Technology, 2019

[16] R. Girshick, J. Donahue, T. Darrell, & J. Malik (2014, Oct.). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* [Online]. (pp. 580-587). Available: https://arxiv.org/abs/1311.2524. [Dec 01, 2020]

[17] R. Girshick (2015). Fast R-CNN. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 1440-1448).

[18] S. Ren, K. He, R. Girshick, & J. Sun (2016, Jan). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks *IEEE Transactions on Pattern Analysis and Machine Intelligence,* [Online]. *39*(6), 1137-1149. Available: https://arxiv.org/abs/1506.01497 [Dec 02, 2020]

[19] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., & Berg, A. (2016, Dec). SSD: Single Shot MultiBox Detector, *Lecture Notes in Computer Science*, [Online]. 21–37. Available: https://arxiv.org/abs/1512.02325. [Dec 15, 2020]

[20] Redmon, J., Divvala, S., Girshick, R.B., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.

[21] International Telecommunication Union (1992), T.81 – DIGITAL COMPRESSION AND CODING OF CONTINUOUS-TONE STILL IMAGES – REQUIREMENTS AND GUIDELINES. [Online]. Available: https://www.w3.org/Graphics/JPEG/itu-t81.pdf [Nov 12, 2020]

[22] P. Svoboda, M. Hradiš, D. Bařina, & P. Zemčík (2016, May.). Compression Artifacts Removal Using Convolutional Neural Networks, *Journal of WSCG*, [Online]. 24(2), 63–72. Available: https://arxiv.org/abs/1605.00366. [Nov 21, 2020]

[23] P. Svoboda, M. Hradiš, L. Maršík, & P. Zemčík (2016, Feb.). CNN for license plate motion deblurring. In *IEEE International Conference on Image Processing (ICIP)* [Online]. (pp. 3832–3836). IEEE Signal Processing Society. Available: https://arxiv.org/abs/1602.07873. [Dec 01, 2020]

[24] M. Hradiš, J. Kotera, P. Zemčík, & F. Šroubek (2015). Convolutional Neural Networks for Direct Text Deblurring. In *Proceedings of BMVC 2015* (pp. 1–13). *The British Machine Vision Association and Society for Pattern Recognition.*

[25] Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) challenge *International Journal of Computer Vision*, 88, 303-338.

[26] R. Padilla, S. L. Netto, & E. A. B. da Silva (2020, Jul.). A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* [Online] (pp. 237-242). Available: https://www.researchgate.net/publication/343194514_A_Survey_on_Performance_Metrics_for_Object-Detection_Algorithms. [Nov 15, 2020]

[27] K. He, X. Zhang, S. Ren, & J. Sun (2016, Dec.). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Online] (pp. 770-778). Available: https://arxiv.org/abs/1512.03385. [Dec 01, 2020]

[28] Andreas Geiger, Philip Lenz, & Raquel Urtasun (2012). Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR).*

[29] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., & Murphy, K. (2017, Apr.). Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. In *CVPR 2017* [Online] (pp. 3296-3297). Available: https://arxiv.org/abs/1611.10012. [Nov 15, 2020]

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, & L. Chen (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520).

[31] Zhou Wang, A. C. Bovik, H. R. Sheikh, & E. P. Simoncelli (2004, Apr.). Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing,* [Online] *13(4),* 600-612. Available: https://ieeexplore.ieee.org/document/1284395. [Dec 01, 2020]