IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 29, 2020, accepted July 17, 2020, date of publication July 22, 2020, date of current version August 5, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3011356*

# Toward Robust Pedestrian Detection With Data Augmentation

**SEBASTIAN CYGERT AND ANDRZEJ CZYŻEWSKI**
Multimedia Systems Department, Faculty of Electronics, Telecommunication, and Informatics, Gdańsk University of Technology, 80-233 Gdańsk, Poland
Corresponding author: Sebastian Cygert (sebastian.cygert@pg.edu.pl)

This work was supported in part by the Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdańsk University of Technology, and in part by the Polish National Centre for Research and Development (NCBR) through the European Regional Development Fund entitled: INFOLIGHT–Cloud-Based Lighting System for Smart Cities under Grant POIR.04.01.04/2019.

**ABSTRACT** In this article, the problem of creating a safe pedestrian detection model that can operate in the real world is tackled. While recent advances have led to significantly improved detection accuracy on various benchmarks, existing deep learning models are vulnerable to invisible to the human eye changes in the input image which raises concerns about its safety. A popular and simple technique for improving robustness is using data augmentation. In this work, the robustness of existing data augmentation techniques is evaluated to propose a new simple augmentation scheme where during training, an image is combined with a patch of a stylized version of that image. Evaluation of pedestrian detection models robustness and uncertainty calibration under naturally occurring corruption and in realistic cross-dataset evaluation setting is conducted to show that our proposed solution improves upon previous work. In this paper, the importance of testing the robustness of recognition models is emphasized and it shows a simple way to improve it, which is a step towards creating robust pedestrian and object detection models.

**INDEX TERMS** Convolutional neural network, pedestrian detection, robustness, style-transfer, data augmentation, uncertainty estimation.

## I. INTRODUCTION

In recent years visual recognition has witnessed a significant progress, mainly due to the introduction of Convolutional Neural Networks (CNN) and the availability of large scale datasets. Even though CNN based models surpassed human performance on some of the benchmarks [1], the application of deep learning methods in safety-critical applications like medicine or autonomous-vehicles has been limited [2]. This is due to the fact that CNNs often fail to generalize outside of the training data distribution.

It is known that models can drastically change their decision due to tiny, invisible to the human eye perturbation of the input [3], [4]. Some works show that models are also vulnerable to small translations and rotations of the input image [5], [6], Gaussian noise, and blur [7], different weather conditions [8], or even different images from a similar distribution of the training set [9]. What is more, current models tend to be overconfident in their outputs [10]. The problem is evident for the distributional shift (occurring when the test-time distribution of data differs from the training distribution, when deep learning models predict wrong output with high confidence [11]. It is of particular importance for models operating in the real-world to be robust to such distributional changes.

Vulnerability to tiny changes in the input might be explained by the fact that neural networks tend to exploit non-robust, high-frequency patterns in the training dataset, which causes them to fail under the distributional shift [12]–[14]. Therefore it is of great importance to test the robustness of the models in the out-of-distribution scenario. A common way to study model robustness in computer vision is evaluation under dataset shift or by adding so-called common corruptions [15] at test-time, which include several types of synthetically generated distortions (e.g., Gaussian noise and blur, JPEG compression, changes in brightness). A popular approach to improve model robustness is employing data augmentation techniques, i.e., using style-transfer data augmentation [16], [17]. Meanwhile, the use of only stylized representation may hurt performance on clean (original) data; hence a popular strategy is to use both clean and stylized samples during training [18], [19]. Inspired by data

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca.

augmentation which works on regions of the image [20], [21] here, we propose to apply style data augmentation only to random patches, which offers competitive accuracy on both clean and corrupted data.

Model robustness and uncertainty calibration are of special importance for safe autonomous driving, i.e. for pedestrian detection, which is studied in this work. However, evaluation of existing models in a realistic scenario where test data comes from different distribution than during training is still lacking. The contribution of this work is as follows:

- we analyze the impact of the distributional shift on the accuracy of pedestrian detection models, i.e. detection models are evaluated in cross-dataset setting, by adding different types of image distortions and when testing on night-time images,
- popular data augmentation methods are evaluated in terms of model robustness and a new simple scheme for data augmentation is proposed and used during training, where an original image is combined with a patch of its stylized version, which offers competitive results
- uncertainty calibration of existing models is evaluated experimentally.

## II. RELATED WORK

**Pedestrian Detection** is an essential topic in the context of autonomous driving. This topic traditionally borrows a lot from standard object detection models, current state-of-the-art models like Faster-R-CNN [22] and Mask R-CNN [23] are used for that purpose. There exist specialized models which modify the loss function to handle occlusions [24], run multi-step prediction for improved localization [25], simultaneously predict the full and visible boxes of pedestrians [26] or utilize low variance in an aspect ratio of visible pedestrians [27]. However, as it was shown a general purpose Faster R-CNN provides very competitive results, [28], [29] so that architecture is also used in this work.

Many large-scale benchmarks were created to facilitate this kind of research. Caltech dataset [30] was one of the first examples with 13674 pedestrians annotated with bounding boxes. The following datasets focused on dense scenes, like CityPersons [28] that have, on average, 6.47 pedestrian per image (comparing to 0.32 in Caltech), increasing the scale and variety of data. EuroCity Persons [31] further increased variety by recording data in European 12 countries in different weather conditions. Recent NightOwls [32] dataset and on the other hand focuses on night-time pedestrian detection.

**Robustness** is of great importance for many visual systems to be deployed in the real-world to improve the model accuracy. It was shown that while modern CNN-based models often obtain very good results on the benchmark it was trained on, they are very vulnerable to tiny changes in the input. In order to measure the model robustness it was proposed to use synthetically generated distortions during testing, so-called Common Corruptions, while a use of those corruptions for training should be avoided [15]. A popular approach to improve model robustness is using small Gaussian noise during training [33]. However, such augmentation may reduce the accuracy on the clean data, so to avoid such side-effect Gaussian noise can be added only to the patches of the input image [20]. Another work uses auxiliary classifiers for training on corrupted samples, which allows to achieve good balance between robustness and clean accuracy [34], however it explicitly requires training on corrupted samples which we want to avoid.

Another line of research showed that CNNs are biased toward texture [16], which may cause a lack of robustness. To increase model attention to the higher-level features (e.g., shape), style transfer [35] data augmentation can also be used during training, which shows promising results in terms of model robustness [18]. Removing part of the image during training has also shown an improvement in model accuracy [21], [36]. Instead of occluding a portion of an image, an approach called CutMix replaces a portion with a patch from a different image [37]. AugMix approach applies different augmentations to the image and interpolates between them to obtain training samples [38]. It is also possible to learn augmentation policy; however, such a process tends to be very costly [39]. Other approaches include using self-supervision [40], adversarial training [41] or using large scale pre-training [42].

### A. UNCERTAINTY ESTIMATION

Providing reliable uncertainty estimates is an essential element of safe autonomous systems [2]. It was shown that the current deep learning models are overconfident in their uncertainty estimates [10] and the effect is more striking under the distributional shift [11]. One of the ways to compute calibration is computing Expected Calibration Error. The previous finding of the poor model calibration was confirmed in the context of object detection from LiDAR data [43]. In this work, calibration is evaluated the context of pedestrian detection under the distributional shift.

## III. PEDESTRIAN DETECTION MODEL

In this section, pedestrian detection model is presented. Section III-A briefly describes the Faster R-CNN and the CSP architectures for pedestrian detection. Section III-B describes used data augmentations and the proposed approach.

### A. PEDESTRIAN DETECTION

A task of pedestrian (object) detection is to return for an image a list of bounding box coordinates, with predicted class and its score. Faster R-CNN is a standard algorithm in generic object detection, which is also commonly used in pedestrian detection [28], [29], belongs to the class of two-stage object detectors and has two main modules: Region Proposal Network (RPN) and a classification layer. Both modules share a common set of convolutional layers, which is also called a backbone network. RPN produces a list of windows (also called anchors) that are likely to contain an object, whereas the classification layer is responsible for classifying each

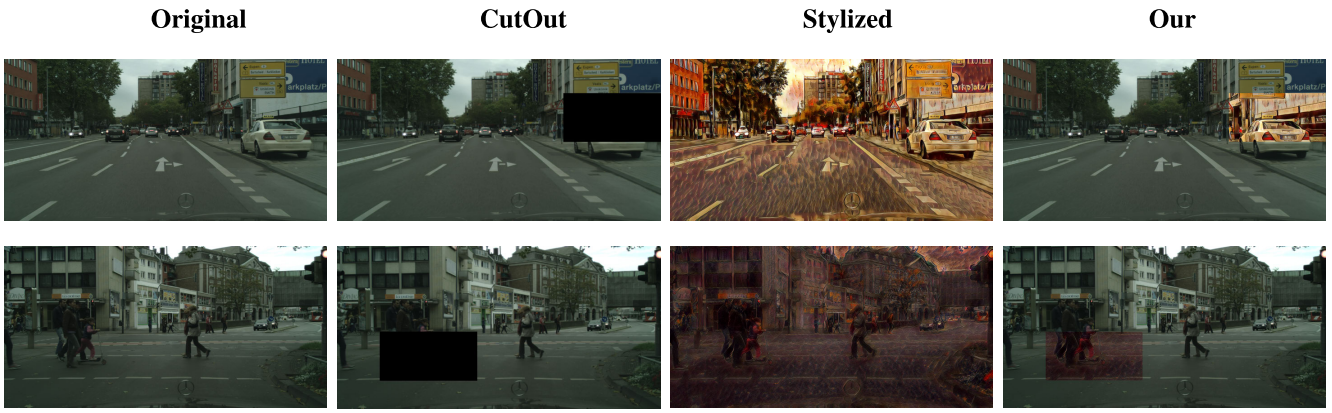


**FIGURE 1.** Different augmentation strategies. The second column shows random region removal with CutOut, and the third column shows stylized version of the original image. The last column shows the proposed augmentation that combines CutOut with style-augmentation.

of the proposed regions into one of the predefined classes (including background). At this step, each anchor returns logit vector $z \in R_k$, where $K$ is the number of classes (in our case of pedestrian detection, there are two classes: pedestrian and background). Further a sigmoid function is applied $p = sigmoid(z)$, which returns a list of predicted class probabilities. Predicted class is the one with the biggest probability, and the probability value for that class is used as the confidence score. Such a model is trained by backpropagation and stochastic gradient descent (SGD) by optimizing a multi-task loss function with standard cross-entropy loss for classification task and L1 smooth loss for regression of bounding boxes localisations.

For the task of pedestrian detection also specialized architectures exist, and a Center-Scale-Prediction [27] (CSP) is a recent model that achieves state-of-the art results. It simplifies the object detection pipeline by simply predicting center of the objects and their scale. Such anchor-free framework does not require defining anchors hyperparameters, i.e. the sizes of anchors or the number of scales (as in Faster R-CNN) and works very well for pedestrian detection.

### B. DATA AUGMENTATIONS

**Style-transfer** is a popular technique that allows transferring style (texture) from one image into another image. While such image synthesis is not perfect, many studies shown that using style-transfer data augmentation can improve the robustness of the model [16], [19]. However, using only stylized data might also decrease accuracy on the original data, so a popular approach is to train a model using 1:1 ratio of stylized and original images [18]. Here, a popular approach in the literature is followed, and as a source of style, texture information from the randomly chosen image from *Kaggle's Painter by Numbers* dataset [44] is used.

#### 1) PROPOSED AUGMENTATION

A problem with style-transfer data augmentation is that whereas it increases the robustness of models, it can decrease the accuracy of clean data since the stylized image differ quite significantly from real images. Instead of mixing original images with their stylized versions here, we tackle this problem from a different angle. There is a growing literature of research which, during training, augments only patches of the image. CutOut, for example, removes random patches from the image that shows positive for model accuracy [21]. In CutMix, on the other hand, random patches are cut and pasted among training images [37]. In this work, a similar strategy is proposed but the patches are mixed between the base image and its stylized version.

Our method works by adding a patch of the stylized image to the original image to the same location. The center of the patch is sampled to be within the image and the method allows for varying the patch size. Details are presented in the Alg. 1 diagram. Fig. 1 shows proposed data augmentation in comparison to other methods. Our motivation comes from the Patch Gaussian augmentation [20], where it was shown that adding Gaussian noise only to patches of the image, can improve both clean accuracy and robustness of the model.

**Algorithm 1** Proposed Data Augmentation

**Input:** Input image $I$, Stylized image $S$,
Image width $I_W$, Image height $I_H$,
Patch width $P_W$, Patch height $P_H$
**Output:** Augmented Image $I_{out}$
*%Compute patch start coordinates*
$x1 \leftarrow random.normal(0, I_W - P_W - 1)$
$y1 \leftarrow random.normal(0, I_H - P_H - 1)$
*%Compute masks*
$mask[H][W] \leftarrow \{0\}$
$mask[y1 : y1 + P_H][x1 : x1 + P_W] \leftarrow \{1\}$
$mask\_inverse[H][W] \leftarrow \{1\}$
$mask\_inverse \leftarrow mask\_inverse - mask$
$I_{out} \leftarrow mask\_inverse * I + mask * S$
*%Compute final image*
**return** $I_{out}$

#### 2) GAUSSIAN AUGMENTATION

Using Gaussian augmentation also proved to be successful in increasing model robustness; therefore it is also used for

our experiments. Two variants (parametrized by $\sigma_{max}$) are evaluated:

- Gaussian augmentation. Firstly for each pixel sample $\sigma$ from uniform distribution - $\sigma \sim U(0, \sigma_{max})$. Add noise to each pixel sampled from $N(0, \sigma)$.
- Patch Gaussian augmentation. Same as above, but the Gaussian noise is added only to the random patch of the image [20].

## IV. EXPERIMENTAL SETUP

In this section, our experimental setup is described: datasets, evaluation metrics, and implementation details.

### A. DATASET

**Citypersons** [28] is a popular and challenging dataset for pedestrian detection. It was built on top of the semantic segmentation dataset CityScapes [45] for autonomous driving. It was recorded in a diverse setting (27 cities in Germany) so that comparing to previous datasets, it contains, on average, 7 pedestrians per image, significantly more than a popular Caltech dataset [30]. In total, it contains 5000 images with around 35000 manually annotated persons with bounding boxes. Areas that contain a dense group of pedestrians in which it is hard to distinguish between them and misleading regions like a pedestrian reflection in the window are marked as ignore regions. Each person is also attributed to occlusion level.

**Eurocity Persons** [31] further improves the diversity of pedestrian detection datasets. It was recorded in 31 cities in 12 European countries. Data were collected during all seasons in changing weather conditions. In total, there are around 238 200 person instances manually annotated in over 47300 images. A subset of 7000 images recorded during the night-time was a novelty at the time of the release of the dataset.

**Nightowls** [32] is a large scale dataset that focuses on night-time pedestrian detection. In comparison to day-time images it is a much more challenging task due to the illumination variation, light reflections, blur artifacts, and changes in contrast. In total, there are 279 000 annotated frames from 3 countries. Night-time pedestrian detection is very important for robust vision applications, that is why the authors showed that pedestrian detectors do not perform well at night, even when they are trained on night-time data.

#### 1) COMMON CORRUPTIONS

Robust perception system in autonomous vision must work well in many different conditions that might occur: night-time, severe rain or snow, fog, noise from sensors, degradation of image quality, and many more. Even though many large scale data benchmarks exist, it is impossible to gather all possible tests, so that is why it is necessary to test models in an out-of-distribution setting when such conditions are synthetically generated. Common corruptions benchmark is widely adopted for testing the robustness of the models [10]. It contains 15 different distortion types grouped in 4 categories: noise (Gaussian noise, shot noise, impulse noise, salt-and-pepper noise), blur (defocus blur, frosted glass blur, motion blur, zoom blur), digital (elastic transformations, pixelation, JPEG lossy compression) and weather corruptions (snow, fog, brightness, contrast) where each corruption has 5 levels of severity. Those corruptions are used exclusively for the test, as it is a common practice adopted by the computer vision community. Fig. 2 present example corruptions.

### B. METRICS

#### 1) DETECTION ACCURACY

The standard metric in pedestrian detection is the log-average miss rate (LAMR). It requires computing the miss-rate (mr) and false positives per image (fppi) that are computed as follows:

$$mr(c) = \frac{fn(c)}{tp(c) + fp(c)} \tag{1}$$

$$fppi(c) = \frac{fp(c)}{\#img} \tag{2}$$

where $fn$ stands for false negatives, $tp$ for true positives, and $fp$ is the number of false positives for detections that have confidence value equal to or bigger than a threshold $c$. A prediction is marked true positive when its overlap with ground truth is greater than the selected threshold, i.e., in pedestrian detection, 0.5 is commonly used as the threshold. To measure the overlap between the predicted bounding box $a$ and its ground truth $b$ an Intersection-over-Union (IOU) is computed as follows:

$$IOU(a, b) = \frac{Area(a) \cap Area(b)}{Area(a) \cup Area(b)} \tag{3}$$

If multiple detections are matched with single ground-truth, then only the detection with the highest confidence is matched, the rest of detections are considered as false positives. Finally, not matched ground truth bounding boxes are considered as false negatives. Threshold $c$ is used for measuring the balance between the number of false positives, false negatives, and true positives. Final LAMR metric is computed by averaging at nine $fppi$ rates spaced equally in the log-space in the range from $10^{-}2$ to $10^{0}$ as it is done in the literature [30]:

$$LAMR(c) = exp(\frac{1}{9}\sum_{f} log(mr(\arg max_{fppi(c)<=f} fppi(c))) \tag{4}$$

For evaluation, the so-called reasonable setup [28] is used. It means that for training and evaluation, only pedestrians whose height is bigger than 50 pixels, and the occlusion level is smaller than 0.35 are used. Note that in generic object detection the most commonly used metric is mean average precision (mAP). However the LAMR metric is preferred in certain applications, such as autonomous driving as there usually exists an upper limit of false positives per image.

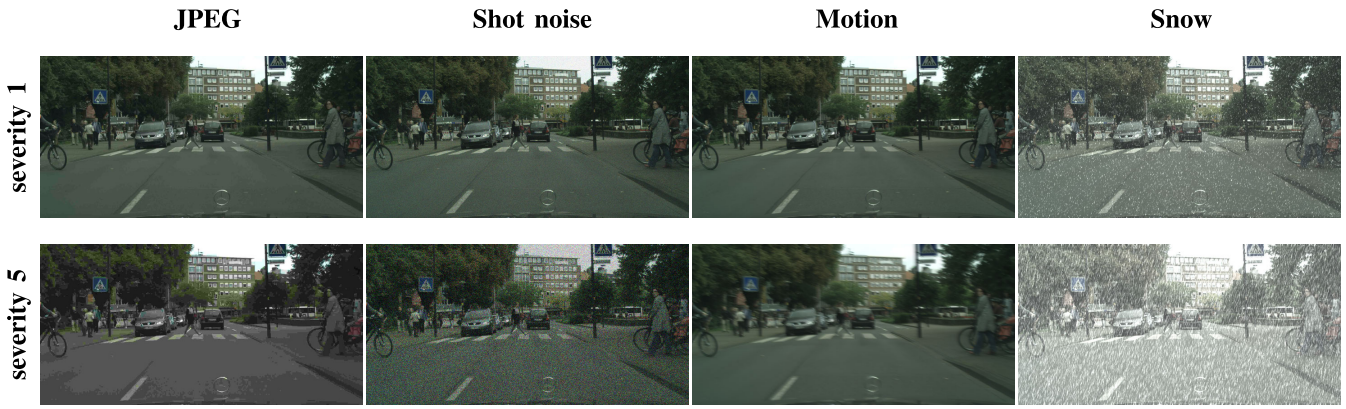


**FIGURE 2.** Examples of different corruption types from Common Corruptions benchmark with different severity. Note that at the lowest severity distortions are barely visible whereas at the highest severity they are clearly visible, however semantics of the images are not changed.

#### 2) UNCERTAINTY ESTIMATION

We are also interested in measuring the classification calibration of the trained models. Intuitively, when a well-calibrated detection model predicts bounding-box with pedestrian class with 80% confidence, it should be accurate in 80% of the cases. The standard way to measure a classification calibration is to compute Expected Calibration Error (ECE) [10]. Firstly predictions are partitioned into M bins based on its confidence, and then the metric is computed as:

$$ECE(c) = \sum_{m=1}^{M} \frac{|B_m|}{n} |acc(B_m) - conf(B_m)| \tag{5}$$

where $B_m$ is the set of indices of samples that prediction confidence falls into the *mth* interval. The lower the score, the better calibration (0 means perfect calibration). It is also possible to compute calibration score for the regression of the bounding box localization; however in our setting there is no measure of bounding box uncertainty, as there is for the classification task (output from the softmax layer).

### C. IMPLEMENTATION DETAILS

MMdetection library [46] is used for the Faster R-CNN model, with ResNet-50 backbone. All models are pre-trained on ImageNet [47]. Stochastic Gradient Descent with an initial learning rate of 0.002 and a momentum of 0.9 is used. The training lasts for 40 epochs, and the learning rate drops to 0.0002 after 25 epochs. All models are trained on the Cityscapes dataset. Data from 3 cities (Darmstadt, Mönchengladbach and Ulm) from the training set are moved into trainval set similar as in the literature [48]. The model with the best accuracy on trainval set is used for testing. Since no ground-truth data is publicly available for all the datasets, results are reported on their validation sets. During the evaluation, each training is repeated 5 times, and the mean accuracy is the final score. All models used the same training settings. In addition to standard random vertical flipping and resizing of the image, each tested model adds its own data augmentation.

For the CSP architecture, public repository published by the authors is used.[1] A training protocol from the original paper is followed, i.e. model is trained for 37.5K iterations and then used for the validation. The only difference is that we employed 1 GPU for training (instead of 2). Also, as significant differences between consecutive runs may occur, each model is trained 5 times and the mean accuracy is reported (similarly as for the Faster R-CNN).

## V. EXPERIMENTS

In this section, results of experiments are presented. Section V-A shows how patch-size affects the accuracy of the model. Section V-B presents the robustness to Common Corruptions benchmark of several augmentation methods, and in section V-C similar experiments are conducted for cross-dataset evaluation, in particular for detecting pedestrians in the night-time. Section V-D shows classification calibration of evaluated models.

### A. PATCH-SIZE SELECTION

First, hyperparameter search for the optimal size of the stylized patch is run. Too small patch size might reduce the positive effects of using style-transfer for model robustness, whereas the too big size of the patch might reduce the clean accuracy. In the experiment the stylized patch is of size $kW$ x $kH$ pixels, where $k \epsilon [0, 1]$. Note that when $k = 0$, it means that style transfer augmentation is not used at all, wheres when $k = 1$, only stylized-images are used.

Figure 3 plots the model accuracy as a function of patch size on the CityPersons dataset. It can be noticed that the model accuracy firstly increases with the size of the patch. However, after the size of the patch is bigger than 0.3 of the image size, then accuracy decreases. This is expected as the model is biased more towards stylized images, and as a result, accuracy on the clean data decreases. All of the remaining experiments are conducted with the selected patch size.

[1] https://github.com/liuwei16/CSP

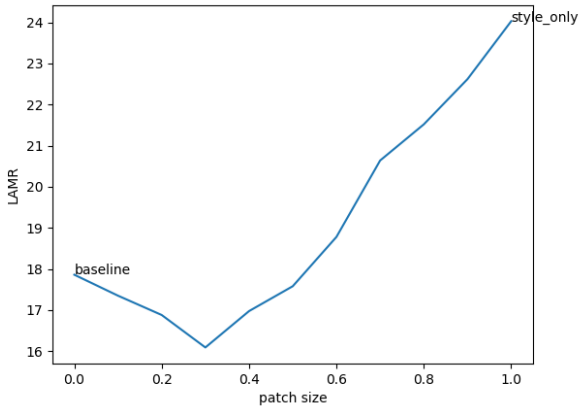


**FIGURE 3.** Log-average miss rate for pedestrian detection accuracy (lower is better) on CityPersons dataset using proposed data augmentation as a function of the patch size. Note that the most left data point corresponds to the baseline trained on original data, while the most right data point is a model trained using only stylized images.

### B. CityPersons AND COMMON CORRUPTIONS

In this section, different data augmentation methods are evaluated with regard to the model robustness. In particular, the following models are evaluated:

- *Baseline* model corresponds to the model trained with standard data augmentation,
- *Sin* [16] is a model trained on both clean data and the stylized version using 1:1 ratio,
- *StyleOnly* model is trained using only stylized data,
- *Our* corresponds to the proposed augmentation model,
- *CutOut* [21] model with the same patch size as *Our* model. It serves as another baseline to the proposed data augmentation and as a sanity check to make sure that similar gains cannot be obtained by simply removing patches of the image,
- *Gaussian* data augmentation with $\sigma_{max} \epsilon \{0.1, 0.5\}$
- PatchGaussian data augmentation with $\sigma_{max} \epsilon \{0.1, 0.5\}$. The same patch size is used for the *Our* and *CutOut* model.

Table 1 shows accuracy of the trained models on the original CityPersons dataset and as well on the Common Corruptions benchmark grouped by distortion category.

First, the results emphasized the importance of robustness testing. While the *Baseline* provides reasonable accuracy on the clean data, it constantly has worse accuracy on all corruption types by a large margin. Further different data augmentation provides the best accuracy on different corruption types. *Our* data augmentation performs the best on clean data, noise corruption (together with *Sin* model), and on weather-related corruption types. *Sin* model also performs the best on the blur corruptions, whereas Gaussian augmentation helped the most on the digital noise. Findings of other authors [20], are confirmed and we show that while Gaussian augmentation improves robustness, it may actually hurt performance on clean data (21.19% LAMR when $\sigma = 0.5$), whereas using only patches of Gaussian noise provides a balance between clean accuracy and robustness. Finally, it can be observed that

**TABLE 1.** Accuracy comparison of Faster R-CNN models trained with different augmentation strategies on clean data (first column) and related to specific corruption types from the Common Corruptions benchmark (the remaining columns). LAMR is reported (lower is better). For models that used Gaussian augmentation, values in noise column are marked in grey colour because the tested corruption type was a part of the training.

| Name | Clean | Noise | Blur | Weather | Digital |
|---|---|---|---|---|---|
| Baseline | 17.86 | 94.72 | 69.61 | 54.05 | 51.78 |
| CutOut | 16.91 | 91.04 | 66.72 | 49.97 | 50.61 |
| StyleOnly | 24.03 | 78.02 | 63.88 | 49.3 | 44.16 |
| Sin | 17.83 | **76.14** | **59.63** | 44.43 | 42.36 |
| Our | **16.09** | **76.17** | 61.46 | **42.74** | 43.12 |
| Gaussian_0.1 | 17.58 | 45.73 | 63.97 | 48.85 | 39.56 |
| Gaussian_0.5 | 21.19 | 39.33 | 60.17 | 49.52 | **38.46** |
| PatchGaussian_0.1 | 16.25 | 52.9 | 66.71 | 50.28 | 46.00 |
| PatchGaussian_0.5 | 16.41 | 45.71 | 63.72 | 48.91 | 41.65 |
| *Combined augmentations* | | | | | |
| Our + PatchGaussian_0.5 | **16.28** | 47.55 | 61.55 | **42.73** | 40.5 |
| Sin + PatchGaussian_0.5 | 18.11 | 43.27 | **59.64** | 44.86 | **37.27** |

combining Style-Transfer using *Our* or *Sin* with *PatchGaussian* provides the best accuracy across all corruption types. However, many of the corruptions still drastically degrade the performance, so there is still large room for improvement in terms of increasing model robustness.

Also, it is interesting to directly compare *Our* and *Sin* data augmentations as they are competitive approaches. For that purpose, those two approaches are directly compared for each corruption type. Table 2 shows that the proposed model provides the most significant gains for fog, brightness, and contrast deformations. This is very interesting in light of findings by Yin *et al.* [4] where they perform Fourier spectral analysis of different distortion types and find that the aforementioned distortions are concentrated in low-frequencies components of images. This means that the proposed data augmentation might be particularly useful for low-frequency distortions types.

Fig. 4 shows example detections for *Baseline* model and model trained with proposed augmentation and Patch Gaussian. In general, data augmentations significantly improve model accuracy, however, there are still many situations when the model lacks robustness. Some types of distortions (especially noise), drastically change the output of the detection model even at the low distortion severity, when the input image is only slightly changed. Even though our best trained models are more robust than the *Baseline* model the problem is still far from being solved.

### C. EVALUATION UNDER THE DISTRIBUTIONAL SHIFT

In the previous section, synthetically generated distortion types were used for the evaluation. However, it is very important to test a model on different dataset because even if the datasets look similar, there still will be a lot of differences regarding data collection protocol (e.g., camera sensor and its placement inside a vehicle, geographical location) which might affect final accuracy. Further, synthetically generated distortions are only approximation of real-world adverse conditions that is why the models are also tested on night-time images.

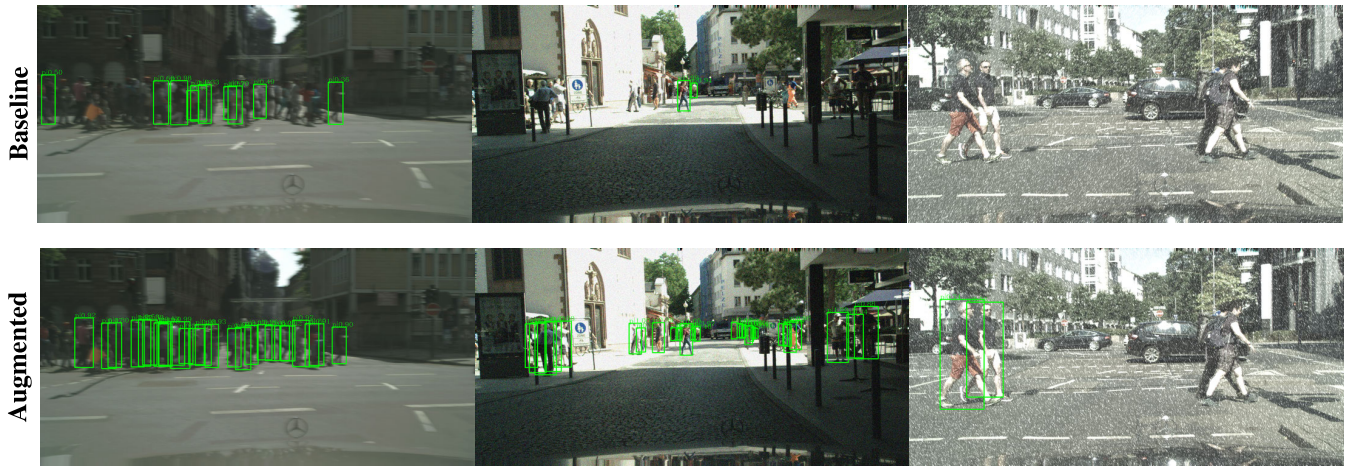


**FIGURE 4.** Detection samples for baseline and augmented (*Our + PatchGaussian_0.5*) models for different corruption types. The first column - motion blur (severity intensity of 4), the second column - Gaussian noise (severity intensity of 2), third column - artificial snow with a severity intensity of 2. Note that the distortion for Gaussian noise is almost impercetible, yet it greatly reduces accuracy of the model. Augmented model is more robust, however in the last column pedestrians on the right are missed by both models.

**TABLE 2.** LAMR for each corruption type of Faster R-CNN models.

| Type | Noise | | | Blur | | | | Weather | | | | Digital | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | shot | impulse | gauss | defocus | glass | motion | zoom | snow | fog | frost | bright | contrast | elastic | pixel | jpeg |
| Baseline | 92.69 | 97.06 | 94.42 | 55.25 | 67.32 | 60.68 | 95.22 | 89.3 | 30.24 | 66.32 | 30.32 | 47.64 | **14.47** | 72.75 | 72.26 |
| Sin | **72.25** | **79.2** | 76.99 | **43.88** | **54.27** | **48.57** | **91.83** | 76.1 | 28.35 | 49.87 | 23.39 | 43.77 | 20.93 | **46.66** | **56.39** |
| Our | 72.4 | 79.39 | **76.73** | 44.85 | 56.89 | 52.31 | 92.54 | **75.52** | **25.97** | **49.39** | **20.06** | **39.12** | 20.84 | 54.17 | 58.37 |

**TABLE 3.** Accuracy comparison of Faster R-CNN models trained with different augmentation strategies on day-time and night-time images from the ECP dataset as well on NightOwls dataset (night-time). LAMR values are reported.

| Name | ECP-day | ECP-night | NightOwls |
|---|---|---|---|
| Baseline | 25.05 | 45.5 | 76.23 |
| CutOut | 21.93 | 41.25 | 70.23 |
| StyleOnly | 26.83 | 36.2 | 58.22 |
| Sin | 21.22 | 30.69 | 56.51 |
| Our | **21.04** | **29.89** | 56.1 |
| Gaussian_0.1 | 23.17 | 32.25 | 56.21 |
| Gaussian_0.5 | 26.11 | 32.22 | **47.55** |
| PatchGaussian_0.1 | 23.6 | 34.24 | 64.4 |
| PatchGaussian_0.5 | 22.31 | 32.61 | 58.87 |
| *Combined augmentations* | | | |
| Our + PatchGaussian_0.5 | **20.35** | **26.13** | 52.34 |
| Sin + PatchGaussian_0.5 | 20.62 | 28.80 | **49.08** |

Table 3 shows the accuracy of the models for different datasets. First, a significant drop in accuracy for all the models can be noticed, i.e., *Baseline* model LAMR increases from 17.86% to 25.05% when tested on ECP-day comparing to CityPersons. It can be explained by the fact that the ECP dataset is more challenging for both but also because of the dataset shift. Accuracy further drops when models were tested on night-time images instead of day-time - for *Baseline* model, the average miss-rate is almost doubled. For data augmentation based on stylization or Gaussian noise the decrease is not that severe and for best performing combination (*Our + PatchGaussian_0.5*) the average miss-rate increases only from 20.35% to 26.13%. Further, the drop in accuracy for NightOwls dataset is bigger than for the night-time images from the ECP dataset. On that benchmark, simple Gaussian augmentation obtained the best result - this might be because the NightOwls dataset contains a lot of noise due to the very low light intensity. It is also worth noting effect of the *Cutout* data augmentation on the accuracy of the clean data, however it only slightly affected the robustness, especially when compared to the *PatchGaussian* and style data augmentations. Finally, the proposed data augmentation provides the same or better accuracy compared to *Sin* across all benchmarks. Fig. 5 shows example detections. Again model accuracy is improved, but the augmented model still lacks robustness for some situations (last column).

### 1) CENTER SCALE PREDICTION

Table 4 shows obtained results on all of the benchmarks. Some interesting observations can be made. Firstly, in most cases, the new architecture provides better accuracy then the Faster R-CNN, i.e. LAMR decreases for the *Baseline* model from 17.86% to 12.37% on the Cityscapes dataset, and from 25.05% to 22.06% on the ECP day-time images. Surprisingly the new model is worse than the Faster R-CNN when testing the *Baseline* model on ECP night-image images. When stylized data augmentation are used the new model is better or on par with Faster R-CNN except for the performance under noise corruption (LAMR increases from 76.14% to 83.24% for the *Sin* data augmentation.

Interestingly, while testing on the Cityscapes dataset (no distributional shift) standard data augmentation already provides a strong baseline, and only *CutOut* augmentation is able to slightly improve over that. The proposed data augmentation obtains the best result (by significant margin) when testing on ECP day-time images. When testing on night-images,


**FIGURE 5.** Detection samples for baseline and augmented model. The augmented model is more accurate on the night-time images, however some pedestrians are still not detected.

**TABLE 4.** Comparison of CSP models accuracy trained with different data augmentations, on various datasets, and on specific corruption types from the Common Corruptions benchmark applied to Cityscapes dataset (columns 2-5). For models that used Gaussian augmentation, values in noise column are marked in grey colour because the tested corruption type was part of the training. LAMR values are reported.

| Name | City | Noise | Blur | Weather | Digital | ECP-day | ECP-night | NightOwls |
|---|---|---|---|---|---|---|---|---|
| Baseline | 12.37 | 95.54 | 63.67 | 52.89 | 49.77 | 22.06 | 60.83 | 65.48 |
| CutOut | **12.17** | 95.06 | 63.39 | 51.52 | 49.83 | 23.31 | 60.52 | 67.13 |
| Sin | 13.85 | **83.24** | **51.45** | 41.23 | 37.63 | 19.48 | 31.25 | **51.16** |
| Our | 12.44 | 84.21 | 54.72 | **40.91** | 41.56 | **18.24** | **31.17** | 52.71 |
| Gaussian_0.5 | 17.99 | 34.64 | 55.74 | 59.59 | **36.22** | 32.11 | 66.32 | 60.72 |
| PatchGaussian_0.5 | 12.47 | 43.64 | 59.31 | 49.95 | 39.99 | 20.06 | 47.20 | 58.42 |
| *Combined augmentations* | | | | | | | | |
| Our + PatchGaussian_0.5 | **12.62** | 44.21 | 53.13 | **40.44** | 37.41 | **18.39** | **28.99** | 52.21 |
| Sin + PatchGaussian_0.5 | 14.12 | 44.27 | **50.69** | 40.72 | **33.73** | 19.15 | 30.08 | **49.88** |

both stylized augmentation provide the best accuracy. For different types of corruptions, the best methods are the same as in the Faster R-CNN model.

Experiments on the CSP architecture confirm that the proposed data augmentation allows to obtain very competitive results across different benchmarks and offers good balance between accuracy on the clean dataset and under distributional shift. Additionally, we show that it is important to test the models under different benchmarks and architectures as the results can largely differ between those.

### D. UNCERTAINTY ESTIMATION

Providing reliable uncertainty estimates is a very important aspect of safe autonomous systems. In this section, the ECE score is measured for all real-world benchmarks. We find that stylized and Gaussian augmentations help to improve prediction confidence with no clear leader between them, so for conciseness Table 5 shows ECE scores of Faster R-CNN models, for the *Baseline*, *Sin*, and for *Our* models.

Ideally, the ECE score would be a constant and small value across different datasets, which would mean that the model "knows what it does not know." However for the *Baseline* the ECE score goes up when testing on night-time images (jump from 0.1468 to 0.1985 on ECP dataset). *Our* model has almost constant calibration error when switching to night-time images on the ECP dataset, whereas *Sin* model has even smaller calibration error. Note from the previous section that the model accuracy drops in that case, which means that the improved calibration cannot be only attributed to the improved model accuracy. All of the models are significantly worse calibrated for the NightOwls dataset because it is more challenging of both and because the distributional shift is greater in this case.

**TABLE 5.** Comparison of ECE for selected Faster R-CNN models on different datasets (lower value means better calibration).

| Name | CityPersons | ECP-day | ECP-night | NightOwls |
|---|---|---|---|---|
| Baseline | **0.1418** | **0.1468** | 0.1985 | 0.3765 |
| Sin | 0.1434 | 0.166 | **0.1429** | 0.3393 |
| Our | 0.1429 | 0.1569 | 0.1572 | **0.331** |

Fig. 6 shows calibration plots. It can be observed that the *Baseline* model has worse uncertainty calibration when switching from day to night-time images, especially in the area of high confidence predictions, which means that the model in under-confident in its predictions. For the proposed model (and other stylized and gaussian augmentations) there is no clear difference in the calibration plot between day and night-time for the ECP dataset.

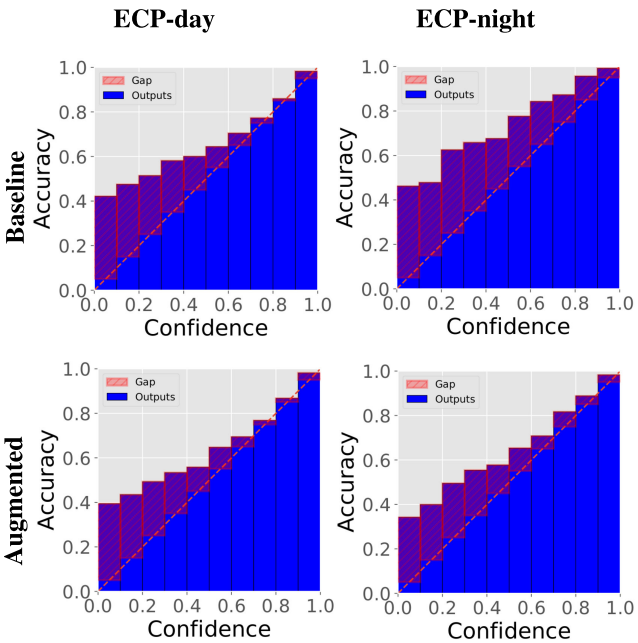


**FIGURE 6.** Calibration plots for selected Faster R-CNN models for day-time and night-time images on ECP dataset. Accuracy near diagonal means perfect calibration.

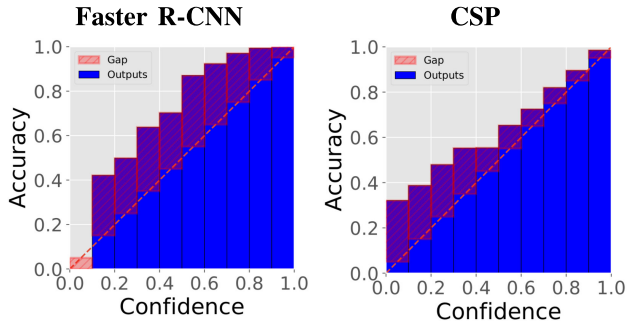


**FIGURE 7.** Calibration plots for Faster R-CNN and CSP architectures on Cityscapes dataset.

CSP architecture shows similar findings - using data augmentation usually improves model calibration, especially for the large distributional shift (night-time images). Interestingly we find however that CSP models are have worse calibration, e.g. for Cityscapes dataset Faster R-CNN *Baseline* model has ECE score of 0.1418 whereas CSP model has a score 0.2446 (Fig. 7).

## VI. CONCLUSION

In this work, the examination was performed of pedestrian detection models in the real-world setting when test-time data come from a different distribution than in training: using cross-dataset evaluation, testing the model by switching illumination conditions (day to night) and through testing it on synthetic distortions. It was confirmed that such a testing is crucial for a realistic evaluation of the model since the accuracy of the baseline model drops drastically. Further, we show that data augmentations in the form of stylized and Gaussian augmentations significantly improve the robustness of the model. A new data augmentation scheme was proposed that uses stylization but only on patches of the original image, and it was shown that such augmentation offers competitive accuracy. Finally, we demonstrated that the use of data augmentations also improves classification calibration of the pedestrian detection models.

Whereas the problem of model robustness is still not solved, this work serves as a step towards that goal. Our work could be combined with self-supervised learning methods, recently gaining much attention and which also could be beneficial to the model robustness and uncertainty.

## REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. Washington, DC, USA: IEEE Computer Society, Dec. 2015, pp. 1026–1034, doi: 10.1109/ICCV.2015.123.

[2] R. McAllister, Y. Gal, A. Kendall, M. van der Wilk, A. Shah, R. Cipolla, and A. Weller, "Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 4745–4753, doi: 10.24963/ijcai.2017/661.

[3] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–10. [Online]. Available: http://arxiv.org/abs/1312.6199

[4] D. Yin, R. G. Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, "A Fourier perspective on model robustness in computer vision," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 13255–13265.

[5] L. Engstrom, B. Tran, D. Tsipras, L. Schmidt, and A. Madry, "Exploring the landscape of spatial robustness," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. Long Beach, CA, USA: PMLR, Jun. 2019, pp. 1802–1811. [Online]. Available: http://proceedings.mlr.press/v108/levine20a/levine20a.pdf

[6] K. Gu, B. Yang, J. Ngiam, Q. Le, and J. Shlens, "Using videos to evaluate image model robustness," in *Proc. Safe Mach. Learn. Workshop ICLR*, 2019. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/papers/Xie_Self-Training_With_Noisy_Student_Improves_ImageNet_Classification_CVPR_2020_paper.pdf

[7] S. Dodge and L. Karam, "A study and comparison of human and deep learning recognition performance under visual distortions," in *Proc. 26th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2017, pp. 1–7.

[8] D. Dai and L. V. Gool, "Dark model adaptation: Semantic image segmentation from daytime to nighttime," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3819–3824.

[9] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, "Do ImageNet classifiers generalize to ImageNet?" in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. Long Beach, CA, USA, Jun. 2019, pp. 5389–5400.

[10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 7, 2017, pp. 1321–1330.

[11] J. Snoek, Y. Ovadia, E. Fertig, B. Lakshminarayanan, S. Nowozin, D. Sculley, J. Dillon, J. Ren, and Z. Nado, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 13969–13980.

[12] A. Ilyas, S. Santurkar, D. Tsipras, L. Engstrom, B. Tran, and A. Madry, "Adversarial examples are not bugs, they are features," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 125–136.

[13] J. Jo and Y. Bengio, "Measuring the tendency of CNNs to learn surface statistical regularities," *CoRR*, vol. abs/1711.11561, 2017. [Online]. Available: http://arxiv.org/abs/1711.11561

[14] R. Geirhos, J. Jacobsen, C. Michaelis, R. S. Zemel, W. Brendel, M. Bethge, and A. F. Wichmann, "Shortcut learning in deep neural networks," *CoRR*, vol. abs/2004.07780, 2020. [Online]. Available: https://arxiv.org/abs/2004.07780

[15] D. Hendrycks and T. G. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, May 2019, pp. 1–16. [Online]. Available: https://openreview.net/forum?id=HJz6tiCqYm

[16] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, May 2019, p. 22.

[17] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1501–1510.

[18] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," *CoRR*, vol. abs/1907.07484, 2019. [Online]. Available: http://arxiv.org/abs/1907.07484

[19] P. T. Jackson, A. Atapour-Abarghouei, S. Bonner, T. P. Breckon, and B. Obara, "Style augmentation: Data augmentation via style randomization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 83–92.

[20] R. G. Lopes, D. Yin, B. Poole, J. Gilmer, and E. D. Cubuk, "Improving robustness without sacrificing accuracy with patch Gaussian augmentation," *CoRR*, vol. abs/1906.02611, 2019. [Online]. Available: http://arxiv.org/abs/1906.02611

[21] T. Devries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *CoRR*, vol. abs/1708.04552, 2017. [Online]. Available: http://arxiv.org/abs/1708.04552

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2017, pp. 2961–2969.

[24] X. Wang, T. Xiao, Y. Jiang, S. Shao, J. Sun, and C. Shen, "Repulsion loss: Detecting pedestrians in a crowd," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7774–7783.

[25] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 643–659.

[26] X. Huang, Z. Ge, Z. Jie, and O. Yoshie, "NMS by representative region: Towards crowded pedestrian detection by proposal pairing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10750–10759.

[27] W. Liu, S. Liao, W. Ren, W. Hu, and Y. Yu, "High-level semantic feature detection: A new perspective for pedestrian detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5187–5196.

[28] S. Zhang, R. Benenson, and B. Schiele, "CityPersons: A diverse dataset for pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3213–3221.

[29] I. Hasan, S. Liao, J. Li, S. U. Akram, and L. Shao, "Pedestrian detection: The elephant in the room," *CoRR*, vol. abs/2003.08799, 2020. [Online]. Available: https://arxiv.org/abs/2003.08799

[30] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 304–311.

[31] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "EuroCity persons: A novel benchmark for person detection in traffic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1844–1861, Aug. 2019.

[32] L. Neumann, M. Karg, S. Zhang, C. Scharfenberger, E. Piegert, S. Mistr, O. Prokofyeva, R. Thiel, A. Vedaldi, A. Zisserman, and B. Schiele, "NightOwls: A pedestrians at night dataset," in *Proc. Asian Conf. Comput. Vis.* Perth, WA, Australia: Springer, 2018, pp. 691–705. [Online]. Available: https://www.springer.com/gp/book/9783030208691

[33] E. Rusak, L. Schott, R. S. Zimmermann, J. Bitterwolf, O. Bringmann, M. Bethge, and W. Brendel, "Increasing the robustness of DNNs against image corruptions by playing the game of noise," *CoRR*, vol. abs/2001.06057, 2020. [Online]. Available: https://arxiv.org/abs/2001.06057

[34] L. Zhang, M. Yu, T. Chen, Z. Shi, C. Bao, and K. Ma, "Auxiliary training: Towards accurate and robust models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 372–381.

[35] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Las Vegas, NV, USA: IEEE Computer Society, Jun. 2016, pp. 2414–2423.

[36] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2020, pp. 1–8.

[37] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6023–6032.

[38] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "Augmix: A simple data processing method to improve robustness and uncertainty," in *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–15.

[39] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation strategies from data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 113–123.

[40] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 15637–15648.

[41] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in *Proc. 6th Int. Conf. Learn. Represent. ICLR*, Vancouver, BC, Canada, Apr./May 2018, pp. 1–10.

[42] D. Hendrycks, K. Lee, and M. Mazeika, "Using pre-training can improve model robustness and uncertainty," in *Proc. Safe Mach. Learn. Workshop ICLR*, 2019, pp. 2712–2721.

[43] D. Feng, L. Rosenbaum, C. Gläser, F. Timm, and K. Dietmayer, "Can we trust you? On calibration of a probabilistic object detector for autonomous driving," *CoRR*, vol. abs/1909.12358, 2019. [Online]. Available: http://arxiv.org/abs/1909.12358

[44] S. Y. Duck. (2016). *Painter by Numbers*. Accessed: May 8, 2020. [Online]. Available: https://www.kaggle.com/c/painter-by-numbers

[45] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[46] K. Chen *et al.*, "Mmdetection: Open MMLab detection toolbox and benchmark," *CoRR*, vol. abs/1906.07155, 2019. [Online]. Available: http://arxiv.org/abs/1906.07155

[47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[48] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. A. Eslami, D. J. Rezende, and O. Ronneberger, "A probabilistic U-net for segmentation of ambiguous images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6965–6975.

**SEBASTIAN CYGERT** received the B.Sc. degree in computer science from the Military University of Technology, Warsaw, in 2012, and the M.Sc. degree from the Warsaw University of Technology, in 2013. He is currently pursuing the Ph.D. degree with the Multimedia Systems Department, Gdańsk University of Technology. His research interests include computer vision and machine learning.

**ANDRZEJ CZYŻEWSKI** received the M.Sc. degree in sound engineering from the Gdańsk University of Technology, in 1982, and the Ph.D. degree from the Cracow Academy of Mining and Metallurgy, in 1992. Since 2002, he has been a Full Professor with the Multimedia Systems Department, Gdańsk University of Technology, where he is currently the Head. He is the author of more than 500 scientific papers in international journals and conference proceedings. He is also the author of more than 20 patents. He has led more than 30 Research and Development projects funded by the Polish Government and participated in five European projects. He has extensive experience in soft computing algorithms and sound and image processing for applications in multimedia technology.