OBJECT DETECTION FOR AUTONOMOUS DRIVING: HIGH-DYNAMIC RANGE VS. LOW-DYNAMIC RANGE IMAGES

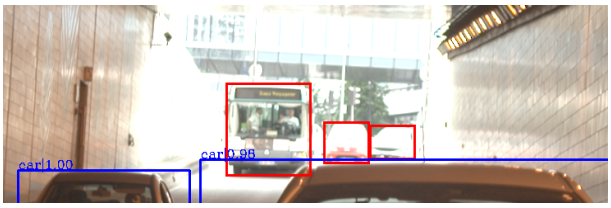
Ismail H. Kocdemir[1,3], *A. Oguz Akyuz*[1,3], *Alper Koz*[2,3], *Alan Chalmers*[4], *Aydin Alatan*[2,3], *Sinan Kalkan*[1,3]

[1] Dept. of Computer Engineering, METU. [2] Dept. of Electrical-Electronics Eng., METU.
[3] Center for Image Analysis (OGAM), METU. [4] WMG, University of Warwick.

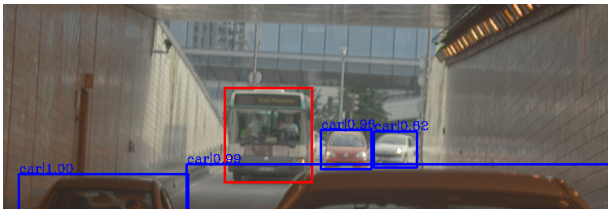
ABSTRACT

An important problem in autonomous driving is to perceive objects even under challenging illumination conditions. Despite this problem, existing solutions use low-dynamic range (LDR) images for object detection for autonomous driving. In this paper, we provide a novel analysis on whether high-dynamic range (HDR) images can provide better performance for object detection for autonomous driving. To this end, we choose a seminal deep object detector and systematically evaluate its performance when trained with (i) LDR images, (ii) HDR images, and (iii) tone-mapped LDR images for scenes with different illuminations. We show that a detector with HDR images pre-processed with normalization and gamma correction can only marginally perform better than a detector with LDR or tone-mapped LDR images. Our analysis of this unexpected finding reveals that a detector with HDR images requires significantly more samples as the space of HDR images is significantly larger than that of LDR images.

***Index Terms*—** Autonomous Driving, Object Detection, High-Dynamic Range (HDR), Low-Dynamic Range (LDR)



(a) LDR image.



(b) Tone-mapped LDR image (using [3]).

Fig. 1. (a) LDR images make it difficult for detection of objects (shown in blue boxes) in adverse illumination conditions e.g. in tunnels. (b) This can be alleviated using HDR or tone-mapped LDR images. Blue: detected, Red: missed objects. (Input from the "tunnel video" of the EU NEVEx Project.)

1. INTRODUCTION

A open problem in autonomous driving is detecting objects under adverse conditions of illumination, encountered e.g. when entering or exiting a tunnel, driving towards the sun or under the headlights of an oncoming car – see e.g. Fig. 1. In addition to various depth and radar sensors, existing systems generally use low-dynamic range (LDR) cameras for visual perception [1, 2], which can be insufficient for providing discriminative details for objects in dark or bright regions of a scene.

In this paper, we propose and investigate using high-dynamic range (HDR) images for better object detection in autonomous driving systems. To be specific, we present and analyze the following approaches for integrating HDR information into the training of an object detector: (i) The LDR image (obtained by a LDR camera.) (ii) The HDR image (either calculated from LDR images or captured by an HDR camera). (iii) A detail-rich LDR image (tone-mapped from HDR by commonly used tone-mapping operators).

Our Contribution and Main Results. Existing studies analyzing the different ways HDR images can be used for object detection are limited. Moreover, to the best of our knowledge, there is no study training object detection networks with real-world HDR images from scratch and comparing it with LDR images in a fair way. Our contribution is to provide a systematic and fair analysis on using LDR, HDR and tone-mapped LDR images for object detection for autonomous driving. Additionally, we introduce novel performance measures for analyzing the performance of detectors for different illumination conditions.

Our results suggest that using tone-mapped LDR images

This project was supported by the Royal Academy of Engineering, UK, through the Transforming Systems through Partnerships programme. Dr. Kalkan was supported by the BAGEP Award of the Science Academy, Turkey.

performs slightly better than raw HDR images. However, applying pre-processing (normalization and gamma correction) on HDR images improves the detection performance, resulting in a detection performance that is only slightly better than detection with the best tone-mapped LDR images. Our analysis on this marginal performance gain reveals that a detector for HDR images requires significantly more samples compared to using LDR images, since the space of HDR images is substantially larger (16 bits vs. 8 bits per pixel per channel).

2. RELATED WORK

Object Detection using deep networks mainly evolved around two approaches: namely one-stage and two-stage methods. Two-stage detectors have two distinct phases: one for region (object) proposal generation, and another stage for classification of these region proposals into objects. The region proposal stage learns to extract features and proposes regions possibly containing objects, and the second stage learns to classify the regions and find tighter bounding boxes around the objects. R-CNN [4], Fast R-CNN [5], Faster R-CNN [6] and Cascade R-CNN [7] are widely used two-stage detectors.

One-stage detectors, on the other hand, unify proposal and classification in a single stage. This is achieved by regularly placing a dense set of "anchor" boxes over an image and directly classifying each box into boxes and regressing their positions. YOLO [8], SSD [9] and RetinaNet [10] are widely used one-stage detectors.

Object Detection from HDR content is limited in the literature. One possible reason is the lack of a general purpose HDR detection dataset at a scale similar to e.g. COCO [11] or Pascal VOC [12] datasets, which provided a significant boost for object detection methods. For this reason, the few studies that perform object detection/recognition from HDR images use either a limited number of tone-mapped images [13] or synthetic data [14]. Mukherjee et al. [13], for example, collect their own dataset and test well-known object detection networks on tone-mapped images. This study is limited in that the authors do not use HDR or tone-mapped LDR images for training the network. Instead, they use a network pre-trained on LDR images and then test this network only on tone-mapped LDR images. In a more related study, Mukherjee et al. [15] generate an HDR dataset from LDR images and train their network on the generated HDR dataset. Next, they test their network on real-world HDR images and measure its performance on the subset of the images where the dynamic range is larger. This study is also limited in that it does not use real-world images for training the network, but only for testing. Furthermore, the subset they use has limited size and does not consider the analysis of the different ranges of dynamic range spectrum, such as lower or medium dynamic range scenes. Weiher [14], on the other hand, applies domain adaptation methods to a synthetic HDR dataset and train a semantic segmentation network on this adapted dataset. They test the network with real-world LDR and HDR images to see whether domain adaptation improves the performance. Due to the limited size of the synthetic dataset, they also pretrain the network on the COCO dataset.

3. METHODOLOGY

To test the hypothesis that HDR content might improve object detection accuracy in autonomous driving settings, we train our object detection network with standard LDR images, tone-mapped LDR images and HDR images, separately, as illustrated in Fig. 2.

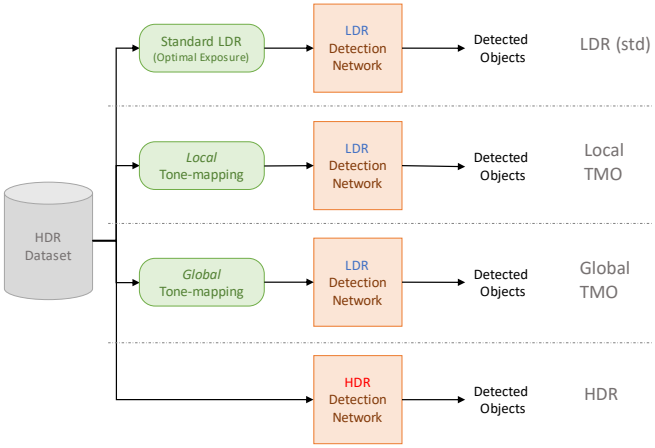


Fig. 2. The different methods we have analyzed for object detection with LDR and HDR images.

3.1. Compared Approaches

Tone-mapping operators (TMOs) aim projecting high-dimensional HDR content into LDR images while preserving the details and color appearance of the original content. For our paper, we have selected widely-used local and global TMOs and compared them with standard LDR and HDR images, namely (see also Fig. 2):

(i) Standard LDR: As the baseline, we take the "standard" LDR images as input to the object detection network. However, since this was not available in the CityScapes dataset, we used the method of optimal exposure compression proposed by [16] to achieve the best exposed LDR image from the HDR one.

(ii) HDR: We directly provided the 16-bit HDR images of the CityScapes dataset as input to the detection network.

(iii) HDR with Gamma: Gamma correction with $\gamma = 2.2$ is performed on the HDR images.

(iv) LDR with global tone-mapping:

- Reinhard: The widely-used photographic tone mapping method by Reinhard et al. [17] in global mode.

- Logarithmic compression: Logarithmic compression on the HDR images – this is provided by the CityScapes dataset.

(iv) LDR with local tone-mapping:

- Reinhard: The tone mapping method by Reinhard et al. [17] in local mode.
- Durand: The tone mapping method by Durand et al. [3]. Target contrast is set to 4. For the rest of the parameters, default values in the PFSTools [18] are used.
- Mantiuk: The tone mapping method by Mantiuk et al. [19]. Scaling factor is set to 0.7 and saturation correction is set to 1.0, as used in the OpenCV implementation [20].
- Fattal: The tone mapping method by Fattal et al. [21]. All parameters are default parameters provided by PFSTools [18].

3.2. Object Detection Network

In all experiments, we use Faster R-CNN [6] as our detector, as it is the seminal method for two-stage detection, providing a good baseline performance without bells and whistles. We follow the general architecture with backbone ResNet50 and feature pyramid networks [22] as commonly performed in the literature [10, 22, 23].

3.3. Dataset

Being the only readily available dataset with HDR images for autonomous driving, we use the CityScapes dataset [24] in this paper. CityScapes provides 16-bit HDR images and the corresponding LDR images obtained by logarithmic compression. The dataset contains 30 object categories, 8 with instance segmentation labels (namely; car, person, bicycle, rider, motorcycle, truck, bus, and train) and 2975 training, 500 validation and 1525 testing images. The dataset does not include bounding boxes for the objects. We used an existing tool for converting instance segmentation masks to object bounding boxes available in the mmdetection toolbox [25]. Since we need the ground-truth labels for extracting the bounding box information, we are unable to use the actual test set for CityScapes since the ground-truth is not publicly available. Instead, we use the validation set as the test set (500 images), and split the training set into training (2625 images) and validation sets (350 images).

4. EXPERIMENTS AND RESULTS

4.1. Implementation and training details

As a detector pre-trained on LDR images can provide bias for CityScape LDR images, we preferred to train Faster R-CNN from scratch. We used a training configuration similar to [25, 26] but slightly adjusted for training from scratch. We also tuned the learning rate for each input type. As the optimizer, we used Stochastic Gradient Descent (SGD), which is decreased by a factor of 10 at epoch 88. We also employed linear warm-up with ratio 0.1 at the beginning for 500 iterations. The networks were trained for 104 epochs on a single GPU with a batch size of 4. We also reduced the size of the images by half while keeping the ratio intact.

Table 1. Overall performance (mAP scores) for different methods in Fig. 2. LR: Learning Rate.

Method		LR	AP@0.5	mAP
Std. LDR		4e-3	55.1	33.1
Local TMO	Reinhard [17]	2e-3	55.7	33.2
	Durand [3]	2e-3	55.7	32.9
	Mantiuk [19]	4e-3	55.4	32.7
	Fattal [21]	2e-3	53.9	32.1
Global TMO	Log comp.	2e-3	53.7	31.9
	Reinhard [17]	2e-3	55.0	32.7
HDR with Gamma		2e-3	**56.1**	**33.3**
HDR		8e-3	55.2	32.9

4.2. Evaluation measures

Average Precision (AP). AP, commonly used in object detection [11, 12], is effectively a measure of the area under the precision-recall curve and AP@0.5 is calculated with 0.50 intersection-over-union (IoU) threshold. We also use the COCO-style mAP [11], which averages AP over 10 IoU thresholds and classes.

AP for diff. illumination intervals. To investigate scenarios where HDR or LDR might be advantageous, we calculate AP for objects (i.e. their bounding boxes) separately for different illumination categories. For this, we use dynamic range (DR) [27], which is the logarithm of the ratio of maximum luminance to the minimum luminance for the pixels in the box: $\text{mAP}_{\text{L-DR}}$ for low DR (0-5^{th} percentile), $\text{mAP}_{\text{L-M-DR}}$ for low-to-medium DR (5-50^{th} percentile), $\text{mAP}_{\text{M-H-DR}}$ for medium-to-high DR (50-95^{th} percentile), and $\text{mAP}_{\text{H-DR}}$ for high DR (95-100^{th} percentile).

4.3. Experiments

All results are obtained on the CityScapes validation set.

Experiment 1: Overall Performance. Table 1 shows the scores calculated over all classes. Overall, we observe that HDR with gamma correction performs slightly better than its closest counterparts, yet we do not observe a strong benefit of using HDR content either in the form of HDR or tone-mapped LDR when compared against the Std. LDR.

Experiment 2: Performance under different dynamic range (DR) values. This experiment provides a deeper analysis by looking at the performances of methods under different illumination categories defined in Section 4.2. Fig. 3 displays the detection performances under four intervals of dynamic

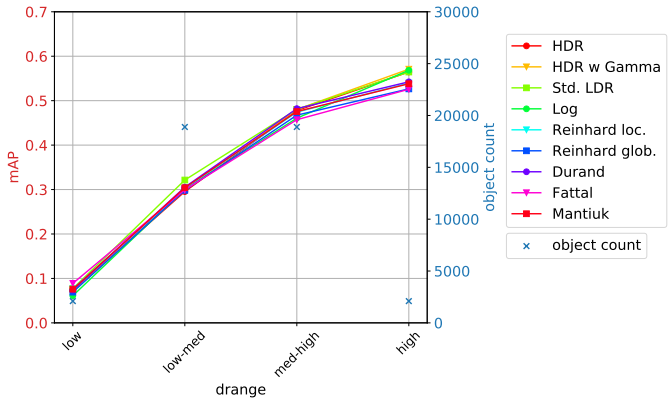


Fig. 3. Performance (mAP scores) under different illumination conditions different methods in Fig. 2. All models are evaluated on the validation set.

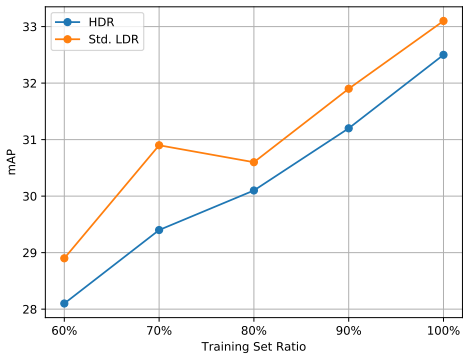
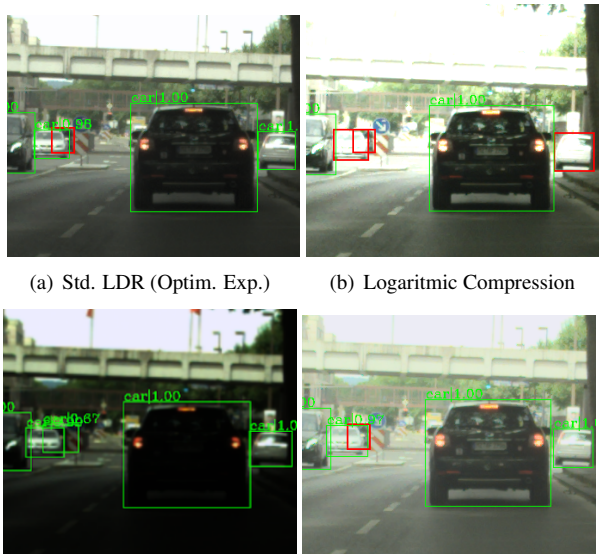


Fig. 4. Providing less data for Std. LDR and HDR. X-axis Percentages ($X\%$) indicate the ratio of the training set used in the experiment. Y-axis indicate the detection evaluation measure (mAP).

range. We observe that no method provides a significant gain over others in any illumination interval. Nonetheless, some methods including HDR with Gamma and Standard LDR seem to perform slightly better than the rest in high dynamic range illumination.

Experiment 3: Does HDR need more data? 16-bit HDR naturally spans a larger space of intensity values and therefore, we hypothesize that the HDR-trained detectors might require more data to obtain the same level of performance as LDR images. To test this hypothesis, we consider training LDR and HDR networks with different amounts of data. In Fig. 4, we see that, indeed, an HDR network provides a comparable level of performance with an LDR network with approx. 10% more data.

Qualitative Results We provide visual detection results in Fig. 5 from a challenging scene from the CityScapes dataset, where HDR content helps detecting the cars further away in the road in an over-exposed region, otherwise undetectable in the LDR content.



(a) Std. LDR (Optim. Exp.) (b) Logaritmic Compression

(c) 16-bit HDR (d) Reinhard (Global)

Fig. 5. Detection results. Missed objects are shown in red.

5. CONCLUSION

In this work, we analyzed the effect of using high dynamic range content in object detection, specifically for autonomous driving. We compared LDR, HDR and tone-mapped LDR images in a systematic way. Our analysis suggests that the improvement obtained when using tone-mapped LDR or HDR images for object detection is rather minimal. Moreover, our deeper analysis with different illumination intervals reveals that, contrary to our expectation, using HDR images performs on par with using the Std. or tone-mapped LDR images in different illumination conditions. To investigate the cause of these unexpected findings, we compared a detector trained with different number of HDR and LDR images, which revealed that a detector using HDR images requires more data as HDR images span a wider range of intensity values. Despite these findings, we showed qualitatively that there are challenging illumination conditions where a detector with HDR images does provide better performance.

6. REFERENCES

[1] B. Wu, F. Iandola, P.H. Jin, and K. Keutzer, "Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," in *CVPR Workshops*, 2017.

[2] M. Hnewa and H. Radha, "Object detection under rainy conditions for autonomous vehicles: A review of state-

of-the-art and emerging techniques," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 53–67, 2020.

[3] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *SIGGRAPH*, 2002.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014.

[5] R. Girshick, "Fast r-cnn," in *ICCV*, 2015.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE PAMI*, vol. 39, no. 6, 2016.

[7] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *CVPR*, 2018.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *CVPR*, 2016.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A.C. Berg, "Ssd: Single shot multibox detector," in *ECCV*, 2016.

[10] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *ICCV*, 2017.

[11] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014.

[12] M. Everingham, L. Van Gool, C. KI Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *IJCV*, vol. 88, no. 2, pp. 303–338, 2010.

[13] R. Mukherjee, M. Melo, V. Filipe, A. Chalmers, and M. Bessa, "Backward compatible object detection using hdr image content," *IEEE Access*, vol. 8, 2020.

[14] M. Weiher, "Domain adaptation of hdr training data for semantic road scene segmentation by deep learning," *Master Thesis, Technical University of Munich*, 2019.

[15] Ratnajit Mukherjee, Maximino Bessa, Pedro Melo-Pinto, and Alan Chalmers, "Object detection under challenging lighting conditions using high dynamic range imagery," *IEEE Access*, vol. 9, pp. 77771–77783, 2021.

[16] K. Debattista, T. Bashford-Rogers, E. Selmanović, R. Mukherjee, and A. Chalmers, "Optimal exposure compression for high dynamic range content," *The Visual Computer*, vol. 31, no. 6-8, pp. 1089–1099, 2015.

[17] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *SIGGRAPH*, 2002.

[18] R. Mantiuk, "Pfstools, `http://pfstools.sourceforge.net/`, v2.1.0," 2017.

[19] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM T. on Applied Perception (TAP)*, vol. 3, no. 3, pp. 286–308, 2006.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[21] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *SIGGRAPH*, 2002.

[22] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017.

[24] 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 *CVPR*, 2016.

[25] K. Chen, J. Wang, J. Pang, and Y. Cao et al., "Mmdetection: Open mmlab detection toolbox and benchmark," 2019.

[26] C. Michaelis, B. Mitzkus, R. Geirhos, and E. et al. Rusak, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," *arXiv:1907.07484*, 2019.

[27] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux, "Performance evaluation of objective quality metrics for hdr image compression," in *Applications of Digital Image Processing XXXVII*, 2014, vol. 9217.