Pattern Recognition Letters 172 (2023) 230–236

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

TMO-Det: Deep tone-mapping optimized with and for object detection

Ismail Hakki Kocdemir[a,*], Alper Koz[b], Ahmet Oguz Akyuz[a], Alan Chalmers[c], Aydin Alatan[d], Sinan Kalkan[a]

[a] Dept. of Computer Engineering & Center for Image Analysis (OGAM), METU, Turkey
[b] Center for Image Analysis (OGAM), METU, Turkey
[c] WMG, University of Warwick, England, United Kingdom
[d] Dept. of Electrical-Electronics Eng. & Center for Image Analysis (OGAM), METU, Turkey

ARTICLE INFO

Article history:
Received 11 October 2022
Revised 15 May 2023
Accepted 28 June 2023
Available online 29 June 2023

Edited by: Jiwen Lu

Keywords:
Object detection
High dynamic range
Low dynamic range
Tone-Mapping
Generative adversarial networks

ABSTRACT

Detecting objects in challenging illumination conditions is critical for autonomous driving. Existing solutions detect objects with standard or tone-mapped Low Dynamic Range (LDR) images. In this paper, we propose a novel adversarial approach that jointly optimizes tone-mapping (mapping High Dynamic Range (HDR) to LDR) and object detection. We analyze different ways to combine the feedback from tone-mapping quality and object detection quality for training such an adversarial network. We show that our deep tone-mapping operator jointly trained with an object detector achieves the best tone-mapping quality as well as detection quality compared to alternative approaches.

1. Introduction

An open problem in computer vision is detecting objects under adverse conditions of illumination, e.g., when entering or exiting a tunnel, driving towards Sun or under the headlights of an oncoming car - see, e.g., Fig. 1. Existing systems generally use Low Dynamic Range (LDR) cameras, in addition to various depth and radar sensors, for visual perception [5,6]. However, LDR cameras may be insufficient for providing discriminative details about objects in dark or bright regions of a scene. In such challenging scenarios, the ability of a camera to capture the darkest and brightest areas in a scene without losing any detail, i.e., the dynamic range of the camera, makes a significant difference [7]. Modern cameras with High Dynamic Range (HDR) capabilities can capture details in a scene with extremely bright and quite dark regions. However, despite their potential, the use of HDR cameras for challenging lighting conditions in object detection, and more importantly in autonomous driving, has not been explored extensively.

A straightforward approach to benefit from HDR content in adverse illumination conditions would be to directly use HDR images as input for training deep object detectors. However, this is prohibitive, as the HDR space is significantly wider than LDR space, and therefore, it requires much more data to train detectors [8]. An alternative is to first tone-map HDR images to LDR images, and then to provide tone-mapped LDR images as input to object detection networks.

Thanks to advances in deep generative modeling, high-quality tone-mapping operators (TMOs) can be obtained with Generative Adversarial Networks (GANs). Although such deep tone-mapping approaches have produced perceptually remarkable LDR images, they rely on classically tone-mapped LDR images for training the generator. Therefore, images generated by classical TMOs or deep TMOs are not optimized to consider performance in downstream vision problems, such as object detection or image classification.

In this paper, to address these issues, we propose a novel approach, called TMO-Det, that jointly optimizes a GAN-based TMO and an object detector. We achieve this by extending a GAN-based TMO with object-detection objectives to consider maximizing detection performance while optimizing tone-mapping quality (Fig. 1).

2. Related work and background

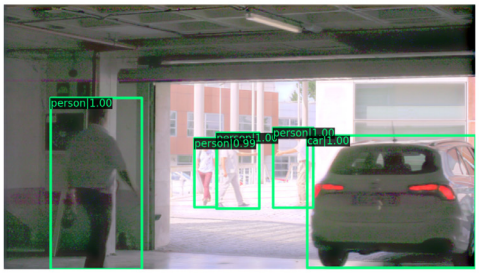
2.1. Deep generative modeling

Generative Adversarial Networks (GANs) [9] are a family of networks that map a latent space (Z) to a target space based on an ad-

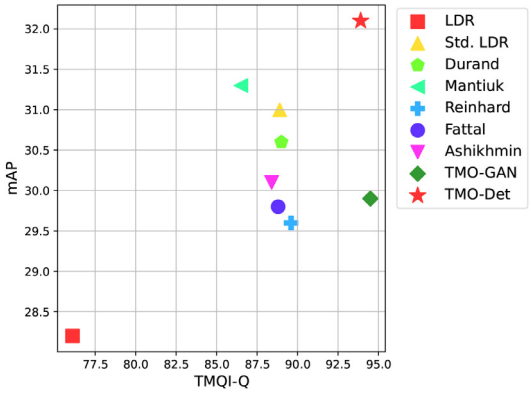
* Corresponding author.
E-mail addresses: ismail.kocdemir@metu.edu.tr (I.H. Kocdemir), koz@metu.edu.tr (A. Koz), aoakyuz@metu.edu.tr (A.O. Akyuz), Alan.Chalmers@warwick.ac.uk (A. Chalmers), alatan@metu.edu.tr (A. Alatan), skalkan@metu.edu.tr (S. Kalkan).

https://doi.org/10.1016/j.patrec.2023.06.017
0167-8655/

(a) RetinaNet results on an LDR image [1].



(b) TMO-Det detection & tone-mapped LDR image output.



(c) Detection vs. HDR quality.

Fig. 1. (a) LDR images make it difficult to detect objects in adverse illumination conditions. (b) TMO-Det is able to jointly tone-map and detect objects that cannot be detected in LDR images. Blue: detected objects. Red: Missed objects. The input is from the OOD dataset [2]. (c) Our method (TMO-Det ★) provides the best detection performance (mAP) and HDR quality (TMQI-Q [3]) compared to RetinaNet [4] trained on hand-designed or learned tone-mapping operators. The results are obtained on the OOD dataset [2].

versarial interplay between two components, namely the discriminator D and the generator G:

$$\mathcal{L}_{GAN}(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log\ D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z))], \tag{1}$$

which is maximized by D and minimized by G. GANs can be easily extended to generate samples that match a condition (called Conditional GAN [10]):

$$\mathcal{L}_{CGAN}(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x,y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z,y),y)], \tag{2}$$

where y is the additional variable on which generation and discrimination are conditioned.

2.2. Deep tone-mapping

Conventional tone-mapping operators that generate LDR images from HDR images are based on global and local characteristics of individual images. With advances in deep learning, recent work has been using generative models (mostly GANs) to learn tone-mapping from the data [11,12]. These approaches generally perform image-to-image translation using conditional GANs, as follows:

$$\mathcal{L}_{CGAN}(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x,T(x))] + \mathbb{E}_{x\sim p_x(data)}[\log(1-D(x,G(x))], \tag{3}$$

where x is an HDR image; T is a classical tone-mapping method, producing the ground-truth tone-mapped image $T(x)$; $G(x)$ is the generated image mimicking $T(\cdot)$; D is the discriminator that discriminates between the images tone-mapped by T and G.

The objective in Eq. 3 has been extended by Rana et al. [11] with (i) a feature-matching loss to penalize G with the discrepancy between the representations of $G(x)$ and $T(x)$ in the intermediate layers of D, and (ii) a perceptual loss to penalize the discrepancy between the representations of $G(x)$ and $T(x)$ in the intermediate layers of a pre-trained backbone network. Panetta et al. [12] report further improvements using an additional gradient-profile loss between the gradient maps of $G(x)$ and $T(x)$.

2.3. Image enhancement

A related line of study pertains to the enhancement of badly illuminated LDR images without the use of HDR content. For example, Li et al. [13] propose a CNN architecture for such a low-light image enhancement problem. Yan et al. [14] present a novel multi-exposure fusion based visual quality enhancement pipeline. Xie et al. [15] also propose a fusion-based approach but additionally makes use of the scene semantics.

2.3.1. Object detection with HDR content

Despite the potential, only a few studies have focused on HDR Object Detection. One possible reason is the lack of a general purpose HDR detection dataset at a scale similar to, e.g., COCO [16] or Pascal VOC [17] datasets, which were a significant resource for object detection methods. The few studies on object detection/recognition from HDR images use either a limited number of tone-mapped images [2] or synthetic data [18]. Mukherjee et al. [2], for example, collect their own dataset and test well-known object detection networks on tone-mapped images. However, the authors do not use HDR or tone-mapped LDR images for training the network but utilize a network pre-trained on LDR images and test the network only on tone-mapped LDR images. A more related study [19] generates an HDR dataset from LDR images and trains a detector on the generated HDR dataset. Then it tests the network on real-world HDR images and measures its performance on the subset of the images where the dynamic range is larger. However, the approach does not use real-world images for training the network, but only for testing. Furthermore, the subset it uses has a limited size and does not consider the analysis of the different ranges of dynamic spectrum such as lower- or medium-dynamic range scenes.

2.4. Comparative summary

There are studies that use deep generative models for developing better tone-mapping operators and studies that use HDR or tone-mapped LDR images for better object detection. However, there is no study that jointly optimizes a deep tone-mapping network with an object detector. This is the topic that is addressed in this paper.

3. Methodology

As illustrated in Fig. 2, our approach augments a GAN-based tone-mapping network with the supervision signal of an object detector. This is achieved by making the generator deliver images

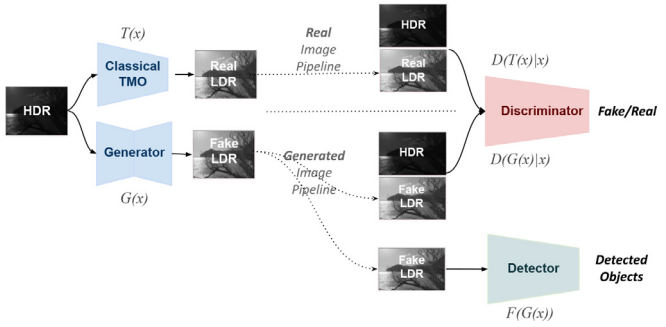


Fig. 2. Overall architecture diagram for the proposed method that combines object detection and tone-mapping objectives.

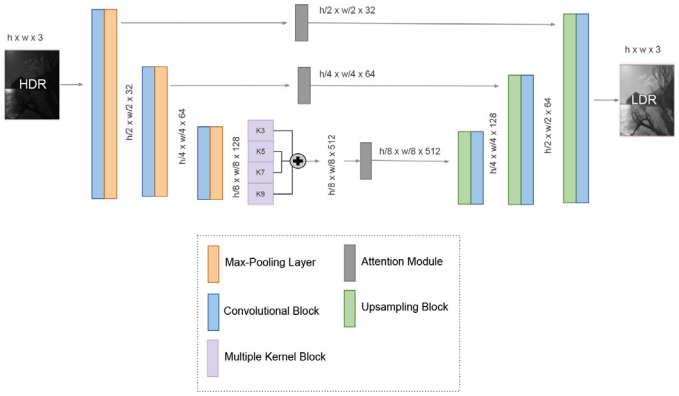


Fig. 3. Overall diagram for the proposed generator architecture.

such that visual similarity to a classical TMO result is enforced with the help of a discriminator, whilst maximizing object detection quality with the help of a detector.

3.1. TMO-Det

In order to jointly optimize the object detection network and deep learning-based TMOs, we modify the objective in the conditional least squares GAN-based TMO framework as follows:

$$\mathcal{L}_{\text{TMO-Det}} = \mathcal{L}_D + \alpha_{\text{det}}\mathcal{L}_{Det} + \alpha_{\text{non-det}}(\mathcal{L}_G + \beta\mathcal{L}_{GPL} + \gamma\mathcal{L}_{FM}), \quad (4)$$

where α_{det}, β and γ are the weights for the detection loss, gradient profile loss and feature matching loss, respectively. $\alpha_{\text{non-det}}$ controls the weight for all losses that are not directly related to detection. The individual loss terms are defined as follows:

$$\mathcal{L}_{Det} = \mathbb{E}_{x\sim p_{data}}[L_{cls}(F(G(x)) + \lambda L_{loc}(F(G(x)))], \quad (5)$$

$$\mathcal{L}_G = \frac{1}{2}\,\mathbb{E}_{x\sim p_{data}}[(1 - D(G(x)|x))^2], \quad (6)$$

$$\mathcal{L}_D = \frac{1}{2}\,\mathbb{E}_{x\sim p_{data}}[(D(T(x)|x) - 1)^2] + \frac{1}{2}\,\mathbb{E}_{x\sim p_{data}}[D(G(x)|x)^2], \quad (7)$$

where x and $T(x)$ represent the HDR and ground-truth tone-mapped LDR images, as illustrated in Fig. 2.

Note on α_{det} and $\alpha_{\text{non-det}}$. To decouple the effect of the detection loss on the generator and the detector, we apply α_{det} in the backward pass between the detector and the generator, by scaling the gradients flowing from the detector to the generator. This way we prevent the detector getting unnecessarily large (or small) updates whilst providing strong influence to the generator from the detection objective.

3.1.1. Architecture details

Generator (G). The architecture of the proposed generator is illustrated in Fig. 3. We use leaky ReLU as the activation function and Instance Normalization for normalizing the activation values. Additionally, we augment the skip connections in the UNet architecture with self-attention [20]. Each feature map in a single level of the network is carried across by a separate attention module, where attention queries $Q^{(i)}$ for layer i are calculated from original image while keys $K^{(i)}$ and values $V^{(i)}$ are calculated from intermediate feature maps as follows:

$$\begin{aligned} Q^{(i)} &= A_q^i(I^{(i)}), \\ K^{(i)} &= A_k^i(G^{(i)}(I)), \\ V^{(i)} &= A_v^i(G^{(i)}(I)), \end{aligned} \quad (8)$$

where i represents the layer index for G; A_q^i, A_k^i, A_v^i are 1×1 convolutional networks, and $I^{(i)}$ is the original image downsampled to the spatial size of the feature map in $G^{(i)}$. The resulting attention values are calculated as follows:

$$O^{(i)} = A_o^i(\text{softmax}(Q^{(i)}K^{(i)^\top})V^{(i)}), \quad (9)$$

where A_o^i is similarly a 1×1 convolutional layer to recover the input channel size. At the innermost layer of the generator, we also convolve the feature maps with multiple kernels with different sizes (Multiple Kernel Block); namely 3, 5, 7 and 9, and concatenate the resulting feature maps.

Discriminator (D). The discriminator is a standard patch discriminator that has a 70×70 receptive field in the final layer [21]. Similar to the generator, we use leaky ReLU as the activation function and Instance Normalization for normalizing the activation values.

Detector (F). We use RetinaNet [4], since it is a prominent and easy-to-adapt one-stage detector (though, our approach can work with any deep object detector). We follow the general architecture with the ResNet50 backbone [22] and Feature Pyramid Networks [23].

4. Experiments and results

4.1. Tone-mapping operators and the compared approaches

We have selected the following approaches for comparison:
(i) Detection with LDR images:

- **LDR**: The middle range of each HDR image, similar to Mukherjee et al. [19]. This middle range has the same range with an LDR image.
- **Std. LDR**: Optimal exposure LDR image. We use the optimal exposure compression method proposed by [24] to achieve the best exposed LDR image from the HDR one, as if the scene is captured by a virtual LDR camera with an optimal exposure setup.

(ii) Detection with HDR images:

- **HDR**: Raw HDR images.
- **HDR with Gamma**: (i) We apply min-max normalization to each image, by subtracting its minimum value and dividing by its pixel range, (ii) we apply gamma encoding (correction) on the normalized image, and (iii) we scale the gamma-corrected image to the LDR range ([0,255]).

(iii) Detection with LDR images obtained by the classical tone-mapping operators:

- **Ashikhmin**: The tone-mapping method by Ashikhmin et al. [25].

- **Reinhard**: The tone-mapping method by Reinhard et al. [26] in local mode.
- **Durand**: The tone-mapping method by Durand et al. [1]. The target contrast is set to 4. For the rest of the parameters, the default values in PFSTools [27] are used.
- **Mantiuk**: The tone-mapping method by Mantiuk et al. [28]. The scaling factor is set to 0.7 and the saturation correction to 1.0, as used in the OpenCV implementation [29].
- **Fattal**: The tone mapping method by Fattal et al. [30]. All parameters are default parameters provided by PFSTools [27].
- **Best TMQI per picture**: For this method, we choose the tone-mapping operator that performs best in terms of the TMQI metric [3]. In this option, different pictures might be tone-mapped with different operators.

(iv) Detection with learning-based tone-mapping:

- **TMO-GAN**: Our proposed architecture, without jointly training with an object detector.
- **TMO-Det**: Our proposed architecture with the detector, which is jointly trained with TMO-GAN. TMO-GAN is pre-trained on the OOD (out-of-distribution) dataset without the detector, and the detector is pre-trained disjointly on the outputs of the trained and frozen TMO-GAN on top of the MS COCO pre-training.

4.2. Dataset

For all experiments in this section, we use the OOD dataset [2] which consists of HDR images with annotated labels for 20 classes from the Pascal VOC dataset [17]. We filter the dataset by removing nearly identical frames from the videos and split the dataset such that we have 1491 training and 380 test images. Additionally, we downsize the images into 1024×576 resolution before performing the experiments.

We form our ground-truth LDR images by selecting the best classical TMO among the ones in Section 4.1 for each picture based on the TMQI (Tone-mapping Quality Index) metric [3].

4.3. Implementation and training details

Initialization. We initialize all RetinaNet architectures at least from their COCO pre-trained versions. For joint training with TMO-GAN, we employ different initializations for the detector and TMO-GAN as mentioned in Section 4.1.

Data Augmentation. We perform the same data augmentation techniques for all experiments. We apply random cropping with a minimum of 0.3 scaling factor, so that the crop contains at least 1 ground truth object bounding box with a minimum of 0.3 Intersection-over-Union with the original box. Furthermore, we apply random horizontal flipping with a probability of 0.5.

4.3.1. Dataset tone-mapped by classical TMOs + detector (disjoint training)

For the detectors, we use Stochastic Gradient Descent (SGD) with a learning rate of 0.001, which is decreased by a factor of 10 at epoch 7. We also employ linear warm-up with ratio 0.1 at the beginning for 500 iterations. The networks are trained for 14 epochs on a single GPU with a batch size of 8.

4.3.2. Dataset tone-mapped by TMO-GAN + detector (disjoint training)

We use Adam [31] with a learning rate of 0.0002 for the generator and the discriminator. The networks are trained for 20 epochs on a single GPU with a batch size of 8. After 20 epochs, the learning rate is decayed linearly to 0 until the 50^{th} epoch. β is set to 0.8 and γ is set to 10. The detector is trained in an identical way to Section 4.3.1.

Table 1
Tone-mapping quality results, comparing hand-crafted TMOs with ours and other deep-learning based approach proposed by Rana et al. [11]. TMQI-Q, TMQI-N and TMQI-S denote the scores for overall quality, naturalness and structural fidelity, respectively [3].

Method	TMQI ↑		
	Q	N	S
LDR	76.1	79.8	4.4
Std. LDR	89.7	90.9	53.1
Ashikhmin	88.4	88.8	47.9
Durand	89.0	**92.2**	45.5
Fattal	88.8	**92.2**	45.4
Mantiuk	86.5	91.6	34.2
Reinhard	89.6	85.9	71.5
DeepTMO [11]	93.4	89.8	74.0
TMO-GAN	**94.5**	90.7	**79.9**

4.3.3. Dataset tone-mapped by TMO-Det (joint training)

We fine tune the pre-trained networks using Stochastic Gradient Descent with a learning rate of 0.0001. We use warm-up for the first 2 epochs. All networks are trained for 15 epochs jointly. The detection pipeline (generator + detector) is trained for 15 more epochs. Cosine scheduling is used for scheduling the learning rate.

β is set to 0.8 whereas γ is set to 10, for both versions (with and without joint training). α_{det} and $\alpha_{non-det}$ are set to 1 for all experiments except the one in which we provide different weights for different objectives using α_{det} and $\alpha_{non-det}$, and analyze the results.

4.4. Evaluation measures

Object Detection. We use the COCO-style mAP (mean Average Precision) as our evaluation measure, as it is commonly used in the object detection benchmarks [16,17]. AP is effectively a measure of the area under the precision-recall curve and is calculated with a certain intersection-over-union (IoU) threshold. The COCO-style mAP [16] averages AP over 10 different IoU thresholds and classes.

To investigate scenarios where HDR or LDR might be advantageous, we also calculate AP by categorizing the objects with respect to the illumination in their bounding boxes. For this, we use *dynamic range* (DR) [32], defined as the logarithm of the ratio of maximum luminance to the minimum luminance for the pixels in the box. AP for different DRs is denoted as follows: $\text{mAP}_{\text{L-DR}}$ 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).

Tone-mapping Quality. For evaluating tone-mapping quality, we use the Tone-mapping Quality Index (TMQI) as our measure [3], which outputs three different scores: (i) TMQI-Q, representing the overall quality, (ii) TMQI-N, representing the naturalness of the tone-mapped image, and (iii) TMQI-S, representing the structural fidelity with respect to the original HDR image.

4.5. Experiment 1: TMO-GAN - TMO quality

In Table 1, we compare TMO-GAN with other TMO operators. We observe that TMO-GAN outperforms all other single TMOs in terms of overall quality (TMQI-Q) by a large margin. Additionally, it surpasses DeepTMO by a considerable margin in all metrics (Q, N and S).

4.6. Experiment 2: TMO-Det, joint training

In these experiments, we jointly train our TMO-GAN with the detector. We perform the same experiments with RetinaNet.

Table 2
Overall performance (mAP scores for detection, and TMQI for TMO quality) for the methods described in Section 4.1, where the detector is chosen as RetinaNet. The best are shown in bold.

TMO	Joint	Real	mAP ↑	TMQI-Q ↑
HDR	-	-	26.3	-
HDR with Gamma	-	-	29.8	-
LDR	-	-	28.2	76.1
Std. LDR	-	-	31.0	88.9
Durand	-	-	30.6	89.0
Reinhard	-	-	29.6	89.6
Fattal	-	-	29.8	88.8
Ashikhmin	-	-	30.1	88.4
Mantiuk	-	-	31.3 ± 0.19	86.5
TMO-GAN	X	-	29.9 ± 0.18	**94.5** ± 0.16
TMO-Det	✓	✓	**32.1** ± 0.09	93.9 ± 0.08

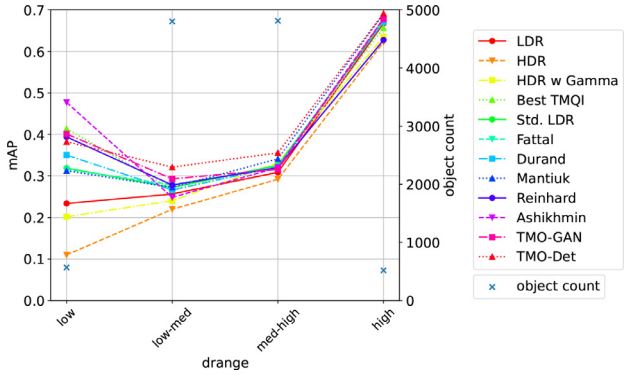


Fig. 4. Overall performance (mAP scores) for RetinaNet under different dynamic range intervals for the methods described in Section 4.1.

Table 2 compares the detection performance and TMO quality of the proposed architecture with classical methods and disjoint training. We report the average of 5 runs for best 3 methods in terms of object detection performance. Based on the result in Table 2, we observe the following:

- TMO-Det achieves the best detection scores while preserving its high-quality tone-mapping in terms of TMQI compared to TMO-GAN. Detection performance is ∼0.8 mAP points higher than the best classical method (Mantiuk [28]).
- Training the detector with real images on top of thegenerated images simultaneously improves the performance. We hypothesize that this makes detector training more stable at the beginning, since the generated images have poorer quality in the initial phase of training. Additionally, using a pre-trained TMO-GAN together with the pre-trained detector can also further improve the detection performance.

We also compare our joint training methodology to other methods under different illumination conditions. Fig. 4 shows the detection performance under four intervals of dynamic range. We observe that TMO-Det performs similarly to other TMOs. Our analysis, using dynamic range, brings out differences amongst methods in adverse conditions: In areas with lowest dynamic range (as per our definition in Section 4.4), the methods seem to differ the most. In particular, HDR without normalization and gamma correction performs worse in low dynamic range areas.

4.7. Experiment 3: The effect of α_{det} and $\alpha_{\text{non-det}}$

In these experiments, we try a range of values for the weights applied on the two different objectives in Eq. 4. We choose RetinaNet as the detector and train it with synthetic (fake) and real images together. We also equip TMO-GAN with all additional features: hard-tanh, attention, and skip-image, and jointly train the overall architecture similar to Experiment 2.

Table 3
Overall performance (mAP scores for detection, and TMQI for TMO quality, reported as the average of 5 runs) for different values of α_{det} and $\alpha_{\text{non-det}}$, where the RetinaNet is chosen as the detector. The best are shown in bold.

TMO	α_{det}	$\alpha_{\text{non-det}}$	mAP ↑	TMQI-Q ↑
HDR with Gamma	-	-	29.8	-
Mantiuk	-	-	31.3	86.5
TMO-GAN	-	-	29.9	**94.5**
TMO-Det	1.0	1.0	32.0	93.9
TMO-Det	1.2	1.0	**32.1**	94.0
TMO-Det	1.5	1.0	31.7	93.8
TMO-Det	2.0	1.0	30.4	93.5
TMO-Det	1.0	2	31.1	94.0
TMO-Det	1.0	4	29.8	94.2
TMO-Det	1.0	8	29.9	94.1

Table 4
Overall performance (mAP scores for detection, and TMQI for TMO quality) for different values of λ_{obj}, where the RetinaNet is chosen as the detector. The best are shown in bold.

TMO	λ_{obj}	mAP ↑	TMQI-Q ↑
HDR with Gamma	-	29.8	-
Mantiuk	-	31.3	86.5
TMO-Det	1.0	32.0	**93.9**
TMO-Det	1.5	**32.1**	93.9
TMO-Det	2.0	**32.1**	94.1
TMO-Det	2.5	31.6	93.8

The results are listed in Table 3. We observe that increasing the influence of the detector on the generator (i.e., higher α_{det}) provides slight improvements. Furthermore, increasing the influence of the discriminator-related objectives (Eq. 4) via $\alpha_{\text{non-det}}$ slightly improves the TMO quality but significantly deteriorates object-detection performance. Finally, when the discriminator-related objectives are decreased below 0.8, the networks diverge.

4.8. Experiment 4: Object-aware patch discriminator

In this step, we propose applying different weights for the discriminator feedback for locations with objects and without objects. As our primary goal is to detect objects, using this approach, we aim to penalize the generator to produce better images in locations that contain objects. We produce binary masks such that a mask pixel is set to 1 if the image pixel belongs to an object or 0 otherwise. Then, we resize this mask to the output of the patch discriminator by using nearest neighbor interpolation. Finally, we use the mask to apply increased weights to the locations that contain objects as follows:

$$\mathcal{L}_G = \frac{\sum_{i,j}\left[\lambda_{obj} M_{i,j} D(I)_{i,j} + (1 - M_{i,j}) D(I)_{i,j}\right]}{\sum_{i,j}\left[\lambda_{obj} M_{i,j} + 1 - M_{i,j}\right]}, \tag{10}$$

where M is the resized binary mask; and D is the discriminator; and I is the input image. i and j are the coordinates of the discriminator output. λ_{obj} designates the weight applied to the locations that contain objects. For the experiments, we use the same settings as in Experiment 2 for combining TMO-GAN and RetinaNet.

As shown in Table 4, we find that setting $\lambda_{obj} = 2$ gives the highest detection score, slightly improving over the default settings. However, we observe that higher values of λ_{obj} do not improve the performance further.

Table 5
Performance of the TMO-GAN and TMO-Det with and without discriminator.

TMO	Disc.?	Det.?	mAP ↑	TMQI-Q ↑
TMO-GAN	✓	X	29.9	**94.5**
TMO-GAN	X	✓	28.2	87.3
TMO-Det	✓	✓	**32.1**	93.9

4.9. Experiment 5: HDR vs. TMO-GAN without the discriminator

With joint training, we effectively use a larger network (Generator + Detector) to detect objects. Here, we design an experiment where we can see how much TMO-GAN + detector (joint training) improves the detection performance on HDR images over just using a larger network. To achieve this goal, we remove the discriminator from the architecture and use only the detection loss (equivalent to setting $\alpha_{\text{non-det}}$ to zero). As we can see from Table 5, the baseline network without the discriminator can improve over HDR (Table 2). However, it falls behind our joint method, which shows the additional improvement provided by joint training (TMO-Det). This result is also in agreement with the results in Experiment 3 where the decrease $\alpha_{\text{non-det}}$ also affects the performance.

4.10. Experiment 6: Detection performance vs. TMO quality

In this experiment, we aim to examine the relation between the detection performance (mAP) and TMO quality metrics (TMQI-Q). To this end, we plot the mAP scores against the TMO quality metric in Fig. 1, for different methods. As can be seen in the figure, our architectures (TMO-GAN and TMO-Det) achieve the top TMQI-Q scores.

5. Conclusion

In this work, we try to improve a GAN-based tone-mapping operator (TMO-GAN) by introducing a detection network into the 2-player adversarial game. In our approach, called TMO-Det, we supervise the generator with the help of an object detection network in addition to the discriminator. We compare the performance of TMO-Det against classical TMOs in terms of image quality and detection performance. We showed that, by jointly training the detection and tone-mapping objectives, we are able to improve detection performance (although the difference between the second best is not very significant) whilst maintaining a good tone-mapping quality in terms of the selected measure [3]. This is significantly better than the classical TMOs.

Although our approach is promising, it has certain limitations, which can provide opportunities for further research. One issue is that it contains multiple sub-networks (i.e., a generator, a discriminator, and a detector) for jointly optimizing tone-mapping and detection qualities. This can incur more GPU memory and time compared to a stand-alone object detector or a stand-alone deep TMO network. Future work can mitigate this limitation by employing more resource efficient and faster layers in these sub-networks. Moreover, our network has many different objectives (tasks), which requires carefully tuning many hyper-parameters to obtain good results. Despite the significant gains we have obtained, we believe that there is potential for even further improvements with multi-task learning approaches.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ismail Hakki Kocdemir reports financial support was provided by Royal Academy of Engineering. Ismail Hakki Kocdemir, Alper Koz, Oguz Akyuz, Aydin Alatan, Sinan Kalkan has patent pending to TURKPATENT.

Data availability

The authors do not have permission to share data.

Acknowledgments

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

References

[1] F. Durand, J. Dorsey, Fast bilateral filtering for the display of high-dynamic-range images, in: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002, pp. 257–266.
[2] R. Mukherjee, M. Melo, V. Filipe, A. Chalmers, M. Bessa, Backward compatible object detection using hdr image content, IEEE Access 8 (2020).
[3] H. Yeganeh, Z. Wang, Objective quality assessment of tone-mapped images, IEEE Trans. Image Process. 22 (2) (2012) 657–667.
[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.
[5] B. Wu, F. Iandola, P. Jin, K. Keutzer, Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving, CVPR Workshops, 2017.
[6] M. Hnewa, H. Radha, Object detection under rainy conditions for autonomous vehicles: a review of state-of-the-art and emerging techniques, IEEE Signal Process. Mag. 38 (1) (2020) 53–67.
[7] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, K. Myszkowski, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting, Morgan Kaufmann, 2010.
[8] I.H. Kocdemir, A.O. Akyuz, A. Koz, A. Chalmers, A. Alatan, S. Kalkan, Object detection for autonomous driving: high-dynamic range vs. low-dynamic range images, in: 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), IEEE, 2022, pp. 1–5.
[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[10] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784 (2014).
[11] A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, A. Smolic, Deep tone mapping operator for high dynamic range images, IEEE Trans. Image Process. 29 (2019) 1285–1298.
[12] K. Panetta, L. Kezebou, V. Oludare, S. Agaian, Z. Xia, Tmo-net: a parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale hdr dataset, IEEE Access 9 (2021) 39500–39517.
[13] C. Li, J. Guo, F. Porikli, Y. Pang, Lightennet: a convolutional neural network for weakly illuminated image enhancement, Pattern Recognit. Lett. 104 (2018) 15–22.
[14] Q. Yan, Y. Zhu, Y. Zhou, J. Sun, L. Zhang, Y. Zhang, Enhancing image visuality by multi-exposure fusion, Pattern Recognit. Lett. 127 (2019) 66–75. Advances in Visual Correspondence: Models, Algorithms and Applications (AVC-MAA)
[15] J. Xie, H. Bian, Y. Wu, Y. Zhao, L. Shan, S. Hao, Semantically-guided low-light image enhancement, Pattern Recognit. Lett. 138 (2020) 308–314.
[16] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. Zitnick, Microsoft coco: common objects in context, ECCV, 2014.
[17] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, IJCV 88 (2) (2010) 303–338.
[18] M. Weiher, Domain adaptation of hdr training data for semantic road scene segmentation by deep learning, Master Thesis, Technical University of Munich (2019).
[19] R. Mukherjee, M. Bessa, P. Melo-Pinto, A. Chalmers, Object detection under challenging lighting conditions using high dynamic range imagery, IEEE Access 9 (2021) 77771–77783.
[20] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, in: International conference on machine learning, PMLR, 2019, pp. 7354–7363.
[21] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
[22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

[24] K. Debattista, T. Bashford-Rogers, E. Selmanović, R. Mukherjee, A. Chalmers, Optimal exposure compression for high dynamic range content, Vis. Comput. 31 (6–8) (2015) 1089–1099.
[25] M. Ashikhmin, A tone mapping algorithm for high contrast images, Rendering Techniques, 2002.
[26] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, in: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002, pp. 267–276.
[27] R. Mantiuk, Pfstools, http://pfstools.sourceforge.net/, v2.1.0 (2017).
[28] R. Mantiuk, K. Myszkowski, H.-P. Seidel, A perceptual framework for contrast processing of high dynamic range images, ACM Trans. Appl. Percept. (TAP) 3 (3) (2006) 286–308.
[29] G. Bradski, The opencv library, Dr. Dobb's J. Softw. Tools (2000).
[30] R. Fattal, D. Lischinski, M. Werman, Gradient domain high dynamic range compression, in: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002, pp. 249–256.
[31] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[32] G. Valenzise, F. De Simone, P. Lauga, F. Dufaux, Performance evaluation of objective quality metrics for hdr image compression, Applications of Digital Image Processing XXXVII, volume 9217, 2014.