https://doi.org/10.2352/lim.2024.5.1.19

# SDR Image Reconstruction for the Improvement of Nighttime Traffic Classification Using a New HDR Traffic Dataset

***Mark Benyamin, Ulrich Schwanecke, Mike Christmann, Rolf Hedtke; Hochschule RheinMain - RheinMain University of Applied Sciences; Wiesbaden, Germany***

## Abstract

*In order to improve traffic conditions and reduce carbon emissions in urban areas, smart mobility and smart cities are becoming increasingly important measures. To enable the widespread use of the cameras required for this, cost and size requirements necessitate the use of low-cost standard dynamic range (SDR) cameras. However, these cameras do not provide sufficient image quality for a reliable classification of road users, especially at night.*

*In this paper, we present a data-driven approach to optimise image quality and improve classification accuracy of a given vehicle classifier at night. Our approach uses a combination of image inpainting and high dynamic range (HDR) image reconstruction to reconstruct and optimise critical image areas. Therefore, we introduce a large HDR traffic dataset with time-synchronised SDR images. We also present an approach to automatically degrade the HDR traffic data to generate relevant and challenging training pairs. We show that our approach significantly improves the classification of road users at night without having to retrain the underlying vehicle classifier. Supplementary information as well as the dataset are published at `https://www.mt.hs-rm.de/nighttime-traffic-reconstruction/`.*

## Introduction

The future of mobility should be sustainable and safe, yet diverse and highly available. To meet all these requirements, an intelligent traffic management system is required. An important component of such is a reliable traffic classification of road users, e.g. for traffic flow, signal control or even autonomous driving. To ensure reliable classification over large areas a high density and number of cameras is required. For this reason, it is necessary to make use of small low-cost cameras.

Although camera-based systems have high classification accuracy, their performance can be affected by environmental factors such as weather and lighting conditions, especially low illumination at night leading to a significant gap between day and night classification accuracy. Bright light sources at night and resulting reflections from vehicle headlights lead to a significant loss of image information (clipping) in the relevant image area, which significantly reduces the classification accuracy. This gap is exacerbated by a distortion of the data caused by the fact that daylight images are predominantly used to train the classifiers, since these can be labelled much more accurately and reliably.

To address this issue, we propose an image-processing approach to close the accuracy gap between daytime and nighttime classification with visible-light cameras without the necessity to retrain a given classifier. Since reducing clipped image areas can significantly improve classification at night, approaches from image inpainting and HDR image reconstruction are used to reconstruct and optimise the traffic image data.

We present a large ground truth (GT) HDR traffic dataset together with a time-synchronous SDR dataset acquired in parallel, and a tool to automatically generate GT and degraded (simulated) SDR patch pairs for training purposes. We adapted a deep learning model from image inpainting to the underlying use case of SDR traffic image reconstruction by using approaches from HDR image reconstruction and achieved a significant improvement in nighttime traffic classification without the need to retrain a given classifier.

## Related Work

### *Image Inpainting*

Great strides have been made in image inpainting, the addition of missing or deteriorated image content, thanks to deep neural networks [1–7]. Liu et al. [1] propose a partial convolution layer with an automatic mask update that adaptively masks convolutional operations based on the validity of surrounding pixels. They use a pixel-wise L1 loss function for both masked and non-masked pixels in combination with a perceptual loss using a pre-trained VGG-16 [8] network to capture high level feature representations.

Other approaches are GAN-based [2, 3, 7], transformer-based [4, 5] or diffusion-based [6]. Suvorov et al. [3] propose a network for large mask (LaMa) inpainting based on a ResNet-like architecture with fast fourier convolutions (FFCs). FFCs include fast Fourier transform (FFT) and inverse FFT allowing for a fast growth of the receptive field.

### *HDR Single-Image Reconstruction*

Various data-driven approaches [9–13] to HDR single-image reconstruction have been published in recent years. Eilertsen et al. [9] propose such a method using a hybrid dynamic range autoencoder. A virtual camera is set up to generate a training dataset of randomly extracted HDR and corresponding SDR patch pairs. Since the process is designed to reconstruct highlight areas only, the virtual camera only generates images that are clipped in the highlights and not in the shadows. To overcome the difficulty of HDR data availability, they pre-trained the network on a large set of simulated HDR images.

Santos et al. [10] introduce a feature masking mechanism inspired by [1]. They use soft RGB clipping masks to represent and suppress clipped image pixels, and adapt the loss function of [1] to the HDR reconstruction task. Similar to [9], Santos et al. propose an inpainting pre-training similar to [1]. The training data is generated by using a patch sampling strategy to select challenging training patches. Guo and Jiang [13] demonstrate the importance of the degradation model by considering the source of noise and compression as well as potential color space discrepancies. Other authors proposed GAN-based reconstruction approaches [11, 12].

## SDR Image Reconstruction

In order to close the accuracy gap between day and night traffic classification with visible light cameras due to weak illumination at night, we propose an SDR image reconstruction approach prior to classification. We adapt and combine an inpainting and an HDR image reconstruction approach to reconstruct light cones within clipped image areas, thus optimising the traffic data for classification.

Since most classifiers are trained on high relative contrast SDR input data and are therefore not familiar with linear or logarithmically encoded HDR data, extending them to the HDR domain would be counterproductive. Furthermore, only the specific reconstruction process is of primary importance for the reconstruction of clipped SDR image information. Thus, we are using SDR data only in combination with the (non-HDR) loss function of [1]. With this approach, any pre-trained traffic classifier can be used without the need for retraining.

### Model Overviews

We utilize the feature masking model of [10] (MaskCNN), as this is one of the best performing HDR reconstruction methods evaluated by Hanji et al. [14]. However, instead of forcing the decoder to operate in HDR domain, we stay in SDR domain. Since GANs are more challenging to train, we choose more classical convolutional neural networks (CNNs). Thus, we only use the generator of [3] (LaMa) as a further model, discarding the discriminator of the adversarial learning approach and use the loss function of [1] for both networks. In addition, we adapt the image inpainting model of [3] to the SDR image reconstruction task, by using image-dependent and soft clipping masks similar to [10] instead of randomly generated binary masks. We utilize the network architectures proposed by [10] and [3] and train them for SDR traffic reconstruction by using corresponding data. Therefore, we call the networks *TrafficMask* in case of [10] and *TrafficLaMa* in case of the generator of [3].

### Soft Clipping Mask

According to [10] we use image-dependent soft clipping masks with values in $[0,1]$. We input the masks together with the image data according to the input and concatenation mechanisms of the networks. The same applies to the inference process as illustrated in Fig. 1 for TrafficLaMa.

### VGG-based Perceptual Loss Function

We utilize the VGG-based perceptual loss from [1] consisting of four types of sub-loss functions: per-pixel loss, perceptual loss, style-loss and total variation (TV) loss. The high-level features for perceptual and style loss are extracted by a pre-trained VGG-16 [8] network. All types of sub-loss functions are applied separately to both valid and invalid (clipped) image regions.

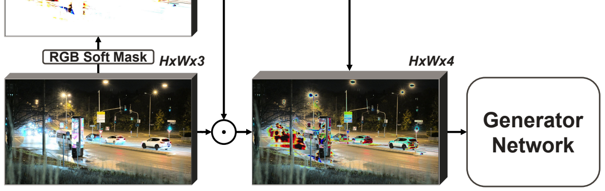

Figure 1: Clipping mask input of TrafficLaMa

## Traffic Dataset

### Image Acquisition

We created a new large HDR traffic dataset mainly consisting of nighttime images taken under difficult lighting conditions, but also including twilight and daytime lighting scenarios. For the acquisition of the HDR data a Sony Venice camera is used. The overall HDR dataset contains almost 3 hours of material resulting in a total of 265,000 frames including various camera angles and settings in different locations.

In parallel, we capture a widely time-synchronous SDR traffic dataset using a Pacidal NMHC2327D, which is a low-cost SDR circuit board network camera actually used for traffic recognition applications, directly set up next to the Sony Venice. In total, more than 15 minutes of different time-synchronised nighttime images were captured, resulting in more than 22,500 HDR and SDR image pairs. This dataset can also be of particular interest for tasks dealing with position shift problems, such as pixel-accurate matching, image alignment, etc. Both datasets will be published. An example image is shown in Fig. 2.

### HDR and SDR Ground Truth Data

Since we follow an SDR reconstruction approach before classification, the target image to be generated should be a reconstructed SDR image with high relative contrast and reduced clipping. Therefore, it is necessary to generate high-quality SDR data from the HDR without causing a loss of image information, apart from the loss of absolute brightness and contrast. For this, we make use of the dynamic local tone mapping method described in [15–17], that performs HDR-to-SDR conversion based on a histogram guided correction mask, in combination with an accurate color volume mapping described in [18].

The resulting SDR data is used as GT in the training process. Both, the resulting HDR (PQ-ST2084 Rec. 2020) and SDR (Gamma BT.709 Rec.709) TIFF datasets consisting of 8032 full HD frames each will be published. Figure 2 shows a corresponding example image.

### SDR Image Degradation and Augmentation

For traffic reconstruction, we generate two different training datasets: The first one is generated semi-automatically by hard clipping the 8032 images of the GT dataset for highlight degradation, and then applying a logistic function with different growth rates for highlight and shadow degradation. The function is applied to image luminance (Y) and the resulting percentage change is determined and applied to the individual RGB channels to achieve a natural clipping behavior of the colors. The real SDR images acquired with the Pacidal NMHC2327D are used as a reference for degradation. An example is shown in Fig. 2. After degradation, a total of 33,028 patch pairs are automatically extracted from the 8032 full HD pairs using a simple static extraction mechanism.

In addition, a large amount of the acquired traffic data contains smooth, unstructured image regions that are of no relevance. Therefore, we develop a method for automated degradation and augmentation to generate relevant and challenging training images. This method is based on the virtual camera presented by [9], but we implemented two additional degradation methods: One based on applying a logistic function for highlight and shadow clipping, the other based on conventional gamma manipulation for shadow clipping only. Furthermore, we select vehicle patches via object recognition using YOLOv8m [19, 20]. Once all vehicles in the image have been extracted, additional areas

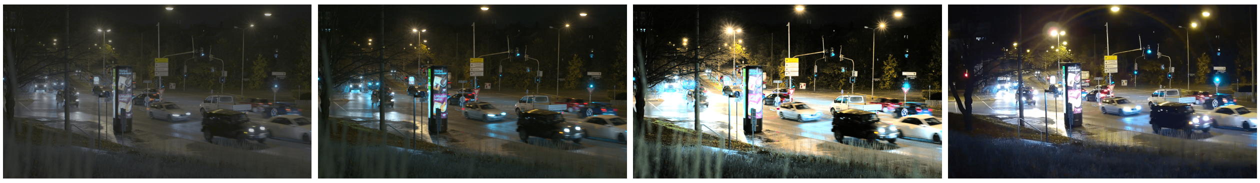

Figure 2: Captured traffic dataset. From left to right: a GT HDR example imagery, a tone-mapped SDR GT, a semi-automatically degraded (simulated) SDR and the real SDR captured by a real traffic camera (Pacidal NMHC2327D)

with a high level of structure and detail can be extracted based on the length of image gradients similar to the patch sampling approach of [10]. In addition, two global patches are extracted that cover the entire right and left portions of the image.

Clipping is performed randomly in terms of clipping method and intensity. No other degradations are applied. Common augmentations are performed. According to this procedure, we automatically generate a very diverse second dataset consisting of 127,677 patch pairs. Patches extracted in varying sizes are always scaled to the target patch size using bilinear interpolation.

## Experiments

### Datasets

For pre-training we use the MIT Places [21] dataset and create a subset of 500,000 training patches. This dataset is corrupted by multiplication with random binary masks generated on-the-fly during training process.

We also use a film and television (FTV) dataset collected on our own as well as our new traffic dataset for fine-tuning purposes. The FTV dataset, which originally consisted of mainly HDR material, is generated according to the approach of dynamic tone and color volume mapping [15–18] in combination with the semi-automatic degradation approach described above, resulting in 123,532 patch pairs.

These datasets consist of 512x512 pixel patches and are equally used for training both networks, TrafficMask and TrafficLaMa. For evaluation under real world conditions, we use a real SDR traffic dataset generated by a traffic detection system including 300 daytime and 300 nighttime images of various types provided with accurate GT labels.

### Training Details

Since HDR traffic data is only available to a very limited extent, we use a multi-stage training strategy according to [9] and [10] by performing pre-training on the image inpainting task using the adjusted MIT Places dataset. We then perform a first fine-tuning stage to adapt the learned representation to the SDR image reconstruction task using the FTV dataset. In the second fine-tuning stage, we use our new traffic dataset. Both fine-tuning sets are used in combination with the corresponding soft clipping masks generated on-the-fly during training. The loss function according to [1] is used for all training sessions without exception. The durations of each training stage can be found in Table 1. The networks are initialized according to [10] and [3] respectively. Training is done on an NVIDIA Quadro P5000 GPU with 16GB VRAM using a batch size of 10 for TrafficMask and a batch size of 6 for TrafficLaMa in each training stage. Apart from the differences in the data, the three training stages and their selected training parameters do not differ from each other. In each training stage for both networks, the optimisation is performed using the Adam optimiser with the default parameters, learning rate is chosen to be $2 \times 10^{-4}$ and gradient clipping is applied.

| Model | Pre-Training | Fine-Tuning 1 | Fine-Tuning 2 | Overall |
|---|---|---|---|---|
| TrafficMask | 2:15:56:20 | 0:10:05:53 | 0:13:17:19 | 3:15:19:32 |
| TrafficLaMa-1 | 1:20:37:39 | 0:13:20:17 | 0:17:39:33 | 3:03:37:29 |
| TrafficLaMa-2 | 1:20:37:39 | 0:13:20:17 | 2:13:54:08 | 4:23:52:04 |

Table 1: Duration of the individual training stages in [d:h:m:s]

### Classification Model for Evaluation

For evaluation, we classify the real SDR traffic dataset with and without applying the trained reconstruction networks to the test dataset. For this purpose, we use a real classification model that is in productive use for actual traffic classification tasks. This model, based on ResNet [22] and EfficientNet [23], was trained to categorise the nine classes *”car or van”*, *”motorcycle”*, *”truck”*, *”bus”*, *”tractor”*, *”other motor vehicle”*, *”non-determinable motor vehicle”*, *”no motor vehicle”* and *”unknown”* using a dataset containing approximately 80% daytime and 20% nighttime images.

### Evaluation Metrics

In addition to a simple ground truth score (GTS), averaging the predicted probabilities of the respective GT classes over all test images, we apply the earth mover's distance (EMD) [24] and the Kullback-Leibler divergence (KLD) [25] to the resulting class probability vectors (distributions) by averaging the resulting EMD and KLD values over all test images as well.

### Qualitative Results

With our reconstruction approach, we achieve a significant reduction of light cones in clipped nighttime traffic images, as shown in Fig. 3. Both, TrafficLaMa and TrafficMask achieve almost the same significant reduction. However, TrafficMask based on [10] tends to introduce slight artefacts in headlight cones, resulting in inaccurate reconstructions. Furthermore, the overall brightness of both results is visibly reduced compared to the untreated input, although TrafficMask leads to a stronger reduction.

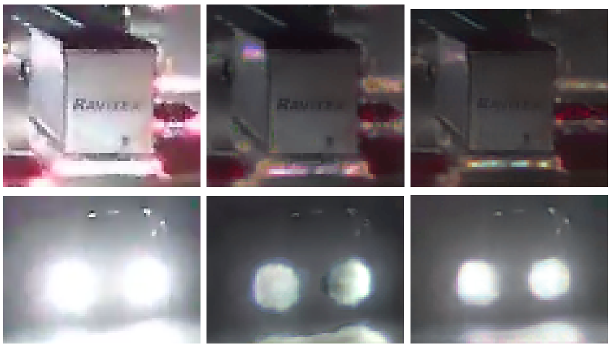

Figure 3: Qualitative results of nighttime SDR traffic reconstruction. From left to right: untreated inputs captured by Pacidal traffic camera, results of TrafficMask, results of TrafficLaMa

|  | Nighttime Classification Results |  |  | Daytime Classification Results |  |  |
|---|---|---|---|---|---|---|
| **Model** | GTS ↑ | EMD ↓ | KLD ↓ | GTS ↑ | EMD ↓ | KLD ↓ |
| None (untreated) | 0.8659 | 0.01212 | 1.34451 | 0.9274 | 0.00765 | 0.60011 |
| TrafficMask | 0.8867 ▲2.4% | 0.01111 ▲8.3% | 1.13047 ▲15.9% | 0.9196 ▼0.8% | 0.00928 ▼21.3% | 0.64368▼7.3% |
| TrafficLaMa-1 | 0.8926 ▲3.1% | 0.01042 ▲14.0% | 1.06115 ▲21.1% | 0.8458 ▼8.8% | 0.01325 ▼73.2% | 1.29657▼116.1% |
| TrafficLaMa-2 | 0.8938 ▲3.2% | 0.00895 ▲26.2% | 1.05438 ▲21.6% | 0.8872 ▼4.3% | 0.00957 ▼25.1% | 0.97042▼61.7% |

Table 2: Evaluation of the quantitative results of nighttime and daytime classification before and after SDR image reconstruction

### *Quantitative Results*

For quantitative evaluation, we input each of the 300 day and 300 night test images in three different versions into the classification model: An untreated one, a version reconstructed by TrafficMask and one reconstructed by TrafficLaMa-1, resulting in three class probability vectors per test image. Both reconstruction networks are fine-tuned using the first semi-automatically generated traffic dataset. The resulting class probability vectors and the GT class (distribution) are compared using the GTS, EMD and KLD metrics. The results are shown in Table 2. As can be seen, all three metrics clearly show a significant improvement in nighttime classification accuracy achieved by each reconstruction model. Furthermore, all metrics clearly show that TrafficLaMa outperforms TrafficMask. The same can be observed with the average discrepancy between day- and nighttime classification accuracy, as shown in Table 3.

|  | Improvement of classification discrepancy |  |  |
|---|---|---|---|
|  | GTS ↑ | EMD ↓ | KLD ↓ |
| Discrepancy | 0.0615 | 0.00447 | 0.7444 |
| TrafficMask | 0.0407▲33.8% | 0.00346▲22.6% | 0.53036▲28.8% |
| TrafficLaMa-1 | 0.0348▲43.4% | 0.00277▲38.0% | 0.46104▲38.1% |
| TrafficLaMa-2 | 0.0336▲45.4% | 0.00130▲70.9% | 0.45427▲39.0% |

Table 3: Improvement of the discrepancy between nighttime and daytime traffic classification

The aim is to reduce this discrepancy using the proposed reconstruction approaches. This goal is clearly achieved by both network approaches. Since the training was designed to improve the nighttime and not the daytime classification accuracy, applying the networks to the daytime images leads to an average degradation of the classification accuracy as shown in Table 2. For the underlying average test image resolution of 100x90 pixels, the real-time requirements are achieved with 11 ms inference speed for TrafficMask and 32 ms for TrafficLaMa using the NVIDIA Quadro P5000 GPU.

### *Ablation Study*

Since TrafficLaMa outperforms TrafficMask, we conduct the ablation study exclusively on TrafficLaMa. However, in addition to the untreated test dataset, we use a version reconstructed by TrafficLaMa and fine-tuned to the semi-automatically generated traffic dataset (TrafficLaMa-1), and a version fine-tuned to the traffic dataset generated by our dataset tool (TrafficLaMa-2). Each of the three dataset versions are again forwarded into the classifier. The averaged results for the 300 night- and daytime images using the metrics are also shown in Table 2 and 3.

Each metric shows that TrafficLaMa-2 outperforms TrafficLaMa-1, thus leading to a further improvement in classification accuracy. Especially through the EMD metric, this trend becomes particularly clear. This, and the fact that we are able to generate our data fully automatically in a relatively short time, is a great advantage considering that the data previously had to be generated in a very time-consuming and laborious process.

## Limitations and Future Work

Although we demonstrate that our approach to SDR image reconstruction achieves a significant improvement overall, this is not achieved in every classification case. In some cases, reconstruction even leads to an accuracy degradation. This can be caused by artefacts in headlight cones, color or by the reduction of the overall brightness or contrast. Fig. 4 shows a failure case, which is classified as a car after reconstruction, while the untreated camera output is correctly classified as a truck.

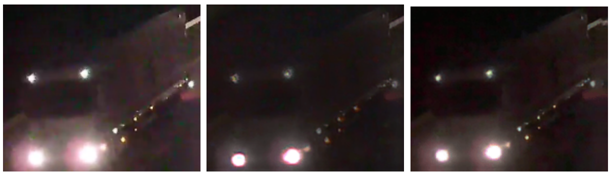

Figure 4: Failure cases of SDR traffic reconstruction. From left to right: untreated input captured by Pacidal traffic camera, result of TrafficMask, result of TrafficLaMa

Various measures could be adopted to improve nighttime reconstruction and classification, such as including classification in an end-to-end learning process, using alternative network approaches such as GAN or transformer architectures, applying video reconstruction and other classification approaches or further using the time-synchronous SDR dataset, which is only used as a reference for SDR degradation. Instead, this dataset could be incorporated directly into the training process using applications such as pixel-accurate matching in order to generate a more realistic training dataset.

Moreover, daytime classification is also currently reduced by the reconstruction process. Since the boundary between day and night is not discrete but rather fuzzy due to dusk and dawn, it is necessary to determine the point at which an improvement in classification occurs to make optimum use of the reconstruction.

## Conclusion

In this paper, we addressed the challenge of improving the accuracy of traffic classification at night with low-cost SDR cameras that struggle to capture high-quality images under low-light conditions. Our data-driven approach combines image inpainting and HDR reconstruction techniques, which we adapt to SDR traffic image reconstruction. We presented a large HDR traffic dataset together with time-synchronised SDR data, which enables effective training of the adapted reconstruction models. Our method involves the automatic degradation of traffic image data to generate sophisticated training image pairs.

The experimental results show a significant reduction of overexposure and a better classification of nighttime traffic without retraining the classifier. While our approach is promising, it also shows its limitations, such as occasionally degraded classification accuracy after reconstruction. Future work may focus on exploring end-to-end learning and alternative network architectures. The use of time-synchronised datasets and methods for reconstructing video images could also help to improve the reconstruction and classification of night-time traffic.

## Acknowledgments

The authors would like to thank their partners at VITRONIC Dr.-Ing. Stein Bildverarbeitungssysteme GmbH and their colleagues at Hochschule RheinMain. This project (HA project no. 1077/21-79) was financed with funds of LOEWE - Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence).

## References

[1] Guilin Liu et al. “Image Inpainting for Irregular Holes Using Partial Convolutions”. In: *European Conference on Computer Vision (ECCV)*. Springer International Publishing, Sept. 2018, pp. 89–105.

[2] Jiahui Yu et al. “Generative Image Inpainting with Contextual Attention”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018, pp. 5505–5514.

[3] Roman Suvorov et al. “Resolution-Robust Large Mask Inpainting With Fourier Convolutions”. In: *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2022, pp. 2149–2159.

[4] Wenbo Li et al. “MAT: Mask-Aware Transformer for Large Hole Image Inpainting”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 10758–10768.

[5] Qiaole Dong, Chenjie Cao, and Yanwei Fu. “Incremental Transformer Structure Enhanced Image Inpainting with Masking Positional Encoding”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 11348–11358.

[6] Chitwan Saharia et al. “Palette: Image-to-Image Diffusion Models”. In: *ACM SIGGRAPH 2022 Conference Proceedings*. SIGGRAPH ’22 15. Vancouver, BC, Canada: Association for Computing Machinery, Aug. 2022, pp. 1–10.

[7] Y. Poirier-Ginter and J. Lalonde. “Robust Unsupervised StyleGAN Image Restoration”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2023, pp. 22292–22301.

[8] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *3rd International Conference on Learning Representations (ICLR 2015)*. May 2015, pp. 1–14.

[9] Gabriel Eilertsen et al. “HDR image reconstruction from a single exposure using deep CNNs”. In: *ACM Transactions on Graphics* 36.6 (Nov. 2017), pp. 1–15.

[10] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. “Single image HDR reconstruction using a CNN with masked features and perceptual loss”. In: *ACM Transactions on Graphics* 39.80 (4 Aug. 2020), pp. 1–10.

[11] Demetris Marnerides, Thomas Bashford-Rogers, and Kurt Debattista. “Deep HDR Hallucination for Inverse Tone Mapping”. In: *Sensors* 21(12).4032 (2021), pp. 1–14.

[12] Chao Wang et al. “GlowGAN: Unsupervised Learning of HDR Images from LDR Images in the Wild”. In: *IEEE/CVF International Conference on Computer Vision (ICCV)* (Oct. 2023), pp. 10475–10485.

[13] Cheng Guo and Xiuhua Jiang. “LHDR: HDR Reconstruction for Legacy Content Using a Lightweight DNN”. In: *Computer Vision – ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part III*. Macao, China: Springer-Verlag, 2023, pp. 306–322.

[14] Param Hanji et al. “Comparison of single image HDR reconstruction methods — the caveats of quality assessment”. In: *ACM SIGGRAPH 2022 Conference Proceedings*. SIGGRAPH ’22 1. Vancouver, BC, Canada: Association for Computing Machinery, Aug. 2022, pp. 1–8.

[15] Lucien Lenzen. “HDR for legacy displays using sectional tone mapping”. In: *International Broadcasting Convention (IBC)*. Jan. 2016, 16 (10.)–16 (10.)

[16] L. Lenzen and M. Christmann. “Subjective viewer preference model for automatic HDR down conversion”. In: *IS&T Electronic Imaging* 29.12 (Jan. 2017), pp. 191–191.

[17] Lucien Lenzen, Rolf Hedtke, and Mike Christmann. “HDR in Consideration of the Abilities of the Human Visual System”. In: *SMPTE Motion Imaging Journal* 128.5 (May 2019), pp. 40–45.

[18] Pascal Kutschbach. “A Color-Volume Mapping System for Perception-Accurate Reproduction of HDR Imagery in SDR production workflows”. In: *SMPTE Motion Imaging Journal* 130.7 (Aug. 2021), pp. 12–21.

[19] J. Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2016, pp. 779–788.

[20] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. *Ultralytics YOLO*. Version 8.0.0. Jan. 2023. URL: `https://github.com/ultralytics/ultralytics`.

[21] Bolei Zhou et al. “Places: A 10 Million Image Database for Scene Recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.6 (July 2018), pp. 1452–1464.

[22] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2016, pp. 770–778.

[23] Mingxing Tan and Quoc Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 6105–6114.

[24] Y. Rubner, C. Tomasi, and L.J. Guibas. “A metric for distributions with applications to image databases”. In: *IEEE International Conference on Computer Vision (ICCV), (Cat. No.98CH36271)*. Jan. 1998, pp. 59–66.

[25] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86.