DB-GAN: Boosting Object Recognition Under Strong Lighting Conditions Supplementary material

Luca Minciullo *
Toyota Motor Europe
luca.minciullo@toyota-europe.com

Fabian Manhardt*
Technical University of Munich
fabian.manhardt@tum.de

Kei Yoshikawa
Toyota Motor Europe
kei.yoshikawa@toyota-europe.com

Sven Meier
Toyota Motor Europe
sven.meier@toyota-europe.com

Federico Tombari†
Technical University of Munich
tombari@in.tum.de

Norimasa Kobori
Woven CORE, Inc.
norimasa.kobori@tri-ad.global

Abstract

This document supplements our main paper entitled DB-GAN: Boosting Object Recognition Under Strong Lighting Conditions *by 1.) providing further details on the dataset that will be released together with our work and show additional qualitative results for detection leveraging TrueBlue images. In addition, we 3.) provide more insights on the GAN's attitude with respect to color variation. In the last section, 4.) we demonstrate the behaviour for the SSD baseline and our proposed method with and without pre-processing.*

1. TrueBlue dataset

TrueBlue consists of 11 image sequences of 3 different scenes, displaying daily household objects. Moreover, we added distractor objects and also a MacBeth Color Checker chart. The employed 3D models will also be made available.

Each scene was illuminated from above by a set of three light sources of different types, *e.g.* LED, incandescent, compact fluorescent, daylight and mixture of different sources. For each scene, we acquired 11 images using a Nikon D750 with Nikon 24-70mm f/2.8 lens at f/8 and ISO 100 with different white balance settings: 2500K, 2700K, 2940K, 3230K, 3570K, 4000K, 4550K, 5260K, 6250K, 7690K, 10000K respectively. The images are processed in-camera to produce a JPEG image in Adobe RGB color space. The images were subsequently re-sampled to 1/4 original resolution using Sinc (lanczos) re-sampling in the GIMP 2.8.16.

Table 1. **Illuminant types used in TrueBlue. LED: Light emitting diode, CF: Compact fluorescent, INC: incandescent**

Illuminant ID	Type	Description	color temperature
1	LED	Philips LED bulb	3000K
2	CF	Carrefour CF bulb	warm white
3	CF	Walimex photo bulb	day light
4	INC	Osram 60W incandescent bulb	n/a
5	sky	100% cloud cover mid-day sky	n/a

Table 1 lists the used illuminant types for acquiring of TrueBlue. Notice that compact fluorescent lights were switched on for 10 minutes prior to acquisition to let them reach steady state temperature.

The light sources included in this dataset were chosen specifically to include near black body emitters (incandescent bulb), approximate broad spectrum (light emitting diode), and narrow spectrum (compact fluorescent), as well as natural light (cloudy sky).

TrueBlue has 3 different setups, *i.e.* arrangements of objects. Each setup always contains 4 objects with 3D model and annotated 2D bounding box, and 2 additional distractor objects. They are laid out in plain manner without inducing occlusions, in an effort to focus on the white balance sensitivity.

1.1. Qualitative Results

In Fig 1, 2 and 3 we illustrate additional qualitative results for both the SSD baseline and our proposed approach, leveraging the GAN as pre-processor. It can be observed

*These authors contributed equally to this work
†Federico Tombari is now working at Google

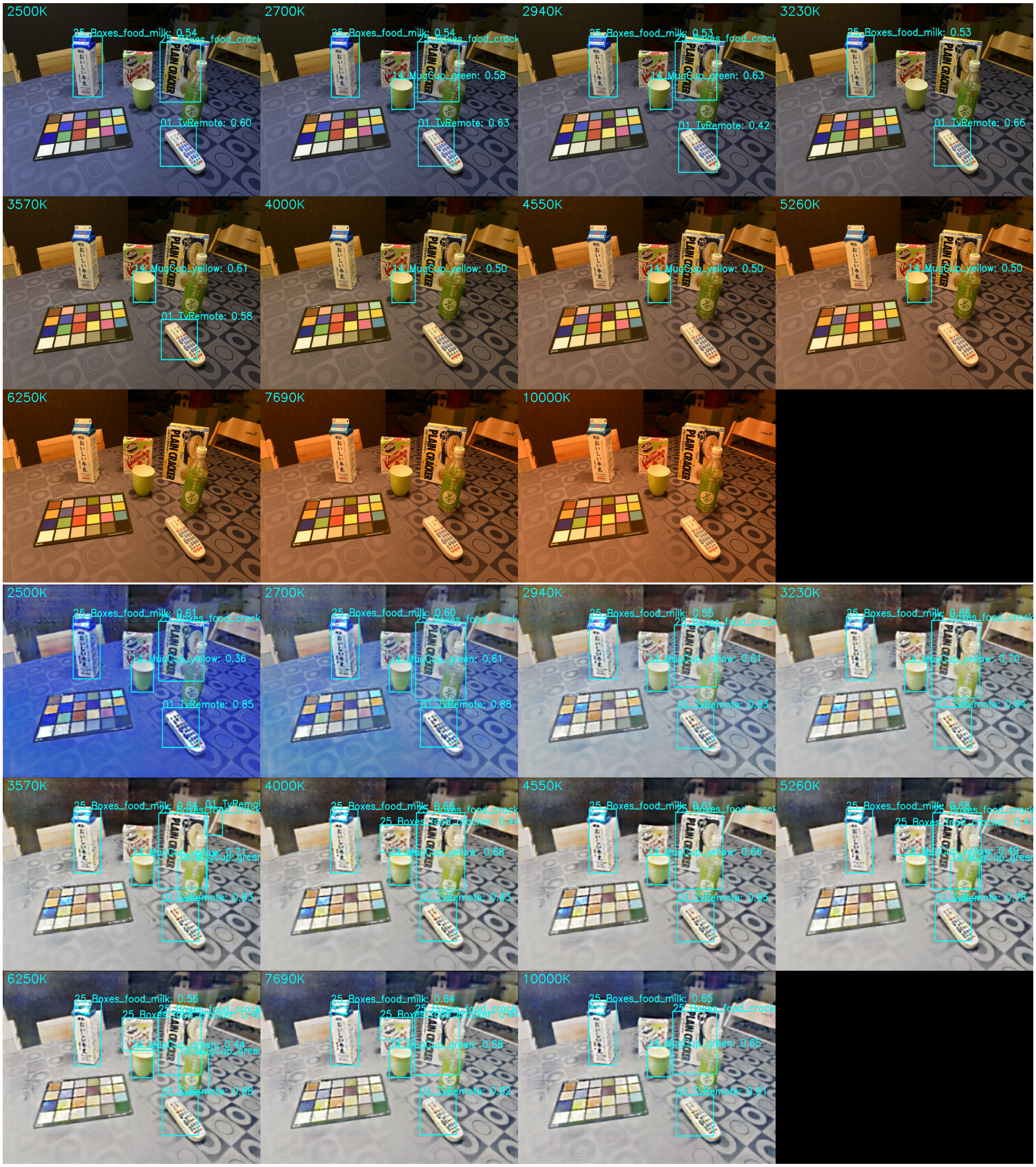

Figure 1. Detections of the baseline SSD(top) and our approach with GAN pre-processing(bottom) on the 1st scene of the TrueBlue dataset. The scene was recorded outdoor.

that when the input image is *too blue* or *too red* many objects are not indeed detected. The GAN pre-processing seems to map input images more to the '*blue*' domain. In practice, we observe that DB-GAN paints blue texture on

Figure 2. Detections of the baseline SSD(top) and our approach with GAN pre-processing(bottom) on the 2nd scene of the TrueBlue dataset. The scene was recorded outdoor.

dark regions of the images and transform red regions into light blue.

Figure 3. Detections of the baseline SSD(top) and our approach with GAN pre-processing(bottom) on the 11th scene of the TrueBlue dataset. The scene was recorded outdoor.

2. TP/FP discussion

Fig 4 reports the average number of true positives and false positives per object for the Toyota Light dataset. We demonstrate our results for the SSD baseline, as well as our enhanced SSD with and without GAN pre-processing GAN

Table 2. **Scene descriptions of TrueBlue**

Scene ID	Type	Illuminant IDs
001	uniform	1
002	uniform	1
003	uniform	2
004	uniform	3
005	uniform	4
006	mixed	2, 4, 2
007	mixed	3, 4, 3
008	mixed	3, 2, 3
009	mixed	2, 3, 2
010	mixed	1, 3, 1
011	uniform	5

Table 3. **DB-GAN detector loss weight tuning study on the BOP Toyota Light dataset.**

λ_5	mAP ↑
4 0.01	**0.715**
0.05	0.662
0.1	0.632
0.5	0.61
1	0.642

pre-processing at test time. Notice that, while the number of true positive dramatically increases, the number of false positives also rises when employing our method, regardless pre-processing the input.

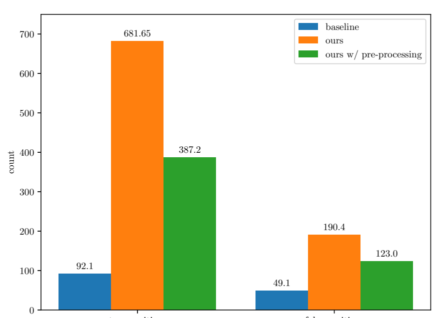

Figure 4. Left: Average number of true positives per object using the baseline SSD, our approach with and without gan pre-processing. Right: Same analysis for false positives. Results from the Toyota Light dataset.

3. SSD loss weight tuning

In the main paper, we report all our results with respect to the empirically found optimal value for the detector loss weight λ_5 is 0.01. We ran a grid-search over the parameter space to find the best working value. The results from the grid search are constituted in Table 3.

Figure 5. **Examples of training images for the GAN.** The input is shown on the left and the output is visualised on the right. We render objects with simulated light onto PHOS background images. Notice that the object is rendered with random ambient and directional light on the input and with static ambient and no directional light on the output.

Table 4. **DB-GAN results using two different background datasets to generate the training images.** The illumination normalization task cannot be replaced by standard auto-encoding, even if the background images have large texture variation.

GAN train background	Auto-Encoder	Toyota Light mAP ↑	TUD Light mAP ↑
PHOS	×	**0.71**	**0.66**
PHOS	✓	0.33	0.12
VOC	✓	0.34	0.50

4. Experiments on using various background datasets

First of all we want to show some examples of the images we used to train our GAN architecture. See Figure

To verify that that the illumination normalisation task contributes to the boost in performance, we tried rendering the object models on Pascal VOC [1] images and trained DB-GAN on the resulting data. Furthermore, we also tried applying the trained GAN as pre-processor to our boosted SSD at inference time. Table 4 shows the results of using only 'correct exposure' images from PHOS or VOC background images to train the GAN, both in input and output. Training DB-GAN as an auto-encoder decreases performance on both datasets. This highlights one important aspect of our method: the normalisation task contributes to the performance increase and cannot be achieved by a higher background diversity.

5. DB-GAN Pre-processing Results at test time

As mentioned in the main paper DB-GAN pre-processing at test time suffers from a domain gap on the Toyota Light dataset. Table 5 shows the results on perform-

Table 5. **DB-GAN results with and without test time pre-processing.** On the TUD Light and TrueBlue datasets DB-GAN preprocessing at test time yields increased performance.

Pre-processing	TUD Light mAP ↑	TrueBlue mAP ↑
✓	**0.69**	**0.82**
×	0.66	0.73

ing GAN pre-processing on test images for the 2D object detection task. We can see that on the TUD Light and True-Blue datasets pre-processing images at test time as well as training time using our DB-GAN yield even better results, showing the effectiveness of our approach.

6. Qualitative Comparison

Here we qualitatively compare the 5 different methods reported in the paper on 2D object detection. We take the Toyota Light dataset for this. Figure 6 shows the detected objects (we use a detection threshold of 0.1 for all SSD instances). We notice that, unlike existing approaches, our method can recover a tight bounding box even with severe low light(1st row), where other image enhancements methods fail. Furthermore, our approach tend to have a much lower number of false positives and can deal with textured (3rd row) as well as texture-less object (1-2 rows).

References

[1] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88**(2), 303–338 (Jun 2010)

[2] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. arXiv preprint arXiv:1906.06972 (2019)

[3] Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6849–6857 (2019)

[4] Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018)

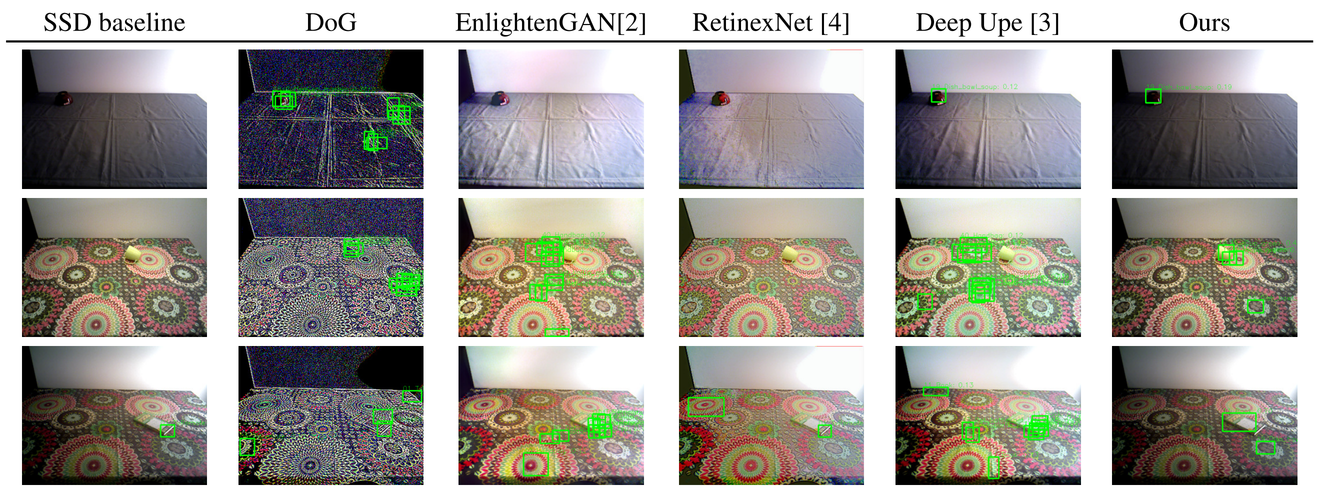

Figure 6. **2D qualitation comparison between the 5 different version of SSD, trained of different pre-processing approaches.** We can see that our approach is able to fit tight bounding boxes on objects that are completely missed by the other approaches.