Supplementary
SIMBAR: Single Image-Based Scene Relighting For Effective Data Augmentation For Automated Driving Vision Tasks

A. Additional Relit Results Comparison

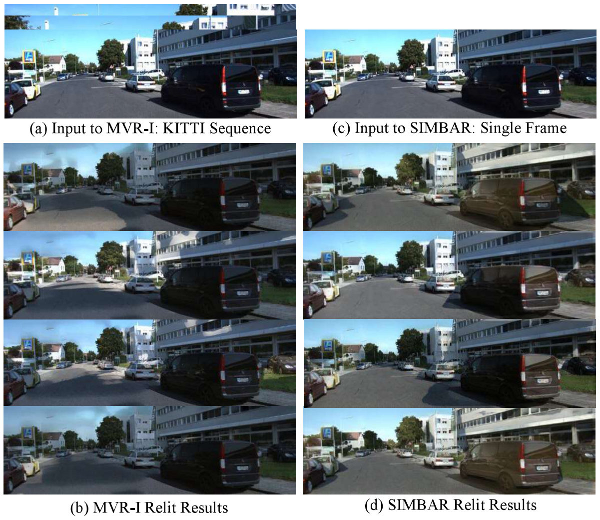

Figure 11. Relit results from MVR-I and SIMBAR on the KITTI traffic intersection sequence 0001.

Figure 12. Relit results from MVR-I and SIMBAR on the KITTI road driving sequence 0002.

Qualitative Comparison: As shown in Fig. 11, MVR-I takes the KITTI sequence 0001 as input, and relights the scene based on the mesh reconstruction from multiview input. This frame is captured around a corner, where the MVR-I method created phantom shadows in the scene. Our SIMBAR method generated physically consistent shadows for the car objects. In Fig. 12, the input data is KITTI sequence 0002 captured at a road intersection, where the ego vehicle is static. MVR-I failed to relight this scene with many hallucinated shadows rendered, due to the insufficient multiview information within the captured sequences. In contrast, taking a single frame, our method SIMBAR provides significantly more realistic and physically consistent relighting results, shown on the right with shadows rendered around the car objects. As highlighted in yellow, SIMBAR relit results also have clear sky.

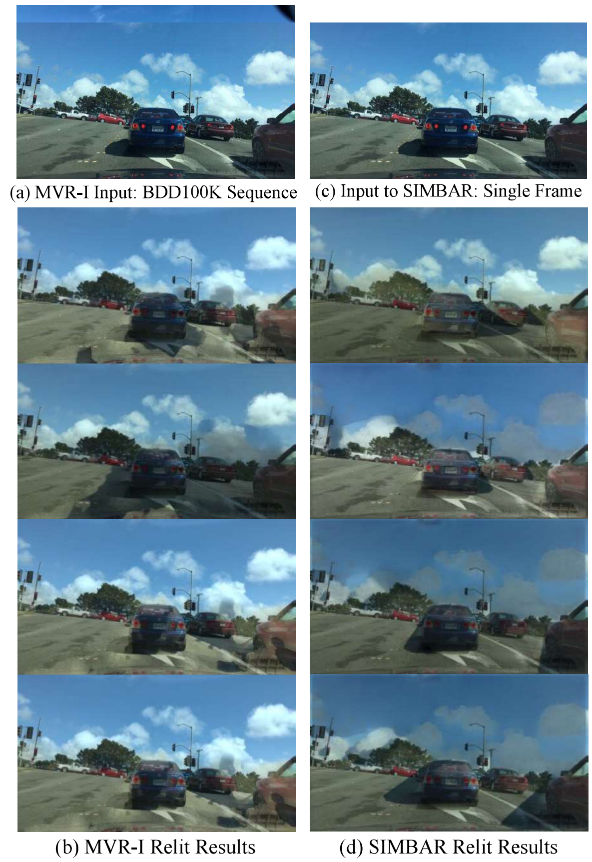

Figure 13. MVR-I and SIMBAR relights the BDD100K traffic intersection.

In addition, SIMBAR can generalize across different datasets with no need to retrain for every new dataset, as indicated in Fig. 13, where MVR-I and SIMBAR relight the same BDD100K sequence, with several cars including the ego vehicle waiting for red lights. In contrast, MVR-I again generated phantom shadows in the scene.

Corner Case Comparison: In Fig. 14. The input data is BDD100K sequence with very small scale objects in the scene, which is a challenging case for relighting. Neither MVR-I nor SIMBAR could relight two cars in the scene

(a) MVR-I Input: BDD100K Sequence (c) SIMBAR Input: Single Frame

(b) MVR-I Relit Results (d) SIMBAR Relit Results

Figure 14. MVR-I and SIMBAR relight the BDD100K road scene with challenging small scale objects.

with new shadows. Particularly, MVR-I failed to relight this scene due to the lacking feature matching between frames during the SFM+MVS based mesh reconstruction process, resulting in noisy mesh, which creates hallucinated shadows in the scene (b). SIMBAR relies on depth backbone for geometry estimation, which has captured some background scene objects (billboard and buildings) and created novel shadows for these objects accordingly.

B. Additional Metrics Evaluation

	RMSE ↓	PSNR ↑	SSIM ↑	MS-SSIM↑	UQI↑
MVR-I	53.70	13.68	0.65	0.71	0.79
SIMBAR	**37.59**	**16.86**	**0.69**	**0.77**	**0.86**

Table 2. Scene relighting evaluation on vKITTI Scene 0001.

Pixel-Based Image Quality Metrics Evaluation Our paper focuses on the benefits the relighting pipeline presents for downstream tasks with the experiments to demonstrate practical applications of the data augmentation. As shown in Table. 2, we also include image quality metrics to highlight the quality of SIMBAR's augmentation and show direct pixel-based evaluation metrics.

The use of SIMBAR demonstrates improvements across the direct image metrics relative to MVR-I. This indicates that SIMBAR is able to not only outperform MVR-I for training downstream models on augmented data, but that the image quality is also improved.

These above metrics could be reference information for measuring MVR/SIMBAR generated data quality. However, such metrics have two limitations - (i) need for ground truth images for different lighting conditions; and (ii) ineffective measurement of perceptual similarity [14]. Section 4 introduces a novel evaluation scheme, that compares deep models trained on datasets augmented with their relit versions against the baseline models on a held-out test set - quantifying the data augmentation effectiveness of SIMBAR for a vision task. Such an evaluation is more reliable in demonstrating real-world applicability of any relighting pipeline.

C. Societal Impact

With the application towards single images, SIMBAR allows for widespread use. The geometry estimation module can be potentially disruptive to people's privacy and safety due to its ability to perform open-world modeling. SIMBAR alleviates this concern by making use of a 3D scene mesh representation that is devoid of any personally identifiable information (PII), as shown in the pipeline schematic in Fig. 2 of an open street scene. Other ways in which SIMBAR can have potential negative social impacts is through the diversification (via relighting) of images scraped off of the internet (as is the case with DIV2K) with PII information in them (such as people's age, relationship, license plates etc. [34]).

We hope to mitigate this concern by incorporating an obfuscation method. Applying an off-the-shelf obfuscation network to SIMBAR's relit results can be challenging. As indicated in Fig. 9, given the wide variety of lighting conditions SIMBAR can generate, an obfuscation network trained with limited lighting conditions might not work well on SIMBAR relit images. Thus, we suggest obfuscating sensitive parts of the image as follows. Step 1: perform object detection on the source images to extract the pixel coordinates with PII. Step 2: obfuscate (using Gaussian Blur [2]) sensitive regions from Step 1 in the relit versions of the source images. This is straightforward since SIMBAR alters scene lighting without altering semantic content.

D. SIMBAR Relights Real-World Datasets

In this section, we present more results on datasets including DIV2K (Fig. 15), BDD100K (Fig. 16), and KITTI Object 2D (Fig. 17). These are provided with individual frames, thus the MVR method cannot be applied to datasets prepared in this manner.

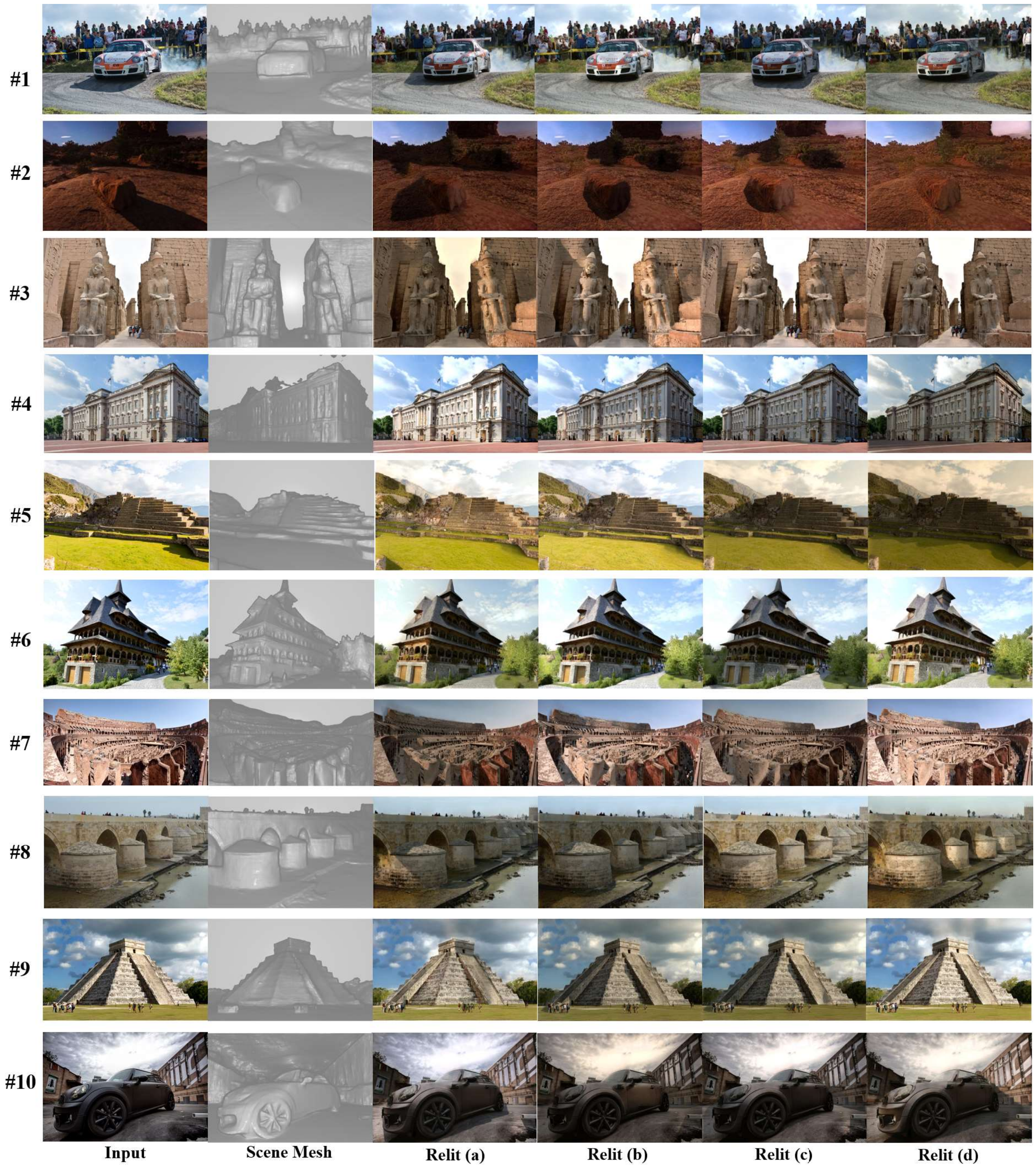

Figure 15. **SIMBAR In The Wild:** SIMBAR relights DIV2K Internet Images with wide variety of outdoor scenes. The 1st column shows the input frames, and 2nd column presents the scene mesh generated for geometry estimation. Novel relit results are presented in Relit (a)(b)(c)(d) with 4 different lighting variations. Note that while the generated mesh is not perfectly consistent with scene geometry, it allows for a general enough understanding of the 3D structure of the scene to generate informative priors to be fed as inputs to the subsequent image relighting module. Overall, SIMBAR In The Wild demonstrates our method's generalization capability without retraining for every new dataset.

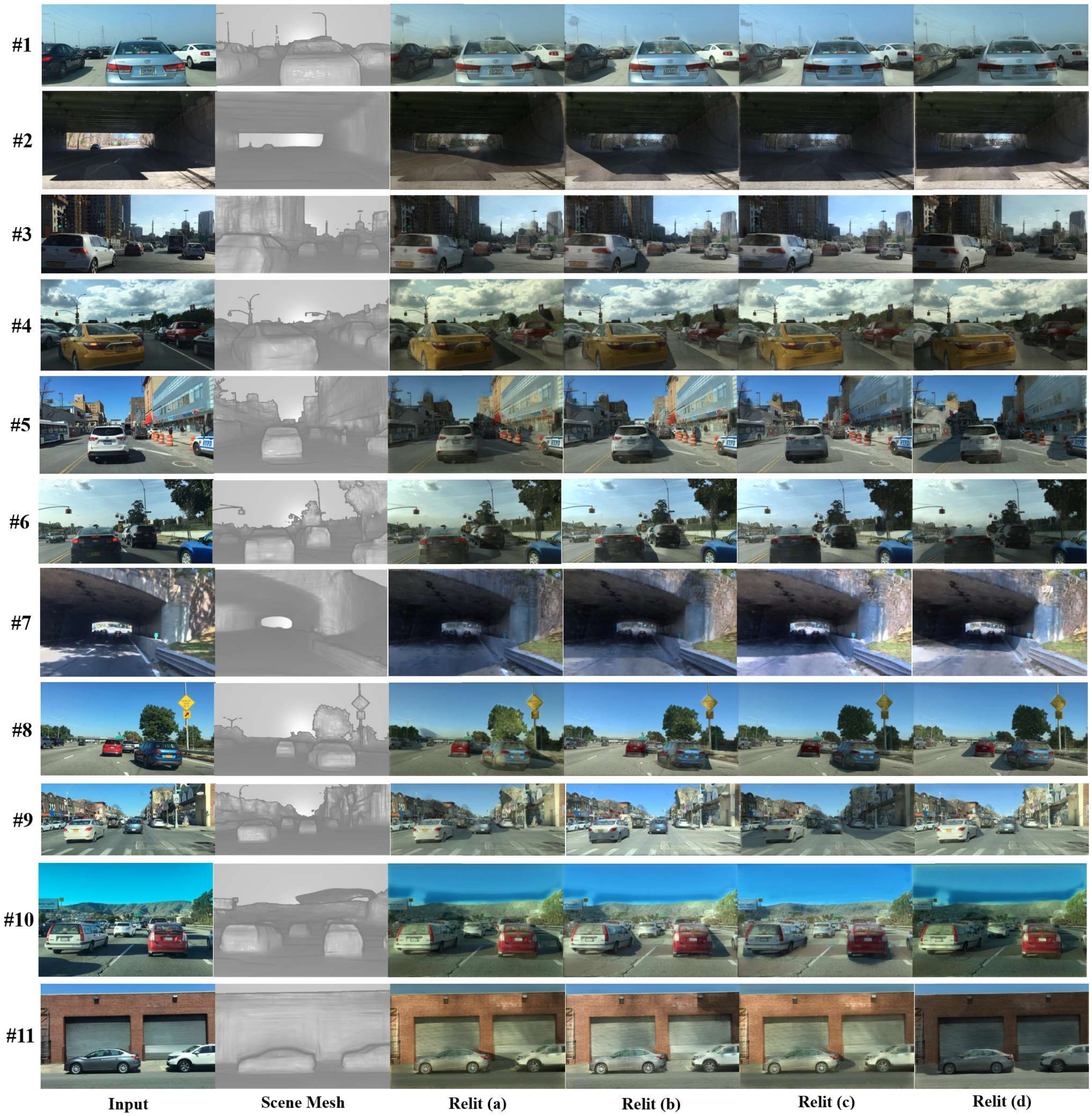

Figure 16. **SIMBAR On BDD100K:** SIMBAR relights BDD100K road driving scenes. BDD100K could be a challenging dataset to be relit due to its set up given its large scale, crowd-sourced data collection methodology,and the existence of edge and corner cases. Once again we see robust mesh generation here including in the very complex case of tunnels(row 2 & 7). The relighting is very effective in these scenes although there is slight warping of the textures. The 3rd row is also interesting because it is an overcast scene so there are no sharp shadows. SIMBAR is able to relight the scene to have sharp shadows but it is not able to raise the overall brightness of the scene. Note that high level scene semantics, such as the clouds, are preserved during relighting, something less controllable methods (e.g. GANs) could struggle to maintain without explicit labels.

Figure 17. **SIMBAR On KITTI:** SIMBAR relights KITTI road driving scenes. Using the #1 column an example, SIMBAR takes the single frame as input, then creates the 3D scene mesh with geometry priors, and finally generates 4 relit results as shown in Relit (a) (b) (c) (d), with different shadow and lighting conditions. Note that, KITTI scenes are collected around noon time with strong contrasted shadow cast on the ground, which are difficult for the relighting networks to remove. As shown in column #2 relit results, there are some shadow residuals left. Sometimes, the textures and colors are slightly washed out from the original during relighting (refer column #7 relit results). Overall, thanks to the depth backbone improvements using dense vision transformer and fine tuned on KTTI dataset, the meshes here have fairly significant complexity and detail leading to robust shadow generation.