BlenderProc: Reducing the Reality Gap with Photorealistic Rendering


*Institute of Robotics and Mechatronics, German Aerospace Center (DLR), {first}.{last}@dlr.de
†Visual Recognition Group, Czech Technical University in Prague, hodantom@cmp.felk.cvut.cz

Abstract—BlenderProc is an open-source and modular pipeline for rendering photorealistic images of procedurally generated 3D scenes which can be used for training data-hungry deep learning models. The presented results on the tasks of instance segmentation and surface normal estimation suggest that our photorealistic training images reduce the gap between the synthetic training and real test domains, compared to less realistic training images combined with domain randomization. BlenderProc can be used to train models for various computer vision tasks such as semantic segmentation or estimation of depth, optical flow, and object pose. By offering standard modules for parameterizing and sampling materials, objects, cameras and lights, BlenderProc can simulate various real-world scenarios and provide means to systematically investigate the essential factors for sim2real transfer.

I. INTRODUCTION

With the advent of deep neural networks, the demand for accurately annotated training images has grown dramatically, resulting in large real and synthetic datasets [1]–[4]. However, the neural networks are robust only when trained on sufficient in-domain data. In dynamic robotic environments and tasks like object pose estimation, the labeling effort often prohibits the use of real annotated data.

Learning in simulation is appealing, especially when considering methods that let the robot automatically generate the required models. On the other hand, the gap between the synthetic training and real test domains (a.k.a. the sim2real gap) prevents algorithms to generalize to the real domain. However, several works [5]–[7] have already shown that more realistic training images lead to better sim2real results.

BlenderProc [8] offers a fully configurable pipeline for procedurally generating scenes and rendering photorealistic training images. The pipeline is built on top of Blender [9], an open-source project which offers a variety of relevant features through a stable API with years of optimization and an active community. The Blender’s physically accurate ray tracer, *cycles*, is used for rendering. A typical run of the pipeline consists of loading a set of objects, sampling object poses with the aid of a physics engine, randomizing object materials, sampling cameras and lights, and rendering that generates modalities such as color, depth, surface normals, semantic segmentation or optical flow. All these contributions go far beyond the Blender feature set and are fully automated using a single, portable config file.

II. RELATED WORK

A crucial feature of BlenderProc is the possibility to render with a ray tracer, which gives significantly better results than any rasterizer can produce because it resembles the actual physical process of light transport. Equally important are material properties which are determined by factors like roughness and specularity. Even if they are not accurately modelled, we show that they can still be used for physically plausible domain randomization. In Table I we compare features with the NVIDIA Deep learning Dataset Synthesizer (NDDS) [10], ViSII - A VRutual Scene Imaging Interface [11], AI Habitat [12] and Stilleben [13].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>real ray tracing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>semantic segm.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>depth rendering</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>optical flow</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>surface normals</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>object pose</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>bounding box</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>physics module</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>camera sampling</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>pbr materials</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>docu for each fct.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

TABLE I: Features in BlenderProc (BP) and other simulators (2020).
In comparison to these four simulators, BlenderProc offers richer features towards simplified sim2real transfer. Real-time capability is often not required and the whole BlenderProc pipeline takes less than 2 seconds for images like in Fig. 1 on a single GPU machine. Large datasets (>100K images) can be generated on an 8-GPU server in a few hours. Furthermore, we already support loaders for a variety of datasets: SceneNet [14], ShapeNet [2], Replica [15], SUNCG [3], T-LESS [16], Linemod [17], Linemod-Occluded [18], MVTec-ITODD [19], HomebrewedDB [20], YCB-Video [21], Rutgers APC [22], Doumanoglou et al. [23], Tejani et al. [24], TUD Light [4], and Toyota Light [4]. Additionally, we are already in the process of integrating 3D-Front and 3D-Future [25, 26].

III. EXPERIMENTS

In the following, we investigate the sim2real capability of photo-realistic data generated by BlenderProc on the task of instance segmentation on the LineMod-Occluded (LM-O) dataset [18]. The dataset provides eight object models with imperfect geometry and texture acquired with KinectFusion-like techniques. We train a Mask R-CNN with Resnet50 backbone on 50K synthetic RGB images that are originally generated for the BOP Challenge 2020 [4].

The objects are placed into an open cube mapped with randomized PBR textures [27]. Object material properties like metallicity, roughness and specularity are also randomized. For details and exact data reproduction we refer to the BlenderProc config file [28]. An example of the synthetic training data is depicted in Fig. 1 on the right.

To assess the sim2real performance of our data, we compare it to the popular render&paste technique using OpenGL renderings on real backgrounds combined with strong domain randomization [29], using 50K images as well. Qualitative test results on LM-O are shown in Fig. 2. The Mask R-CNN models were pretrained on COCO and fine-tuned on the respective datasets. As found by Hinterstoisser et al. [30], when fine-tuning on the domain-randomized OpenGL data, freezing the complete backbone is necessary to avoid overfitting to the synthetic domain. This is accounted to the domain gap of low-level image statistics between real and OpenGL data. Interestingly, training on synthetic data generated by BlenderProc allows unfreezing the backbone which even slightly improves results. With default hyperparameters from [31], we achieve a clear improvement from 26.2 (OpenGL+DR) to 33.8 (BlenderProc) Mask mAP50 (+7.6).

Furthermore, we evaluated the reconstruction of surface normals from color images. We tested here a typical U-Net architecture, which was trained on simulated images from BlenderProc in SUNCG scenes and we have shown some results here on the Replica dataset, see Fig. 3.

IV. CONCLUSION

In this work, we have shown sim2real results on instance segmentation using BlenderProc, that compare favorably against the standard render&paste approach using OpenGL renderings on real images and strong image-based domain randomization. Additionally, we provided qualitative surface normal reconstruction results.

The transfer from simulation to the real world has become significantly easier with BlenderProc. Physically-plausible domain randomization, like the randomization of material properties and light, enables sim2real transfer even with imperfect models like the LineMOD objects. This is crucial to replace accurate, manual modelling with automatized real2sim methods performed by robots themselves to close the real2sim2real cycle.
REFERENCES


