Novel Synthetic Data Tool for Data-Driven Cardboard Box Localization

Peter Kravár[1](✉), Lukáš Gajdošech[2,3], and Martin Madaras[2,3]

[1] Faculty of Informatics, Masaryk University, Brno, Czech Republic
xkravar@fi.muni.cz

[2] Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia
{gajdosech,madaras}@fmph.uniba.sk

[3] Skeletex Research, Bratislava, Slovakia

Abstract. Application of neural networks in industrial settings, such as automated factories with bin-picking solutions requires costly production of large labeled datasets. This paper presents an automatic data generation tool with a procedural model of a cardboard box. We briefly demonstrate the capabilities of the system, and its various parameters and empirically prove the usefulness of the generated synthetic data by training a simple neural network. We make sample synthetic data generated by the tool publicly available.

Keywords: Synthetic Data · Neural Applications · Intelligent Robotics

1 Introduction

Automatic detection and localization of bins on a conveyor belt is an essential task in automated factories. This detection must be robust to guarantee the safe operation of robotic arms. It includes handling edge cases such as missing edges, occlusion, and variance in the materials and shapes of the bins. Moreover, in a specific scenario of package delivery factories, bins are made from a non-rigid cardboard material. These boxes are prone to various deformations, and their paper flaps are semi-randomly opened while being filled by workers and robots.

Analytical detection algorithms lack robustness and are hard to modify for new cases [5]. On the other hand, machine learning-based methods require data. Capturing real RGB-D samples in various scenarios in factories is costly. Therefore, the generation of synthetic data is recently a popular research topic [1,6], outlined by the boom of commercial solutions such as NVIDIA Omniverse™[1].

[1] https://www.nvidia.com/en-us/omniverse/solutions/digital-twins/.

Supported by the TERAIS project in the framework of the program Horizon-Widera-2021 of the European Union under the Grant agreement number 101079338.

L. Iliadis et al. (Eds.): ICANN 2023, LNCS 14254, pp. 565–569, 2023.
https://doi.org/10.1007/978-3-031-44207-0_50

Following our previous work [3], in this short submission we propose a novel data-generation tool for the automated generation of training data containing cardboard boxes. We evaluate the results of a neural network trained upon this novel data against a baseline synthetic generator, which has no automatic parametrization and cannot produce boxes with paper flaps.

2 Generating Data

This project aimed to create a high-level system for generating synthetic datasets of 3D bin scans using Blender 3D compiled into a python module (`bpy`)[2]. We accomplished this by wrapping Blender's functionality into high-level classes representing respective parts of the 3D scanning pipeline. Our pipeline simulates the real scanning process of a structured light scanner. Render settings, scanner parameters and the behavior of random parameter generation are fully customizable by the user. The output of our system comes in the form of structured point cloud data. The camera transformation matrix and the volume box of the generated cardboard box are also exported and used as ground truth data.

2.1 Parametric Cardboard Box

Variety in synthetic data can be achieved by randomizing parameters of appropriate parametric model [2]. To generate virtual cardboard boxes, we have created a parametric model which approximates the most significant box features, see Fig. 2a for visual illustration. By changing the parameters, we are able to obtain a wide variety of virtual cardboard boxes. The box parameters are:

1. **Size** - box dimensions
2. **Flap Length** - flap dimensions
3. **Flap Width** - flap taper
4. **Open** - flap open angle
5. **Thickness** - cardboard thickness
6. **Bevel** - roundness of box edges

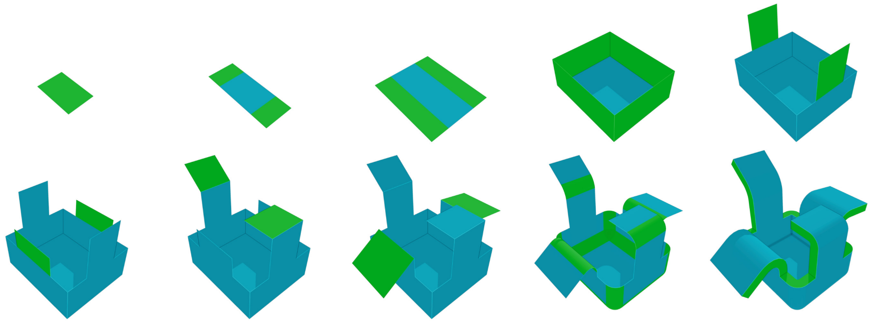
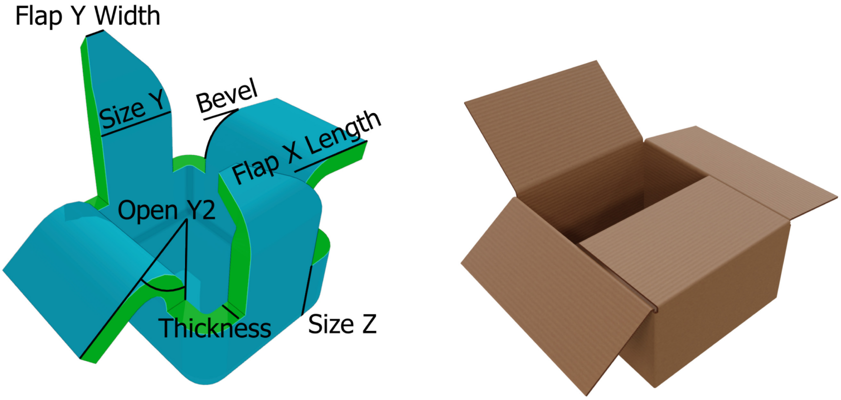


Fig. 1. Box creation process, operations are exaggerated for visual clarity.

[2] https://docs.blender.org/api/current/info_advanced_blender_as_bpy.html.

We have approximated a generic cardboard box as an object created using the corresponding sequence of steps as shown in Fig. 1. The steps include a series of extrusions, rounding corner edges, and adding thickness. The parametric model is implemented using Blender's Geometry Nodes system [4].

In real production, a box is assembled by folding a sheet of cardboard. The resulting object can therefore be closely approximated in 2D. Such 2D representation can serve as a UV map without visible seams, used for procedural shading of the parametric cardboard box, Fig. 2b shows the resulting rendered image.



(a) parameters of the model (b) procedurally shaded cardboard box

Fig. 2. Illustration of box parameters and the resulting rendered image.

2.2 Generation Parameters

The camera location was generated as a random unit vector in the positive XYZ part of a sphere scaled by uniformly distributed random distance in the $(1m, 1.7m)$ interval. The rotation of the scanner was then calculated such that the camera would point at world origin. Generation of boxes utilized random distributions for multiple parameters, ex. a single dimension was randomized as:

$$\text{Size}_X = 0.25 + min(max(-\sigma \times \gamma,\ \mathcal{N}(\mu,\ \sigma^2)), \sigma \times \gamma).$$

For our experiments, we set $\sigma = 0.1$ and $\gamma = 2.0$, each constant is in SI units.

3 Experiment

We have verified the added value of the proposed generator by training a neural network for 6D pose estimation of the cardboard boxes [3]. We have created two sets of synthetic training data, each consisting of 496 samples. The first set was generated using a baseline generator, without the automated box parametrization and flaps, see Fig. 3a. The second set is generated using our novel tool. The data, together with loading scripts in Python is publicly available[3].

3.1 Metrics

Translation of the box origin is evaluated using Euclidean distance: $e_{\mathrm{TE}}(\hat{\boldsymbol{t}}, \boldsymbol{t}) = \| \boldsymbol{t} - \hat{\boldsymbol{t}} \|_2$. For rotation, we use model-independent angle distance between rotational axes calculated from corresponding rotation matrices as: $e_{\mathrm{RE}}(\hat{R}, R) = \min_{\hat{R}' \in \{\hat{R}_1, \hat{R}_2\}} \operatorname{arcos}((\mathrm{Tr}(\hat{R}' R^{-1}) - 1)/2)$, where Tr is the matrix trace operator.

Table 1. Comparison of network's performance using different training data.

Training Data	val $\overline{e_{\mathrm{TE}}}$ (mm)	val $\overline{e_{\mathrm{RE}}}$ (rad)	test $\overline{e_{\mathrm{TE}}}$ (mm)	test $\overline{e_{\mathrm{RE}}}$ (rad)
Baseline Synthetic	35.603	1.336	14.237	1.121
Novel Synthetic	4.326	0.240	12.161	0.787

3.2 Evaluation

Table 1 compares networks trained over the two synthetic datasets. The validation set consists of 100 synthetic samples from the proposed generator and a test set of 22 real-world samples captured by PhoXi 3D Scanner[4]. Figure 3 shows qualitative examples of the predictions. Note that it has only the 3D point cloud on the input, without any information about the dimensions of the boxes.

We conclude that the novel generator helped the network to generalize and learn to ignore paper flaps, showing promise in improving synthetic data tools for more successful training. Future work includes expanding this tool for additional possible variances, such as bins from semi-transparent plastic materials with a simulation of physical phenomena like light caustics in photo-realistic textures.

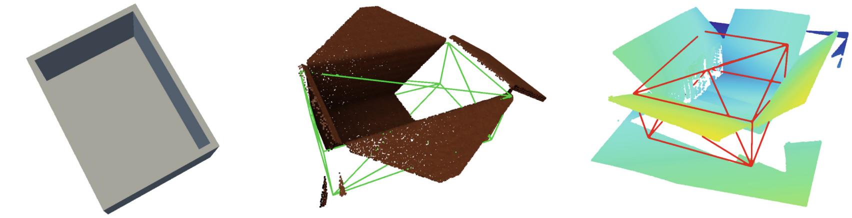

(a) baseline generator (b) validation sample (c) test sample

Fig. 3. Sample from the baseline generator and network predictions.

[3] http://www.st.fmph.uniba.sk/~gajdosech2/icann2023-dataset/.
[4] https://www.photoneo.com/phoxi-3d-scanner/.

References

1. Chen, K., et al.: Sim-to-real 6D object pose estimation via iterative self-training for robotic bin picking. In: European Conference on Computer Vision (ECCV), pp. 533–550 (2022)
2. Fedorova, S., et al.: Synthetic 3D data generation pipeline for geometric deep learning in architecture. In: The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences (ISPRS Congress), pp. 337–344 (2021)
3. Gajdošech, L., Kocur, V., Stuchlík, M., Hudec, L., Madaras, M.: Towards deep learning-based 6D bin pose estimation in 3D scan. In: VISAPP, pp. 545–552 (2022)
4. van Gumster, J., Lampel, J.: Procedural modeling with blender's geometry nodes. In: SIGGRAPH Labs. N (2022). https://doi.org/10.1145/3532725.3538516
5. Katsoulas, D.: Localization of piled boxes by means of the hough transform. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 44–51. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45243-0_7
6. Periyasamy, A.S., Schwarz, M., Behnke, S.: SynPick: a dataset for dynamic bin picking scene understanding. In: IEEE CASE, pp. 488–493 (2021)