Simulation-Based Engineering Lab
University of Wisconsin-Madison
Technical Report TR–2023–13

Instance Performance Difference (IPD)

Bo-Hsun Chen and Dan Negrut

Department of Mechanical Engineering, University of Wisconsin – Madison

September 25, 2023

Contents

Algorithm 1: Calculation of Instance Performance Difference (IPD)

Input: YOLOv5 as performance algorithm $H(image)$, labeled real image dataset S_{real}, labeled synthetic image dataset S_{synth}

Performance task: rock detection
$\boldsymbol{P}_{real} \leftarrow \phi, \boldsymbol{P}_{synth} \leftarrow \phi$;
for *each pair of images* $(I_{real} \in S_{real}, I_{synth} \in S_{synth})$ **do**
  predict bboxes $\boldsymbol{B}_{predict,real} \leftarrow H(I_{real})$; `//` m_r `predicted and` n_r `GT bboxes`
  predict bboxes $\boldsymbol{B}_{predict,synth} \leftarrow H(I_{synth})$; `//` m_s `predicted and` n_s `GT bboxes`
  Get IOU table $T_{real} \in \mathbb{R}^{n_r \times m_r}$ based on $\boldsymbol{B}_{predict,real}$;
  Get IOU table $T_{synth} \in \mathbb{R}^{n_s \times m_s}$ based on $\boldsymbol{B}_{predict,synth}$;
  for *each pair of rocks* $(x_{real} \in I_{real}, x_{synth} \in I_{synth})$ **do**
    $\boldsymbol{P}_{real}$ append $\max\{T_{real}[x_{real}, :]\}$;
    $\boldsymbol{P}_{synth}$ append $\max\{T_{synth}[x_{synth}, :]\}$;
  end
end
return $Wasserstein_1(\boldsymbol{P}_{real}, \boldsymbol{P}_{synth})$

1 Introduction

To verify synthesized images, many past works trained NNs on synthetic images, tested on real photos, and took the NN testing performance as evaluation [1,2]. However, it is unfair to judge the similarity between synthetic and real images. If the synthetic image domain covers much broader than the real photo domain, the method above could have high performance though the synthetic domain is not similar to the real domain. In the essence of similarity, a fairer method is to compare the performance difference. For example, assume a synthetic image is very similar to the real corresponding photo. If a rock in the real photo detected by a perception algorithm badly, the same rock in the synthetic image should be also detected badly. Conversely, if the real rock is detected well, the synthetic one should also be detected well.

2 Method

From the stance of simulation and the principle above, the *instance performance difference (IPD)* was proposed to judge simulation performance more fairly, which is extended from the contextualized performance difference metric [3] and similar to the metric in [4]. Unlike targeting the performance value of an algorithm tested on the real dataset and trained on the synthetic dataset, IPD focuses on the performance similarity of an instance tested on between the synthetic and real datasets, making it a fairer metric to evaluate the simulation.

The algorithm to compute IPD is described in Algorithm 1. The performance algorithm here for evaluation is YOLOv5, which is used in the rock detection task. The intersection over union (IOU) is set as the performance metric. Steps to calculate IPD between two datasets are as follows. First, for each pair of the real photo and synthetic image of the same configuration, the synthetic and real rocks were paired and indexed, shown as the rock indices in Fig. 1. Notice that since some rocks were hidden in real images while appeared in synthetic images, and the simulation setup was

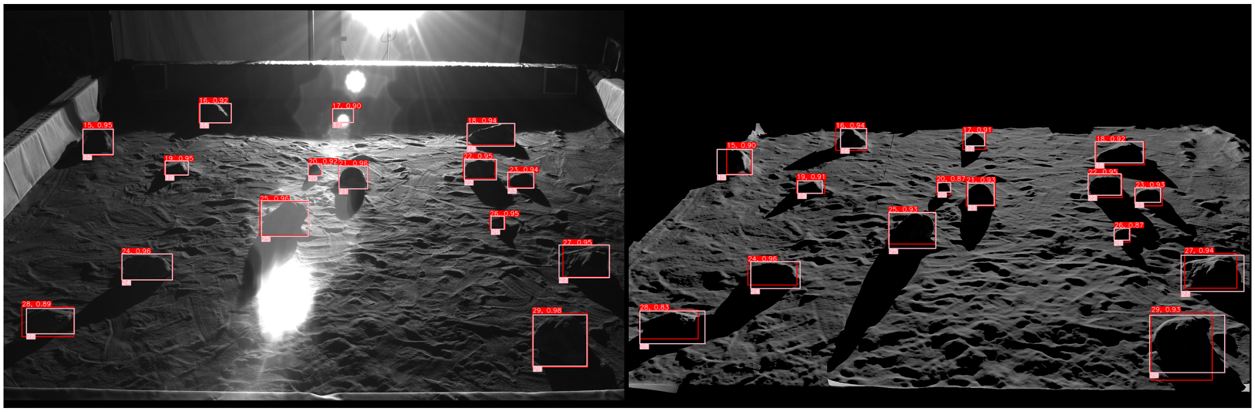

Figure 1: Indexed pairs of rocks from the (left) real and (right) synthetic images, respectively.

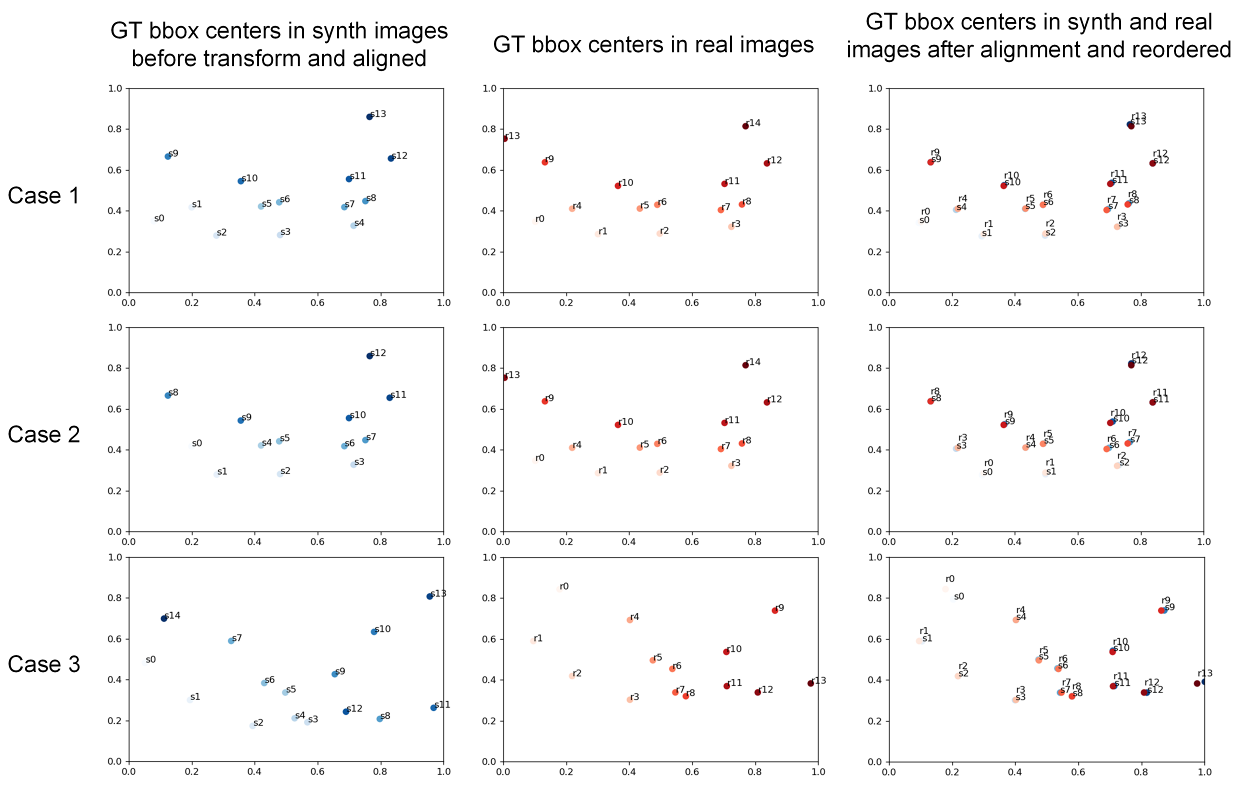

Figure 2: The point-set registration problem before and after solved of some cases shown for demonstration. (GT: ground-true, bbox: bounding box, synth: synthetic)

not perfectly as same as the reality, the number, order, and position of rocks in the pair of real and synthetic images were scarcely aligned.

This formed a 2D point-set registration problem, where the points are the ground-true bounding box centers. A modified RANSAC algorithm was utilized to solve this problem. It iteratively found the affine transformation matrix between all permutations of randomly selected three bounding box centers from the two respective datasets, and tried to achieve the smallest average Euclidean distance between the two aligned point sets after applying the transformation. The 2D point-set registration problem before and after solved is illustrated in Figure 2.

Next, a trained YOLOv5 predicted bounding boxes for rocks in both real and synthetic datasets. For each rock in either real or synthetic dataset, the predicted bounding box of the largest IOU with the ground-true label of the rock was chosen as the predicted label of the rock, and this IOU value was the performance-value of the rock. The ground-true bounding boxes were manually labeled in real photos and automatically generated by the program in synthetic images. Finally, the performance-value difference between the pair of two rocks (one from the synthetic image and the other one from the paired real photo) was calculated and averaged over all pairs of rocks. This averaged value is defined as the IPD between the two datasets.

References

[1] A. Tsirikoglou, J. Kronander, M. Wrenninge, and J. Unger, "Procedural Modeling and Physically Based Rendering for Synthetic Data Generation in Automotive Applications," 2017. [Online]. Available: https://arxiv.org/abs/1710.06270

[2] A. Prakash, S. Boochoon, M. Brophy, D. Acuna, E. Cameracci, G. State, O. Shapira, and S. Birchfield, "Structured Domain Randomization: Bridging the Reality Gap by Context-Aware Synthetic Data," in *IEEE International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 2019, pp. 7249–7255.

[3] A. Elmquist, R. Serban, and D. Negrut, "A performance contextualization approach to validating camera models for robot simulation," *arXiv e-prints*, p. arXiv:2208.01022, Aug. 2022.

[4] K. Hagn and O. Grau, "Improved Sensor Model for Realistic Synthetic Data Generation," in *ACM Computer Science in Cars Symposium*. Ingolstadt, Germany: Association for Computing Machinery, Nov. 2021, event-place: Ingolstadt, Germany.