Object pose estimation is a very important problem in domains such as robotics manipulation and augmented reality; however, 3D bounding box annotation of objects is a very expensive and laborious task and ground-truth labels are often just an approximation. In this work, we have created a large-scale object pose estimation dataset which makes use of domain randomization techniques such as use of extreme lighting and flying distractors for both single objects as well as multiple object interactions. The ground truth annotations are precise and are created by an proprietary version of the NVIDIA Deep Learning Data Synthesizer.

NVIDIA Deep Learning Data Synthesizer

- A plugin for Unreal Engine 4 to create synthetic dataset. It is very fast (50-100HZ) and provides the following annotations:
  - 3D object pose via 3D bounding box for each annotated object
  - Projected 3D bounding boxes
  - 2D bounding boxes
  - Instance segmentation mask
  - Class segmentation mask
  - Depth map
  - Percentage of object visibility from camera

It provides various randomization such as:
- Random camera movement
- Random object pose
- Random texture
- Camera path following via anchor points

Consider the following Domino Sugar Box

Visualizing the 3D cuboids using NVIDIA Dataset Utilities toolset

Comparison with other Object Pose Estimation Datasets