# Exploring the Sim2Real Gap using Digital Twins

Sruthi Sudhakar[1], Jon Hanzelka[2], Josh Bobillot[2], Tanmay Randhavane[2], Neel Joshi[2], Vibhav Vineet[2]

Columbia University · Microsoft

2021 ICCV OCTOBER 11-17 VIRTUAL

## Highlights

**Problem**: Creating synthetic datasets is time-consuming and sometimes ineffective
- Synthetic data useful when there is limited resources (real data, time, budget)
- Models trained on synthetic data often under-perform on real world test data
- Little understanding of what causes the Sim2Real domain gap

**Research question**: What aspects of realism in synthetic data have the largest affect on performance?
- Can we isolate which model quality factors cause performance drops?

**Contributions:**
- A new YCB-Real and digital twin YCB-Synthetic dataset
- Discover which factors of variation in 3D model quality affect performance
- A cost-benefit analysis between artist time for correcting an artifact and trained model accuracy

## Creating Synthetic Data

**Asset Digitization:** Creating the building blocks of synthetic scenes
- Automated photogrammetry to reconstruct textured 3D mesh
- Capture 360° HDR image of environment to accurately represents color and exposure values

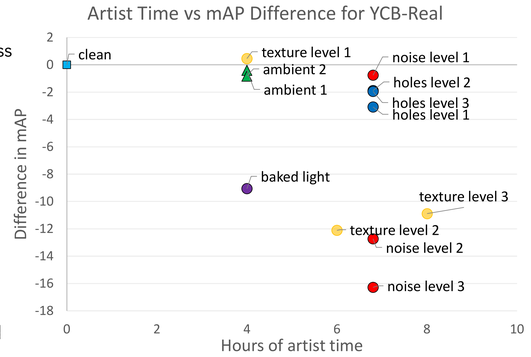
**Scene Composition:** Combining objects and environment
- Artist configures scene layout based in the specific goals for the synthetic dataset
- Scene rendered using a ray-traced engine producing highly photorealistic images and near-perfect annotations.

**Model Quality Factors:** Issues that arise during object digitization
- **Holes in Geometry**: 3D meshes may have holes due to the capture process or surface properties of object
- **Texture Artifacts**: Image anomalies (i.e. blurry images and misalignments) result in texture artifacts (doubling of projected details in the texture)
- **Lighting Artifacts**: Lack of consistent diffuse lighting (shadows, light directionality, and specular/reflective properties) get embedded in the resulting texture
- **Mesh Noise**: High-frequency noise due to lacking imagery or surface properties

## YCB Digital Twin Dataset

**YCB-Real**:
- 16 physical objects matching synthetic counterparts
- ~100 images per object, manually annotated with class label and bounding boxes
- Scene lit using area light for ambient light and two floor lamps for directional lighting
- Data captured using digitized lighting, staging table, Android phone, and various camera positions

**YCB-Synthetic**:
- 20 daily life objects from YCB Object and Model Set
- 1000 images per object rendered using cleaned and polished photogrammetry models
- For each model quality factors, we artificially add corruption back into a clean version of the model and render 1000 images per object





## Experimental Setup

**Dataset**: YCB Synthetic, YCB Real, YCB-In-The-Wild, YCB-Video
**Task**: Object Detection, Semantic Segmentation
**Classes:** 21 YCB Objects with 16 real counterparts
**Models:** Faster and Mask R-CNN pertained on COCO

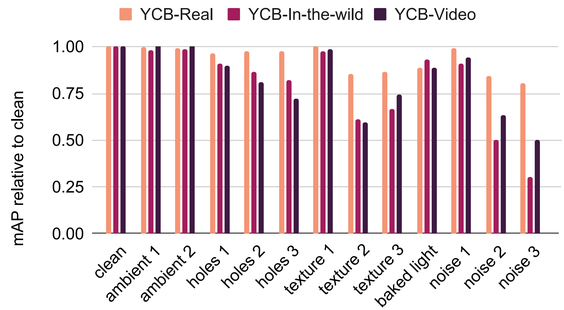
**Train**: Fine-tune a model per each of 12 corrupted synthetic datasets (and one on clean dataset)
**Test:** Test each corrupted model on all 3 real datasets to understand if that model quality factor causes a performance drop

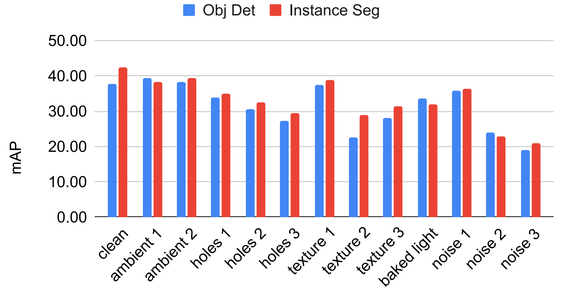
## Findings

Worst performing artifacts are Noise 2 & 3 while ambient lighting has little to no impact

"Best bang for one's buck" is fixing baked lighting issues and significant texture and noise errors

Trends hold true across all 3 real datasets and both tasks.




Dataset Released! Use this highly controlled training and test environments to: study how effects of corruptions vary based on object properties, evaluate domain generalization/ adaptation methods, build new/more complex environments.