# Dex-NeRF: Using a Neural Radiance field to Grasp Transparent Objects

Jeffrey Ichnowksi*, Yahav Avigal*, Justin Kerr, Ken Goldberg
*equal contribution

https://sites.google.com/view/dex-nerf

## Robots struggle to manipulate transparent objects

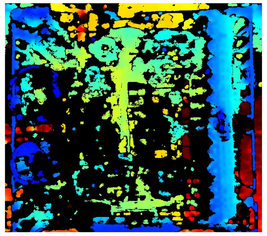
Many manipulation tasks require interacting with transparent objects.

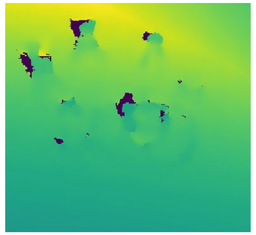
**Dishwasher**

**Lab Equipment**

But depth sensors cannot provide the geometric information around transparent surfaces that is required to plan grasps.

**RealSense**



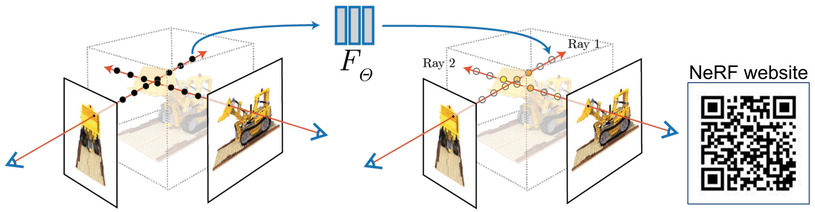


## Neural Radiance Fields (NeRF) (prior work)

Given many images of a scene, NeRF learns to output a color and density (translucency) for an input position and viewing angle

$$(x,y,z,\theta,\phi) \rightarrow \text{NN} \rightarrow (RGB\sigma)$$

NeRF generates new images by sampling the network along camera rays

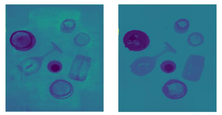



NeRF website

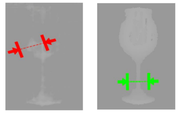
## Transparency-Aware Depth Rendering

To robustly recover depth, instead of compositing samples along camera rays Dex-NeRF truncates when the density estimation reaches a threshold

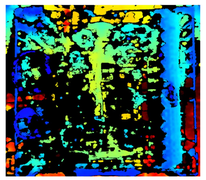

Real Image

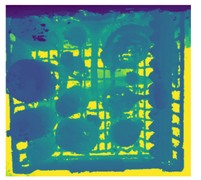

Depth (vanilla-NeRF)

Depth (Dex-NeRF)


Dex-Net Grasp (vanilla-NeRF)

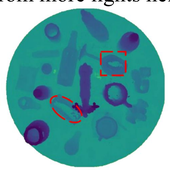

Dex-Net Grasp (Dex-NeRF)


Depth (RealSense)

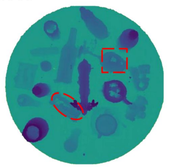

Depth (Dex-NeRF)

Additional specularities from more lights helps reconstruct surfaces

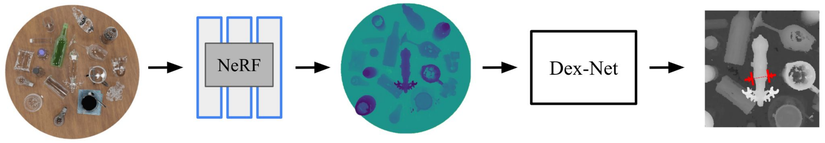

(a) RGB Scene Single Light Source


(b) Depth Rendering Single Light Source


(c) RGB Scene Multiple Light Sources


(d) Depth Rendering Multiple Light Sources

## Experimental Setup




Scene Imaging

NeRF Training

Depthmap Rendering

Grasp Planning

Grasp Execution

### Experiment Categories

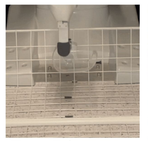
- *Physical*: images taken by handheld camera, grasp evaluated on YuMi robot
- *Simulated*: images rendered in Blender, grasps evaluated in physics simulator

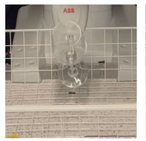
## Physical Results

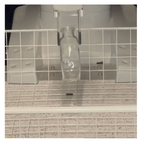
We evaluated Dex-NeRF on a YuMi robot on 6 scenes: 5 with singulated objects and 1 with clutter.

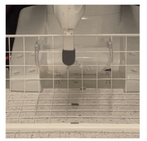
| Object | PhoXi | NeRF | Dex-NeRF |
|---|---|---|---|
| Tape Dispenser | 0/10 | 0/10 | **10/10** |
| Wineglass | 0/10 | 0/10 | **9/10** |
| Flask | 0/10 | 1/10 | **9/10** |
| Safety Glasses | 0/10 | 0/10 | **10/10** |
| Bottle | 0/10 | **10/10** | **10/10** |
| Lion Figurine | 0/10 | 3/10 | **10/10** |

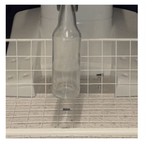
Dex-NeRF achieves 90-100% success on all objects while grasps planned with a PhoXi depth camera and vanilla NeRF fail

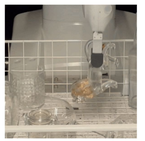

Tape Dispenser


Wineglass

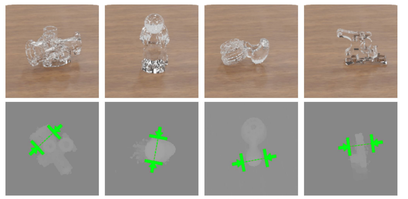

Flask


Safety Glasses


Bottle


Lion Figurine

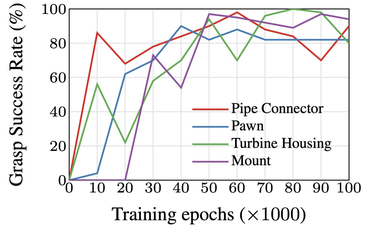
View synthesis typically requires over 200k epochs, but grasping succeeds much earlier since NeRF learns geometry before appearance


Pipe Connector

Pawn

Turbine Housing

Mount

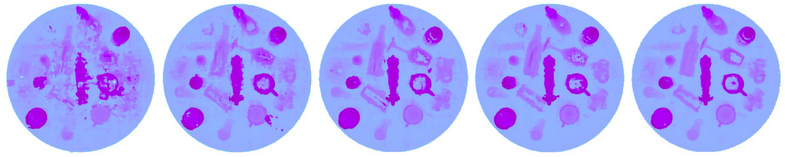



## Towards Practical Usage

Simplifying NeRF results in significant speedup, though still impractical

| Network Name | Time | Success (%) |
|---|---|---|
| Original | 253 min | 84 |
| **½ hidden units + ½ samples** | **47 min** | **96** |

Simulation experiments with top-down camera arrays show promise for work-cell setups


9 Cameras

16 Cameras

25 Cameras

36 Cameras

49 Cameras

Future work: use pre-conditioning and depth priors to speed training