

Nathan Silberman  
4Catalyzer  
Lead Deep Learning Scientist



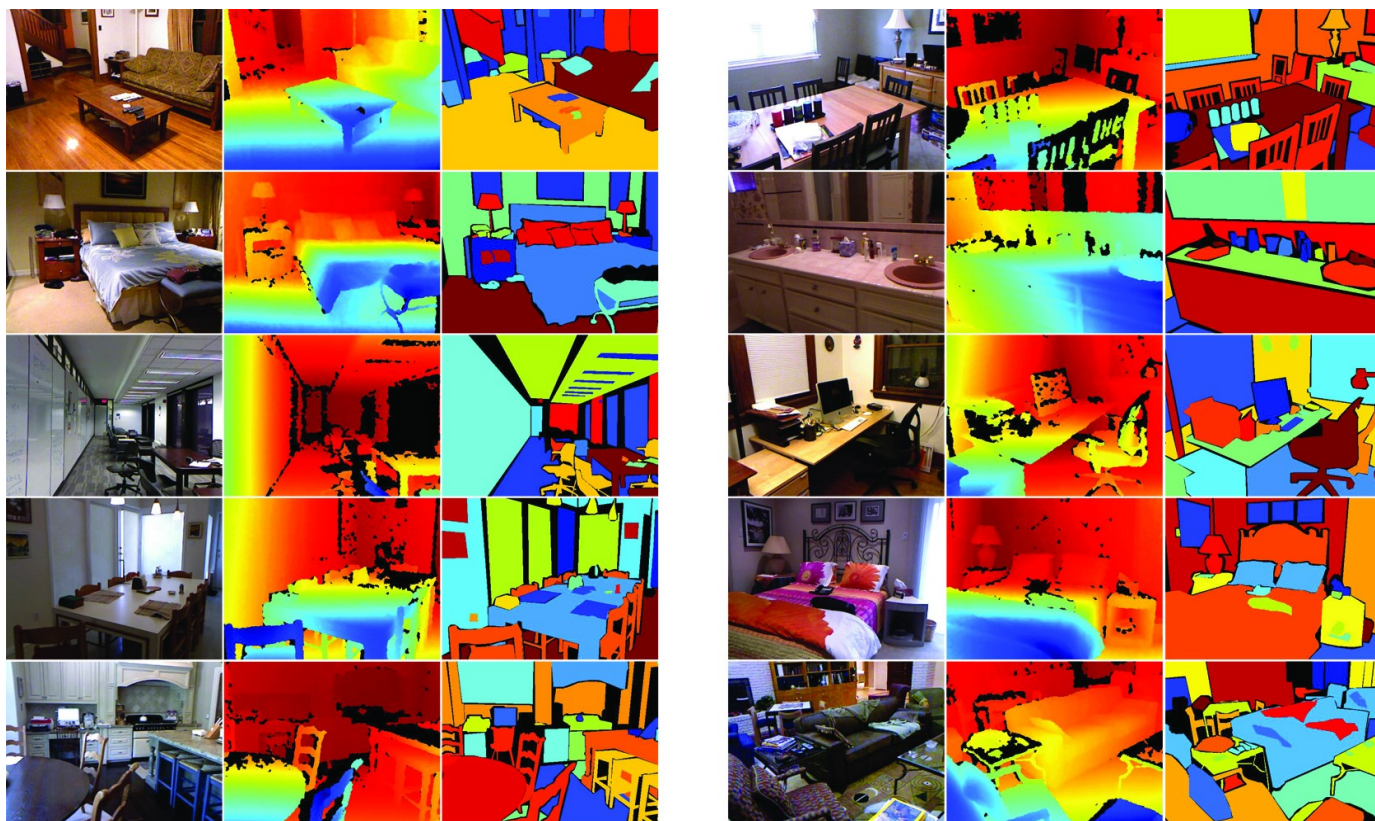
- [Home](#)
- [Publications](#)
- [Datasets](#)
- [Code](#)
- [About](#)

## NYU Depth Dataset V2

[Nathan Silberman](#), [Pushmeet Kohli](#), [Derek Hoiem](#), [Rob Fergus](#)

If you use the dataset, please cite the following work:

**Indoor Segmentation and Support Inference from RGBD Images**  
ECCV 2012 [\[PDF\]](#)[\[Bib\]](#)



Samples of the RGB image, the raw depth image, and the class labels from the dataset.

### Overview

The NYU-Depth V2 data set is comprised of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft [Kinect](#). It features:

- 1449 densely labeled pairs of aligned RGB and depth images
- 464 new scenes taken from 3 cities
- 407,024 new unlabeled frames
- Each object is labeled with a class and an instance number (cup1, cup2, cup3, etc)

The dataset has several components:

- Labeled: A subset of the video data accompanied by dense multi-class labels. This data has also been preprocessed to fill in

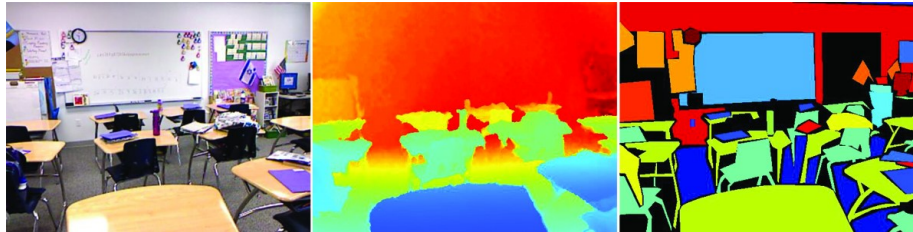
missing depth labels.

- **Raw:** The raw rgb, depth and accelerometer data as provided by the Kinect.
- **Toolbox:** Useful functions for manipulating the data and labels.

## Downloads

- [Labeled dataset \(~2.8 GB\)](#)
- [Raw dataset, Single File \(~428 GB\)](#)
- [Raw dataset, Multi-part \(~428 GB\)](#)
- [Toolbox](#)

## Labeled Dataset

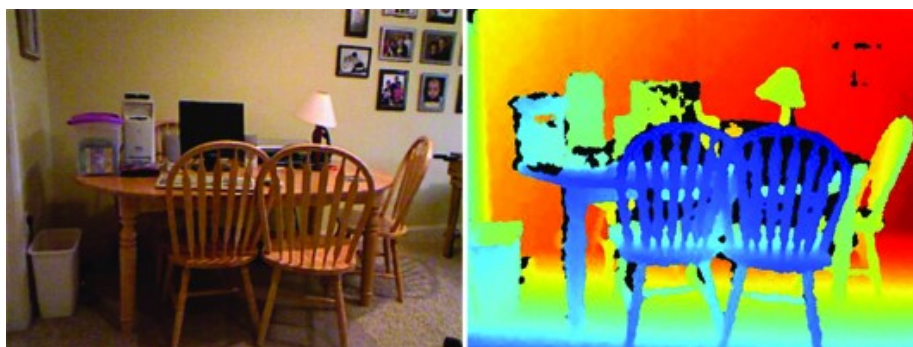


Output from the RGB camera (left), preprocessed depth (center) and a set of labels (right) for the image.

The labeled dataset is a subset of the Raw Dataset. It is comprised of pairs of RGB and Depth frames that have been synchronized and annotated with dense labels for every image. In addition to the projected depth maps, we have included a set of preprocessed depth maps whose missing values have been filled in using the [colorization scheme of Levin et al.](#) Unlike, the Raw dataset, the labeled dataset is provided as a Matlab .mat file with the following variables:

- **accelData** –  $N \times 4$  matrix of accelerometer values indicated when each frame was taken. The columns contain the roll, yaw, pitch and tilt angle of the device.
- **depths** –  $H \times W \times N$  matrix of in-painted depth maps where  $H$  and  $W$  are the height and width, respectively and  $N$  is the number of images. The values of the depth elements are in meters.
- **images** –  $H \times W \times 3 \times N$  matrix of RGB images where  $H$  and  $W$  are the height and width, respectively, and  $N$  is the number of images.
- **instances** –  $H \times W \times N$  matrix of instance maps. Use `get_instance_masks.m` in the Toolbox to recover masks for each object instance in a scene.
- **labels** –  $H \times W \times N$  matrix of object label masks where  $H$  and  $W$  are the height and width, respectively and  $N$  is the number of images. The labels range from  $1..C$  where  $C$  is the total number of classes. If a pixel's label value is 0, then that pixel is 'unlabeled'.
- **names** –  $C \times 1$  cell array of the english names of each class.
- **namesToIds** – map from english label names to class IDs (with  $C$  key-value pairs)
- **rawDepths** –  $H \times W \times N$  matrix of raw depth maps where  $H$  and  $W$  are the height and width, respectively, and  $N$  is the number of images. These depth maps capture the depth images after they have been projected onto the RGB image plane but before the missing depth values have been filled in. Additionally, the depth non-linearity from the Kinect device has been removed and the values of each depth image are in meters.
- **rawDepthFileNames** –  $N \times 1$  cell array of the filenames (in the Raw dataset) that were used for each of the depth images in the labeled dataset.
- **rawRgbFileNames** –  $N \times 1$  cell array of the filenames (in the Raw dataset) that were used for each of the RGB images in the labeled dataset.
- **scenes** –  $N \times 1$  cell array of the name of the scene from which each image was taken.
- **sceneTypes** –  $N \times 1$  cell array of the scene type from which each image was taken.

## Raw Dataset



Output from the RGB camera (left) and depth camera (right). Missing values in the depth image are a result of (a) shadows caused by the disparity between the infrared emitter and camera or (b) random missing or spurious values caused by specular or low albedo surfaces.

The raw dataset contains the raw image and accelerometer dumps from the kinect. The RGB and Depth camera sampling rate lies between 20 and 30 FPS (variable over time). While the frames are not synchronized, the timestamps for each of the RGB, depth and accelerometer files are included as part of each filename and can be synchronized to produce a continuous video using the `get_synched_frames.m` function in the Toolbox.

The dataset is divided into different folders which correspond to each 'scene' being filmed, such as 'living\_room\_0012' or 'office\_0014'. The file hierarchy is structured as follows:

```
/
../bedroom_0001/
../bedroom_0001/a-1294886363.011060-3164794231.dump
../bedroom_0001/a-1294886363.016801-3164794231.dump
...
../bedroom_0001/d-1294886362.665769-3143255701.pgm
../bedroom_0001/d-1294886362.793814-3151264321.pgm
...
../bedroom_0001/r-1294886362.238178-3118787619.ppm
../bedroom_0001/r-1294886362.814111-3152792506.ppm
```

Files that begin with the prefix a- are the accelerometer dumps. These dumps are written to disk in binary and can be read with file `get_accel_data.mex`. Files that begin with the prefix r- and d- are the frames from the RGB and depth cameras, respectively. Since no preprocessing has been performed, the raw depth images must be projected onto the RGB coordinate space into order to align the images.

## Toolbox

The matlab toolbox has several useful functions for handling the data.

- `camera_params.m` - Contains the camera parameters for the Kinect used to capture the data.
- `crop_image.m` - Crops an image to use only the area when the depth signal is projected.
- `fill_depth_colorization.m` - Fills in the depth using Levin et al's Colorization method.
- `fill_depth_cross_bf.m` - Fills in the depth using a cross-bilateral filter at multiple scales.
- `get_accel_data.m` - Returns the accelerometer parameters at a specific moment in time.
- `get_instance_masks.m` - Returns a set of binary masks, one for each object instance in an image.
- `get_rgb_depth_overlay.m` - Returns a visualization of the RGB and Depth alignment.
- `get_synched_frames.m` - Returns a set of synchronized RGB and Depth frames that can be used to produced RGBD videos of each scene.
- `get_timestamp_from_filename.m` - Returns the timestamp from the raw dataset filenames. This is useful for sampling the RAW video dumps at even intervals in time.
- `project_depth_map.m` - Projects the Depth map from the Kinect on the RGB image plane.

## Raw Dataset Parts

If you don't want to download the entire RAW dataset in a single file, different parts of the dataset can be downloaded individually:

Basements <a href="#">zip md5</a>	Bedrooms (6/7) <a href="#">zip md5</a>	Home Offices <a href="#">zip md5</a>	Misc (2/2) <a href="#">zip md5</a>
Bathrooms (1/4) <a href="#">zip md5</a>	Bedrooms (7/7) <a href="#">zip md5</a>	Kitchens (1/3) <a href="#">zip md5</a>	Offices (1/2) <a href="#">zip md5</a>
Bathrooms (2/4) <a href="#">zip md5</a>	Bookstore (1/3) <a href="#">zip md5</a>	Kitchens (2/3) <a href="#">zip md5</a>	Offices (2/2) <a href="#">zip md5</a>
Bathrooms (3/4) <a href="#">zip md5</a>	Bookstore (2/3) <a href="#">zip md5</a>	Kitchens (3/3) <a href="#">zip md5</a>	Office Kitchens <a href="#">zip md5</a>
Bathrooms (4/4) <a href="#">zip md5</a>	Bookstore (3/3) <a href="#">zip md5</a>	Libraries <a href="#">zip md5</a>	Playrooms <a href="#">zip md5</a>
Bedrooms (1/7) <a href="#">zip md5</a>	Cafe <a href="#">zip md5</a>	Living Rooms (1/4) <a href="#">zip md5</a>	Reception Rooms <a href="#">zip md5</a>
Bedrooms (2/7) <a href="#">zip md5</a>	Classrooms <a href="#">zip md5</a>	Living Rooms (2/4) <a href="#">zip md5</a>	Studies <a href="#">zip md5</a>
Bedrooms (3/7) <a href="#">zip md5</a>	Dining Rooms (1/2) <a href="#">zip md5</a>	Living Rooms (3/4) <a href="#">zip md5</a>	Study Rooms <a href="#">zip md5</a>
Bedrooms (4/7) <a href="#">zip md5</a>	Dining Rooms (2/2) <a href="#">zip md5</a>	Living Rooms (4/4) <a href="#">zip md5</a>	
Bedrooms (5/7) <a href="#">zip md5</a>	Furniture Stores <a href="#">zip md5</a>	Misc <a href="#">zip md5</a>	

## Correction to Segmentation Results

Due to a bug in the segmentation evaluation, the segmentation results in the paper were about 2% higher across the board:

Features	Weighted Score	Unweighted Score
RGB Only	50.3	44.0
Depth Only	53.7	43.0
RGBD	60.1	47.9
RGBD + Support	60.7	48.8
RGBD + Support + Structure classes	61.1	49.1