

Pose Tracking: Object Pose Estimation and Tracking

CS 6334 Virtual Reality

Professor Yapeng Tian

The University of Texas at Dallas

A lot of slides of course lectures borrowed from Professor Yu Xiang's VR class

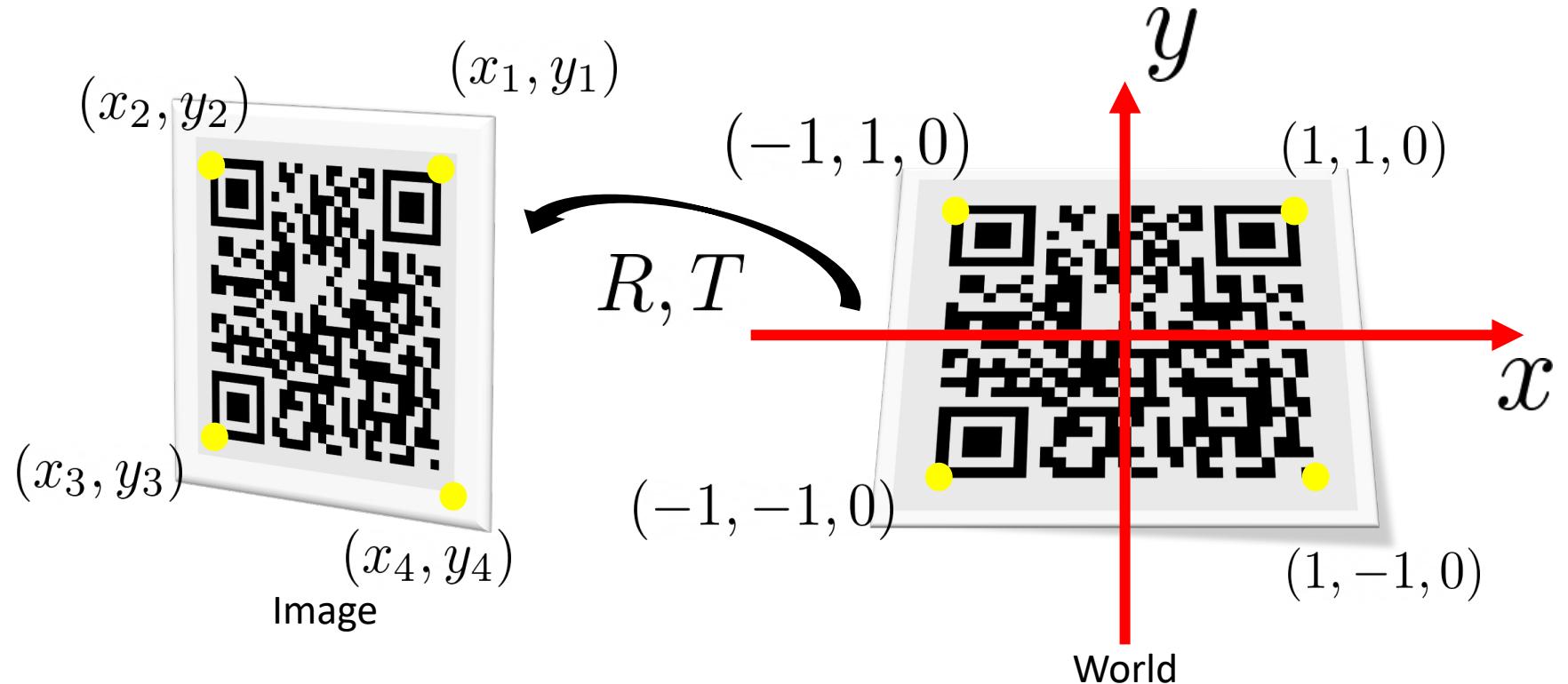
Tracking in VR

- Tracking the user's sense organs
 - E.g., Head and eye
 - Render stimulus accordingly
- Tracking user's other body parts
 - E.g., human body and hands
 - Locomotion and manipulation
- Tracking the rest of the environment
 - Augmented reality
 - Obstacle avoidance in the real world



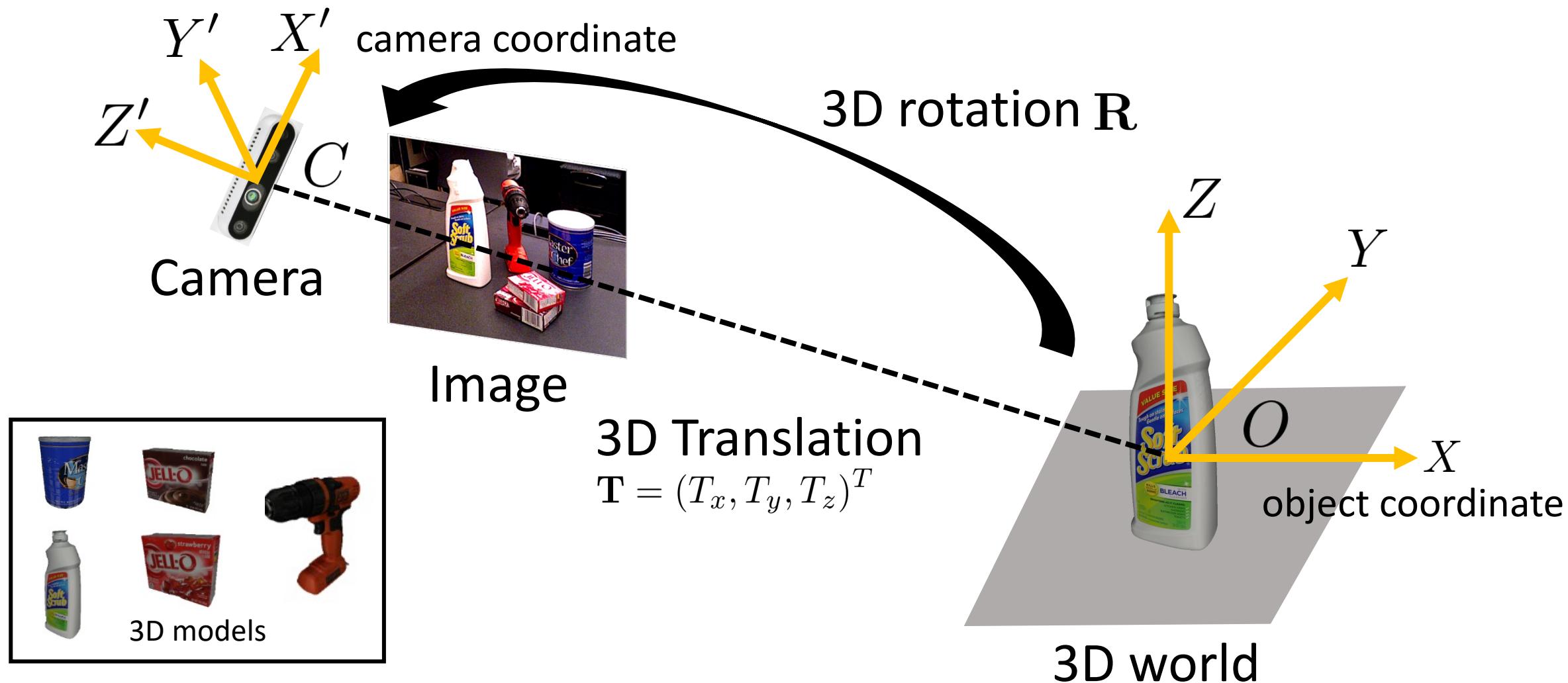
Tracking Objects in the Real World

- AR tags



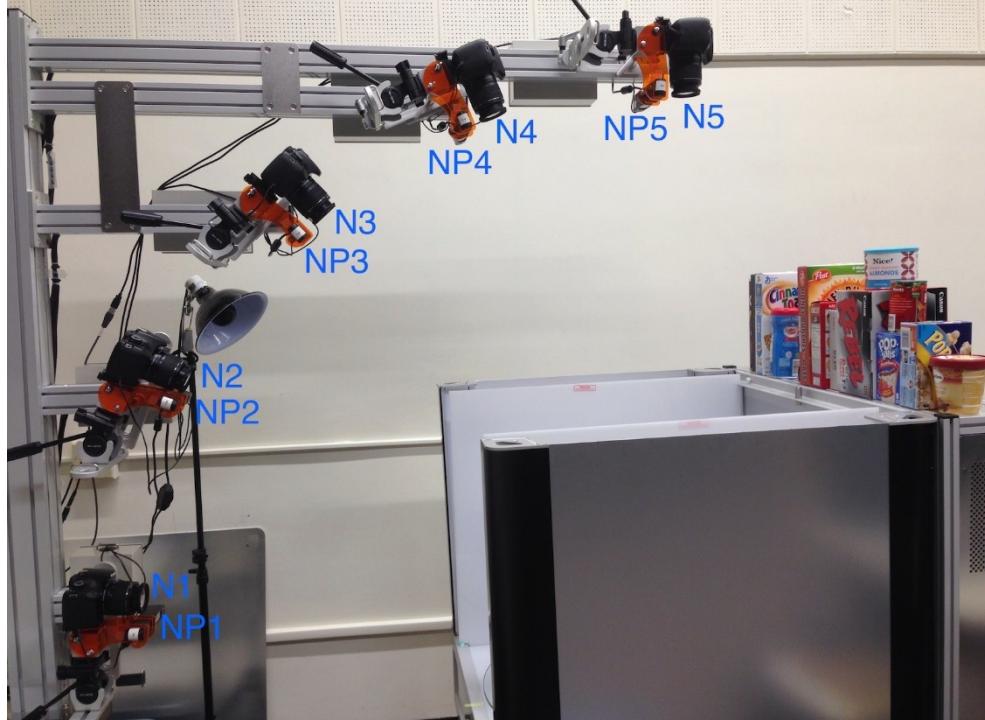
How about tracking general objects in the world?

6D Object Pose Estimation



Building 3D Object Models

- 3D reconstruction from multiple images



Berkeley Instance Recognition Dataset. Singh et al., ICRA, 2014

Building 3D Object Models

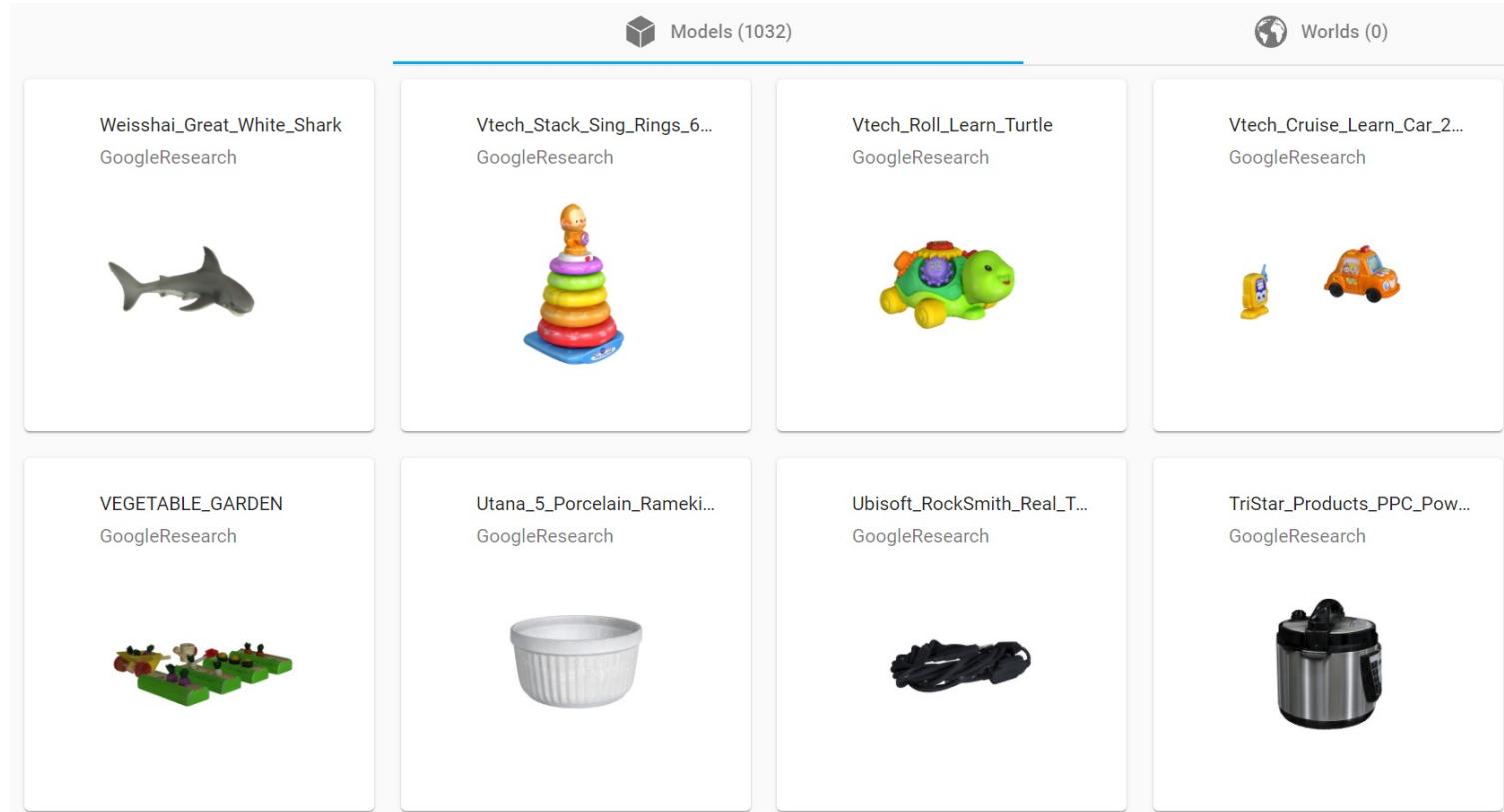
- A 3D reconstruction example



<https://blog.kitware.com/3d-reconstruction-from-smartphone-videos/>

Building 3D Object Models

- 3D Scanning



<https://app.ignitionrobotics.org/GoogleResearch/fuel/collections/Google%20Scanned%20Objects>

Building 3D Object Models

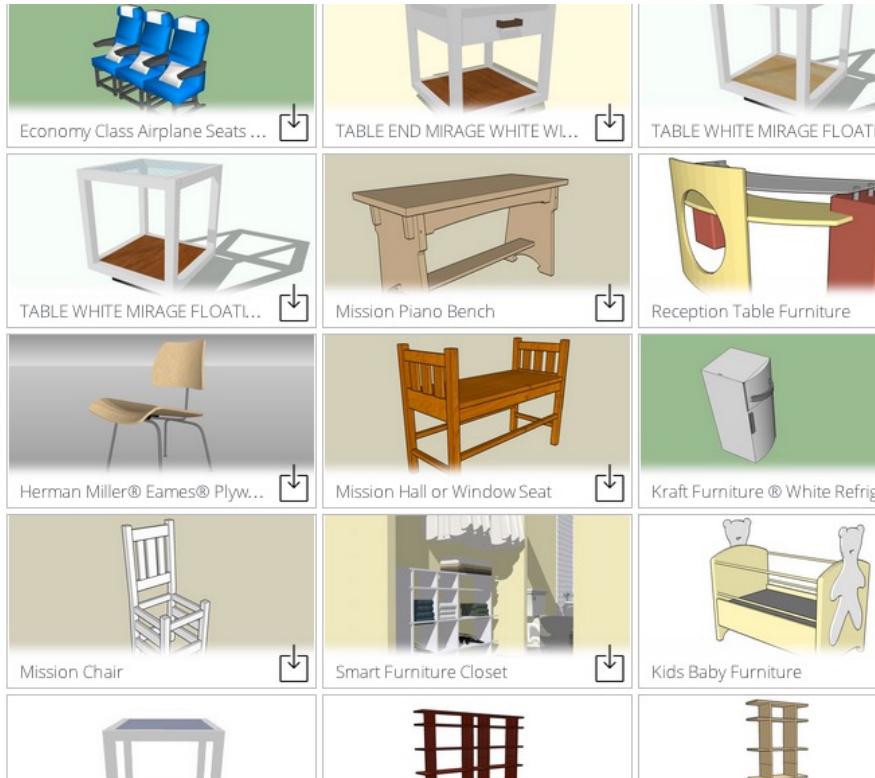
- 3D Scanning



<https://3dscanexpert.com/shining-3d-einscan-pro-3d-scanner-review/>

Building 3D Object Models

- 3D Computer-Aided Design (CAD) models



Trimble 3D Warehouse
<https://3dwarehouse.sketchup.com>

10/19/22

A screenshot of a taxonomy search interface. The search term 'bench' is entered, with a definition: 'a long seat for more than one person'. Below is the 'ImageNet MetaData' link. A dropdown menu shows 'Choose taxonomy: ShapeNetCore'. A scrollable list of categories includes: bathtub, bathing tub, bath, tub(0,856); bed(13,233); bench(5,1813); bicycle, bike, wheel, cycle(0,59); birdhouse(0,73); bookshelf(0,452); bottle(6,498); bowl(1,186); bus, autobus, coach, charabanc, double-decker, j; cabinet(9,1571); camera, photographic camera(4,113); can, tin, tin can(2,108); cap(4,56); car, auto, automobile, machine, motorcar(18,353); chair(23,6778); clock(3,651); computer keyboard, keypad(0,65); dishwasher dish washer dishwashing machine.

Yapeng Tian

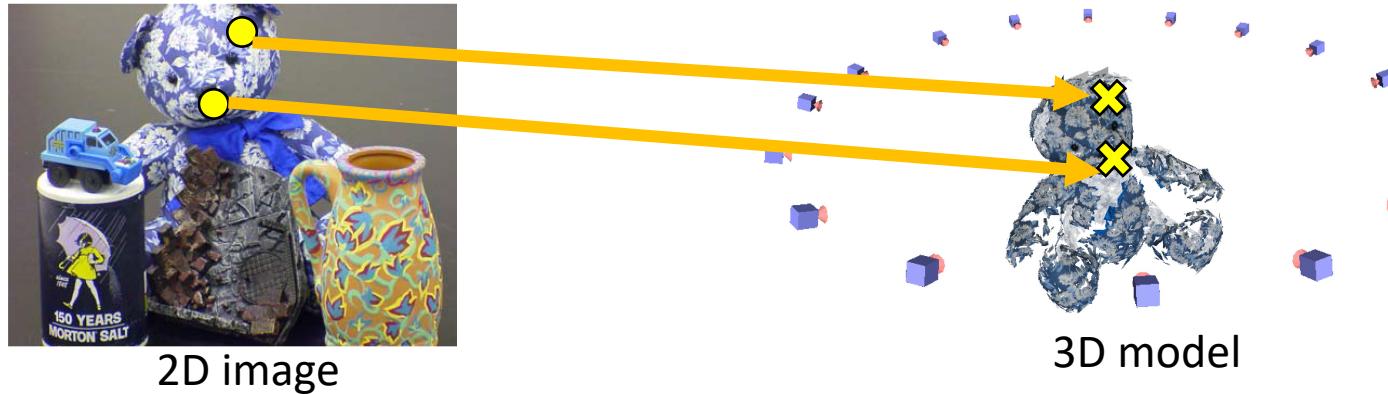
A screenshot of the ShapeNet website. It shows a grid of 3D models of various types of benches, labeled as 'Synset Models'. The top navigation bar includes 'Synset Models', 'TreeMap', 'Stats', and 'Measures'. Below it, it says 'Displaying 1 to 160 of 1813' with a page number selector from 1 to 12. The grid contains 16 rows of 10 models each, with labels like 'bench', 'park bench', and 'park bench' below them.

ShapeNet
<https://www.shapenet.org/>

9

6D Object Pose Estimation

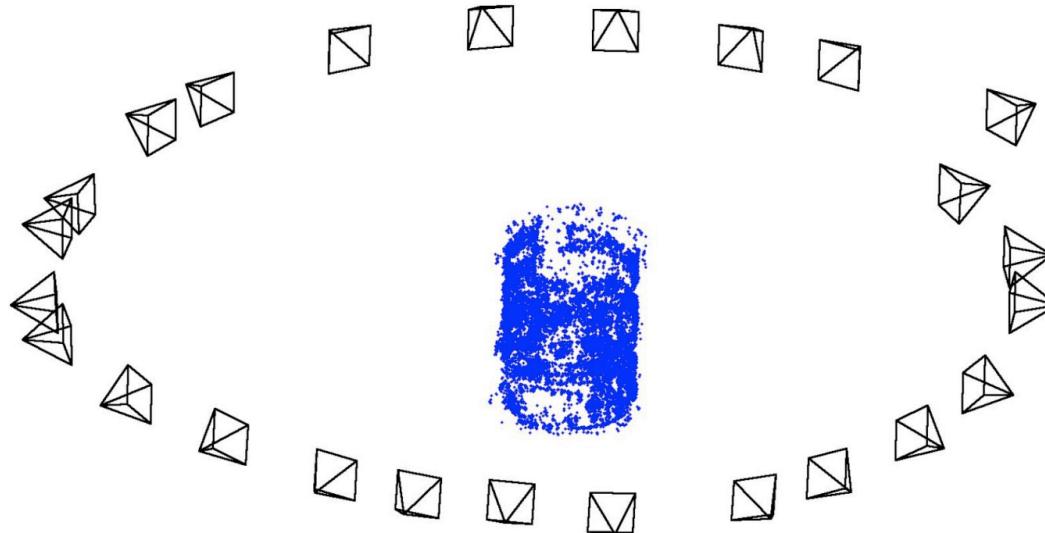
- Feature matching-based methods



Rothganger et al., IJCV, 2006

A Case Study for Feature Matching

- 3D Models of Objects using Structure from Motion
 - 3D points with SIFT descriptors (each 3D point can have a list of descriptors or use the mean of the descriptors)



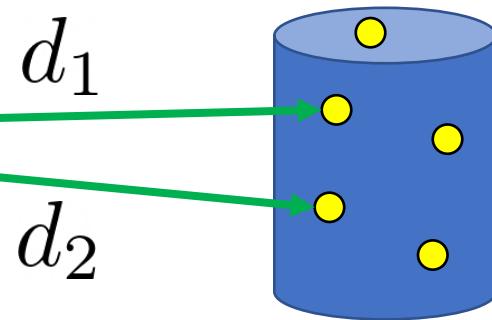
Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR'10.

A Case Study for Feature Matching

- Ratio test



Query Image



3D Model

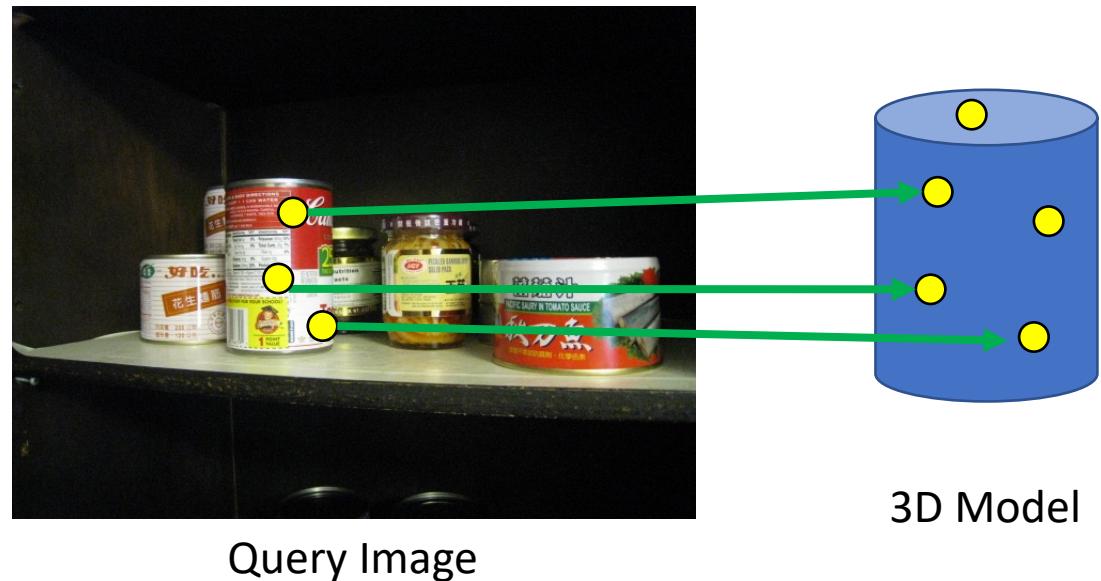
Distance to closest 3D point

$$\text{ratio} = \frac{d_1}{d_2} < 0.8$$

Distance to second
closest 3D point

A Case Study for Feature Matching

- 3D-2D correspondences from feature matching $(\mathbf{X}_i, \mathbf{x}_i)_{i=1}^N$



Option 1: minimizing reprojection error
• Levenberg-Marquardt

$$g(\mathbf{R}, \mathbf{T}) = \sum_{i=1}^N \|P(\mathbf{X}_i, \mathbf{R}, \mathbf{T}) - \mathbf{x}_i\|^2$$

Option 2: solve the PnP problem
• EPnP

Random Sample Consensus (RANSAC)

- An iterative method for parameter estimation from a set of observed data that contains **outliers**

RANSAC Algorithm {

1. Selects N data items as random
2. Estimates parameter \vec{x}
3. Finds how many data items (of M) fit the model with parameter vector \vec{x} within a user given tolerance. Call this K .
4. If K is big enough, accept fit and exit with success.
5. Repeat step 1 until 4 (as L times)
6. Algorithm will be exit with fail

}

Sample N 3D-2D correspondences $(\mathbf{X}_i, \mathbf{x}_i)_{i=1}^N$

Estimate (\mathbf{R}, \mathbf{T})

Find how many $(\mathbf{X}_i, \mathbf{x}_i)$ obeys (\mathbf{R}, \mathbf{T})

A Case Study for Feature Matching

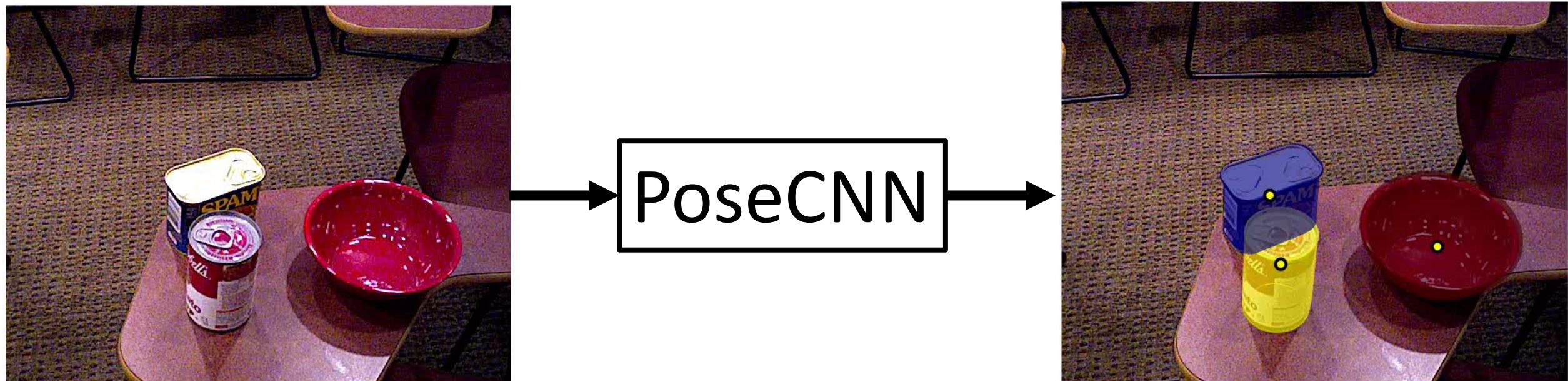
- Pose estimation examples



3D models

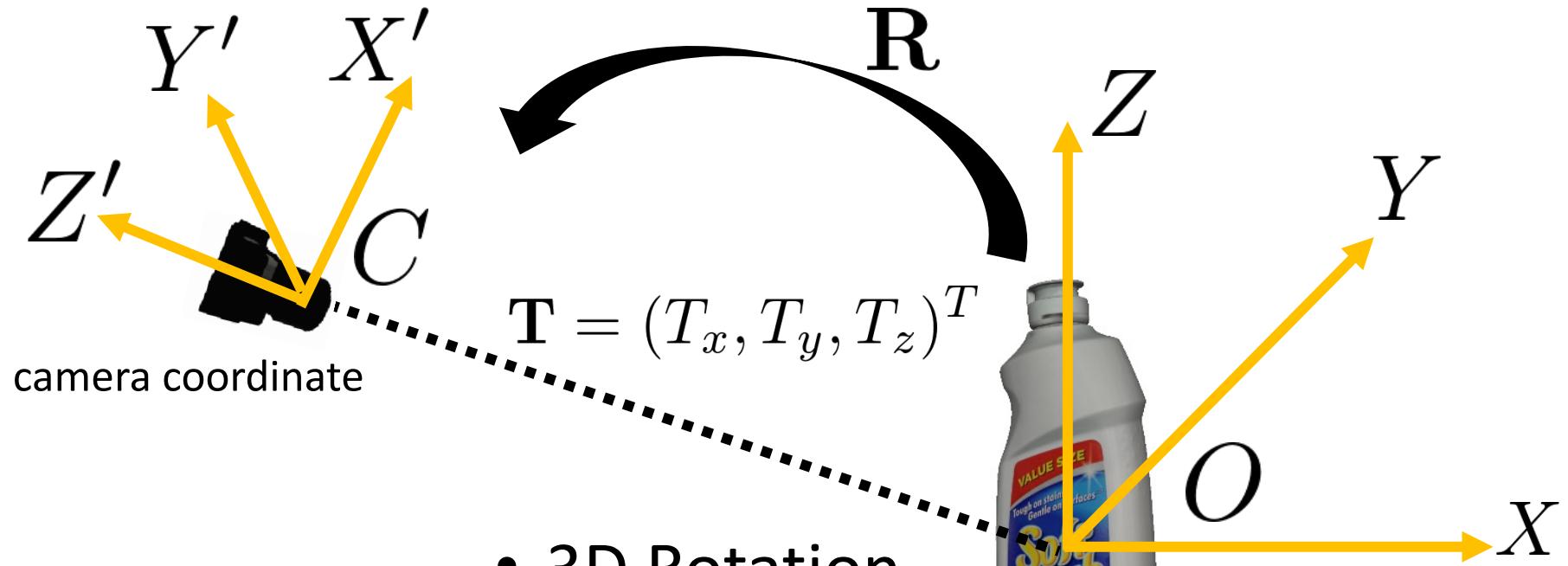
Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR'10.

PoseCNN



Y. Xiang, T. Schmidt, V. Narayanan and D. Fox. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. In RSS'18.

PoseCNN: Decouple 3D Translation and 3D Rotation



- 3D Translation



2D center

$$\mathbf{c} = (c_x, c_y)^T$$

Distance T_z

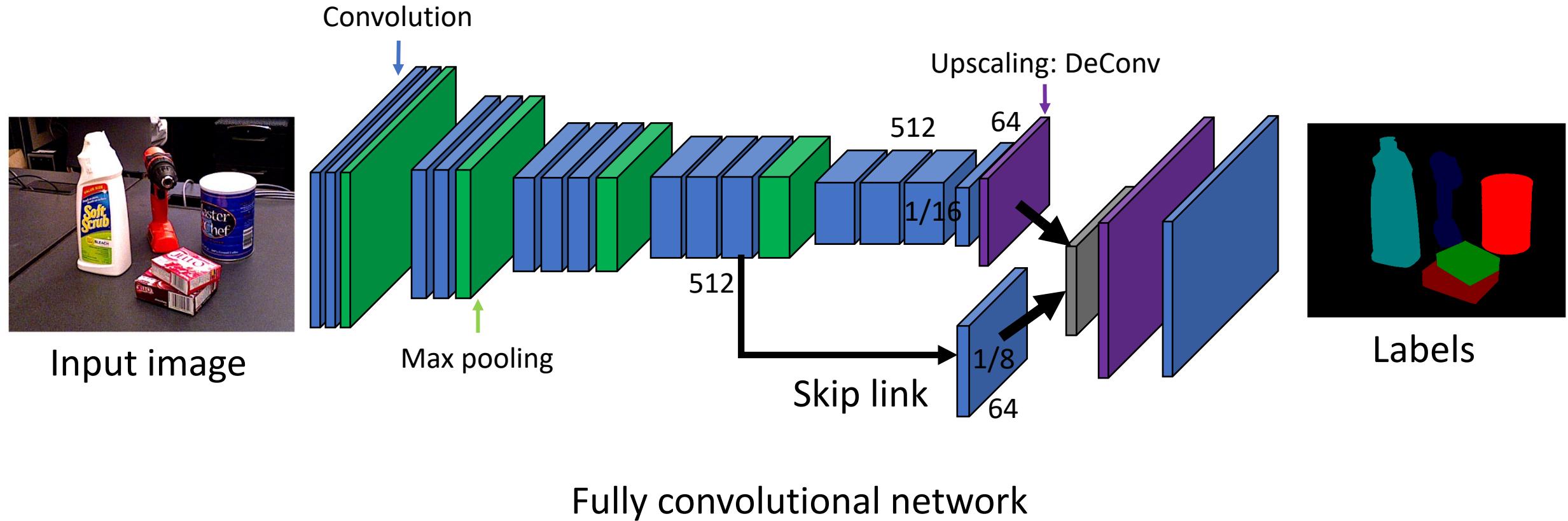
2D Center Localization

- 3D Rotation



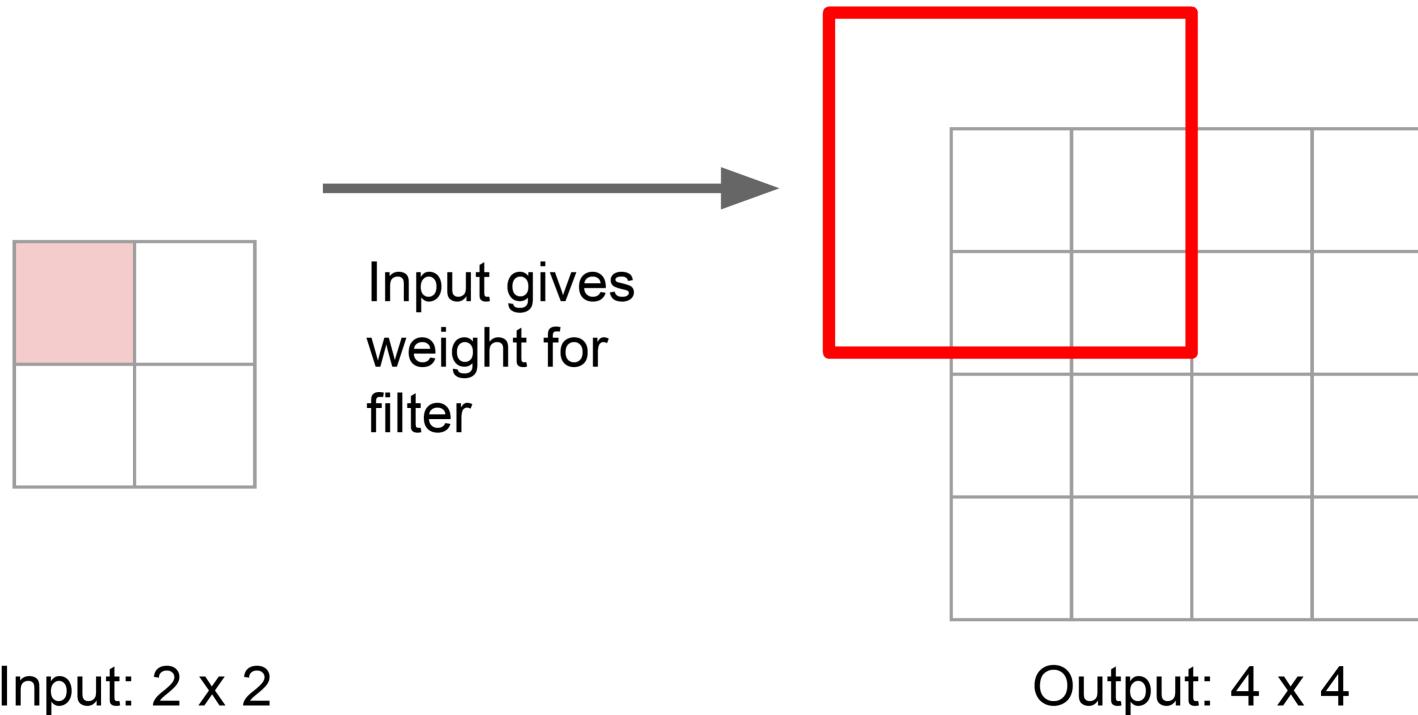
3D Rotation Regression

PoseCNN: Semantic Labeling



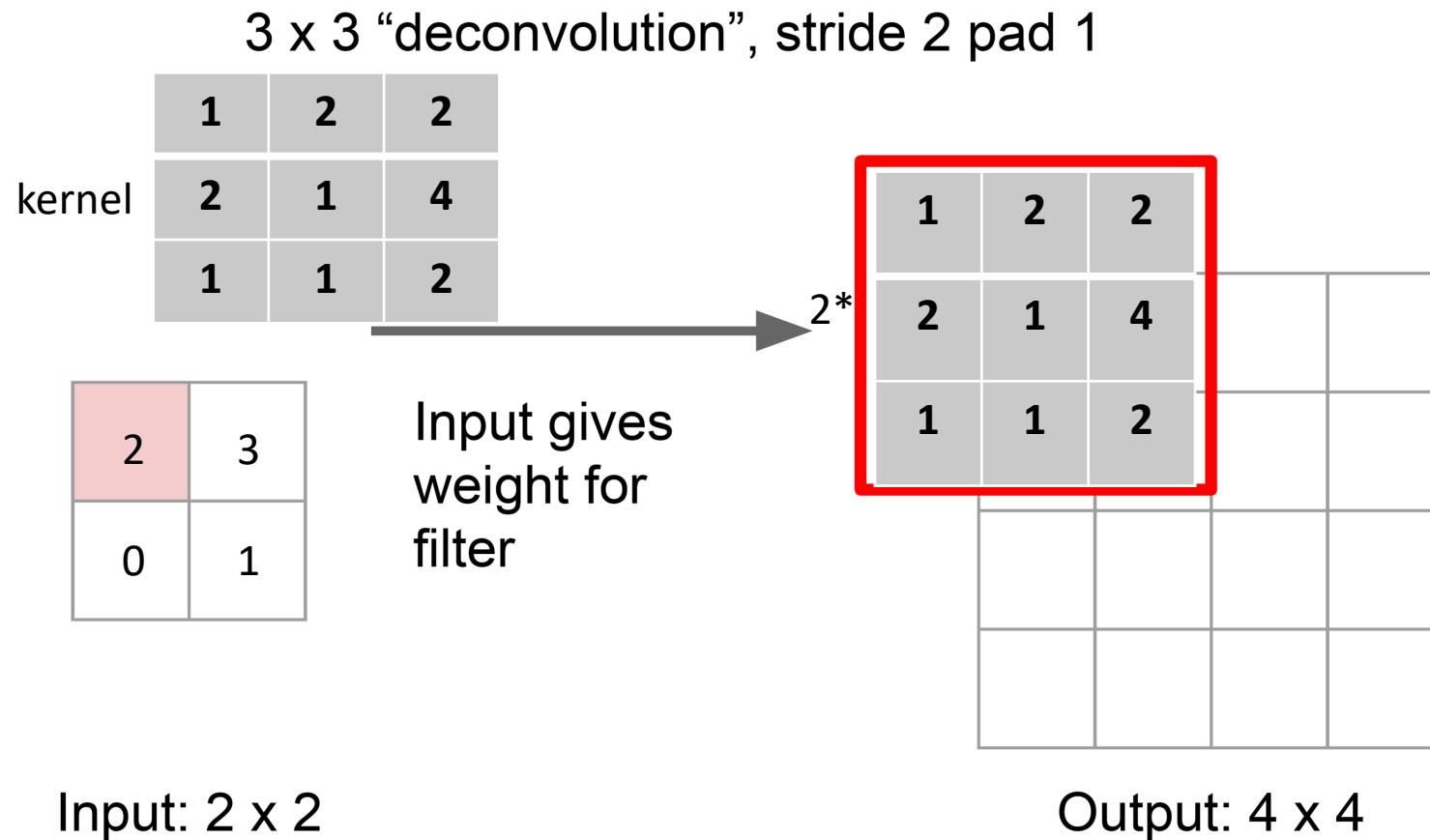
Deconvolution

3 x 3 “deconvolution”, stride 2 pad 1

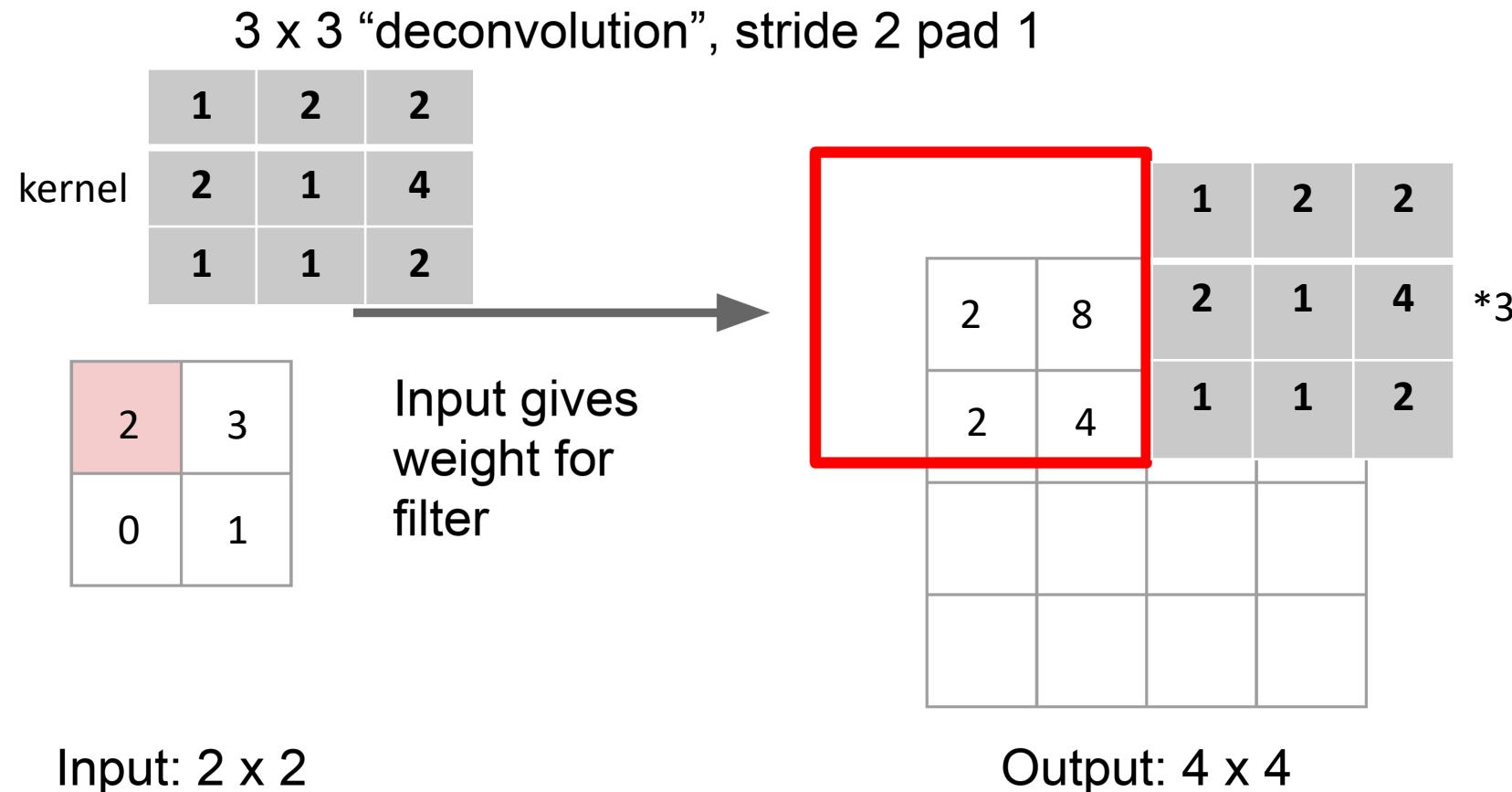


Credit: Andrej Karpathy & Justin Johnson

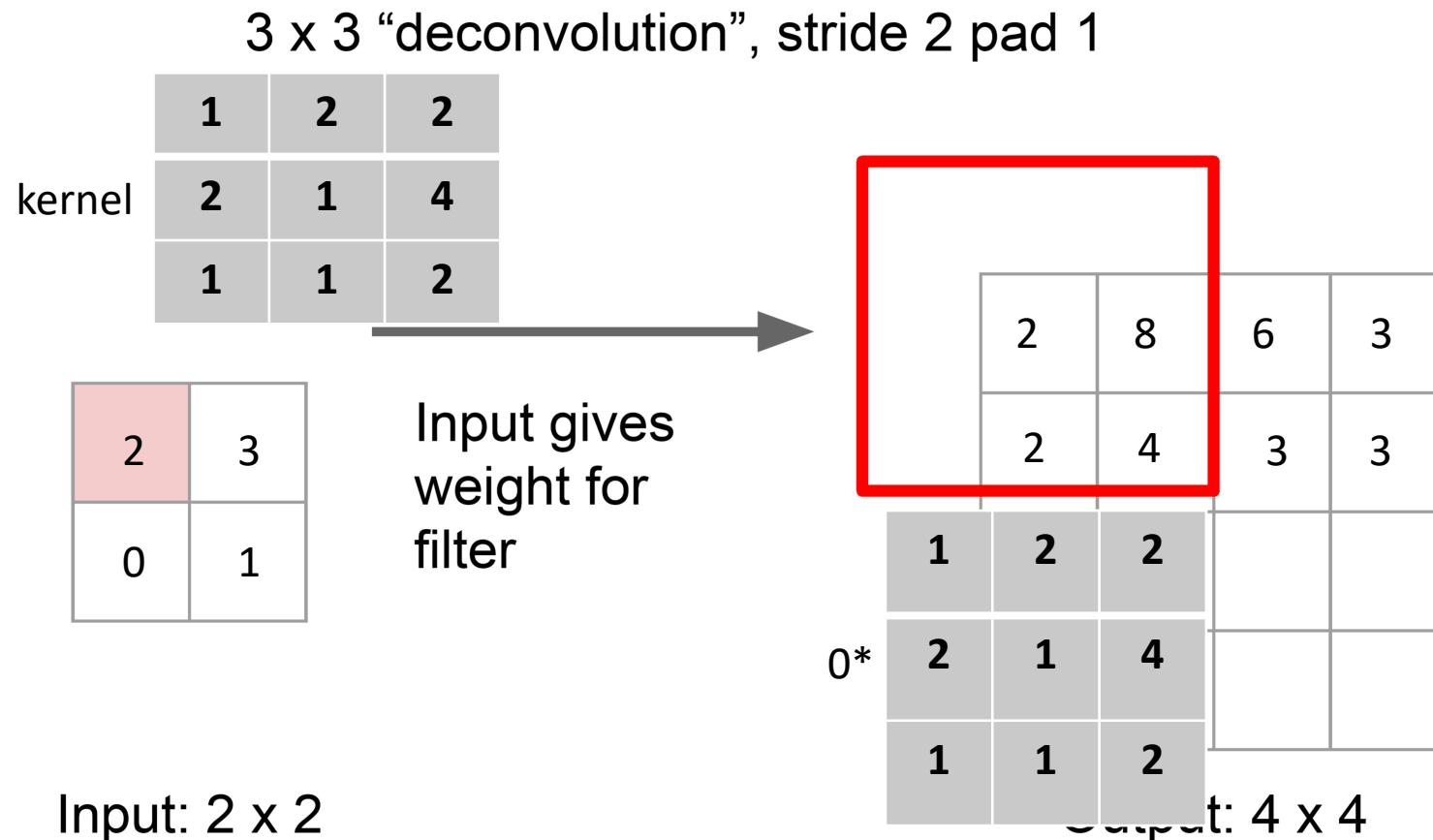
Deconvolution



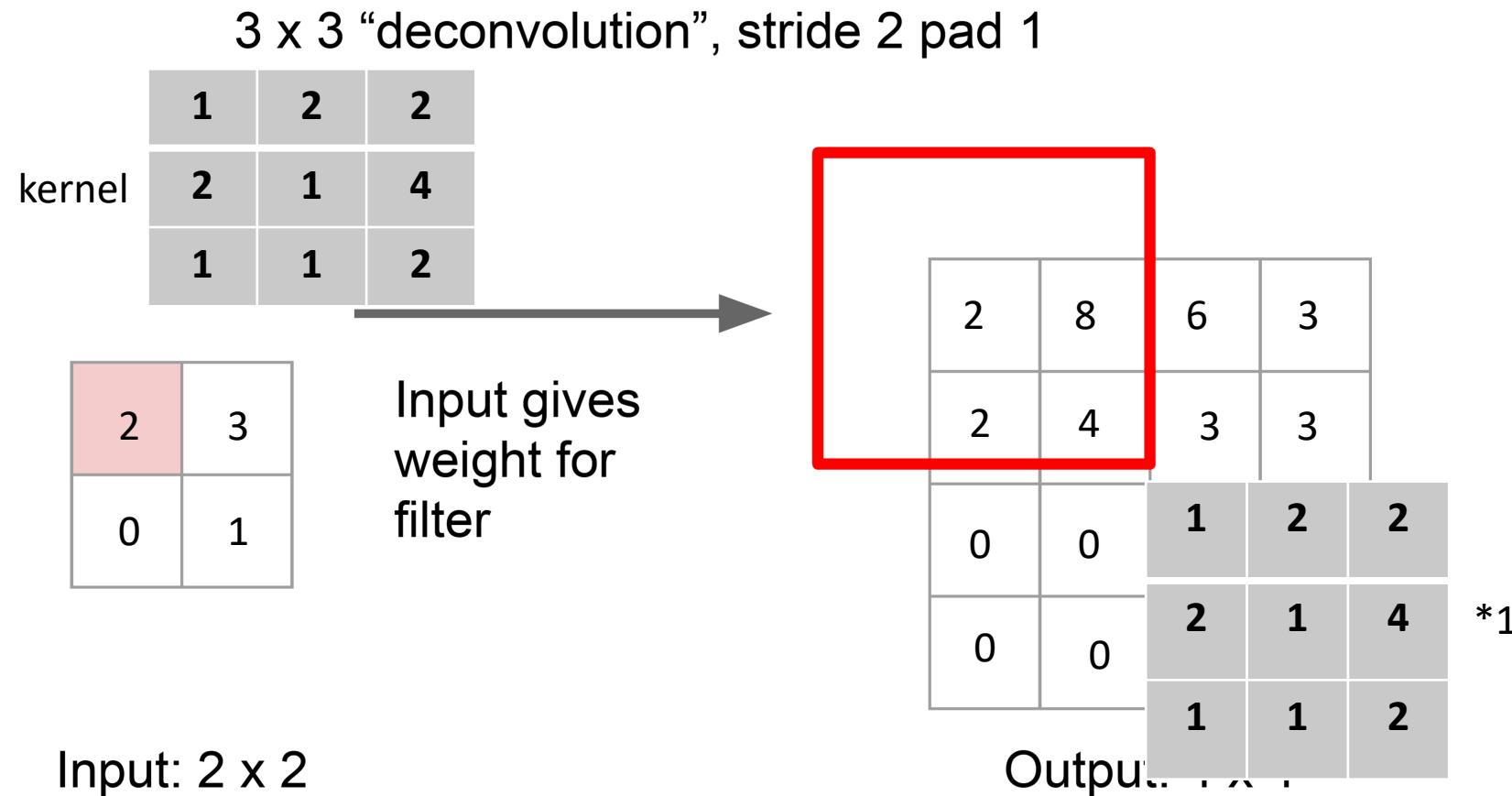
Deconvolution



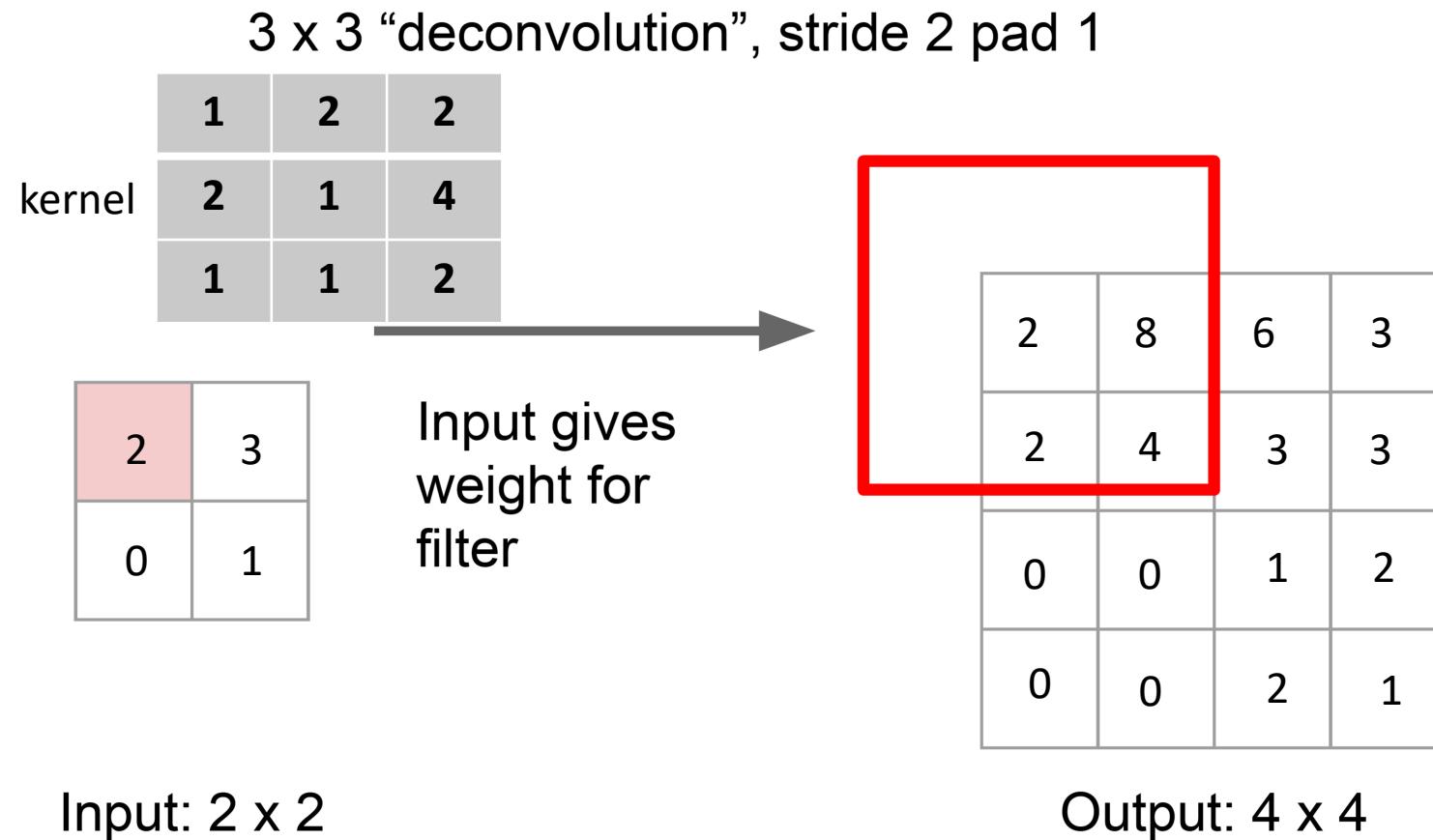
Deconvolution



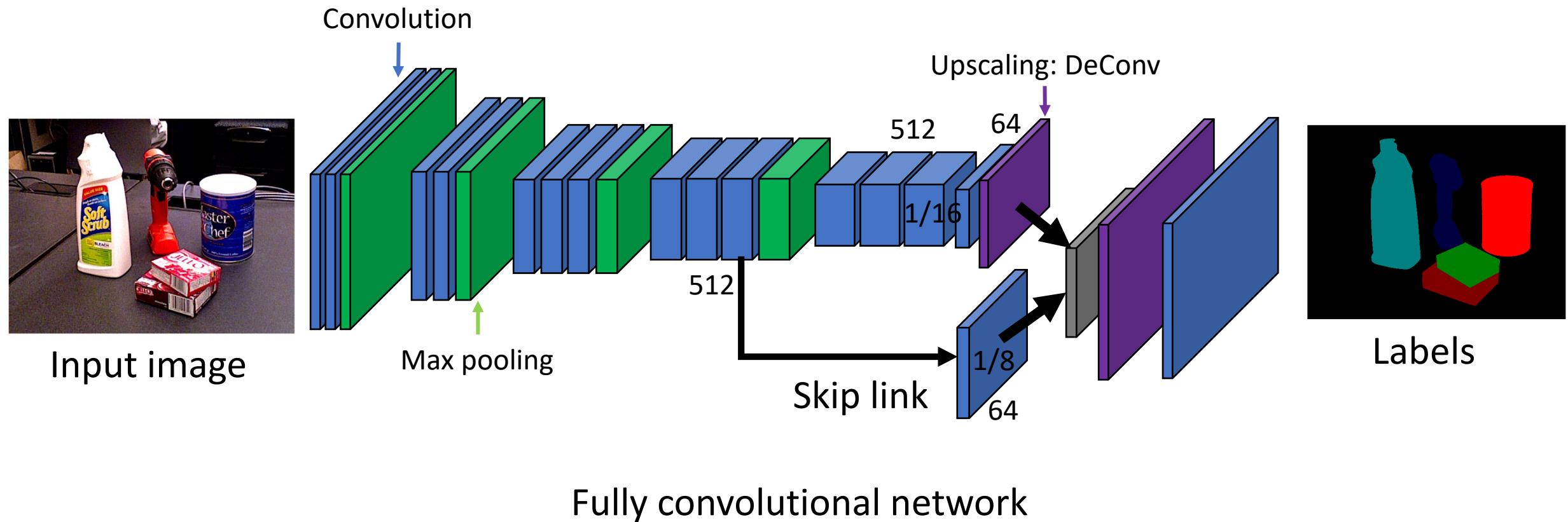
Deconvolution



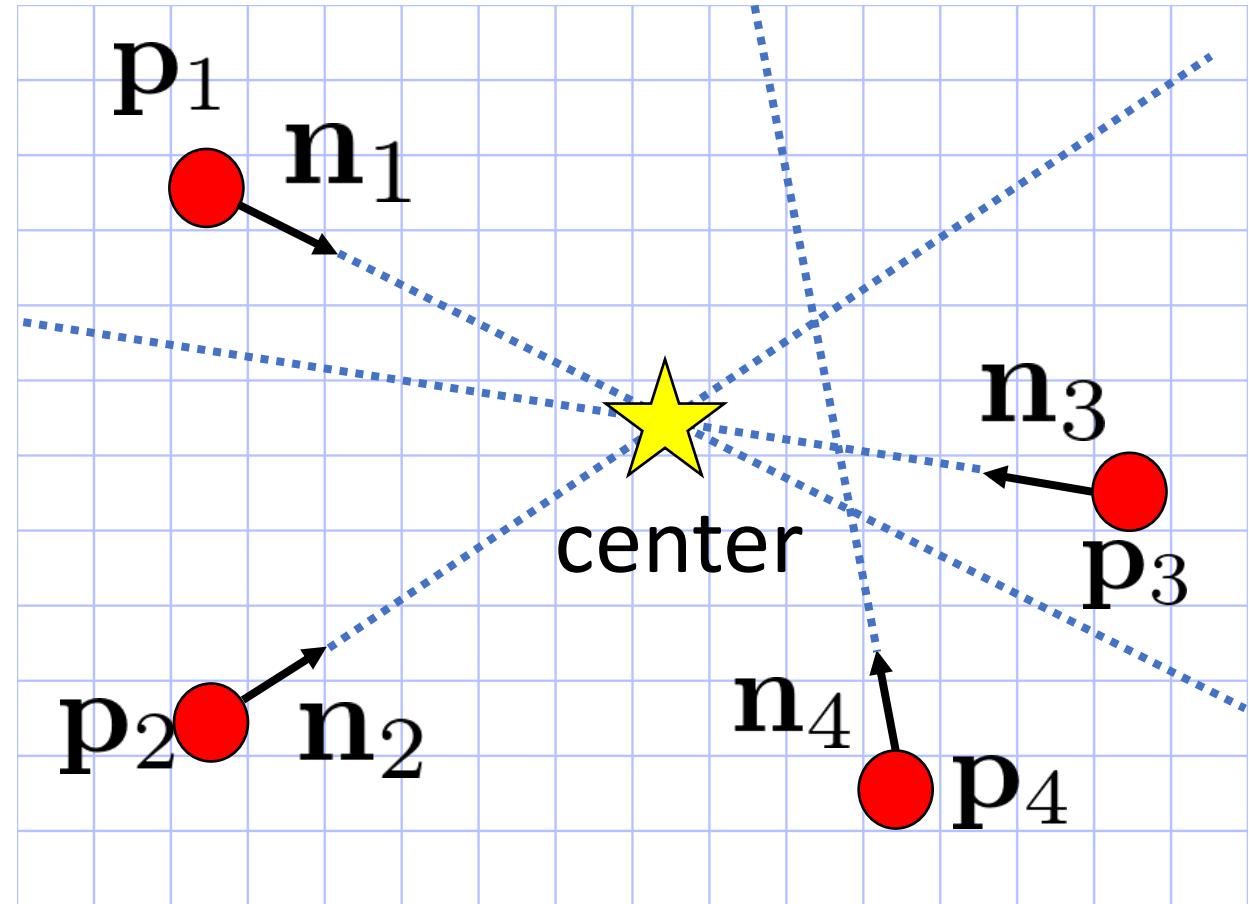
Deconvolution



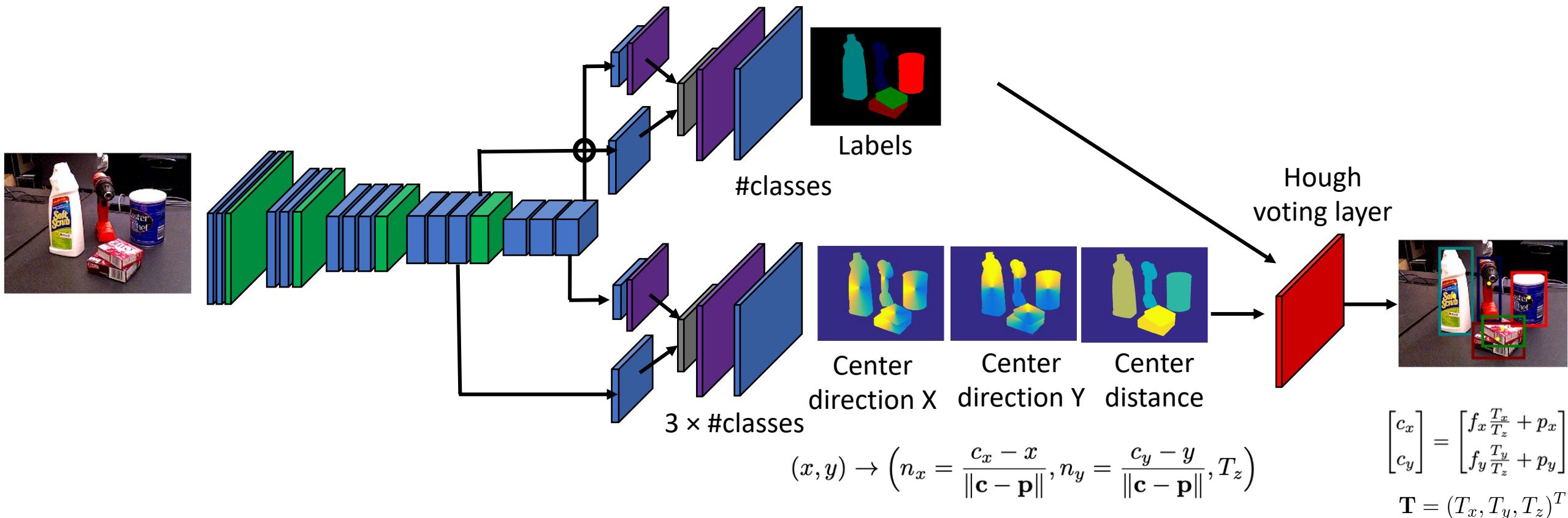
PoseCNN: Semantic Labeling



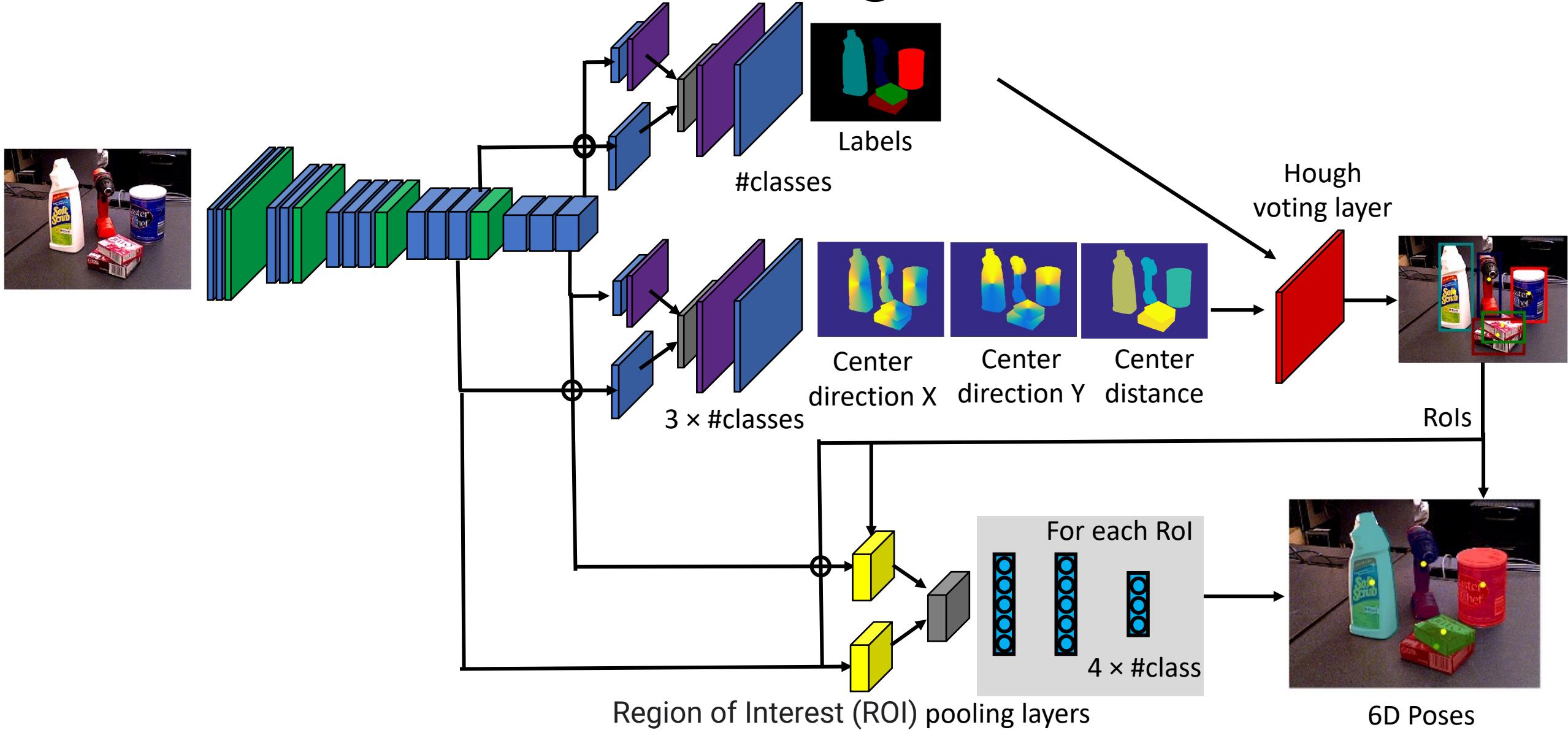
PoseCNN: 2D Center Voting for Handling Occlusions



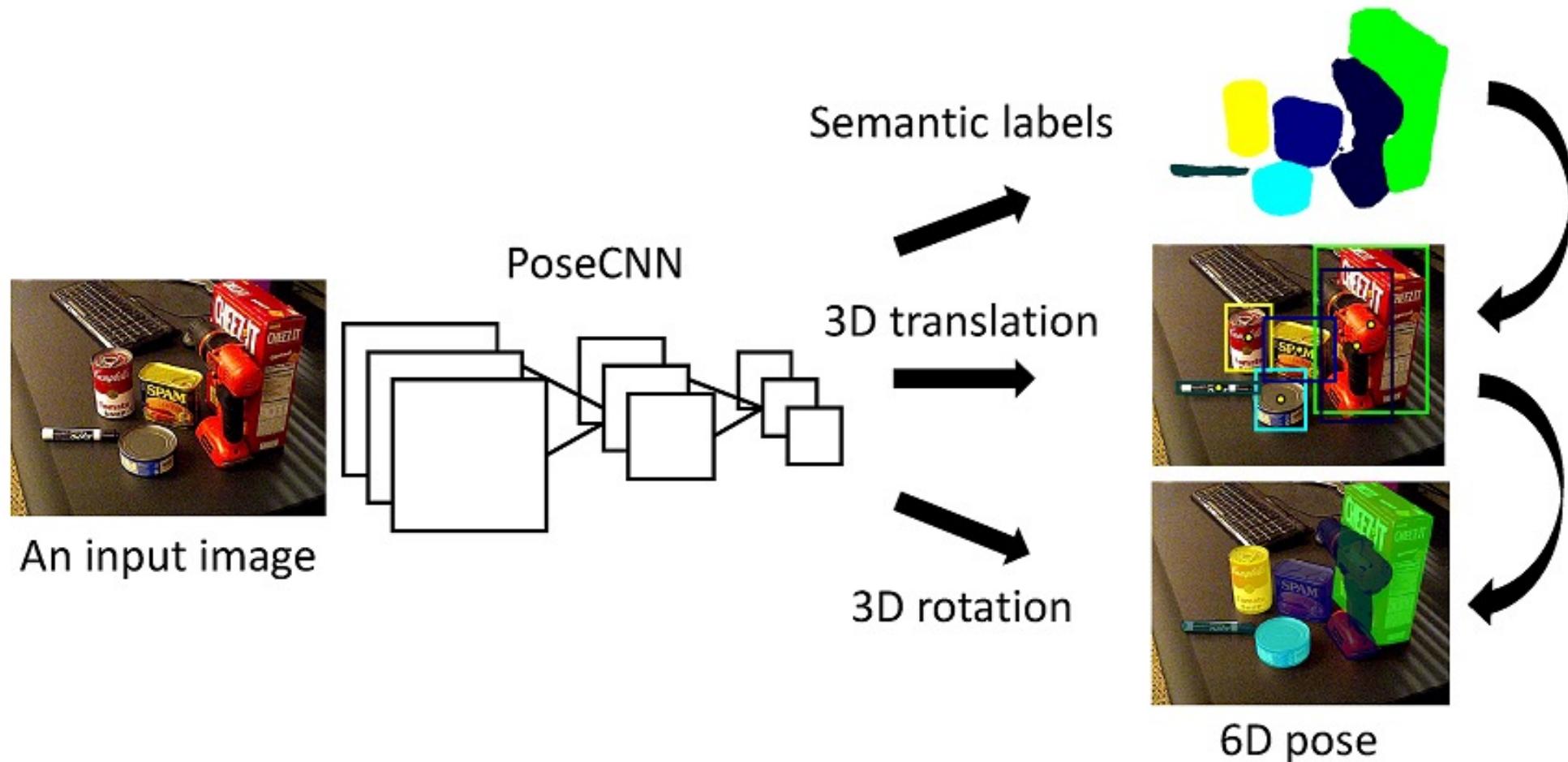
PoseCNN: 3D Translation Estimation



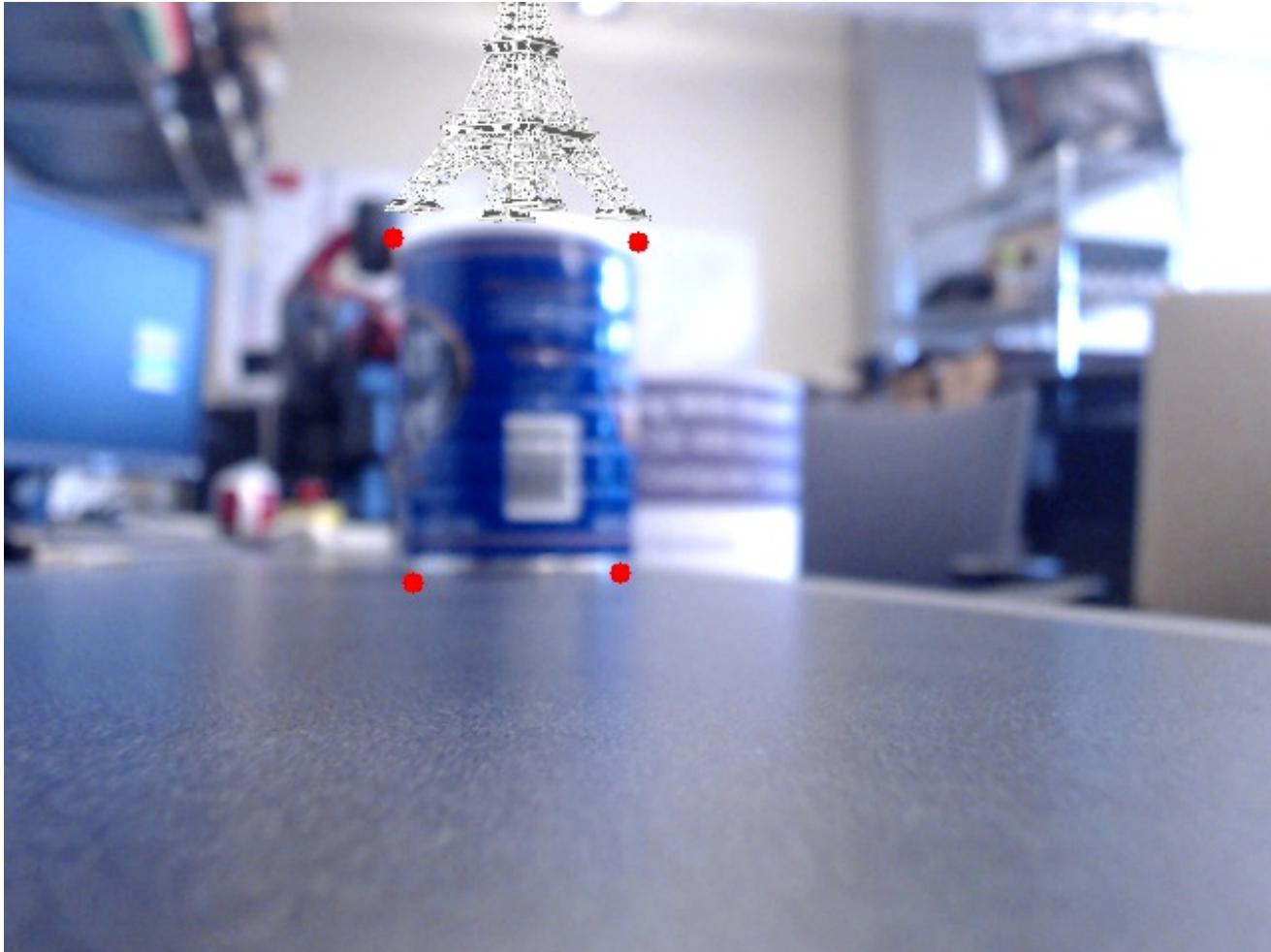
PoseCNN: 3D Rotation Regression



PoseCNN



AR Demo with 6D Pose Estimation



DeepIM, Li et al., IJCV'19



Credit: Lirui Wang

Further Reading

- Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR'10.
- PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. Xiang et al., RSS'18.