PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation

Yu Xiang\textsuperscript{1}, Tanner Schmidt\textsuperscript{2}, Venkatraman Narayanan\textsuperscript{3} and Dieter Fox\textsuperscript{1,2}
\textsuperscript{1}NVIDIA Research, \textsuperscript{2}University of Washington, \textsuperscript{3}Carnegie Mellon University
6D Object Pose Estimation for Robotic Manipulation

- Camera
- Input image
- 3D location
- 3D orientation
- Known 3D model

Challenge:
- Texture-less objects
- Symmetric objects
- Occlusions
Our Contribution: A Generic Convolutional Neural Network for 6D Object Pose Estimation

PoseCNN

✓ Texture-less objects
✓ Symmetric objects
✓ Occlusions
PoseCNN: Decouple 3D Translation and 3D Rotation

• 3D Translation \( \mathbf{T} = (T_x, T_y, T_z)^T \)

2D center
\[ \mathbf{c} = (c_x, c_y)^T \]
Distance \( T_z \)

2D Center Localization

• 3D Rotation \( \mathbf{R} \)

3D Rotation Regression
PoseCNN: Semantic Labeling

Input image

Fully convolutional network

Labels

- Long et al., CVPR, 2015
- Xiang & Fox, RSS, 2017
PoseCNN: 2D Center Voting for Handling Occlusions
PoseCNN: 3D Translation Estimation

- Labels
- Center direction X
- Center direction Y
- Center distance

Hough voting layer

3 × #classes

#classes
PoseCNN: 3D Rotation Regression

- Labels
  - #classes
  - Center direction X
  - Center direction Y
  - Center distance

- Hough voting layer

- RoIs
  - RoI pooling layers
  - 4 × #class

- For each RoI

- 6D Poses
PoseCNN: 3D Rotation Regression Loss Functions

Pose Loss (non-symmetric)

\[ P\text{Loss}(\tilde{q}, q) = \frac{1}{2m} \sum_{x \in M} \| R(\tilde{q})x - R(q)x \|^2 \]

Shape-Match Loss for symmetric objects (symmetric)

\[ S\text{Loss}(\tilde{q}, q) = \frac{1}{2m} \sum_{x_1 \in M} \min_{x_2 \in M} \| R(\tilde{q})x_1 - R(q)x_2 \|^2 \]
Our YCB-Video Dataset

21 YCB Objects

92 Videos, 133,827 frames
Results on the YCB-Video Dataset

3D Translation Error Analysis

3D Rotation Error Analysis
Conclusion

- **PoseCNN**
  - An end-to-end neural network for 6D pose estimation
  - Handle texture-less objects, symmetric objects and occlusions
  - Code and dataset are available online

Thank you!