Introduction 000000000	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ⊙	End ○

Object Pose Estimation

Tom Sanitz

16. November 2013

Introduction • 00000000	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ○	End o
Table of (Contents				

Introduction

- Related topics
- 6D pose estimation
- Synthetic dataset generation
- Object Pose Estimation (DOPE) [7]
- Oatasets and Evaluation
- Other approaches
- 6 Live demo

Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End o
Related t	copics				

2D image domain

- O Classification
 - Many advances in last 10 years (CNNs / Transformers)
 - Important for many downstream tasks
- **2** Semantic Segmentation
 - Similar to image classification
 - Per pixel classification



Tom Sanitz

Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End o
Related	topics				

2D Multi Object

- Object Detection
 - 2D localisation of multiple objects in (rotated) bounding boxes + class id
 - Examples: YOLO [5], Faster R-CNN [6], ...
- **2** Instance Segmentation
 - Hybrid between Object Detection and Semantic Segmentation
 - Example: Mask R-CNN [3]



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ○	End O
Related 1	topics				

3 dimensional domain

Object Detection

• Predicts 3d center of the box, width, height and length (+ vertical rotation)



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ○	End ○
6D Pose E	stimation				



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End O
6D Pose E	stimation				

What do we want to achieve

- Prediction of 3D position + 3D rotation
- But 3D position of what? Relative to where?



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End O
6D Pose I	Estimation				

Defining a coordinate system

- 3D Representation of the object (e.g pointcloud)
- Used to define object intrinsic pose



Introduction 0000000●0	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End ○
6D Pose E	stimation				

- Representation of object pose in camera frame
- Finding transformation of objects frame to camera frame, see robotics lecture



Introduction DOPE Datasets and Evaluation Other approaches Live demo End

Synthetic Dataset generation

Synthetic Datasets

Allows the creation of massive amounts of training data at minimal cost.

- When more data is required
- When real data is to expensive to annotate

Problems with synthetic data

Potentially a bad representation of real data: Reality gap

- Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects [7]
- Published at "Conference on Robot Learning (CoRL) 2018" by Tremblay et al.(NVIDIA)
- Simple network architecture
- Trained only on synthetic data
- Achieved state-of-the-art performance

Introduction	DOPE 0●00	Datasets and Evaluation	Other approaches	Live demo ○	End O

Synthetic Dataset generation

Reality gap solutions in DOPE

- Photo realistic renders in UE 4
- Domain randomization

photorealistic



Introduction	DOPE ○○●○	Datasets and Evaluation	Other approaches	Live demo ⊙	End O

Synthetic Dataset generation

Domain Randomization

- Object of interest in front of random backgrounds
- Adding various distractors
- randomize and overlay textures, lighting, poses and noise
- Force the network to learn general features

domain randomized



Introduction	DOPE 000●	Datasets and Evaluation	Other approaches	Live demo ○	End ○
Network ar	chitecture				

DOPE [7] - Fully convolutional Network architecture

- Predict 9 belief maps of 2D keypoints per objects
- 2 Learn 8 vector fields for vertices-centroid assignment
- **(**) Use perspective-*n*-point (PnP) on the peaks to estimate 6D Pose



Introduction	DOPE	Datasets and Evaluation	Other approaches	Live demo	End
000000000	0000	●0000		O	○
Common	Ohiects				

YCB Objects

- Common household objects
- Meant to standardise evaluation methods between research
- Different materials and shapes
- Also used in the YCB-Video dataset [8]



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ○	End ○
Evaluation	Dataset				

YCB Video Dataset [8]

- Dataset including accurate 6D poses for YCB objects
- Originally 21 YCB objects included
- 92 Videos with 133.827 frames



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo ⊙	End ○
Evaluation I	Metrics				

Average Distance (ADD)

- Average 3D Euclidean distance of all model points
- ADD pass rate = Percentage of predictions(p) with ADD value <= threshold (t)



Introduction	DOPE 0000	Datasets and Evaluation ○○○●○	Other approaches	Live demo ○	End ○
Evaluation I	Metrics				

• Note the difference between synthetic and domain randomization



Introduction	DOPE	Datasets and Evaluation	Other approaches	Live demo	End
000000000	0000	0000●		○	O
Evaluation N	/letrics				

- Why does DOPE perform worse than PoseCNN [8] here?
- Answer: Reality gap for metallic objects



Introduction	DOPE 0000	Datasets and Evaluation	Other approaches ●○	Live demo ○	End ○
	and Calut	and			

- RGB-D Based Solutions
 - What if we include depth information?
 - PVN3D: A Deep Point-Wise 3D Keypoints Voting Network [4]



Introducti	ion DOP	Datasets and Evaluation 0 00000	Other approaches ○●	Live demo ○	En O
And	many more				
	 EfficientPos RNN Pose ROPE [2] 	se [1] [9]			
=	Google Scholar	6 dof pose estimation	٩		
+	Artikel	Ungefähr 16.800 Ergebnisse (0,12 Sek.)			
	Beliebige Zeit Seit 2023 Seit 2022 Seit 2019 Zeitraum wählen	Pvn3d: A deep point-wise 3d keypoints voting netwing Y He, W Sun, H Huang, J Liu Proceedings of the, 2020 - ope 60oF object pose estimation from a single RGBD image. Unlik directly regressing pose of objects and then estimate the 6D pr ☆ Speichern 99 Zitieren Zitiert von: 361 Åhnliche Artikel Alle	ork for 6dof pose estimation enaccess.thecvf.com e previous methods that ose parameters within a least e 10 Versionen ⊗⊳	[PDF] thecvf.com	
	2020 — Suche	A hybrid approach for 6DoF pose estimation R König. <u>B Drost</u> - European Conference on Computer Vision, 2021 We propose a method for 6DoF pose estimation of rigid object deep learning based instance detector to segment object instances	0 - Springer s that uses a state-of-the-art in an RGB image,	[PDF] arxiv.org	
	sortieren Nach Datum sortieren	6dof pose estimation of transparent object from a CXu J Chen M Yao J Zhou L Zhang Y Liu - Sensors 2020 - md	single rgb-d image	[PDF] mdpi.com	
	Т	om Sanitz	Object Pose Estin	mation	2

Introduction 000000000	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo •	End o
DOPF Dem	າດ				

- Showcasing live DOPE demo
- Inference on the YCB tomato soup object
- Using rgb only on the intel realsense D435 camera

Introduction 000000000	DOPE 0000	Datasets and Evaluation	Other approaches	Live demo O	End ●

Thank you for your attention!

References

- Yannick Bukschat and Marcus Vetter. "EfficientPose: An efficient, accurate and scalable end-to-end 6D multi object pose estimation approach". In: arXiv preprint arXiv:2011.04307 (2020).
- [2] Bo Chen, Tat-Jun Chin, and Marius Klimavicius. "Occlusion-robust object pose estimation with holistic representation". In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
- [3] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn". In: Proceedings of the IEEE international conference on computer vision. 2017.
- [4] Yisheng He, Wei Sun, Haibin Huang, Jianran Liu, Haoqiang Fan, and Jian Sun. "Pvn3d: A deep point-wise 3d keypoints voting network for 6dof pose estimation". In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [6] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: Advances in neural information processing systems (2015).
- [7] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. "Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects". In: <u>Conference on Robot Learning (CoRL)</u>. 2018. URL: https://arxiv.org/abs/1809.10790.
- [8] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes". In: arXiv preprint arXiv:1711.00199 (2017).
- [9] Yan Xu, Kwan-Yee Lin, Guofeng Zhang, Xiaogang Wang, and Hongsheng Li. "Rnnpose: Recurrent 6-dof object pose refinement with robust correspondence field estimation and pose optimization". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

ADD vs. ADD-S

ADD

- R: Ground truth rotation
- T: Ground truth translation
- \tilde{T} and \tilde{R} the predicted values
- M are the set of 3D model points, m the number of points
- Compute the mean of pairwise distances

$$ADD = \frac{1}{m} \sum_{\mathbf{x} \in \mathcal{M}} \| (\mathbf{R}\mathbf{x} + \mathbf{T}) - (\tilde{\mathbf{R}}\mathbf{x} + \tilde{\mathbf{T}}) \|$$

ADD vs. ADD-S

ADD-S

- For symmetric objects, matching can be ambiguous
- Solution in DOPE: closest point distance

$$ADD = \frac{1}{m} \sum_{\mathbf{x} \in \mathcal{M}} \| (\mathbf{R}\mathbf{x} + \mathbf{T}) - (\tilde{\mathbf{R}}\mathbf{x} + \tilde{\mathbf{T}}) \|$$