

# Objektposenschätzung

Abschlussvortrag zum Proseminar (Bachelor)

Bearbeiter:

Robert Jeutter

Betreuer:

M. Sc. Benedict Stephan

Betreuernder Hochschullehrer:

Prof. H.-M. Groß

Datum:

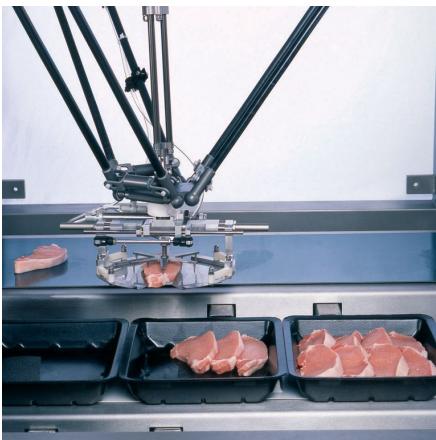
25.01.2022

# Gliederung

- 1. Motivation**
- 2. Probleme**
- 3. Vergleichbarkeit**
- 4. Übersicht unterschiedlicher Verfahren**
- 5. Teaser**
- 6. BundleTrack**
- 7. Vergleich & Fazit**

# Motivation

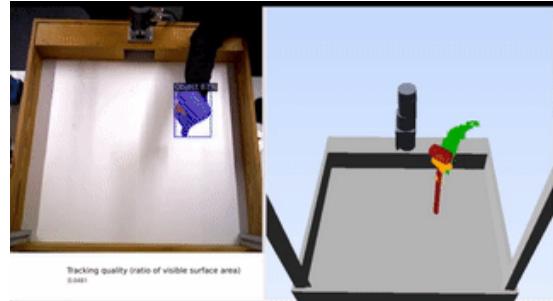
- Objekte erkennen, sortieren und zuordnen
- Vergleich mit Computer-Modellen
- Interaktion mit Umwelt



[Pick&Place]



[Metra]



[ObjectTracking]



[Care-o-bot]

# Probleme

- viele einzelne Objekte
- schlecht/nicht texturierte Objekte
- Verdeckungen
- unbekannte Objekte

# Gliederung

1. Motivation
2. Probleme
3. Vergleichbarkeit
4. Übersicht unterschiedlicher Verfahren
5. Teaser
6. BundleTrack
7. Vergleich & Fazit

# Verfahren vergleichbar machen

**Modell:** 2D/3D, CAD, Schablone, Kategorie, ohne

**Input:** RGB (2D), RGB-D (3D), Punktwolken (3D)

**Datensätze:** NOCS, YCB, LineMOD, LineMOD-Occlusion, ShapeNet

**Genauigkeit**

**Ressourcen**

**Laufzeit**



YCB <https://www.researchgate.net/profile/Jens-Lundell-2/publication/336147823>

# Datensatz: Bildstrom vs. Standbild

## Bildstrom

- Variabler, meist geringerer Rechenaufwand
- Bei korrekter Instanziierung meist bessere/robustere Ergebnisse
- Optimierung durch vorherige Ergebnisse; Abgleich von Objekten und Hintergrund
- Fehlerfortpflanzung möglich
- Durch FPS variabel einstellbar

## Standbild

- Konstanter höherer Rechenaufwand
- Ergebnisse unabhängig voneinander
- Keine Optimierung durch vorherige Ergebnisse
- Keine Fehlerfortpflanzung

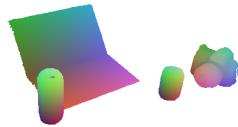
# Datensatz: Vergleich

	RGB	Tiefe	Maske	Modell	Koordinaten	Metadaten
NOCS REAL275	✓	✓	✓	✗	✓	✓
LineMOD	✓ <small>(registriert)</small>	✓	✓	✓ 3D	✗	✓
YCB-Video	✓	✗	✗	✓ 3D	✗	✓
3DMatch - Sun3D	✓	✓	✗	✗	✗	✓
3DMatch – Analysis by Synthesis	✗	✓	✗	✗	✗	✓

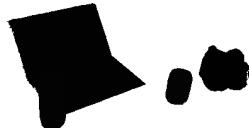
# Datensatz: NOCS REAL275



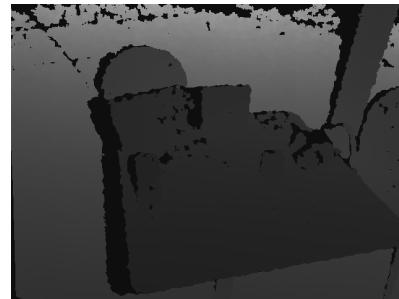
RGB-Bild



Koordinaten



Maske



Tiefenbild

# Berechenbare Genauigkeit

- **5°5cm** Prozentsatz der Schätzungen mit Orientierungsfehler<5° und Translationsfehler<5cm
- **IoU25** (Intersection over Union) Prozentsatz, in dem Überschneidung von Vorhersage und 3D Bounding Box größer ist als 25%
- **AUC** (Area Under Curve)

- **ADD**  $\frac{1}{m} \sum_{x \in M} \| (Rx + T) - (\bar{Rx} + \bar{T}) \|$

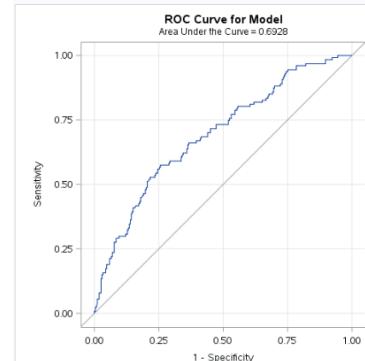
- **ADD-S**  $\frac{1}{m} \sum_{x_1 \in M} \min_{x_2 \in M} \| (Rx_1 + T) - (\bar{Rx}_2 + \bar{T}) \|$

R – reale Rotation

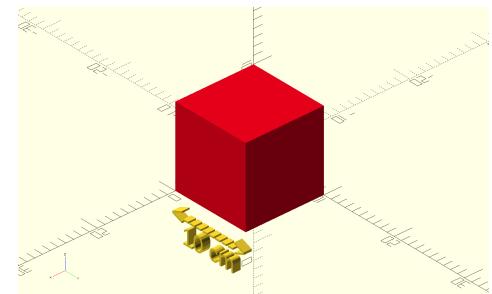
T – reale Translation

$\bar{R}$  – geschätzte Rotation

$\bar{T}$  – geschätzte Translation



[AUC Beispiel]



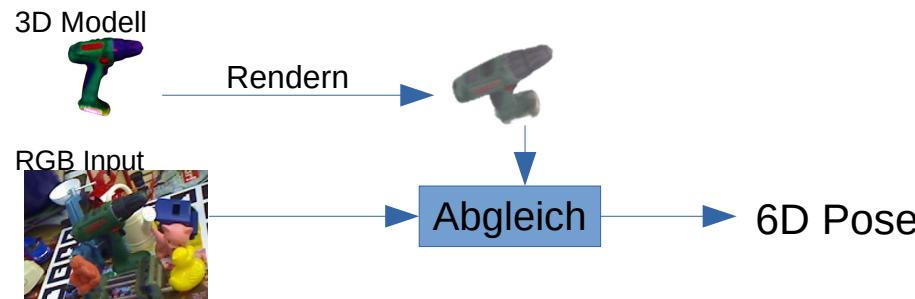
[5°5cm Beispiel]

# Gliederung

1. Motivation
2. Probleme
3. Vergleichbarkeit
4. Übersicht unterschiedlicher Verfahren
5. Teaser
6. BundleTrack
7. Vergleich & Fazit

# Übersicht unterschiedlicher Verfahren

3D CAD  
Modell

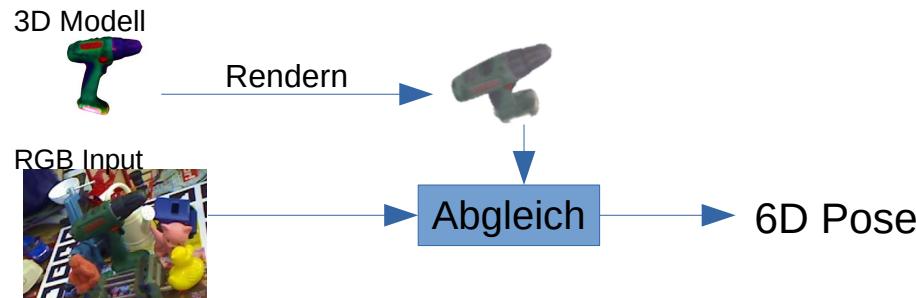


Kategorie  
Modell

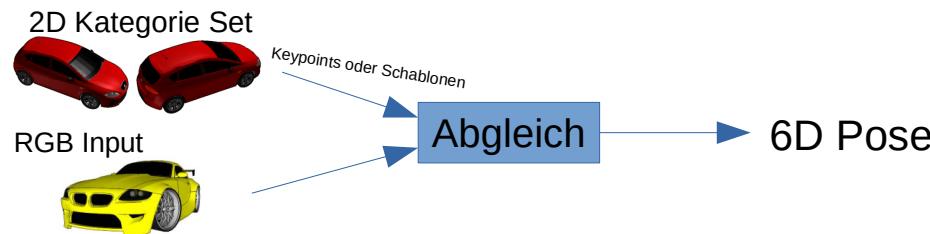
ohne  
Modell

# Übersicht unterschiedlicher Verfahren

3D CAD  
Modell



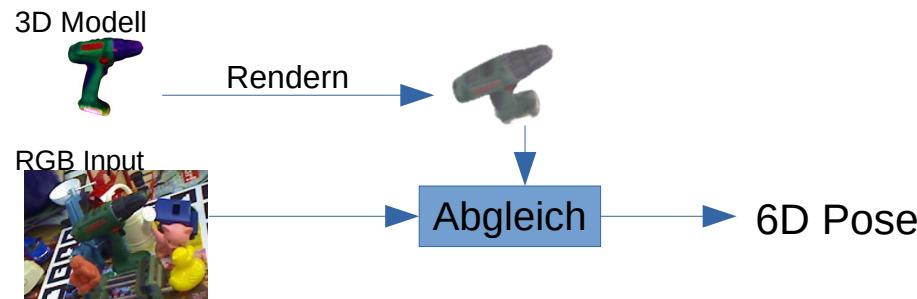
Kategorie  
Modell



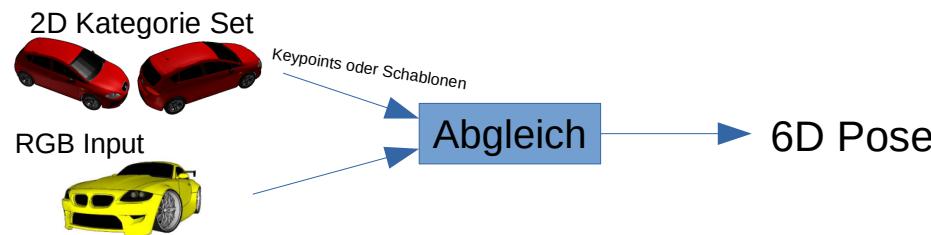
ohne  
Modell

# Übersicht unterschiedlicher Verfahren

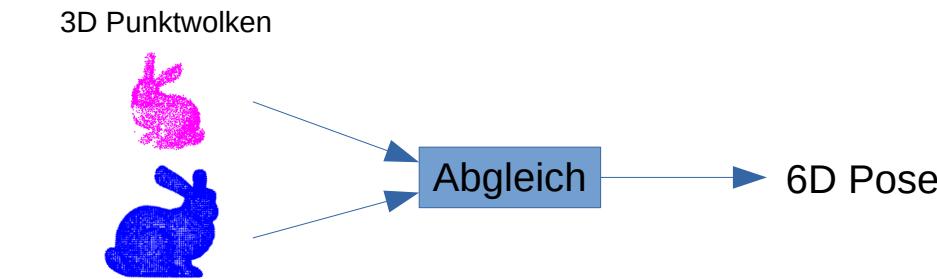
3D CAD  
Modell



Kategorie  
Modell

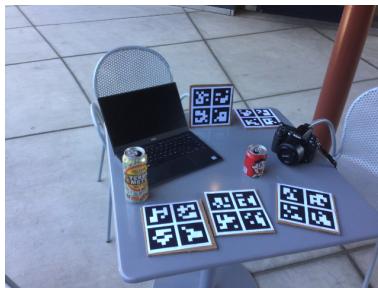


ohne  
Modell

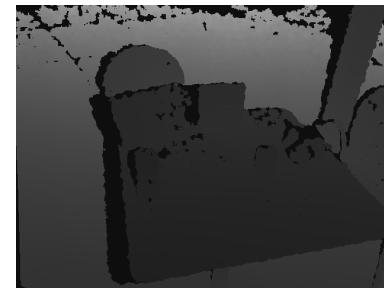


# Übersicht unterschiedlicher Verfahren

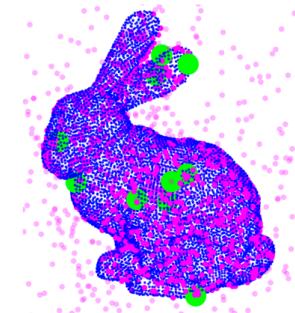
Farbbild-basiert



Tiefenbild



3D Pointcloud



3D CAD  
Modell

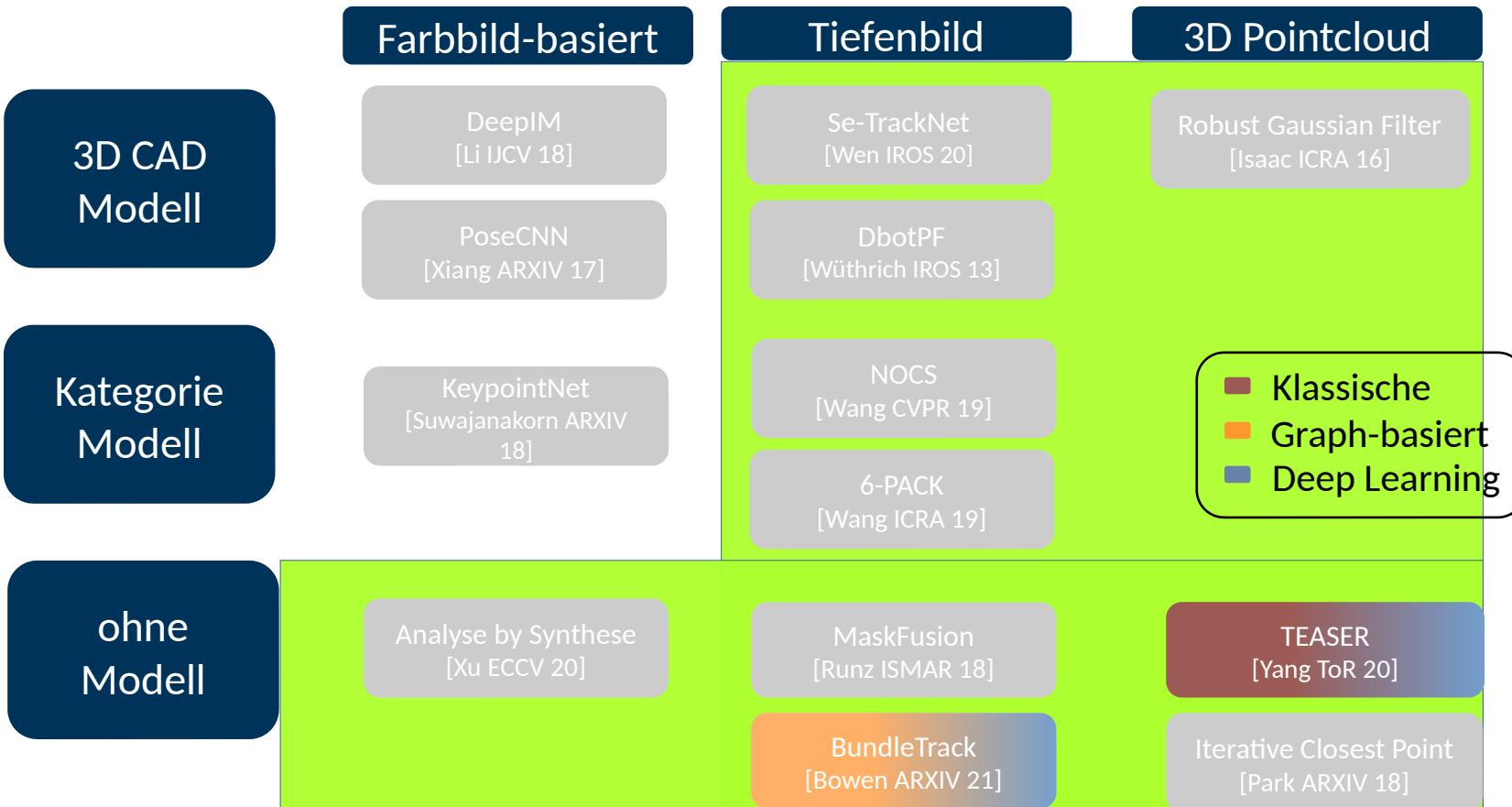
Kategorie  
Modell

ohne  
Modell

# Übersicht unterschiedlicher Verfahren

	Farbbild-basiert	Tiefenbild	3D Pointcloud
3D CAD Modell	DeepIM [Li IJCV 18]	Se-TrackNet [Wen IROS 20]	Robust Gaussian Filter [Isaac ICRA 16]
Kategorie Modell	PoseCNN [Xiang ARXIV 17]	DbotPF [Wüthrich IROS 13]	
ohne Modell	KeypointNet [Suwajanakorn ARXIV 18]	NOCS [Wang CVPR 19]	
	Analyse by Synthese [Xu ECCV 20]	6-PACK [Wang ICRA 19]	<ul style="list-style-type: none"><li>■ Klassische</li><li>■ Graph-basiert</li><li>■ Deep Learning</li></ul>
		MaskFusion [Runz ISMAR 18]	TEASER * [Yang ToR 20]
		BundleTrack [Bowen ARXIV 21]	Iterative Closest Point * [Park ARXIV 18]

# Übersicht unterschiedlicher Verfahren



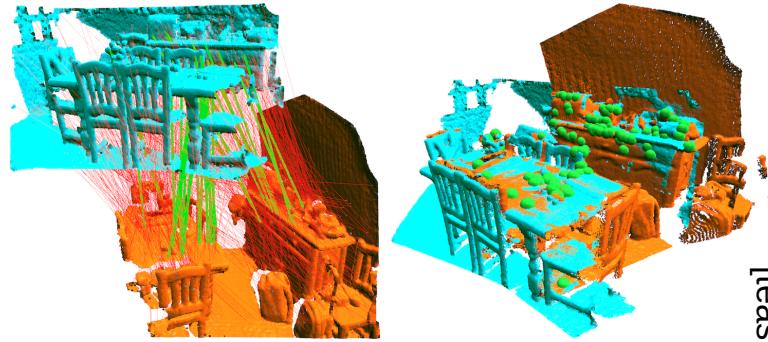
# Übersicht unterschiedlicher Verfahren

	Farbbild-basiert	Tiefenbild	3D Pointcloud
3D CAD Modell	DeepIM [Li IJCV 18]	Se-TrackNet [Wen IROS 20]	Robust Gaussian Filter [Isaac ICRA 16]
Kategorie Modell	PoseCNN [Xiang ARXIV 17]	DbotPF [Wüthrich IROS 13]	
ohne Modell	KeypointNet [Suwajanakorn ARXIV 18]	NOCS [Wang CVPR 19]	
	Analyse by Synthese [Xu ECCV 20]	6-PACK [Wang ICRA 19]	<ul style="list-style-type: none"><li>■ Klassische</li><li>■ Graph-basiert</li><li>■ Deep Learning</li></ul>
		MaskFusion [Runz ISMAR 18]	TEASER [Yang ToR 20]
		BundleTrack [Bowen ARXIV 21]	Iterative Closest Point [Park ARXIV 18]

# Teaser (Truncated Least Squares Estimation And Semidefinite Relaxation)

- **Modell:** ohne
- **Input:** RGB-D, Punktfolge
- **Datensatz:** Stanford 3D Scanning
- **Genauigkeit**
  - 20,0% 5°5cm
  - 99,9% IoU25
  - 57,91% ADD
  - 81,17% ADD-S
- **Ressourcen:** i7-8850H CPU mit 32GB RAM
- **Laufzeit:** 16,9 FPS (Teaser++)

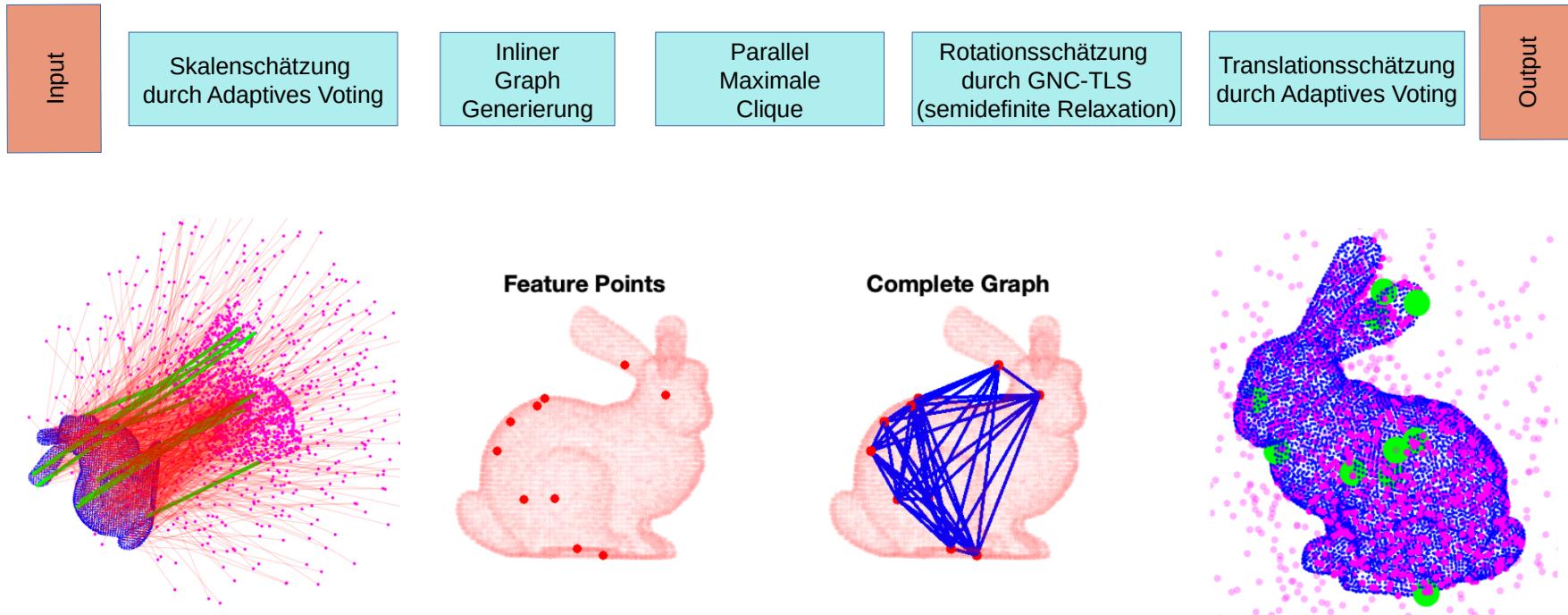
Original in Matlab implementiert,  
Teaser++ in C++ implementiert



`mples/python_3dsMOOTH$ python python_3dsMOOTH.py`

[teaser\_python\_3dsMOOTH]

# Teaser

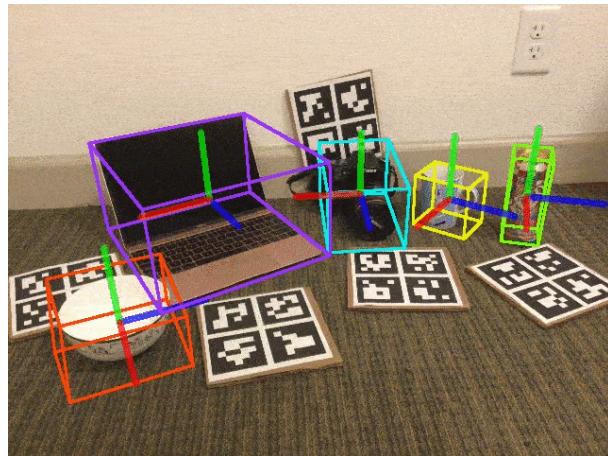


# Übersicht unterschiedlicher Verfahren

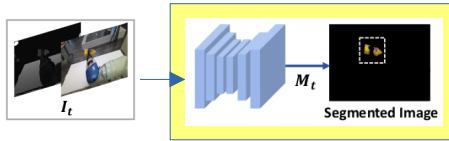
	Farbbild-basiert	Tiefenbild	3D Pointcloud
3D CAD Modell	DeepIM [Li IJCV 18]	Se-TrackNet [Wen IROS 20]	Robust Gaussian Filter [Isaac ICRA 16]
Kategorie Modell	PoseCNN [Xiang ARXIV 17]	DbotPF [Wüthrich IROS 13]	
ohne Modell	KeypointNet [Suwajanakorn ARXIV 18]	NOCS [Wang CVPR 19]	<ul style="list-style-type: none"><li>■ Klassische</li><li>■ Graph-basiert</li><li>■ Deep Learning</li></ul>
	Analyse by Synthese [Xu ECCV 20]	6-PACK [Wang ICRA 19]	
		MaskFusion [Runz ISMAR 18]	
		BundleTrack [Bowen ARXIV 21]	
		TEASER [Yang ToR 20]	
		Iterative Closest Point [Park ARXIV 18]	

# BundleTrack

- **Modell:** ohne
- **Input:** RGB-D
- **Datensatz:** NOCS, YCB, Youtube-VOS
- **Genauigkeit**
  - 87,5% 5°5cm
  - 99,9% IoU25
  - 87,34% ADD
  - 92,54% ADD-S
- **Ressourcen:** 3Ghz CPU + 2080Ti GPU
- **Laufzeit:** 10 FPS

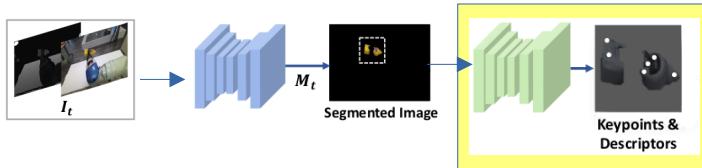


# BundleTrack



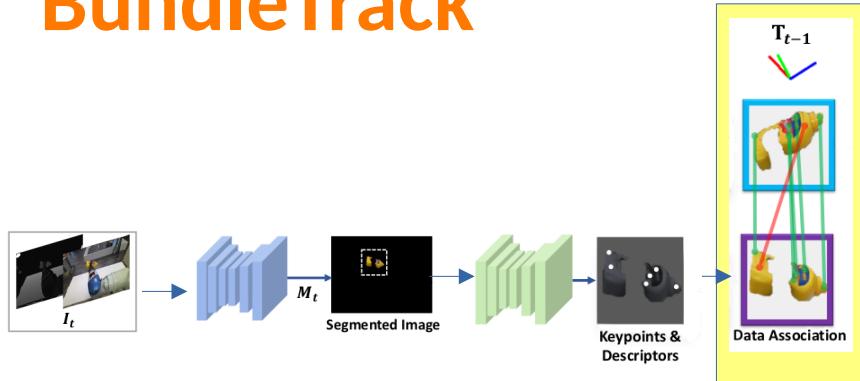
1) Bildsegmentierung mit VOS liefert Objektmaske

# BundleTrack



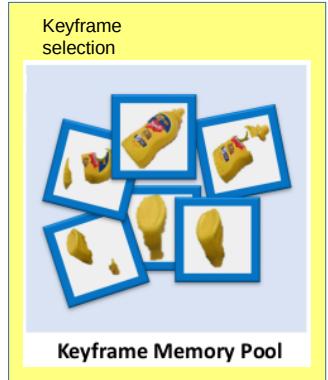
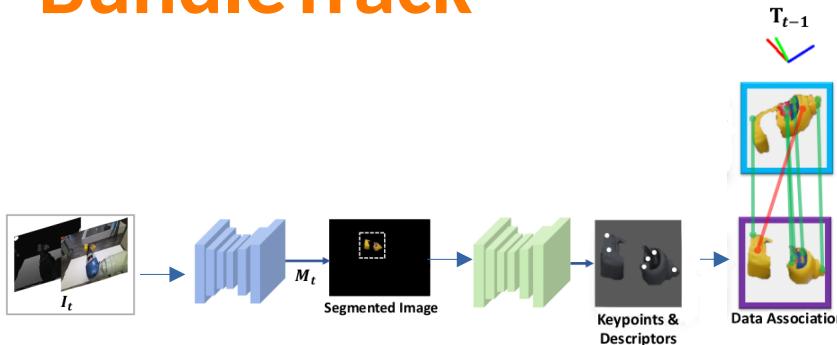
- 1) Bildsegmentierung mit VOS liefert Objektmaske
- 2) Keypoints und Deskriptoren mit LF-Net erkennen

# BundleTrack



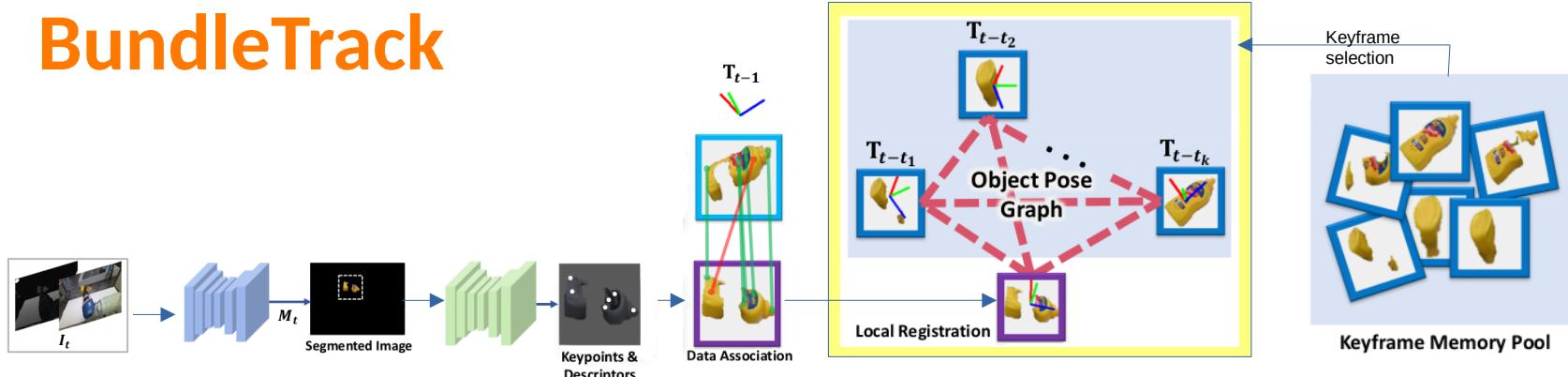
- 1) Bildsegmentierung mit VOS liefert Objektmaske
- 2) Keypoints und Deskriptoren mit LF-Net erkennen
- 3) Keypointabgleich und grobe Registrierung zwischen Bildern

# BundleTrack



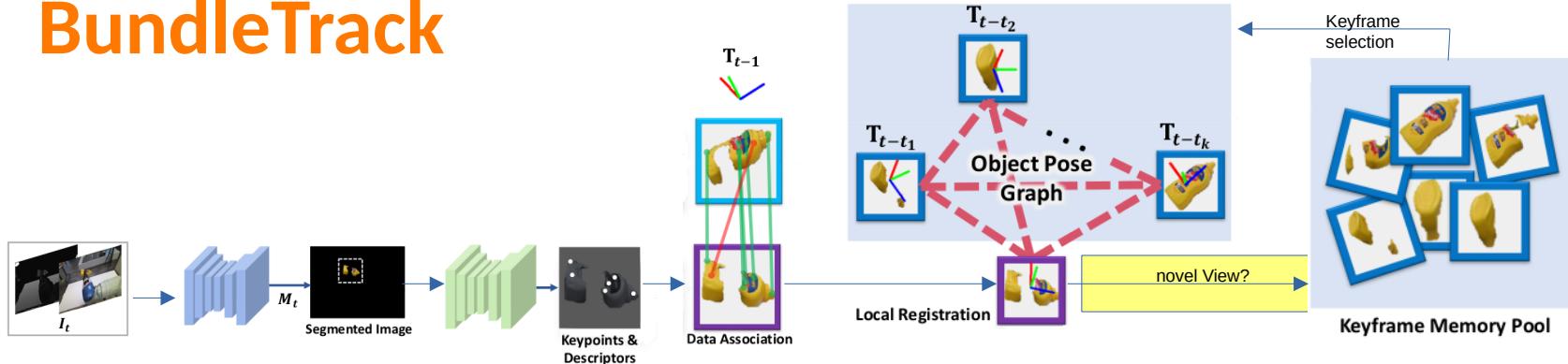
- 1) Bildsegmentierung mit VOS liefert Objektmaske
- 2) Keypoints und Deskriptoren mit LF-Net erkennen
- 3) Keypointabgleich und grobe Registrierung zwischen Bildern
- 4) Keyframes aus Speicherpool für Pose-Graph-Optimierung wählen

# BundleTrack



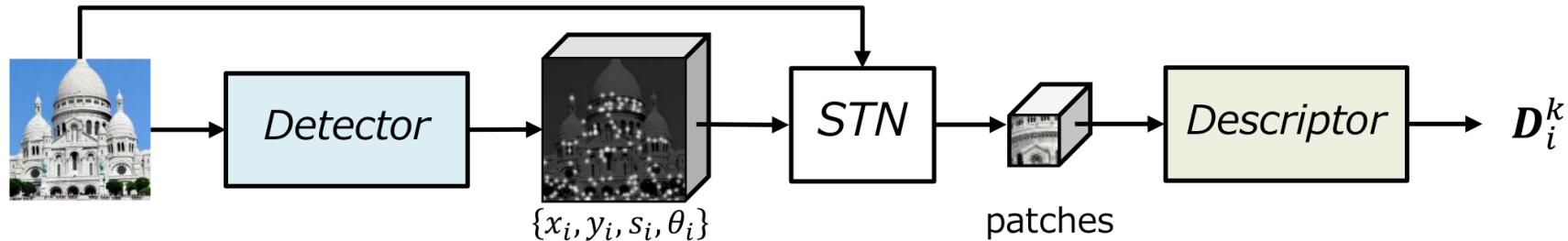
- 1) Bildsegmentierung mit VOS liefert Objektmaske
- 2) Keypoints und Deskriptoren mit LF-Net erkennen
- 3) Keypointabgleich und grobe Registrierung zwischen Bildern
- 4) Keyframes aus Speicherpool für Pose-Graph-Optimierung wählen
- 5) Pose-Graph-Optimierung für Pose

# BundleTrack



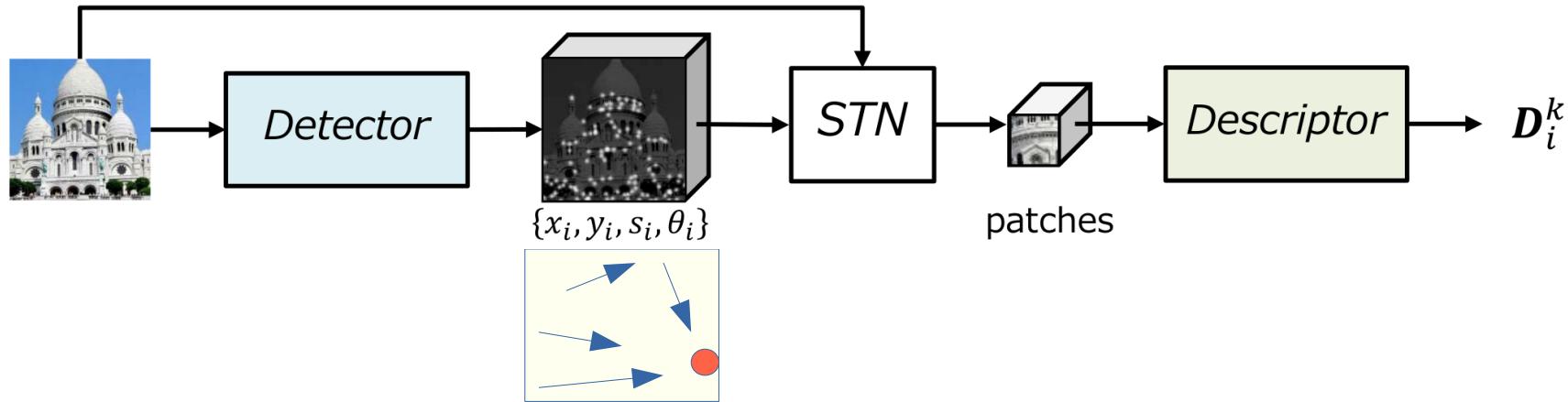
- 1) Bildsegmentierung mit VOS liefert Objektmaske
- 2) Keypoints und Deskriptoren mit LF-Net erkennen
- 3) Keypointabgleich und grobe Registrierung zwischen Bildern
- 4) Keyframes aus Speicherpool für Pose-Graph-Optimierung wählen
- 5) Pose-Graph-Optimierung für Pose
- 6) neue Ansicht in den Speicherpool

# BundleTrack - LF-Net



1) Detektor erzeugt eine Skalen-Raum-Score-Karte

# BundleTrack - LF-Net

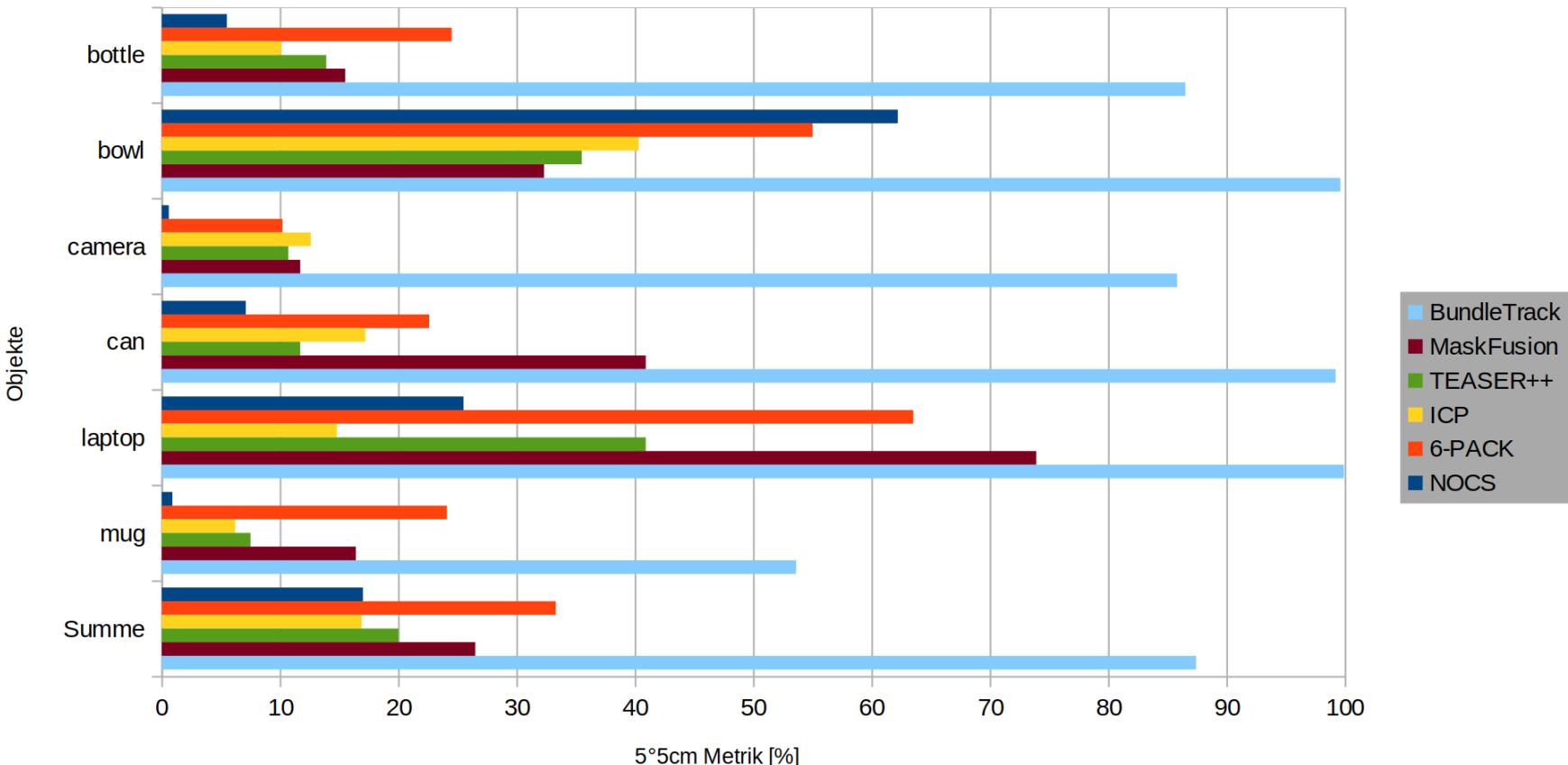


- 1) Detektor erzeugt eine Skalen-Raum-Score-Karte
- 2) Bildausschnitte um Keypoints mit Sampler Netz (STN) erzeugt Patches
- 3) Deskriptor erzeugt für jeden Patch einen Deskriptor

# Gliederung

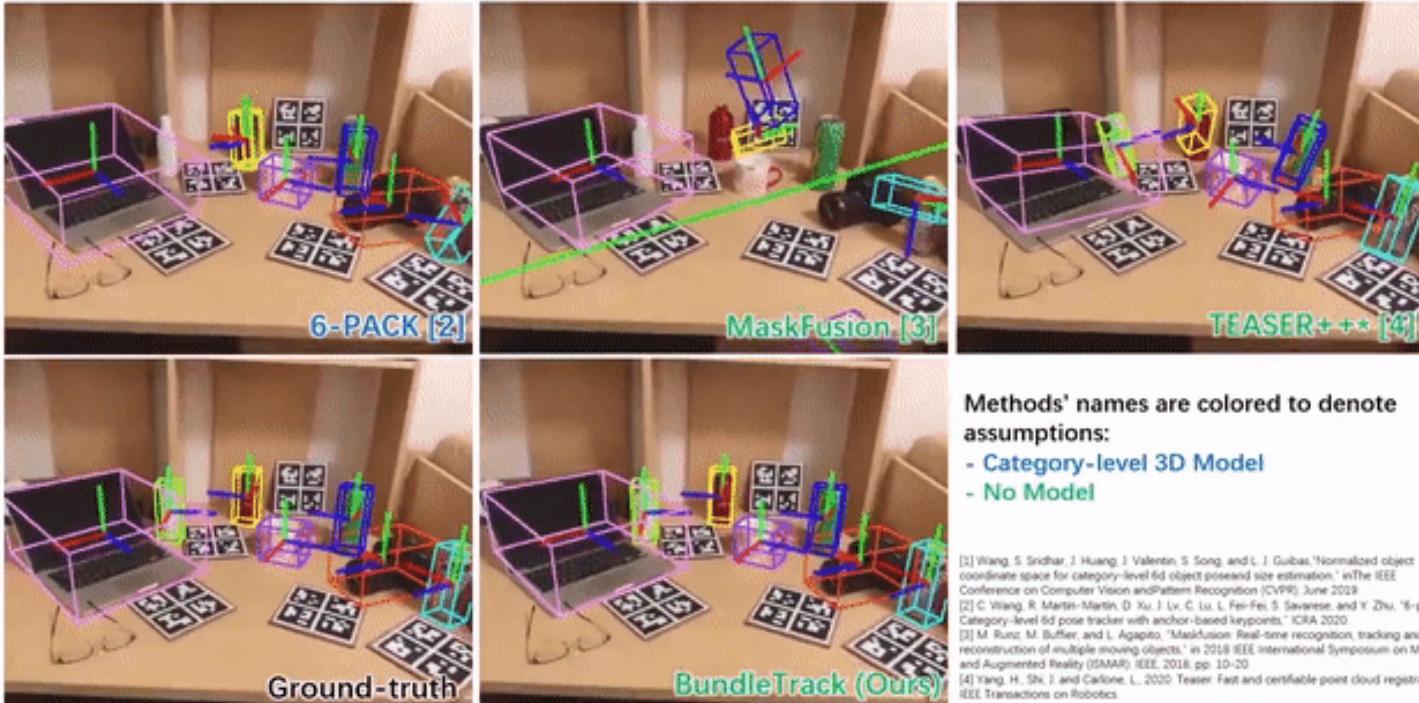
1. Motivation
2. Probleme
3. Vergleichbarkeit
4. Übersicht unterschiedlicher Verfahren
5. Teaser
6. BundleTrack
7. Vergleich & Fazit

# Vergleich auf NOCS Datensatz



# Vergleich

## Results on the NOCS dataset [1]



Methods' names are colored to denote assumptions:

- Category-level 3D Model
- No Model

- [1] Wang, S. Sridhar, J. Huang, J. Valentim, S. Song, and L. J. Guibas, "Normalized object coordinate space for category-level 6d object pose and size estimation," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.  
[2] C. Wang, R. Martín-Martín, D. Xu, J. Lv, C. Lu, L. Fei-Fei, S. Savarese, and Y. Zhu, "6-pack: Category-level 6d pose tracker with anchor-based keypoints," ICRA 2020.  
[3] M. Kunz, M. Butler, and L. Agapito, "Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects," in 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, 2018, pp. 10–20.  
[4] Yang, H., Shi, J., and Carlone, L., 2020. Teaser: Fast and certifiable point cloud registration. IEEE Transactions on Robotics.

# Fazit

- Mehrere Objekte gleichzeitig möglich
- Texturierung oder Verdeckung kein Problem
- Neue Objekte können ohne Vorwissen erkannt werden



- Mobiler Einsatz ausbaufähig
- Grundlagen sind geschaffen

[Metra]

# Quellen

- [Xu ECCV 20] Chen, Xu, et al. "Category level object pose estimation via neural analysis-by-synthesis,, European Conference on Computer Vision. Springer, Cham, 2020
- [Yang ToR 20] Yang, Heng, Jingnan Shi, and Luca Carlone. "Teaser: Fast and certifiable point cloud registration." IEEE Transactions on Robotics 37.2 (2020): 314-333.
- [Issac ICRA 16] Jan Issac u. a. „Depth-based object tracking using a Robust Gaussian Filter“. In:2016 IEEE International Conference on Robotics and Automation.
- [Li IJCV 18] Yi Li u. a. „DeepIM: Deep Iterative Matching for 6D Pose Estimation“. In:International Journal of Computer Vision128.3 (Nov. 2018), S. 657-678. issn: 1573-1405.doi:10.1007/s11263-019-01250-9.url:arxiv.org/abs/1804.00175
- [Runz ISMAR 18] Runz, Martin, Maud Buffier, and Lourdes Agapito. "Maskfusion: Real-time recognition, tracking and reconstruction of multiple moving objects." 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2018.
- [Suwanjanakorn ARXIV 18] S. Suwanjanakorn und et al. Discovery of latent 3d keypoints via end-to-end geometric reasoning. Website. 2018. url:<https://arxiv.org/abs/1807.03146>.
- [Wen IROS 20] Wen, Bowen, et al. "se (3)-tracknet: Data-driven 6D pose tracking by calibrating image residuals in synthetic domains." 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020.
- [Wang ICRA 19] Wang, Chen, et al. "6-pack: Category-level 6d pose tracker with anchor-based keypoints." 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020.
- [Wang CVPR 19] Wang, He, et al. "Normalized object coordinate space for category-level 6d object pose and size estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [Bowen ARXIV 21] Wen, Bowen, and Kostas Bekris. "BundleTrack: 6D Pose Tracking for Novel Objects without Instance or Category-Level 3D Models." arXiv preprint arXiv:2108.00516 (2021).
- [Wüthrich IROS 13] Wüthrich, Manuel, et al. "Probabilistic object tracking using a range camera." 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013.
- [Xiang ARXIV 17] Yu Xiang u. a. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Clutte-red Scenes. 2017. <url:arxiv.org/abs/1711.00199>
- [Park ARXIV 18] J. Park Q.-Y. Zhou und V. Koltun. Open3d: A modern library for 3d data processing. 2018. url:arxiv.org/abs/1801.09847
- [PwP] Papers with Code, Website, url:paperswithcode.com/task/6d-pose-estimation-1

# Grafiken

[3dmatch] Abgerufen 11.01.22, <https://3dmatch.cs.princeton.edu>

[5°5cm Beispiel] selbst erstellt in OpenSCAD

[AUC Beispiel] Abgerufen am 27.12.21, <https://www.listendata.com/2014/08/learn-area-under-curve-auc.html>

[BundleTrackShow] Abgerufen am 12.12.21, <https://github.com/wenbowen123/BundleTrack>

[Care-o-bot] Abgerufen 30.11.21, <https://www.theoldrobots.org/images5/Care-O-bot-3m.JPG>

[Linemod] Abgerufen 2.12.21, <https://bop.felk.cvut.cz/datasets/>

[Metra] Aushändigung des Fachgebiets

[ObjectTracking] Aushändigung des Fachgebiets

[Pick&Place] Abgerufen 30.11.21, <https://rivasrobotics.com/wp-content/uploads/2017/05/Pick-and-place-robot.jpg>

[PoseCNN] Yu Xiang u. a. PoseCNN: A Convolutional Neural Network [...] 2017. [url:arxiv.org/abs/1711.00199](https://arxiv.org/abs/1711.00199)

[TEASER\_Python] Erstellt mit Python & Anaconda, <https://github.com/MIT-SPARK/TEASER-plusplus>

[YCB] Abgerufen 30.11.21, <https://www.researchgate.net/profile/Jens-Lundell-2/publication/336147823>

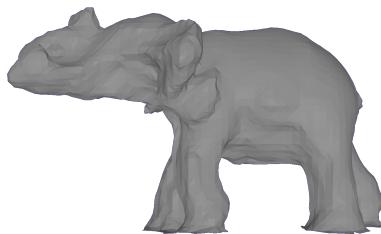
Gesamtes Seminar verfügbar unter:

**GitHub.com/WieErWill/Objektposeabschätzung**



# Folien die nicht mehr in die Präsentation passen

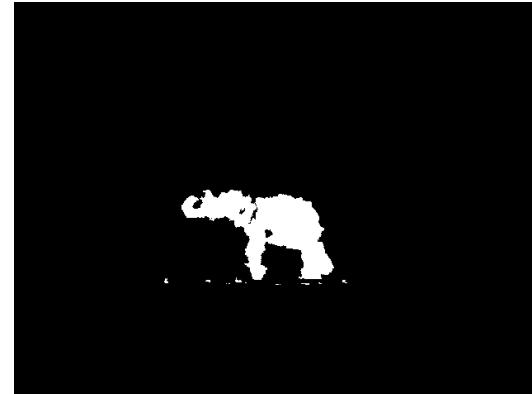
# Beispiel Datensatz: Linemod



3D Objekt



RGB-Bild (registriert)



Maske



Tiefenbild

Elephant\_1\_light  
image size: 640 480  
rotation:

0.999993	0.00335704	0.000844505
-0.00335104	0.99997	-0.00683047
-0.000867251	0.00682782	0.999976
center: -0.0399028	-0.0418303	-0.732117
extent: 0.296402	0.14491	0.174695

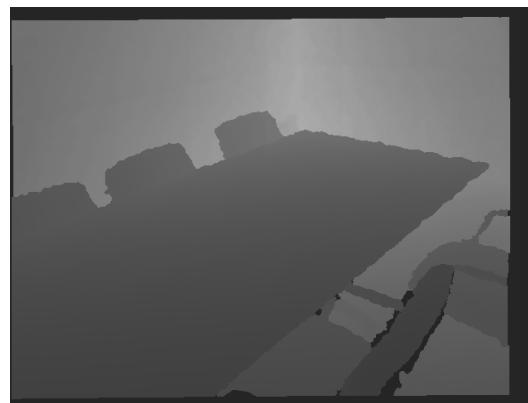
# Beispiel Datensatz: 3dMatch

Pose:

7.57396000e-01	2.17607000e-01	-6.15629000e-01	1.24883400e+00
-6.12240000e-02	-9.15013000e-01	-3.98753000e-01	8.89882000e-01
-6.50081000e-01	3.39704000e-01	-6.79705000e-01	1.26297900e+00
0.00000000e+00	0.00000000e+00	0.00000000e+00	1.00000000e+00



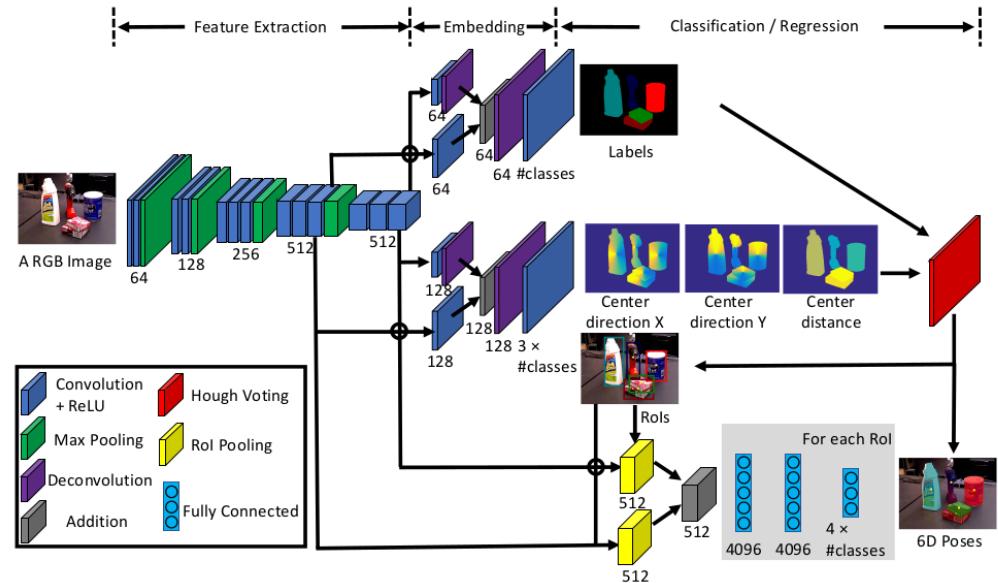
RGB



Tiefenbild

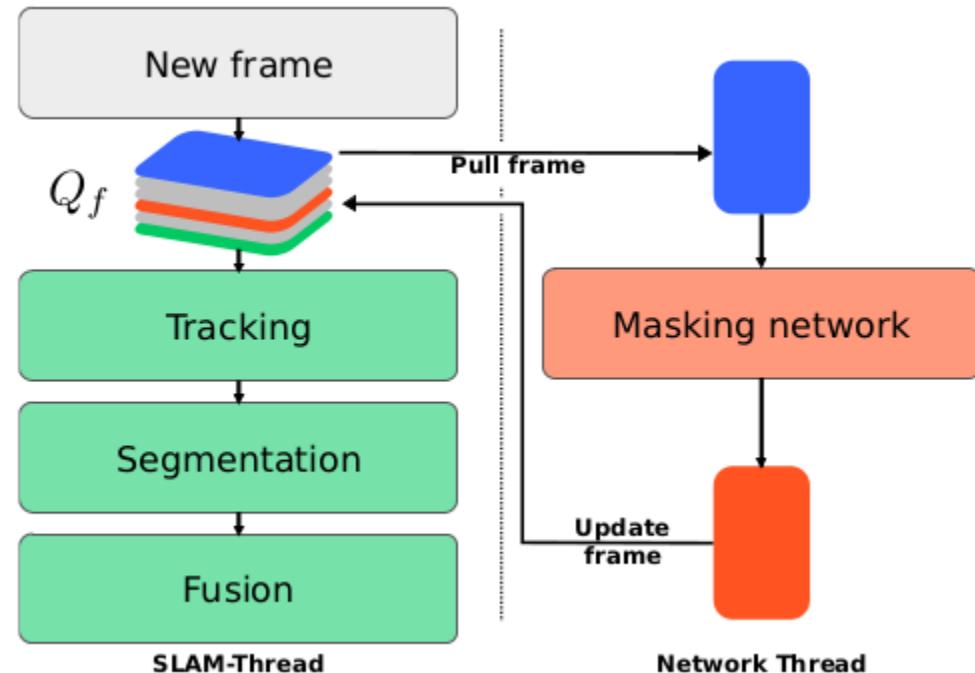
# PoseCNN

- **Modell:** 3D Modell
- **Input:** RGB, RGB-D
- **Datensatz:** YCB, LineMOD, LineMOD-Occlusion
- **Genauigkeit**
  - 53,7% ADD (RGB)
  - 75,9% ADD-S (RGB)
  - 79,3% ADD (RGB-D)
  - 93,0% ADD-S (RGB-D)



# MaskFusion

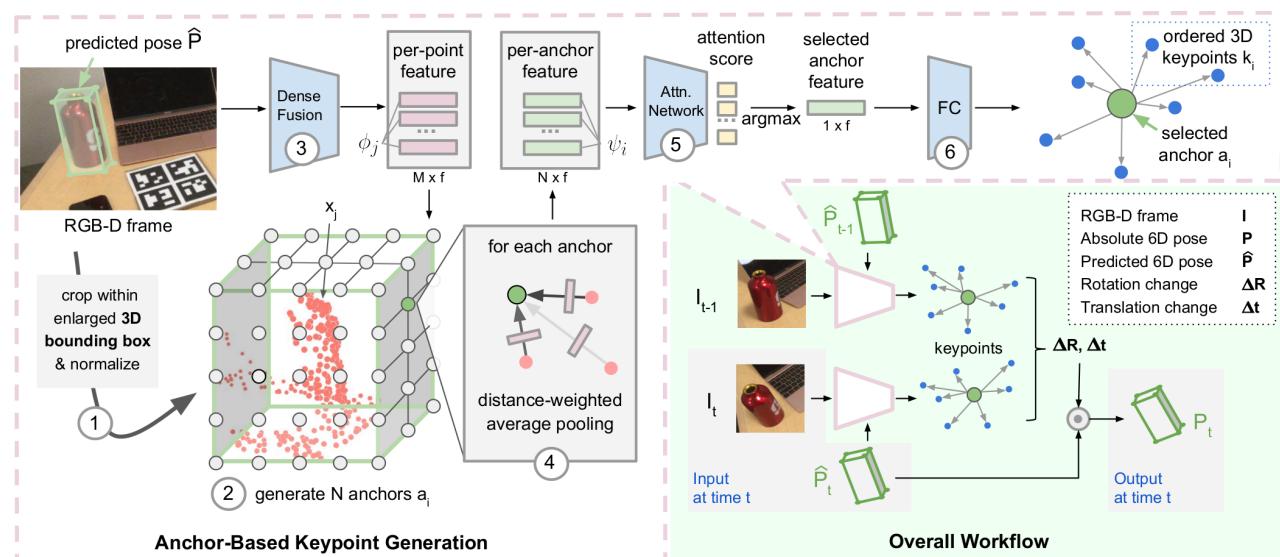
- **Modell:** ohne
- **Input:** RGB-D
- **Datensatz:** MS-COCO
- **Genauigkeit**
  - 26,5% 5°5cm
  - 64,9% IoU25
- **Ressourcen:** 2 GPU
- **Laufzeit:** 20-30 Hz



# 6-PACK

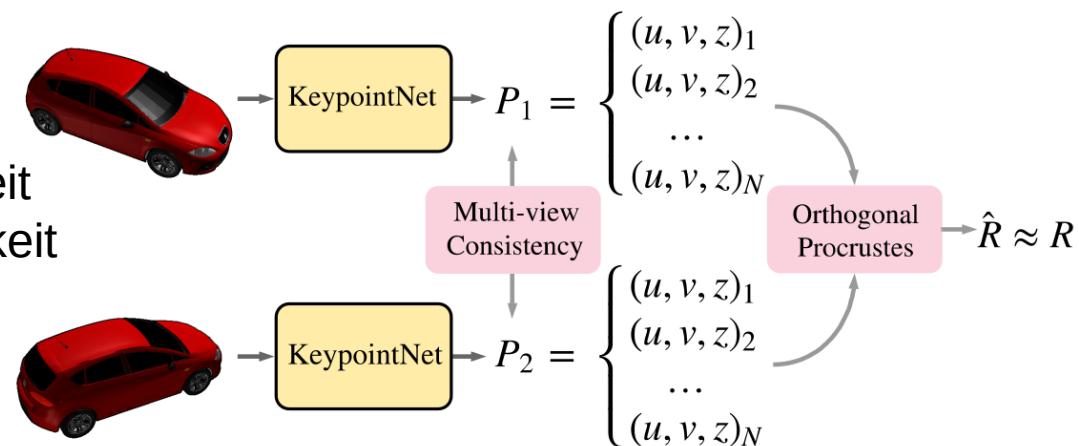
- **Modell:** Kategorie 3D Modell
- **Input:** RGB-D
- **Datensatz:** NOCS, ShapeNetCore
- **Genauigkeit**
  - 33,3% 5°5cm
  - 94,2% IoU25

- **Ressourcen:** GTX 1070 GPU, i7-6700K CPU
- **Laufzeit:** 10 Hz



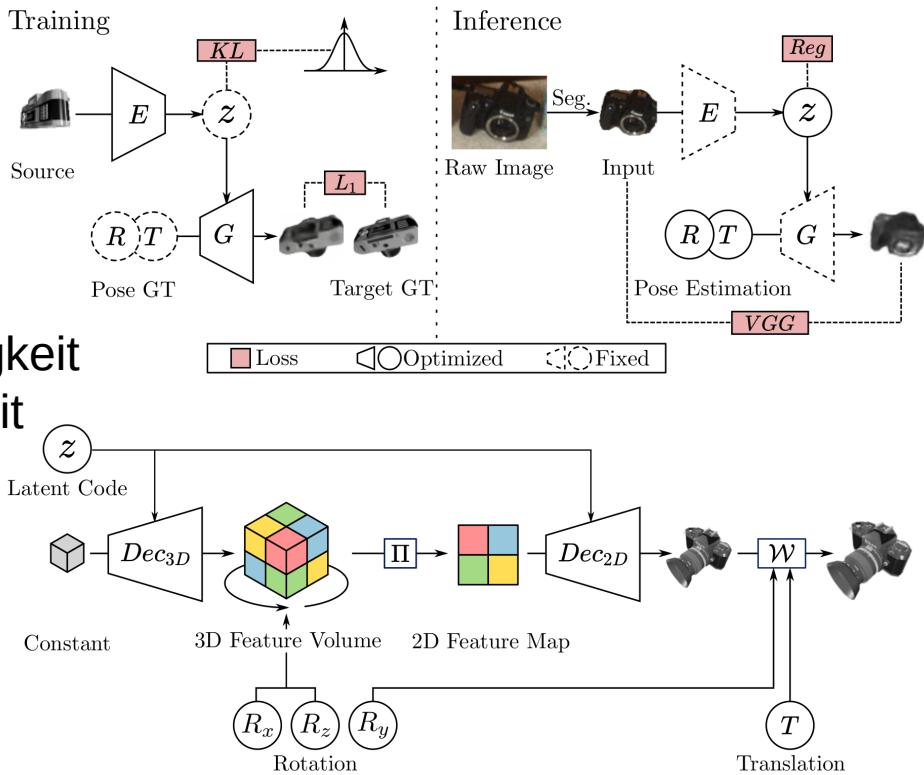
# KeypointNet

- **Modell:** Kategorie 3D Modell
- **Input:** RGB
- **Datensatz:** ShapeNetCore, Pascal3D+
- **Genauigkeit**
  - 96,2% mittlere Genauigkeit
  - 99,0% mediane Genauigkeit



# Analysis-by-Synthesis

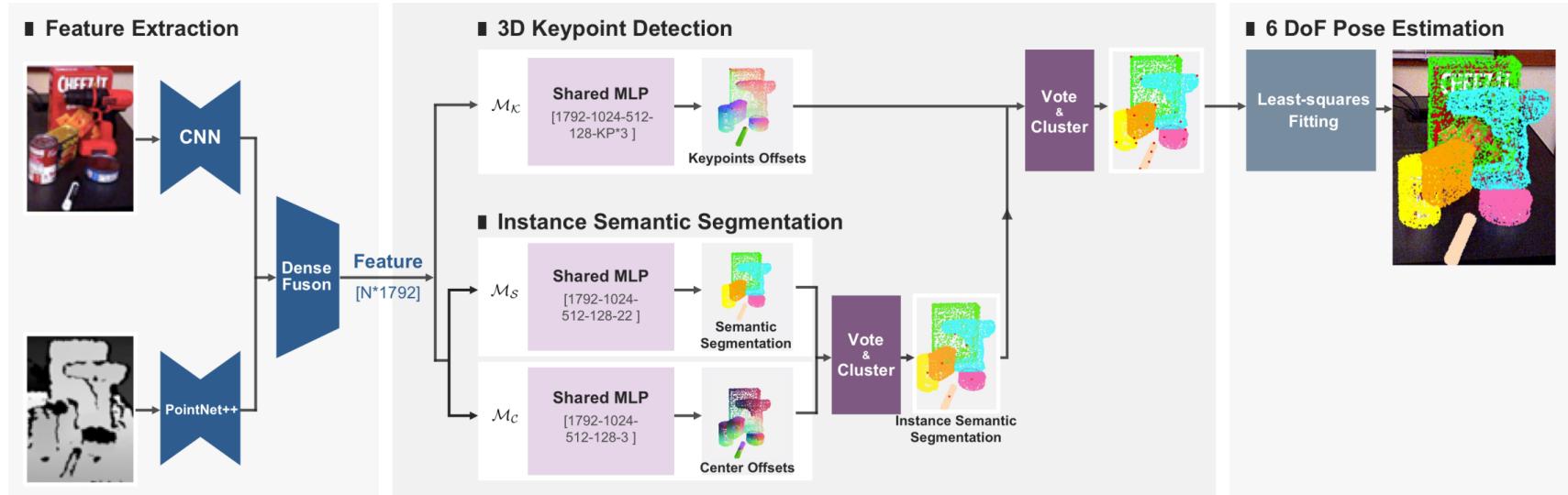
- **Modell:** ohne
- **Input:** RGB, RGB-D
- **Datensatz:** ohne
- **Genauigkeit**
  - 95% durchs. Translationsgenauigkeit
  - 90% durchs. Rotationsgenauigkeit



# PVN3D

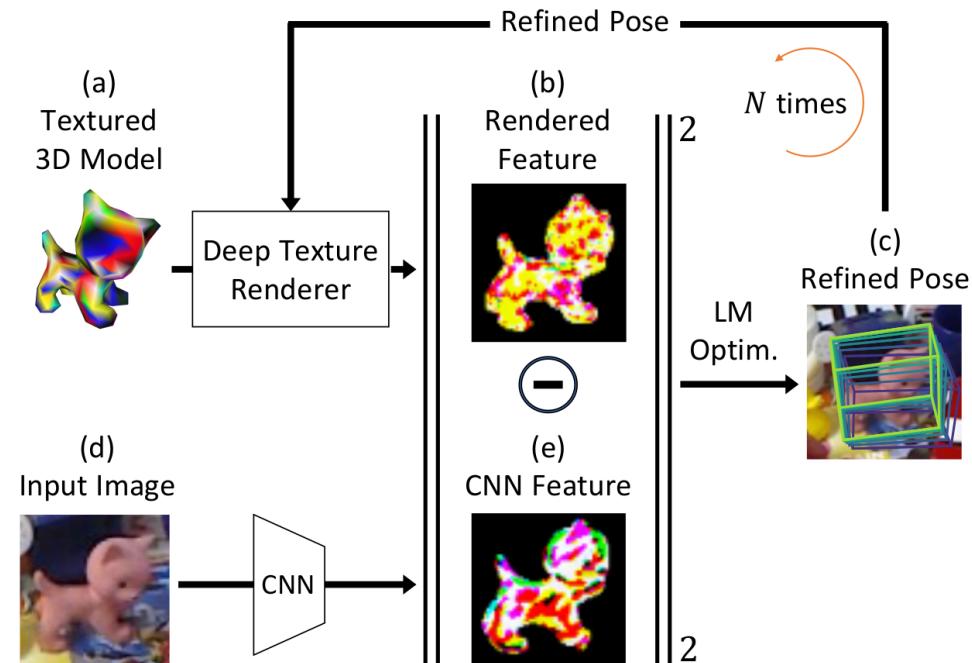
- **Modell:** 3D Modelle
- **Input:** RGB-D
- **Datensatz:** YCB, LineMOD

- **Genauigkeit**
  - 95,5% ADD (YCB)
  - 91,8% ADD-S (YCB)
  - 99,4% ADD-S (Linemod)

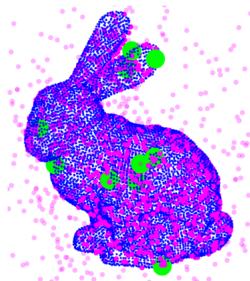
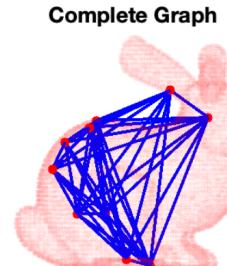
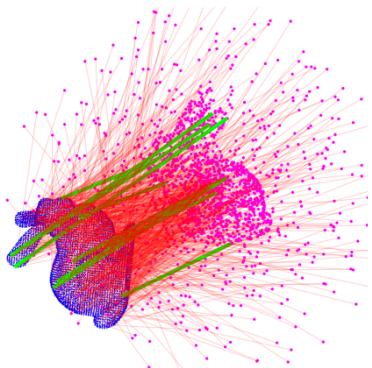
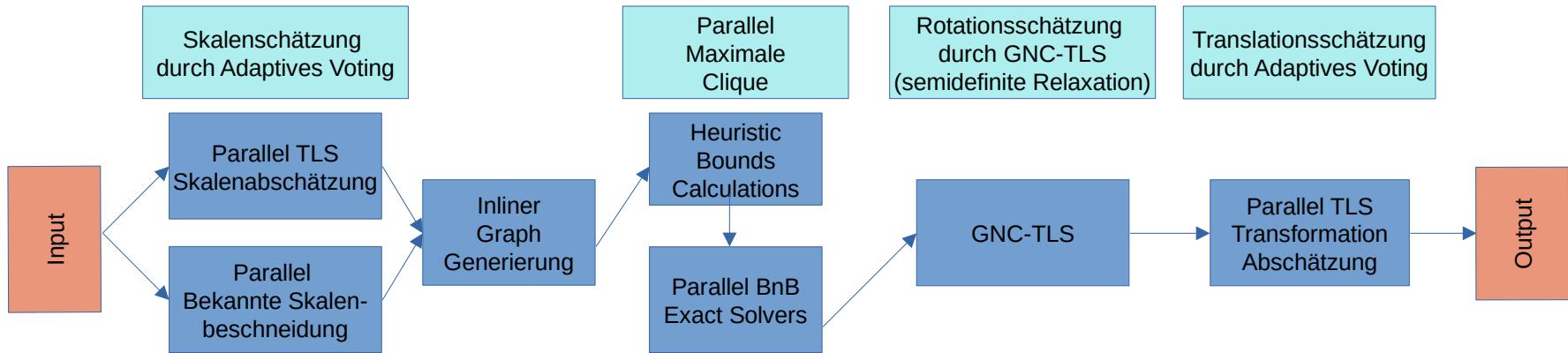


# RePOSE: Pose Refinement via Deep Texture Rendering

- **Modell:** 3D Modell
- **Input:** RGB
- **Datensatz:** YCB, LineMOD, Occlusion LineMOD
- **Genauigkeit**
  - 80,8% ADD
  - 86,7% ADD-S
- **Ressourcen:** RTX2080 Super GPU, Ryzen7-3700X CPU
- **Laufzeit:** 92 FPS



# Teaser Ablauf



# Vergleich auf NOCS Datensatz

Ergebnisse auf NOCS Datensatz	Metrik	bottle	bowl	camera	can	laptop	mug	$\Sigma$
NOCS	5°5cm	5,5	62,2	0,6	7,1	25,5	0,9	17,0
	IoU25	48,7	99,6	90,6	77,0	94,7	82,8	82,3
6-Pack	5°5cm	24,5	55,0	10,1	22,6	63,5	24,1	33,3
	IoU25	91,1	100,0	87,6	92,6	98,1	95,2	94,2
ICP	5°5cm	10,1	40,3	12,6	17,2	14,8	6,2	16,9
	IoU25	29,9	79,7	53,1	40,5	50,9	27,7	47,0
TEASER++	5°5cm	13,9	35,5	10,7	11,7	40,9	7,5	20,0
	IoU25	100,0	99,9	99,9	100,0	99,9	99,9	99,9
MaskFusion	5°5cm	15,5	32,3	11,7	8,8	73,9	16,4	26,5
	IoU25	51,4	71,4	60,8	49,7	99,9	56,2	64,9
BundleTrack	5°5cm	86,5	99,6	85,8	99,2	99,9	53,6	87,4
	IoU25	100,0	99,9	99,9	100,0	99,9	99,9	99,9

- 3D Modell
- Kategorie Modell
- Ohne Modell