IEEE International Conference on Image Processing (ICIP) 2020
[TEC-05] Machine Learning for Image and Video Processing III

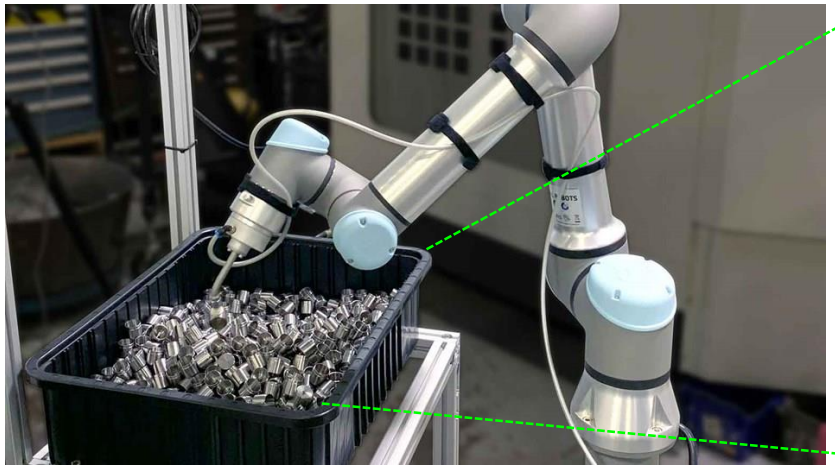
Segmenting Unseen Industrial Components in a Heavy Clutter Using RGB-D Fusion and Synthetic Data

Seunghyeok Back, Jongwon Kim, Raeyoung Kang,
Seungjun Choi and Kyoobin Lee
Gwangju Institute of Science and Technology (GIST)

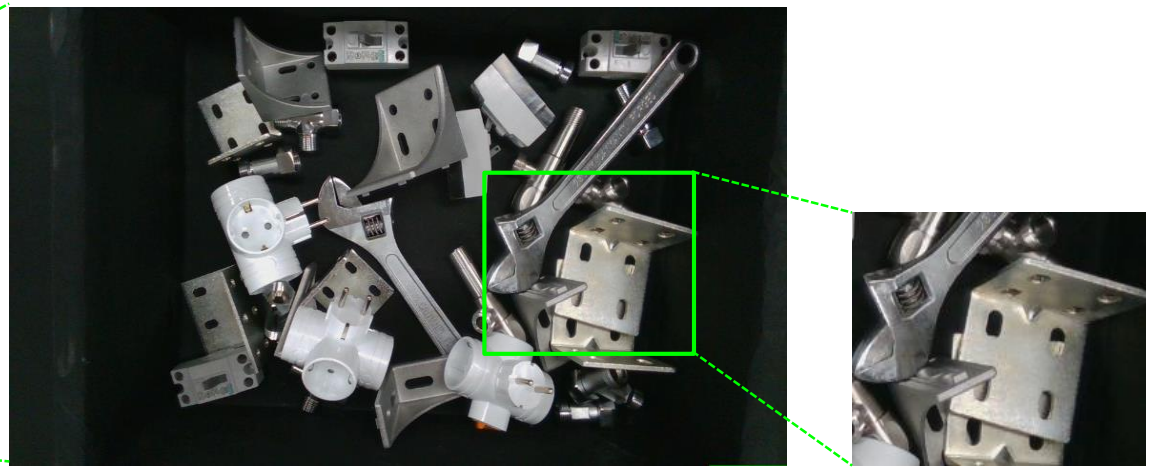



Unseen Industrial Component Instance Segmentation

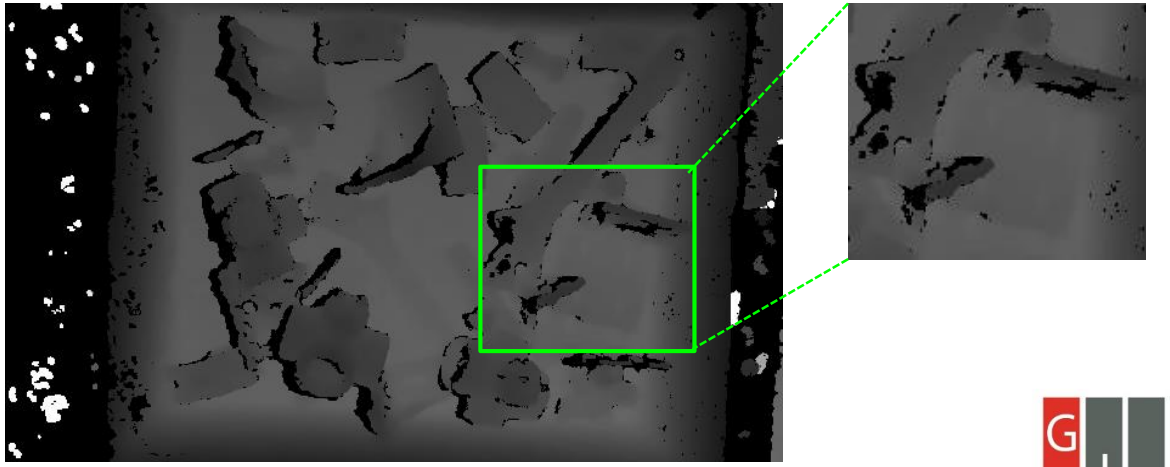
https://www.universal-robots.com/about-universal-robots/news-centre/universal-robots-launches-actinav-autonomous-bin-picking/



Industrial Bin Picking



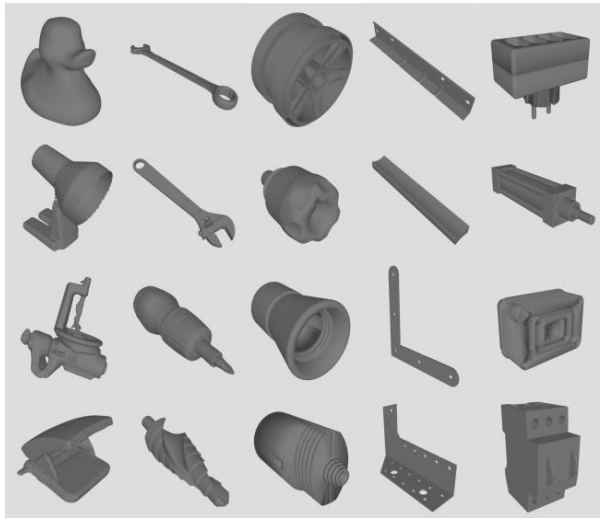
RGB



Depth

- Data collecting and labeling
- Texture-less, reflective
- Unstructured environments
- Heavy occlusion

Unseen Industrial Component Instance Segmentation

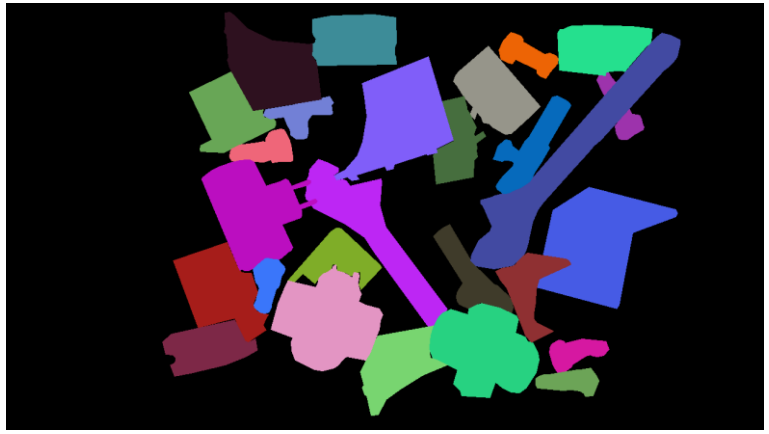


Synthetic Dataset

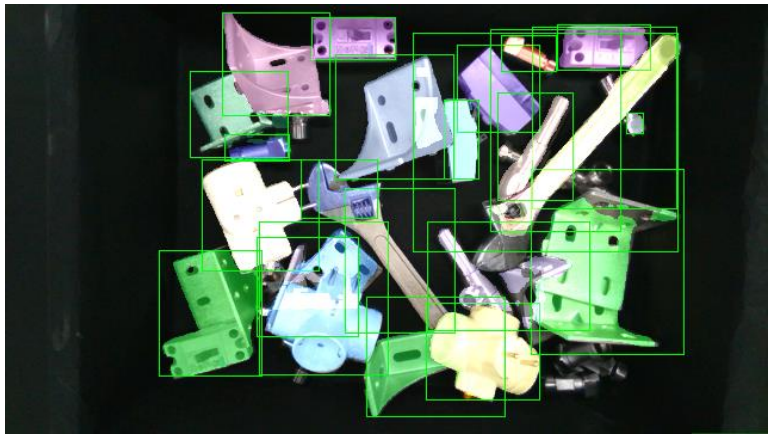
Train / Val, Known

Real-world Dataset

Test, Unseen

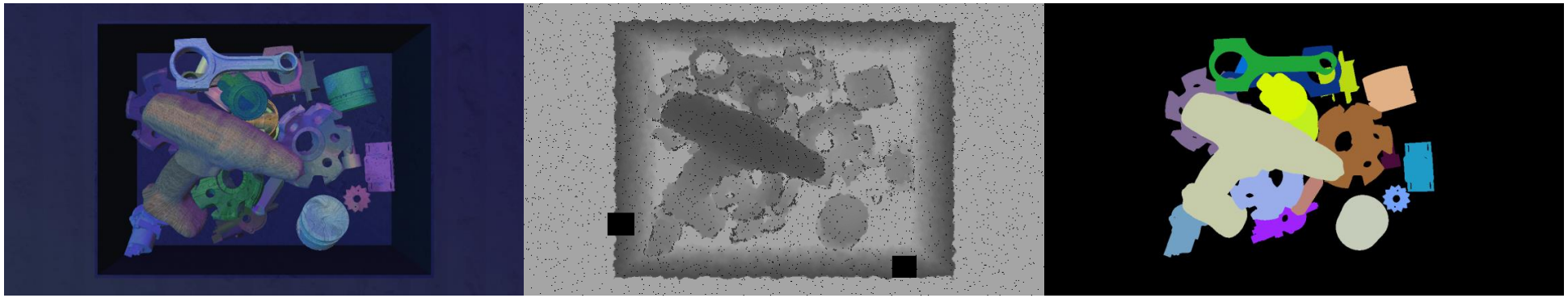


Category-agnostic Instance Mask

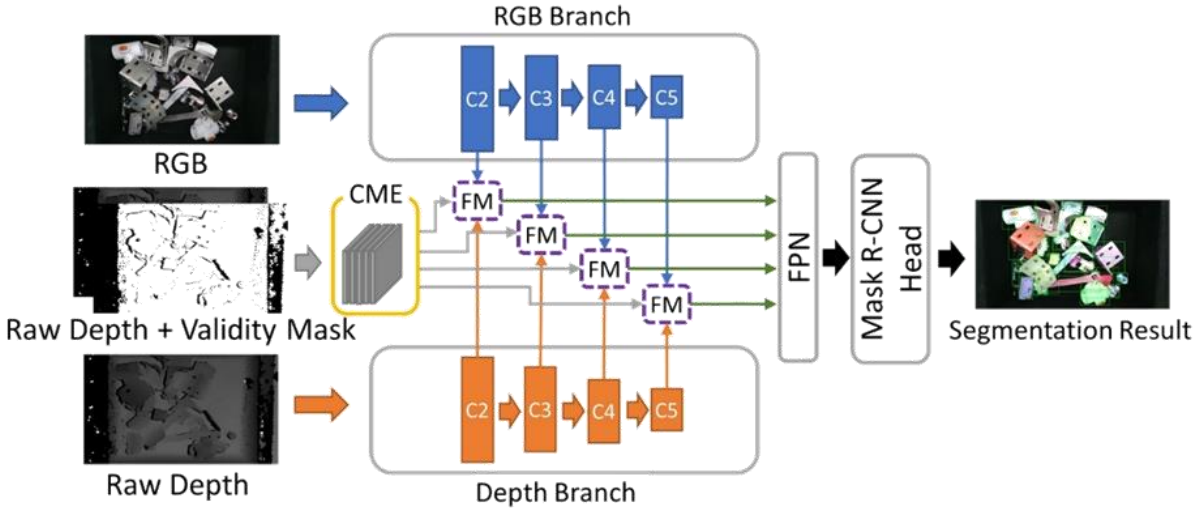


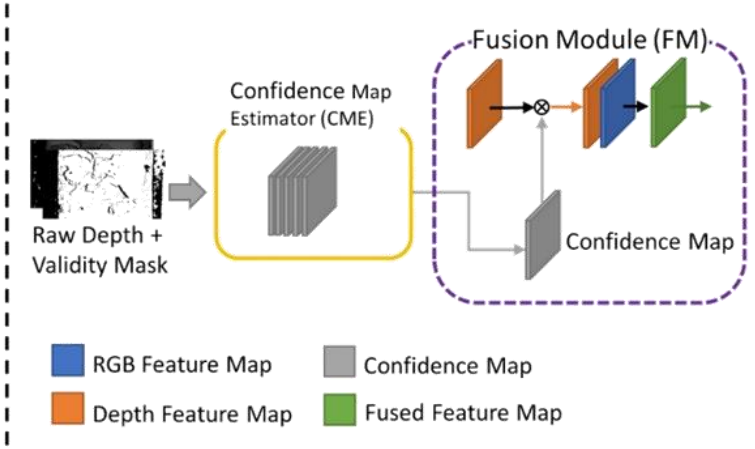
Inference Results

Unseen Industrial Component Instance Segmentation



Synthetic data generation pipeline





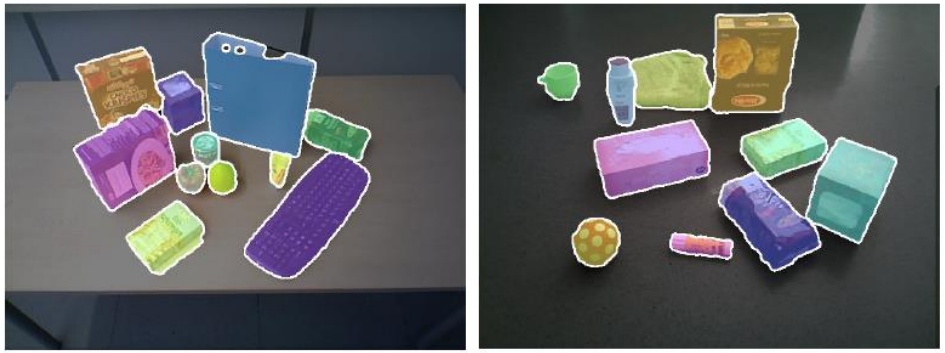
RGB-D fusion Mask R-CNN with a confidence map estimator

Unseen Object Instance Segmentation
(category-agnostic instance segmentation)

Synthetic Dataset

Synthetic Dataset

Inference Results

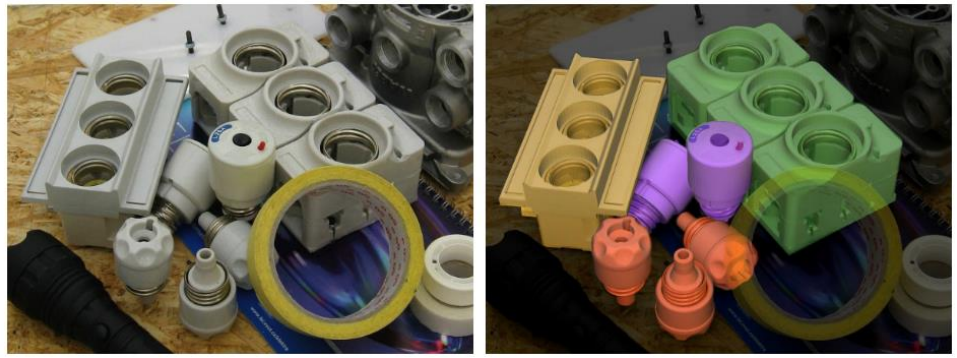


Inference Results

[1] Danielczuk, Michael, et al. (2019)

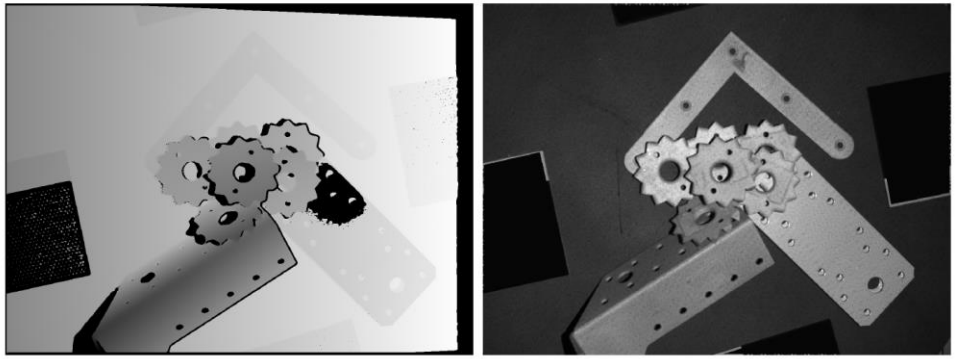
[2] Xie, Christopher, et al. (2020)

Instance Segmentation of Industrial Components



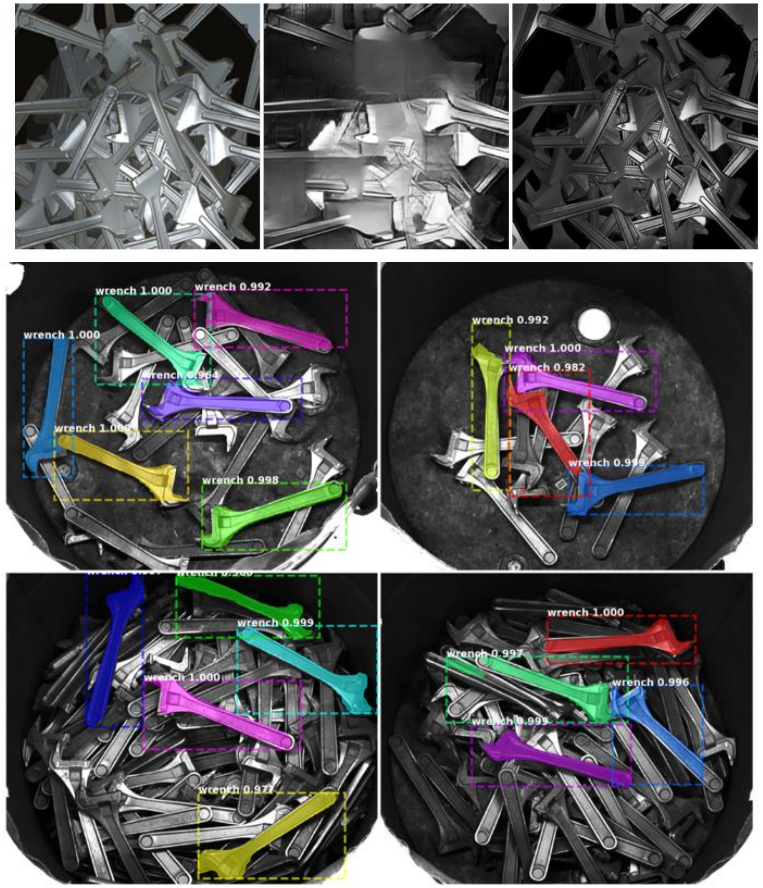
T-LESS

[3] Hodan, Tomáš, et al. (2017)



ITODD

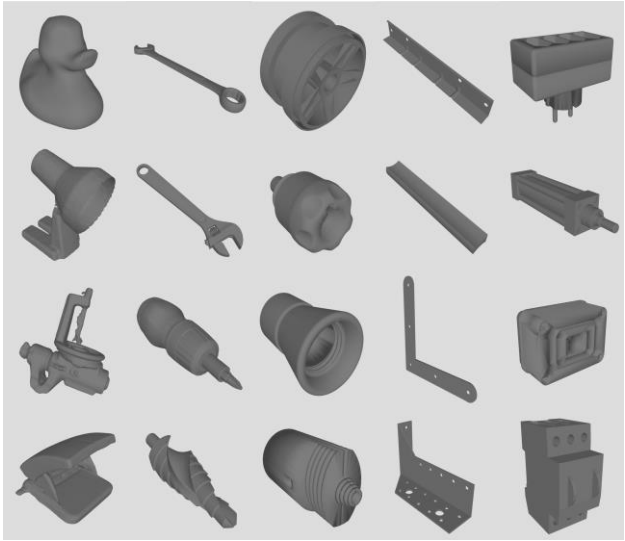
[3] Drost, Bertram, et al. (2017)



Fastener Detection

[5] Lee, Yu-Hui, et al. (2019)

Synthetic Dataset Generation



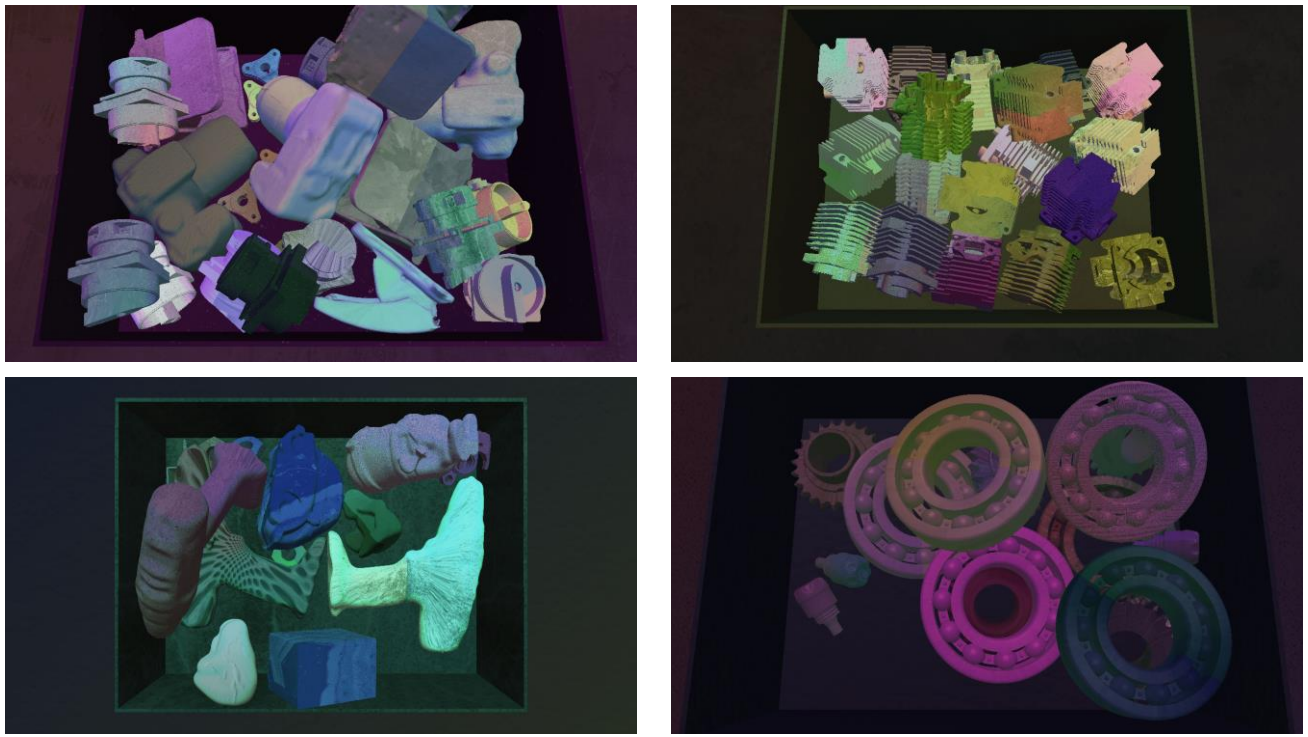
CAD models for train/val

Synthetic Dataset

Industrial object for test

- 149 3D CAD models from public datasets (Industrial & Household objects)
- Generate synthetic RGB-Depth-Mask pairs in a bin clutter environment
- 30,000 training image pairs, 5,000 validation image pairs

Synthetic Dataset Generation

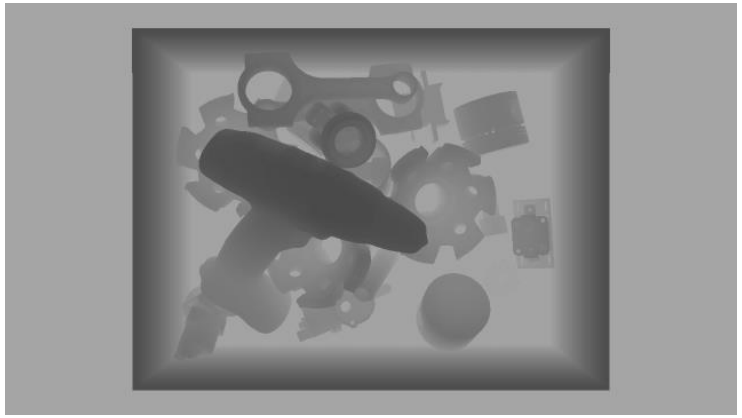


Examples of synthetic RGB images

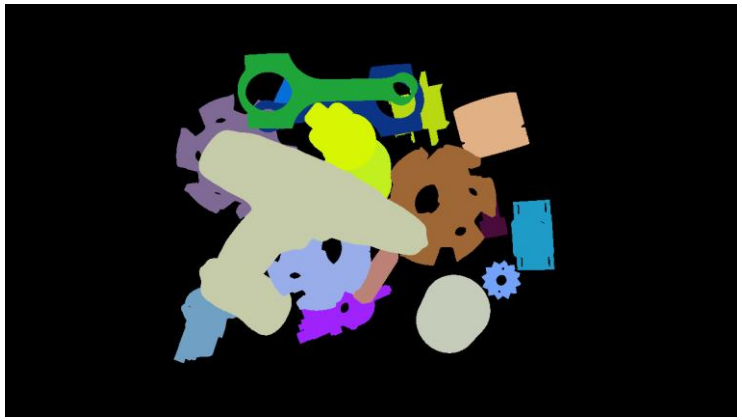
- Domain randomization [6] to learn domain-irrelevant features (shape, edge ..)
- Varies the object's position, pose, texture, and color, and configuration of the light and camera

Synthetic Dataset Generation

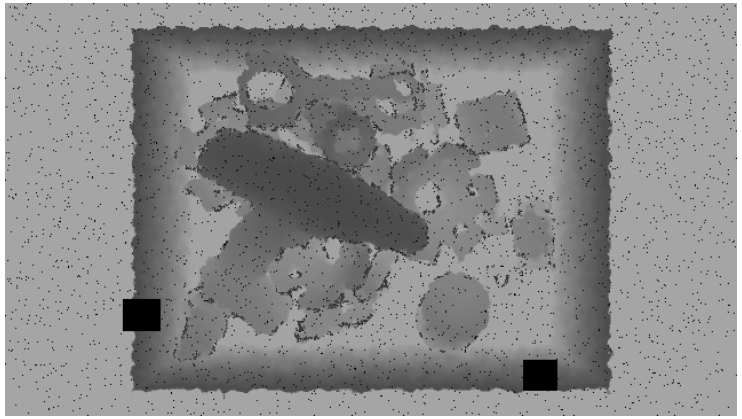
RGB



Synthetic Depth (Initial)



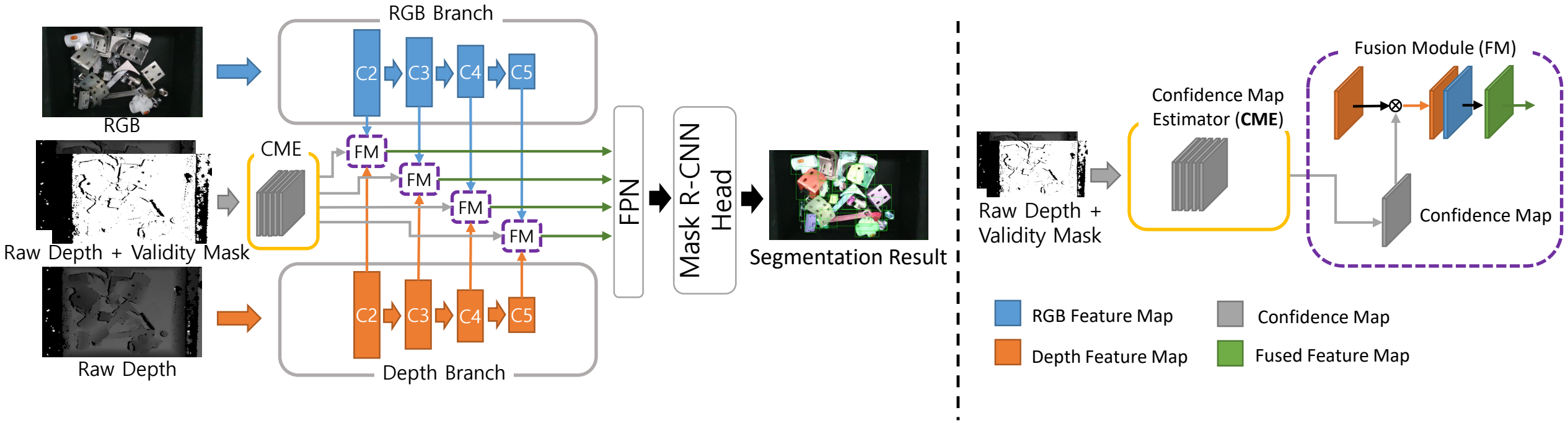
Instance Mask



Synthetic Depth (augmented)

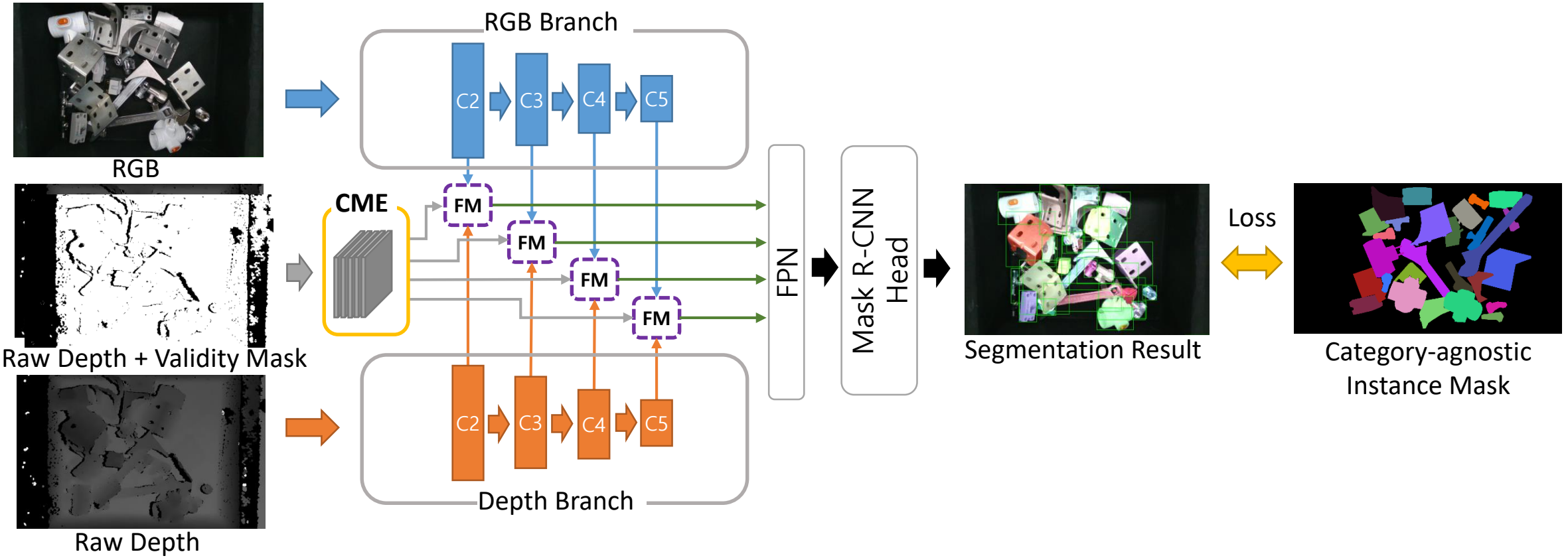
- Instance Mask
- Synthetic Depth (+3D Perlin [7], Sparse Edge, S&P Noise)

RGB-D Fusion Mask R-CNN



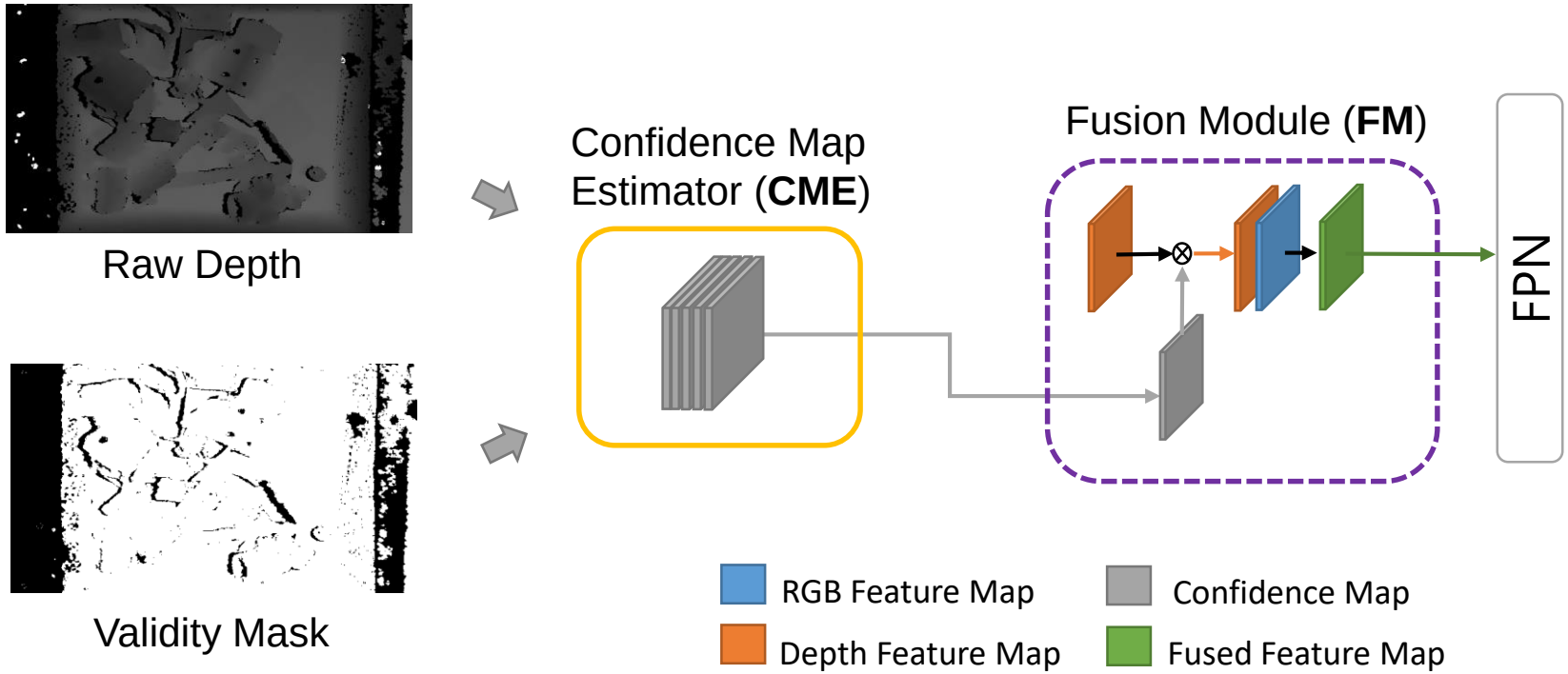
- Exploiting depth maps with RGB can improve the segmentation performance
- We propose a RGB-D Fusion Mask R-CNN with a confidence map estimator (CME)

RGB-D Fusion Mask R-CNN



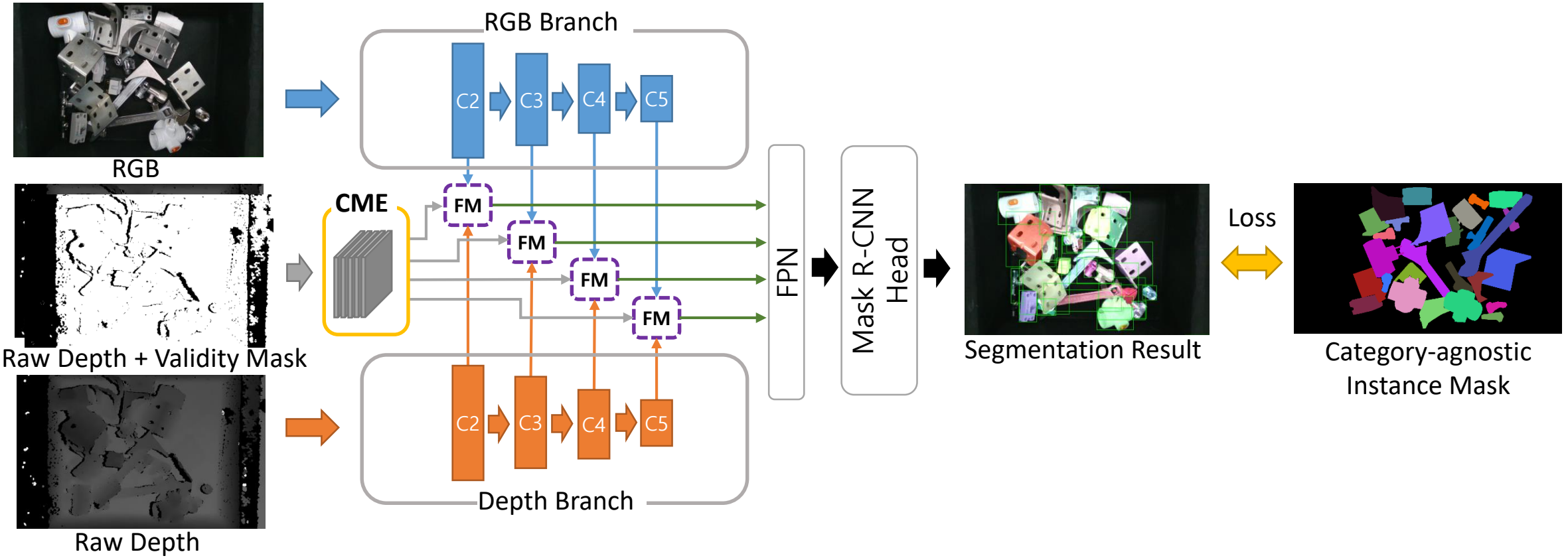
- Category-agnostic instance segmentation using Mask R-CNN with FPN
- RGB and depth branch to handle different modalities with separate networks (ResNet-50)
- Fusion module (FM) & Confidence Map Estimator (CME) for RGB-D Fusion

Confidence Map Estimator



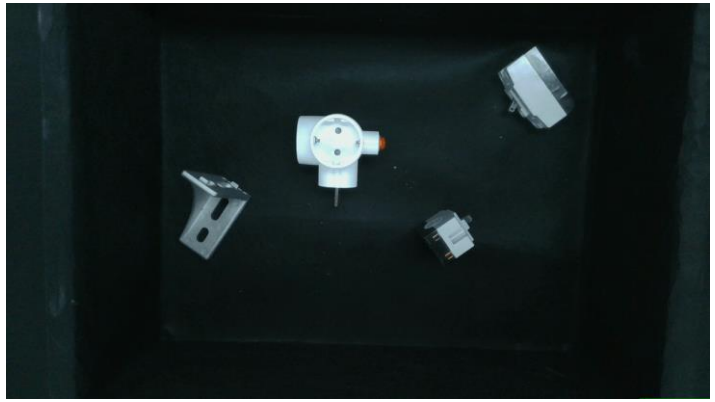
- The confidence map estimator (CME) assigns pixel-wise reliability (confidence map)
- The confidence map apply a spatial attention on depth map feature in the fusion module (**FM**)
- RGB and depth feature maps are fused and fed into the FPN

RGB-D Fusion Mask R-CNN



- Our model can fuse the RGB and depth features in four different scales

Evaluation Settings



low-occlusion

low-occlusion

high-occlusion

- Test on the unseen real-world objects / trained using only synthetic dataset
- The real-world test set captured with commodity-level RGB-D sensor (Realsense D415)
  - 60 low-occlusion images (no. object $<$ 15)
  - 40 high-occlusion images (no. object $\geq$ 15)

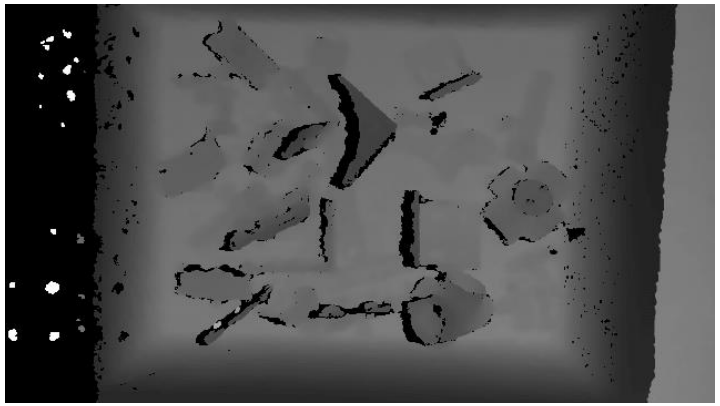
Performance Comparison

Method	RGB	Depth		RGB-D Fusion Strategy	Metrics		
		filled	raw		AP_{50}	AP	AR
Mask R-CNN	✓				67.1	54.9	65.0
		✓			61.2	52.6	62.4
			✓		51.9	45.7	56.6
	✓	✓		Early Fusion	63.3	54.5	64.5
	✓		✓	Early Fusion	60.8	53.7	62.3
	✓	✓		Late Fusion	67.9	55.8	63.2
	✓		✓	Late Fusion	67.5	55.5	65.5
Ours	✓		✓	Confidence Map	**69.0**	**57.7**	**66.1**

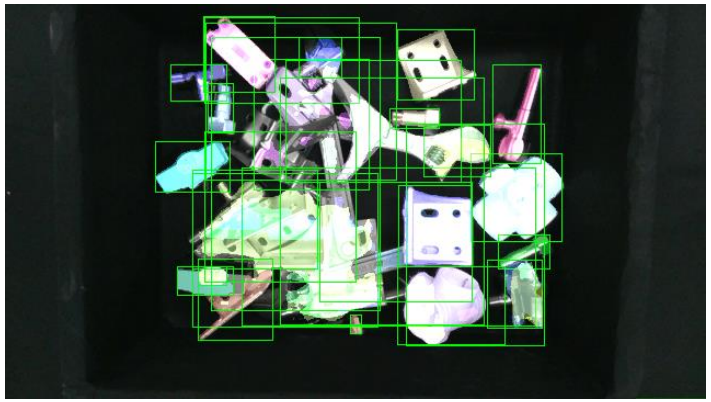
Performance comparison of the Mask R-CNN baselines and our network on real data

- Mask R-CNN with RGB inputs show reasonable generality over unseen objects
- RGB-D Fusion Mask R-CNN with a confidence map estimator achieved the best performance
- Our model can effectively exploit a raw depth map along with RGB

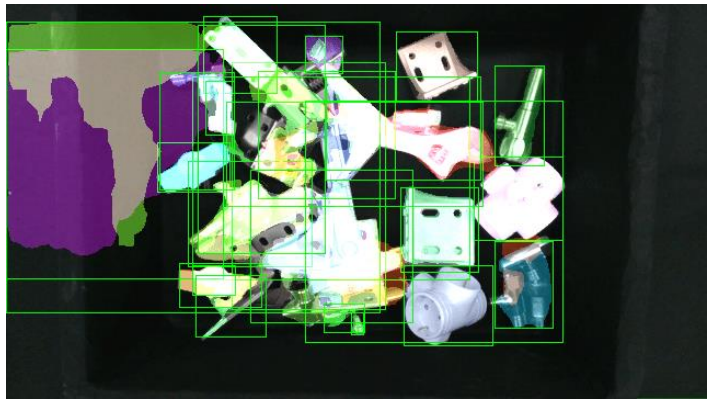
Performance Comparison



Raw depth

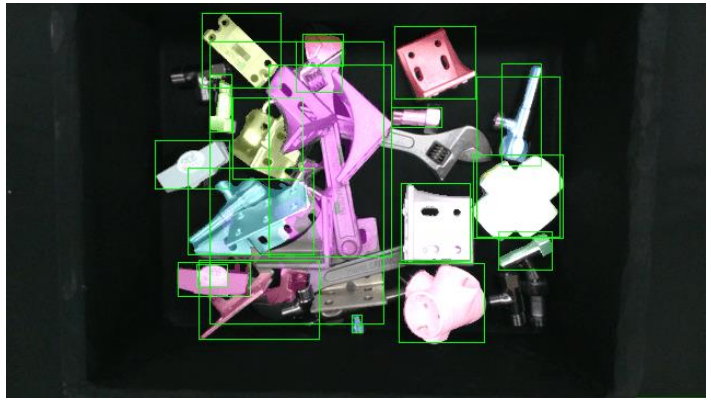


RGB only

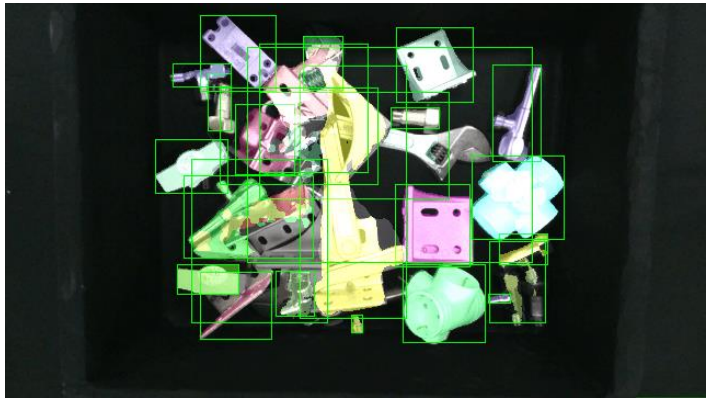


RGB + Raw depth (Late fusion)

Filled depth [8]

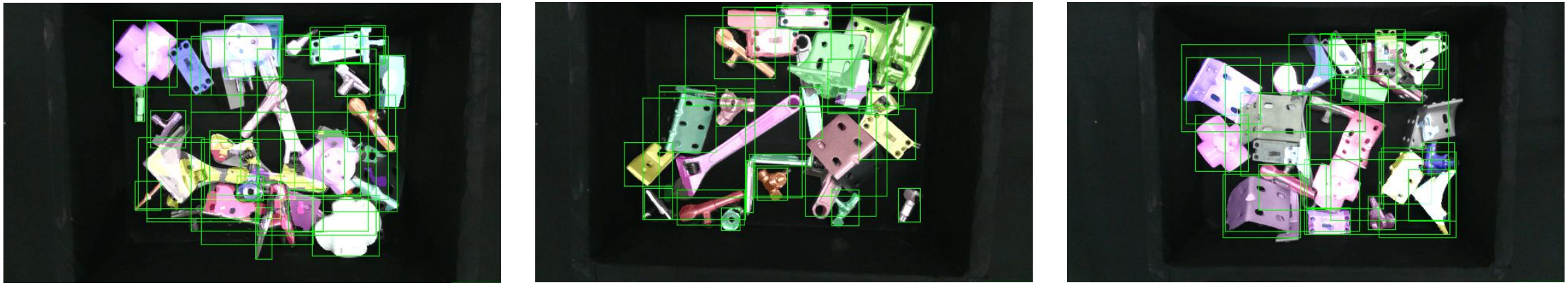


RGB + Filled depth (Late fusion)

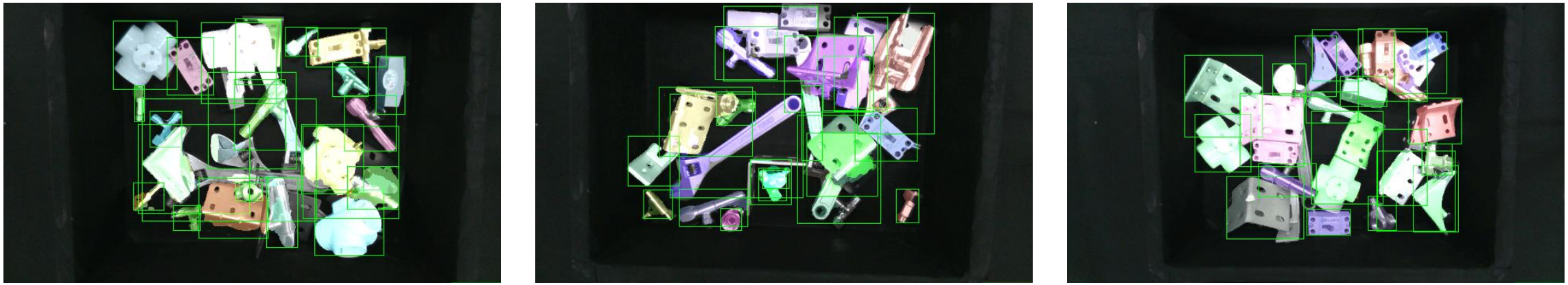


RGB + Raw depth (Confidence map)

Performance Comparison



RGB Only



RGB + Raw Depth (Confidence Map Estimator)

Demo

Thank you !

References

[1] Danielczuk, Michael, et al. "Segmenting unknown 3d objects from real depth images using mask r-cnn trained on synthetic data." 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.

[2] Xie, Christopher, et al. "The best of both modes: Separately leveraging rgb and depth for unseen object instance segmentation." Conference on Robot Learning. 2020.

[3] Hodan, Tomáš, et al. "T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects." 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017.

[4] Drost, Bertram, et al. "Introducing mvtec itodd-a dataset for 3d object recognition in industry." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.

[5] Lee, Yu-Hui, et al. "Automatic Generation of Photorealistic Training Data for Detection of Industrial Components." 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019.

[6] Tobin, Josh, et al. "Domain randomization for transferring deep neural networks from simulation to the real world." 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017.

[7] Zakharov, Sergey, et al. "Keep it unreal: Bridging the realism gap for 2.5 d recognition with geometry priors only." 2018 International Conference on 3D Vision (3DV). IEEE, 2018.

[8] Gastal, Eduardo SL, and Manuel M. Oliveira. "Domain transform for edge-aware image and video processing." ACM SIGGRAPH 2011 papers. 2011. 1-12.