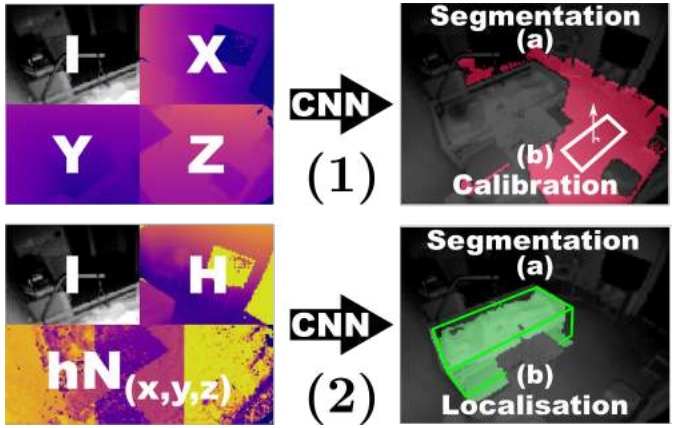
UCLouvain

ToF²Net

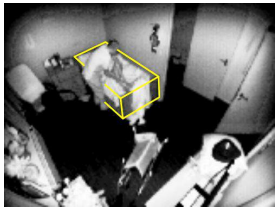


ISP GROUP
Image & Signal Processing

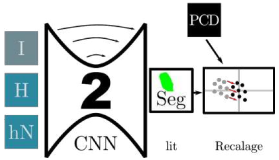
Joint use of semantic and geometric information for object localization using time-of-flight cameras

V. Joos A. Vanderschueren C. De Vleeschouwer

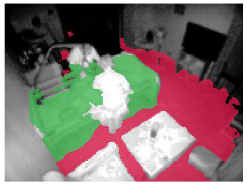
Overview



Background



Method



Results

Conclusion & Further work

Background

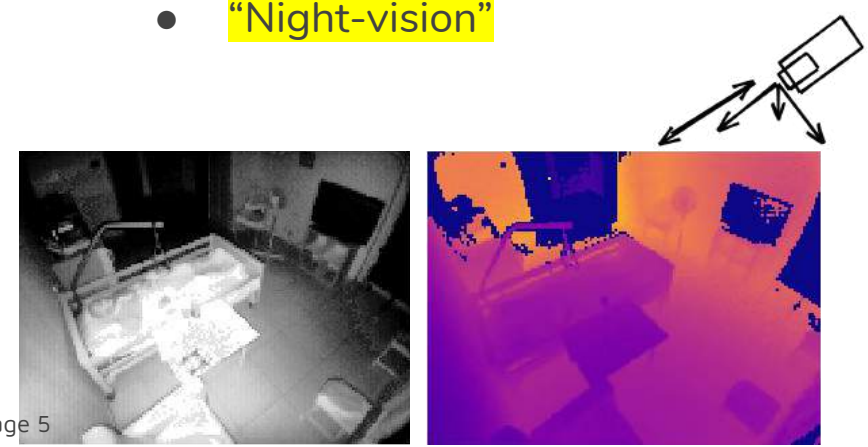
Context

- Kaspard's Objective
  - Detect bed exits & returns
  - Detect fall-related incidents
- Problem
  - Manual bed localization is costly
  - Beds can move (hospitals)
- Our solution
  - Automatic bed/object localization

Time-of-Flight Cameras – Indoor Scenes

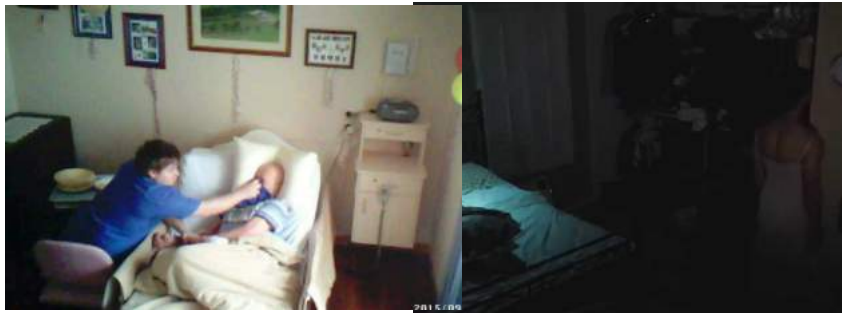
ToF Sensor

- Low-res (120x160)
- Spatial information: distance to camera along x,y,z
- Noisy distance values
- "Night-vision"

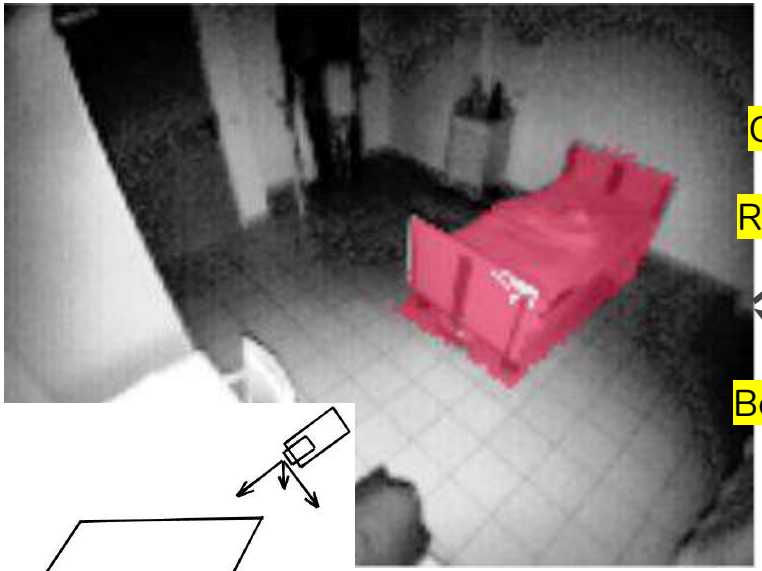


RGB Sensor

- High-res (1080x1920 or up)
- Noisy in poor lighting conditions
- No spatial information
- Need for external light source

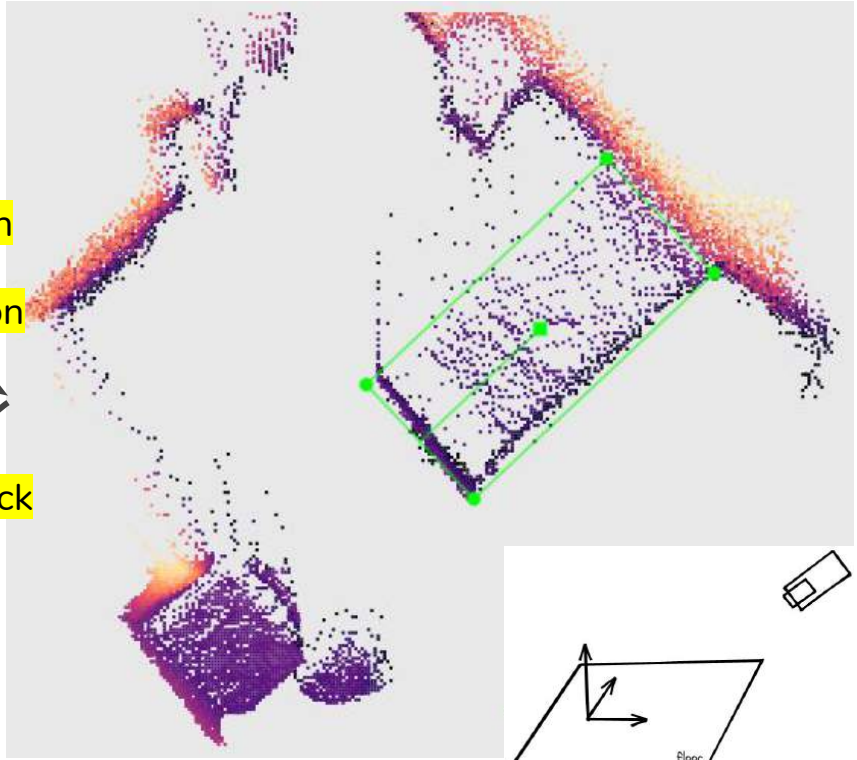


Segmentation & Object Localization



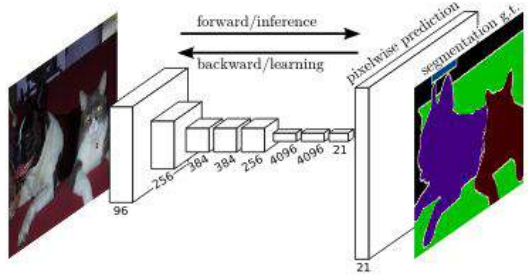
Calibration
+
Registration

Bound check

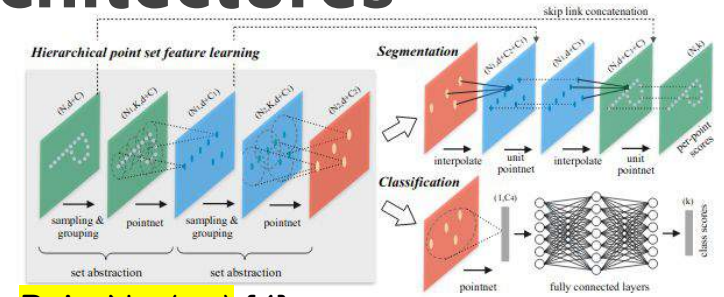


Segmentation Network Architectures

Fully Convolutional Networks [1]
2014

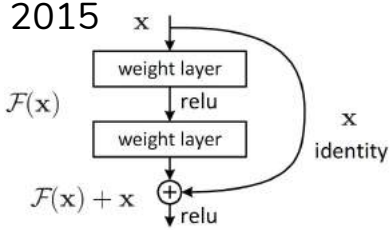


U-Net [2]
2015

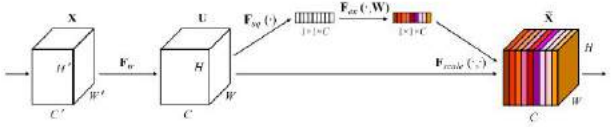


PointNet(++) [4]
2016-2017

ResNet [3]
2015

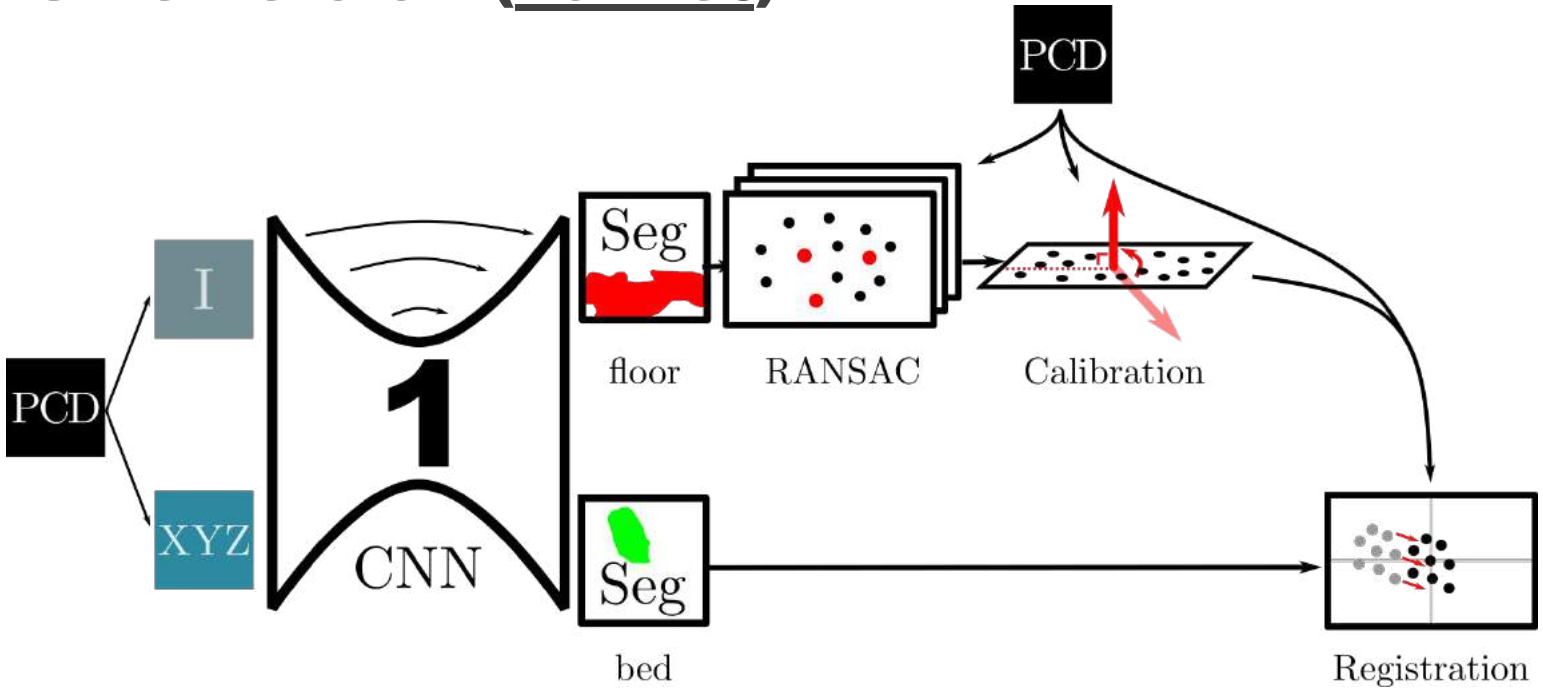


Squeeze-and-Excite [5]
2018

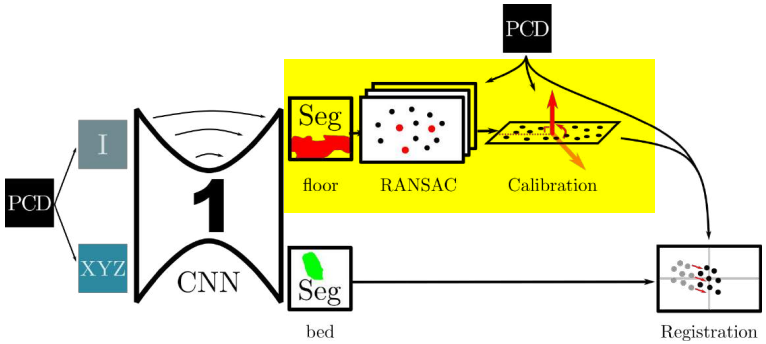


Method

Naive version (TofNet)



Calibration



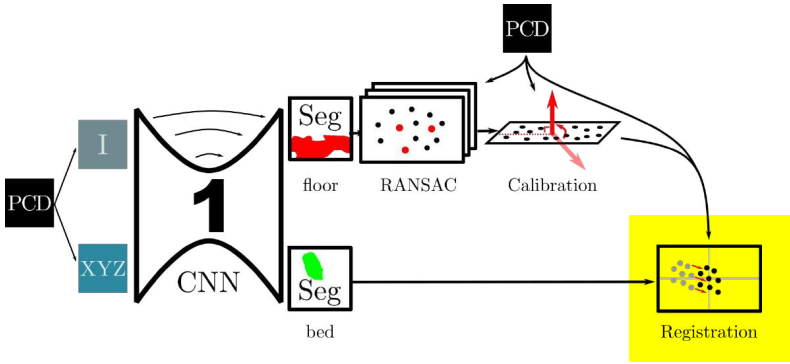
RANSAC – Random Sample Consensus

N times
- Sample K points
- Fit floor model (SVD) using these points
- Compute distance metric for inlier points

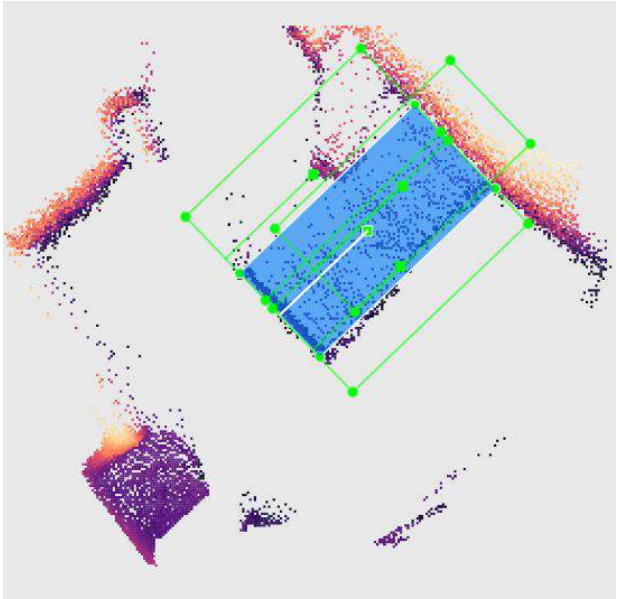
➔ Minimize distance metric

➔ Find new basis with estimated normal vector such that
- Z-axis = height
- Floor points are in z=0

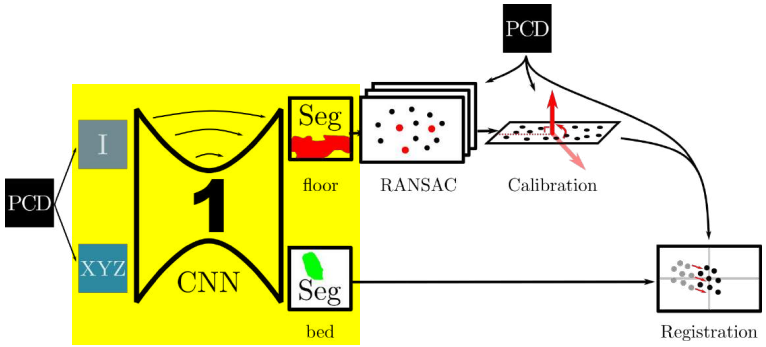
Registration

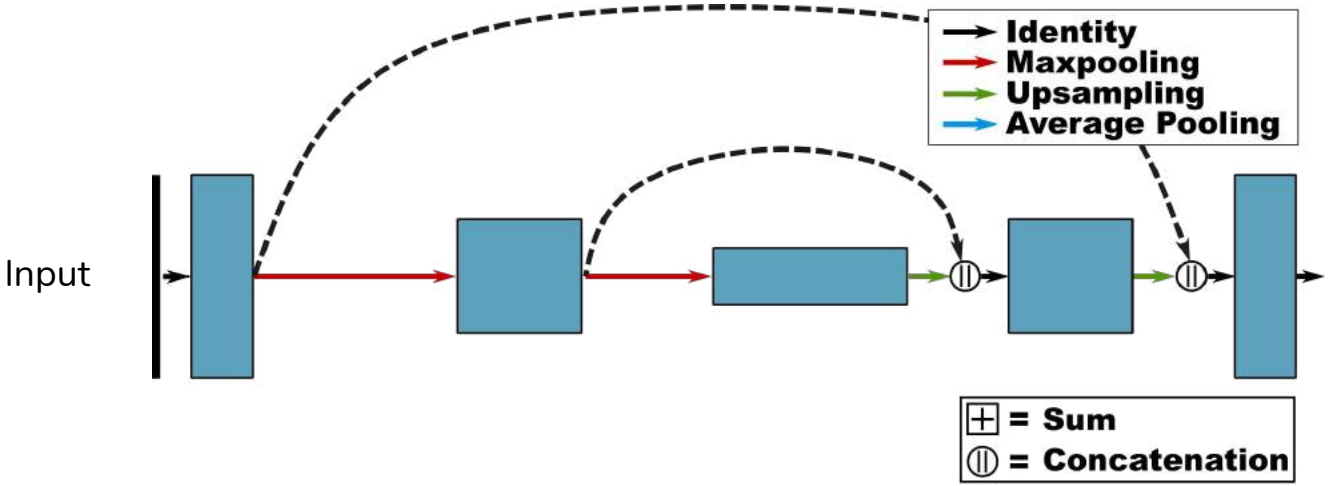


- First guess :
  - SVD for angle
  - Min + Max-Min/2 (in x and y)
- Fitness Brute-force optimization :
  - Metric : Overlap between points and rectangle (= *bed 'model'*)
  - Test all possibilities in a neighbourhood around first guess
- Purposely simple : analyze use of segmentation for localization

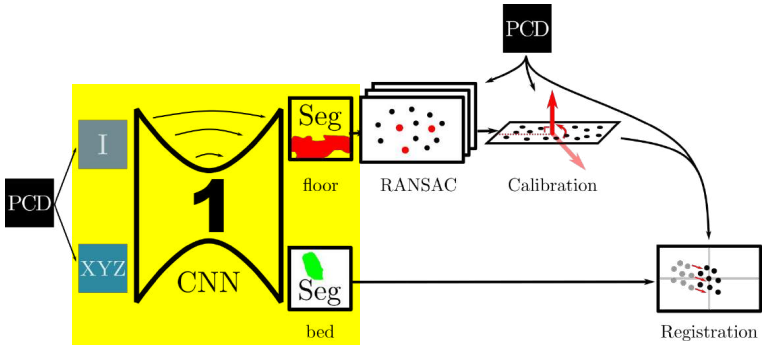


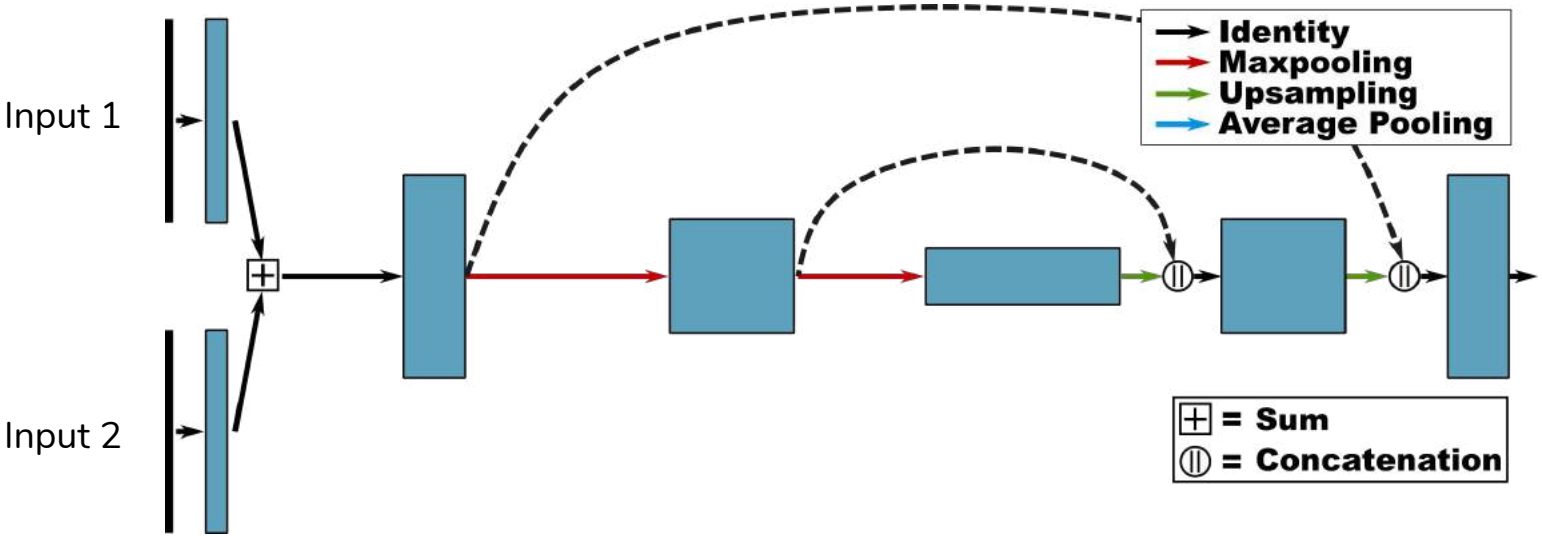
Segmentation CNN



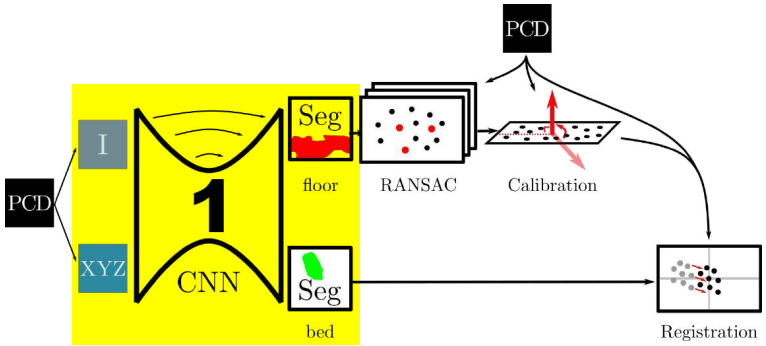


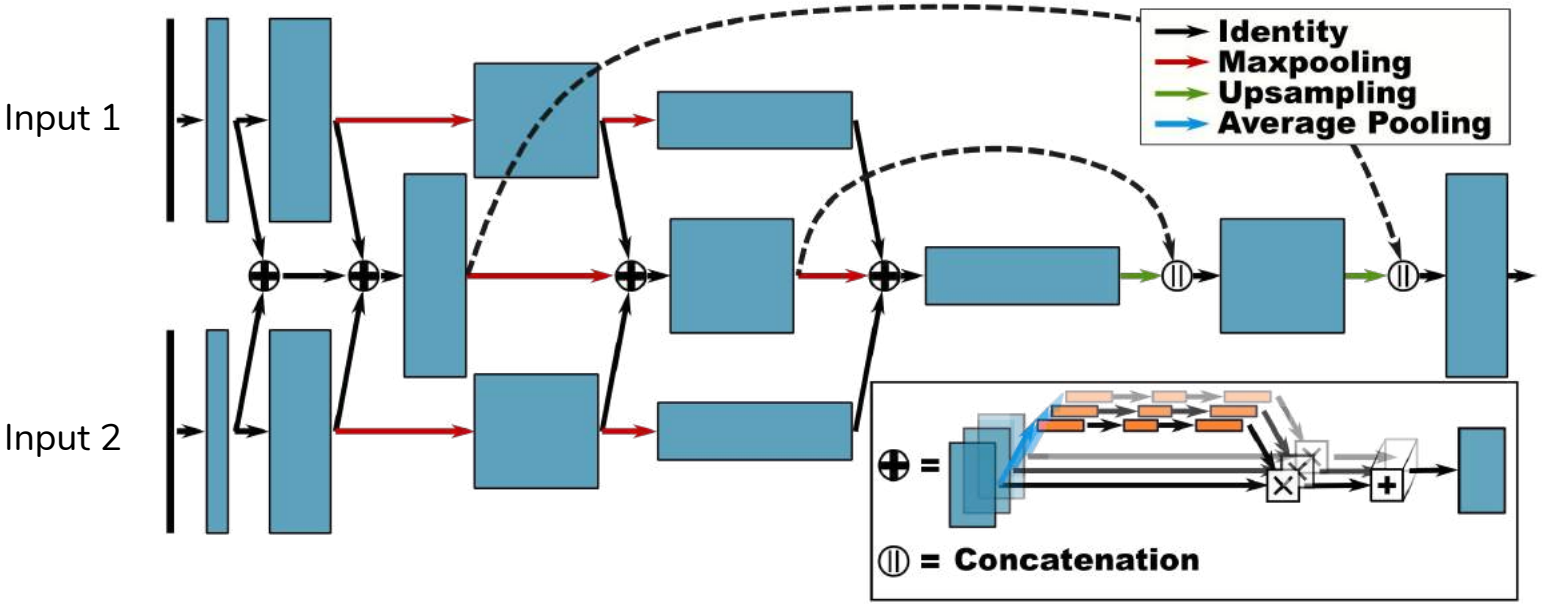
Direct vs Continuous Fusion



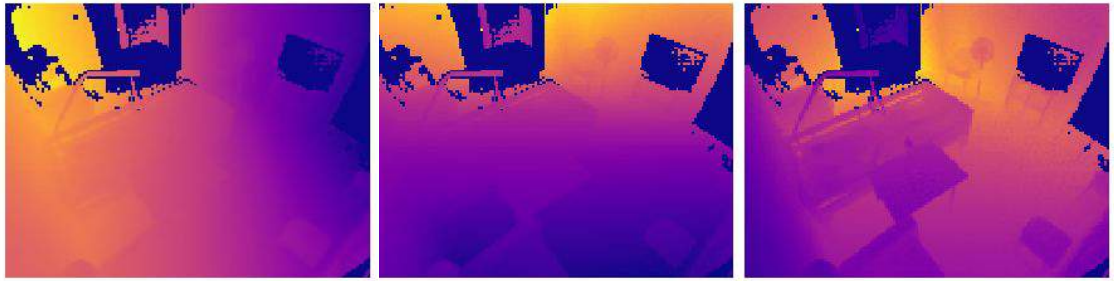


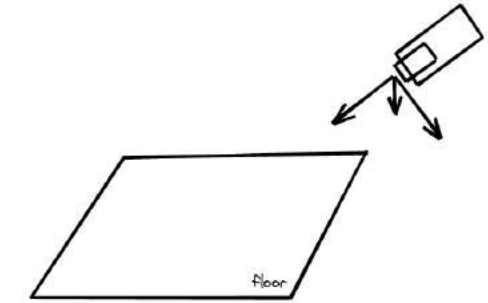
Direct vs Continuous Fusion [6]

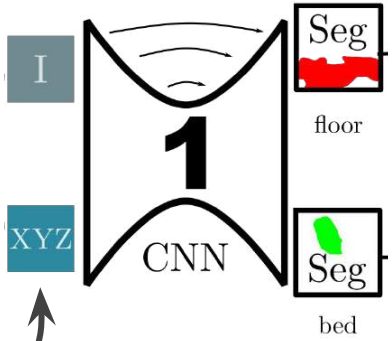




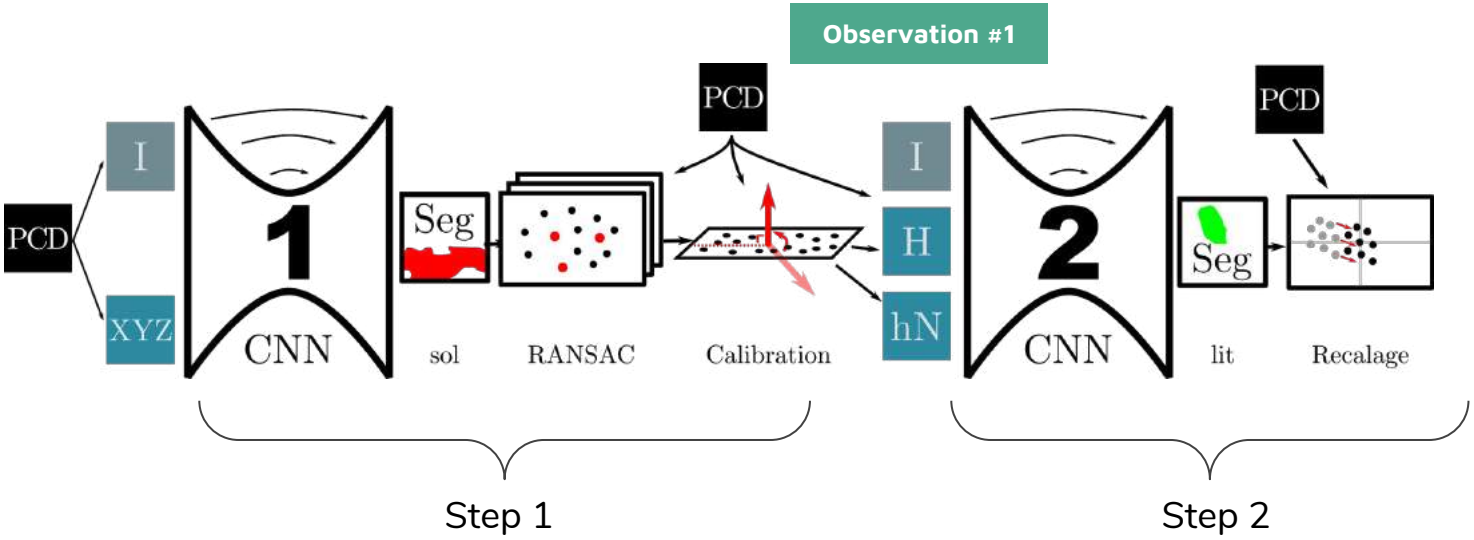
Naive Version – Inputs



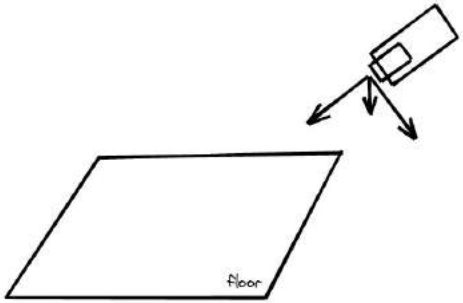


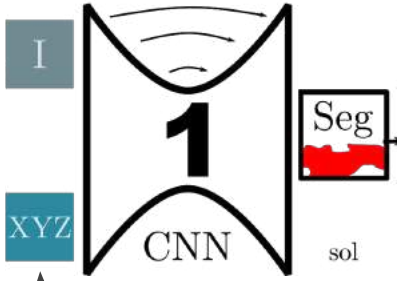


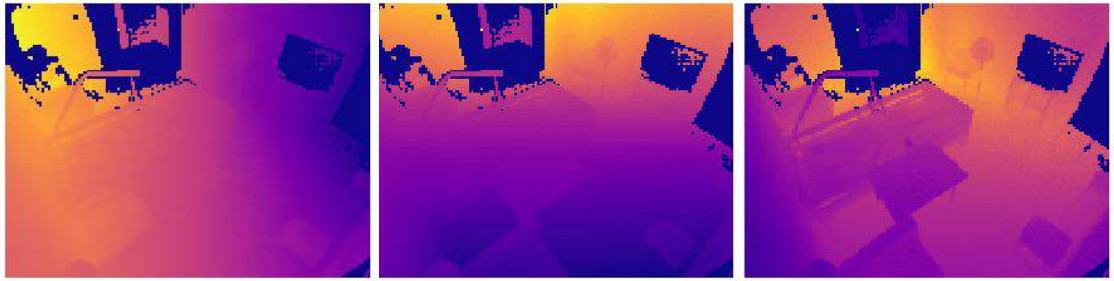
2-Step Version (Tof²Net)



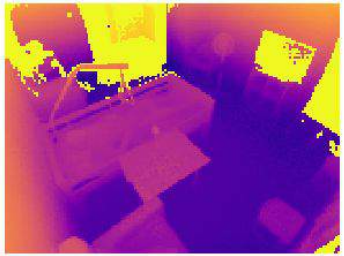
Step 1 – Inputs

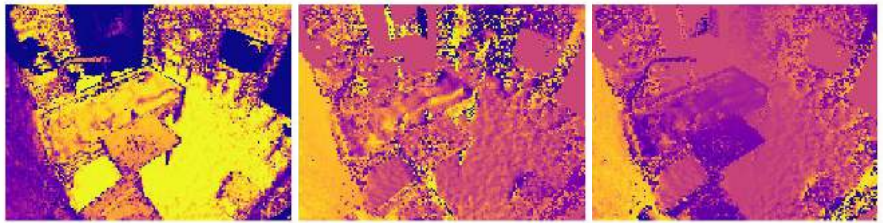


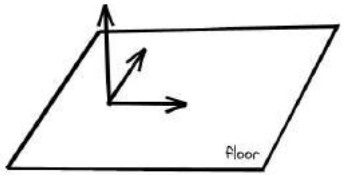


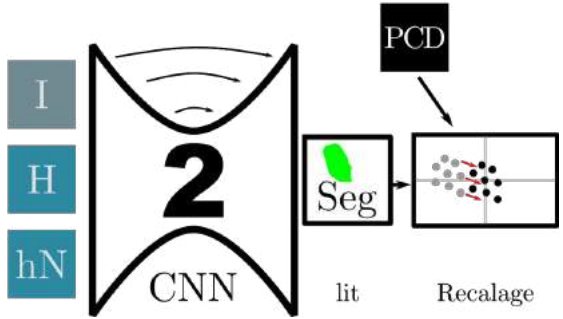


Step 2 – Inputs







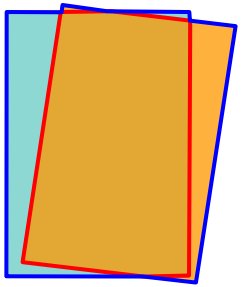


Observation #2

Results

Methodology & Metrics

- 2245 images
  - 8 institutions
  - 51 rooms
- K-Fold cross-validation : 5 folds, for each iteration :
  - 1 training set (3 folds)
  - 1 validation set (1 fold)
  - 1 test set (1 fold)
- Intersection-over-Union (IoU) :
  - Segmentation IoU – pixels
  - Localization IoU – bbox
- Angle difference
  - Calibration “metric”



$$\frac{\text{intersection}}{\text{union}}$$

Quantitative Results - Step 1

Method	Input(s)	Fusion	**Floor** Segmentation			|ΔAngle| (°)
			IoU (%)	Rec. (%)	Prec. (%)	
RANSAC	XYZ	-	-	-	-	13.2
PointNet++	XYZ + I	-	72.7	85.7	82.8	3.37
Ours	I	-	71.6	82.2	78.6	4.74
Ours	I + XYZ	Direct	**82.5**	**90.7**	**90.4**	**1.87**
Ours	I + XYZ	Cont.	81.5	**90.7**	89.4	1.90
Ours	I + XYZ + N	Cont.	80.6	89.0	89.9	1.89

Quantitative Results - Step 1

Method	Input(s)	Fusion	**Floor** Segmentation			|ΔAngle| (°)
			IoU (%)	Rec. (%)	Prec. (%)	
RANSAC	XYZ	-	-	-	-	13.2
PointNet++	XYZ + I	-	72.7	85.7	82.8	3.37
Ours	I	-	71.6	82.2	78.6	4.74
Ours	I + XYZ	Direct	**82.5**	**90.7**	**90.4**	**1.87**
Ours	I + XYZ	Cont.	81.5	**90.7**	89.4	1.90
Ours	I + XYZ + N	Cont.	80.6	89.0	89.9	1.89

Quantitative Results - Step 1

Method	Input(s)	Fusion	**Floor** Segmentation			|ΔAngle| (°)
			IoU (%)	Rec. (%)	Prec. (%)	
RANSAC	XYZ	-	-	-	-	13.2
PointNet++	XYZ + I	-	72.7	85.7	82.8	3.37
Ours	I	-	71.6	82.2	78.6	4.74
Ours	I + XYZ	Direct	**82.5**	**90.7**	**90.4**	**1.87**
Ours	I + XYZ	Cont.	81.5	**90.7**	89.4	1.90
Ours	I + XYZ + N	Cont.	80.6	89.0	89.9	1.89

Quantitative Results - Step 1

Method	Input(s)	Fusion	**Floor** Segmentation			|ΔAngle| (°)
			IoU (%)	Rec. (%)	Prec. (%)	
RANSAC	XYZ	-	-	-	-	13.2
PointNet++	XYZ + I	-	72.7	85.7	82.8	3.37
Ours	I	-	71.6	82.2	78.6	4.74
Ours	I + XYZ	Direct	**82.5**	**90.7**	**90.4**	**1.87**
Ours	I + XYZ	Cont.	81.5	**90.7**	89.4	1.90
Ours	I + XYZ + N	Cont.	80.6	89.0	89.9	1.89

Quantitative Results - Step 1

Method	Input(s)	Fusion	**Floor** Segmentation			|ΔAngle| (°)
			IoU (%)	Rec. (%)	Prec. (%)	
RANSAC	XYZ	-	-	-	-	13.2
PointNet++	XYZ + I	-	72.7	85.7	82.8	3.37
Ours	I	-	71.6	82.2	78.6	4.74
Ours	I + XYZ	Direct	**82.5**	**90.7**	**90.4**	**1.87**
Ours	I + XYZ	Cont.	81.5	**90.7**	89.4	1.90
Ours	I + XYZ + N	Cont.	80.6	89.0	89.9	1.89

Quantitative Results - Step 2

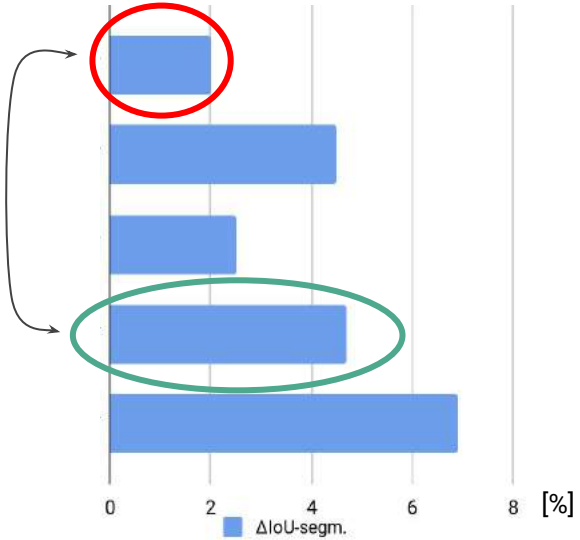
Method	Input(s)	Fusion	**Bed** Segmentation	**Bed** Registration
			IoU (%)	IoU^b (%)
PointNet++	XYZ + I	-	43.3	47.1
ToF-Net	I	-	67.2	69.3
ToF-Net	I + XYZ	Direct	69.2	74.0
ToF-Net	I + XYZ	Cont.	71.7	<u>75.1</u>
ToF-Net	I + XYZ + N	Cont.	69.7	73.2
ToF²-Net	I + H	Direct	<u>71.9</u>	74.3
ToF²-Net	I + H + hN	Cont.	**74.1**	**75.5**

Quantitative Results - Step 2

Method	Input(s)	Fusion	**Bed** Segmentation IoU (%)		**Bed** Registration IoU^b (%)
PointNet++	XYZ + I	-	43.3	Observation #3	47.1
ToF-Net	I	-	67.2		69.3
ToF-Net	I + XYZ	Direct	69.2		74.0
ToF-Net	I + XYZ	Cont.	71.7		<u>75.1</u>
ToF-Net	I + XYZ + N	Cont.	69.7		73.2
ToF²-Net	I + H	Direct	<u>71.9</u>		74.3
ToF²-Net	I + H + hN	Cont.	**74.1**		**75.5**

Quantitative Results - Step 2

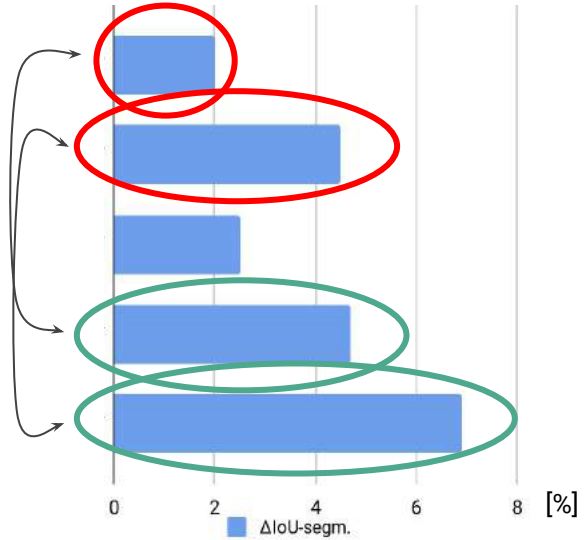
Method	Input(s)	Fusion	**Bed** Segmentation IoU (%)	**Bed** Registration IoU^b (%)
ToF-Net	I	-	67.2	69.3
ToF-Net	I + XYZ	Direct		
ToF-Net	I + XYZ	Cont.		
ToF-Net	I + XYZ + N	Cont.		
ToF²-Net	I + H	Direct		
ToF²-Net	I + H + hN	Cont.		



Observation #1

Quantitative Results - Step 2

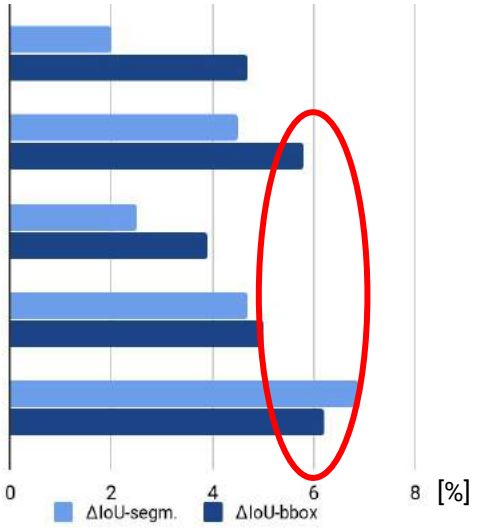
Method	Input(s)	Fusion	**Bed** Segmentation IoU (%)	**Bed** Registration IoU^b (%)
ToF-Net	I	-	67.2	69.3
ToF-Net	I + XYZ	Direct		
ToF-Net	I + XYZ	Cont.		
ToF-Net	I + XYZ + N	Cont.		
ToF²-Net	I + H	Direct		
ToF²-Net	I + H + hN	Cont.		



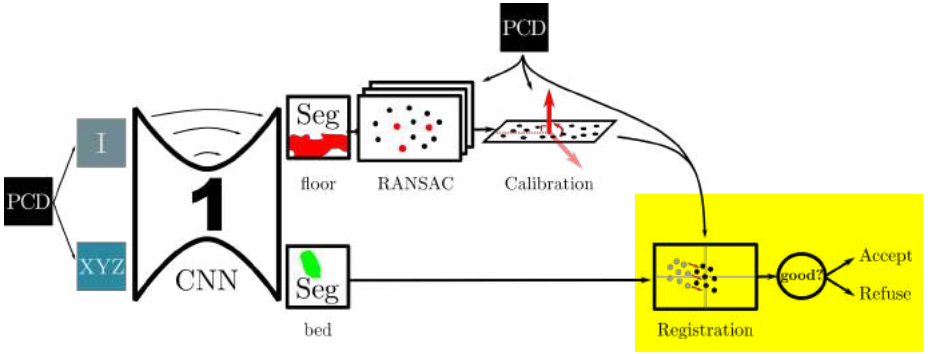
Observation #1

Quantitative Results - Step 2

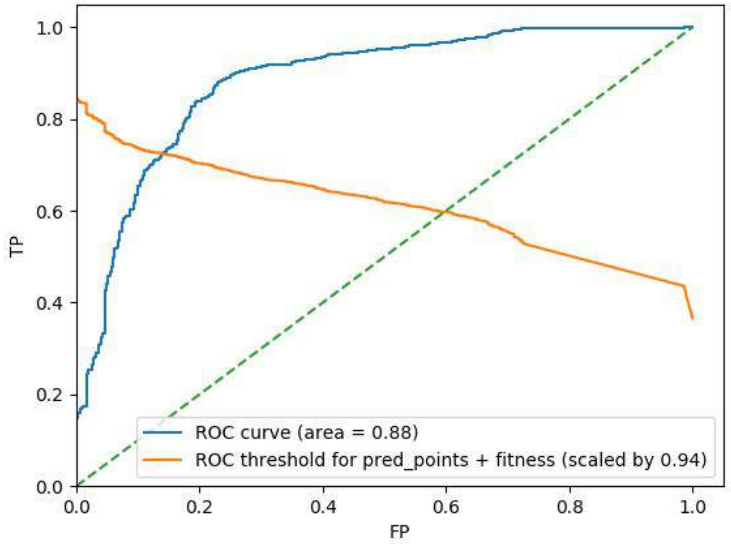
Method	Input(s)	Fusion	**Bed** Segmentation IoU (%)	**Bed** Registration IoU^b (%)
ToF-Net	I	-	67.2	69.3
ToF-Net	I + XYZ	Direct		
ToF-Net	I + XYZ	Cont.		
ToF-Net	I + XYZ + N	Cont.		
ToF²-Net	I + H	Direct		
ToF²-Net	I + H + hN	Cont.		



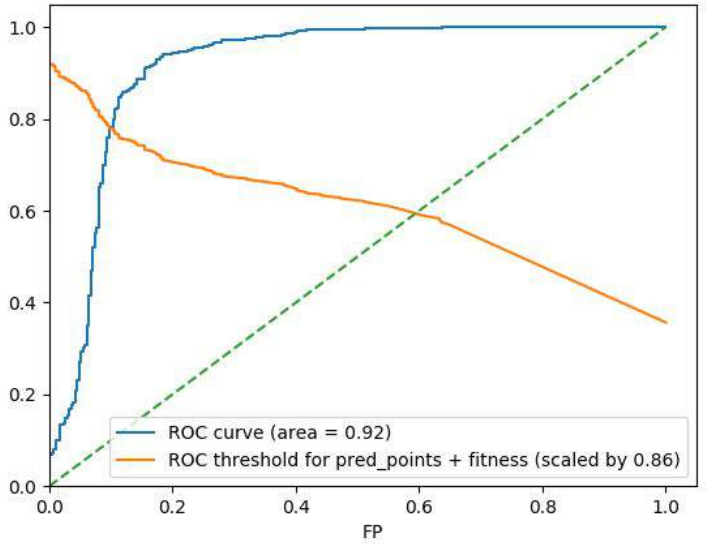
Bad localization detection



1-step (AUC=0.88)

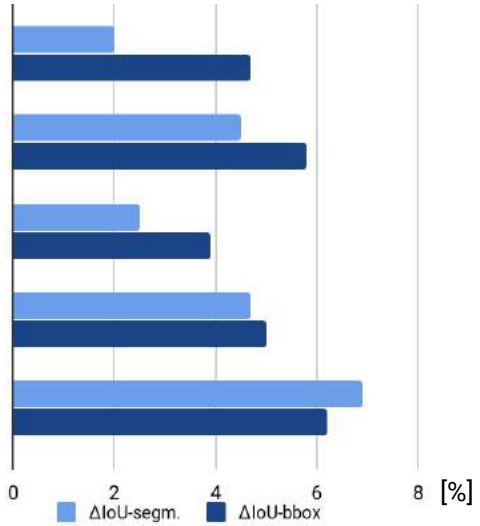


2-step (AUC=**0.92**)



Quantitative Results - Step 2

Method	Input(s)	Fusion	**Bed** Segmentation IoU (%)	**Bed** Registration IoU^b (%)
ToF-Net	I	-	67.2	69.3
ToF-Net	I + XYZ	Direct		
ToF-Net	I + XYZ	Cont.		
ToF-Net	I + XYZ + N	Cont.		
ToF²-Net	I + H	Direct		
ToF²-Net	I + H + hN	Cont.		



Qualitative Results – Step 2

Naïve

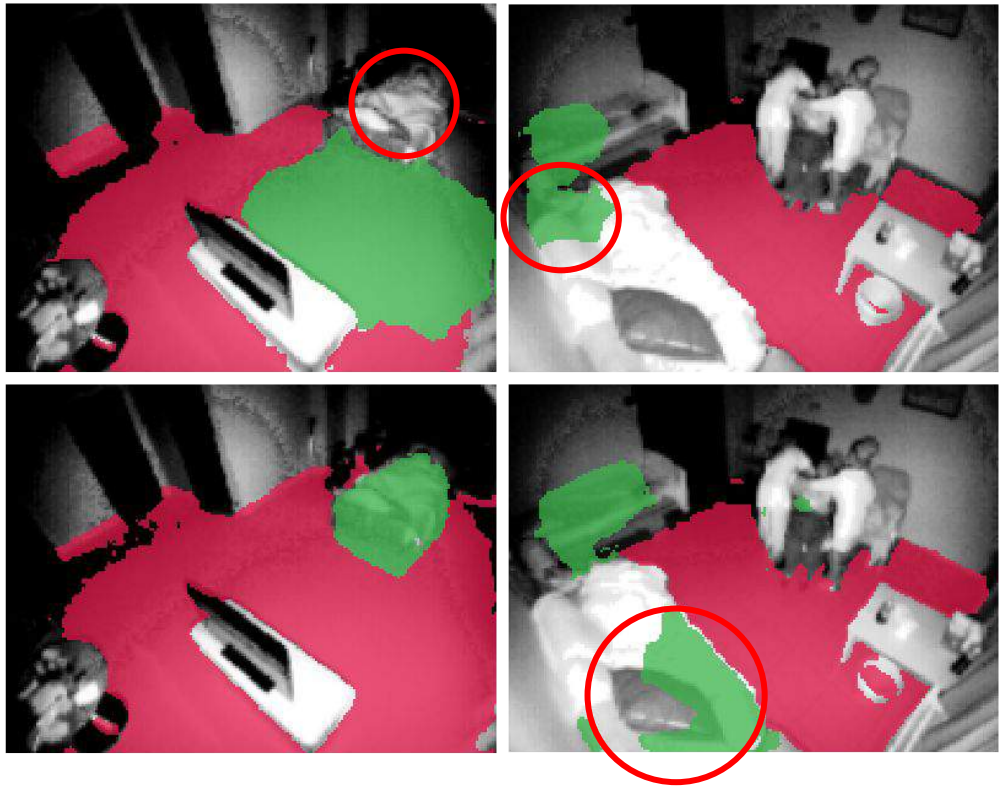
Observation #4

2-Step

Qualitative Results – Step 2

Naïve

2-Step



Conclusion
&
Further work

Conclusion & Further work

Observation #1	Height encoding offers a clear advantage over raw depth information
Observation #2	Adding the *estimated* normals only helps in the height-centric viewpoint
Observation #3	Hierarchical Neural Networks do not perform well on single-viewpoint real-world spatial images
Observation #4	The borders of the naïve/1-step method are less clearly defined, which doesn't help registration
Future work	Instance segmentation can be subjected to the same approach

Sources

[1] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 3431-3440.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science, 9351, 234–241.

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR.

[4] Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. Advances In Neural Information Processing Systems.

[5] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. CVPR, 7132–7141.

[6] Hu, X., Yang, K., Fei, L., & Wang, K. (2019). ACNET: Attention Based Network to Exploit Complementary Features for RGBD Semantic Segmentation. ICIP, 1440–1444.

Questions?