MixedFusion: 6D Object Pose Estimation from Decoupled RGB-Depth Features

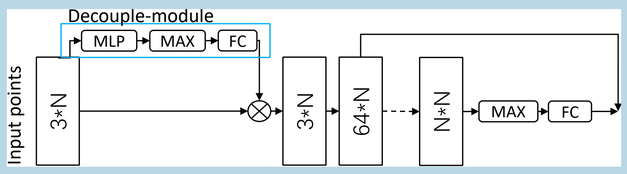
Hangtao Feng, Lu Zhang, Xu Yang, Zhiyong Liu
fenghangtao2018, zhanglu2016, xu.yang, zhiyong.liu}@ia.ac.cn

Problem

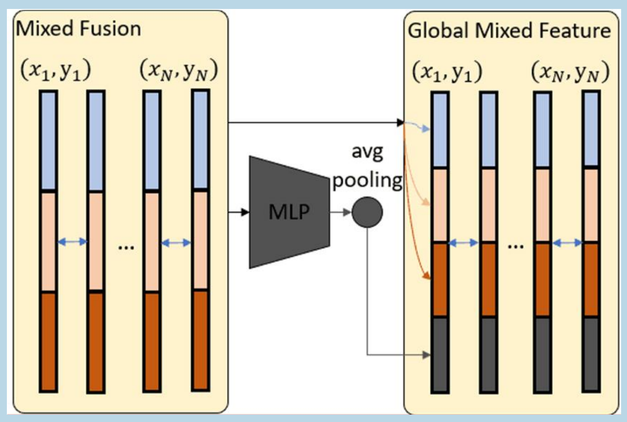
Estimating the 6D pose of objects is an important process for intelligent systems to achieve interaction with the real-world. As the RGB-D sensors become more accessible, the fusion-based methods have prevailed, since the point clouds provide complementary geometric information with RGB values. However, due to the difference in feature space between color image and depth image, the network structures that directly perform point-to-point matching fusion do not effectively fuse the features of the two. We argue that the spatial correspondence of color and point clouds could be decoupled and reconnected, thus enabling a more flexible fusion scheme.

Contributions

Contribution 1 An decouple-module is added to expand the point cloud receptive field, break the point-to-point correspondence, and realize the decoupling of point cloud features and RGB features(MAX represents max-pooling layer and FC represents fully connected layers).



Contribution 2 We design a mixed fusion module, and it concatenates the point embeddings and the color embeddings. We also concatenate the mixed fusion feature embeddings and the global feature embeddings as the input feature map of the posenet [2] to fully mine the performance of the network.



References

[1] Xiang, Yu and Schmidt, Tanner and Narayanan, Venkatraman and Fox, Dieter: *Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes*, arXiv preprint arXiv:1711.00199

[2] Wang, Chen and Xu, Danfei and Zhu, Yuke and Martín-Martín, Roberto and Lu, Cewu and Fei-Fei, Li and Savarese, Silvio: *Densefusion: 6d object pose estimation by iterative dense fusion*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2019)

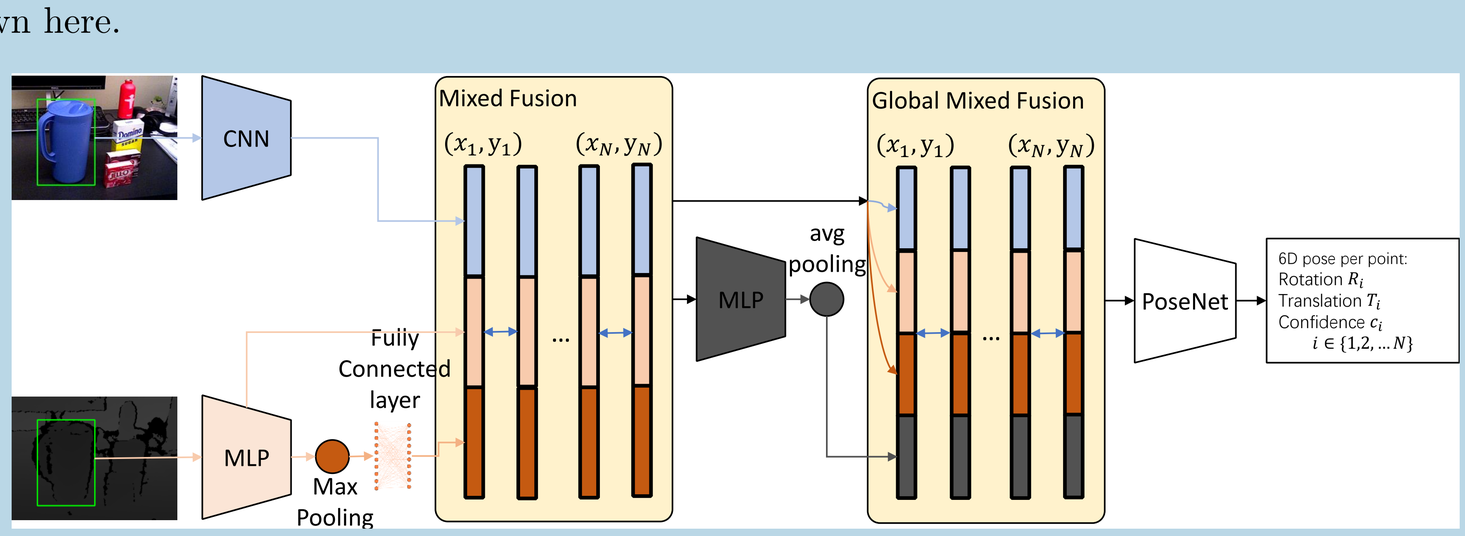
[3] Hinterstoisser, Stefan and Lepetit, Vincent and Ilic, Slobodan and Holzer, Stefan and Bradski, Gary and Konolige, Kurt and Navab, Nassir: *Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes*. Asian conference on computer vision (2012)

Acknowledgements

This work is fully supported by National Key Research and Development Plan of China grant 2017YFB1300202, NSFC, China grants U1613213, 61627808, and the Strategic Priority Research Program of Chinese Academy of Science under Grant XDB32050100.

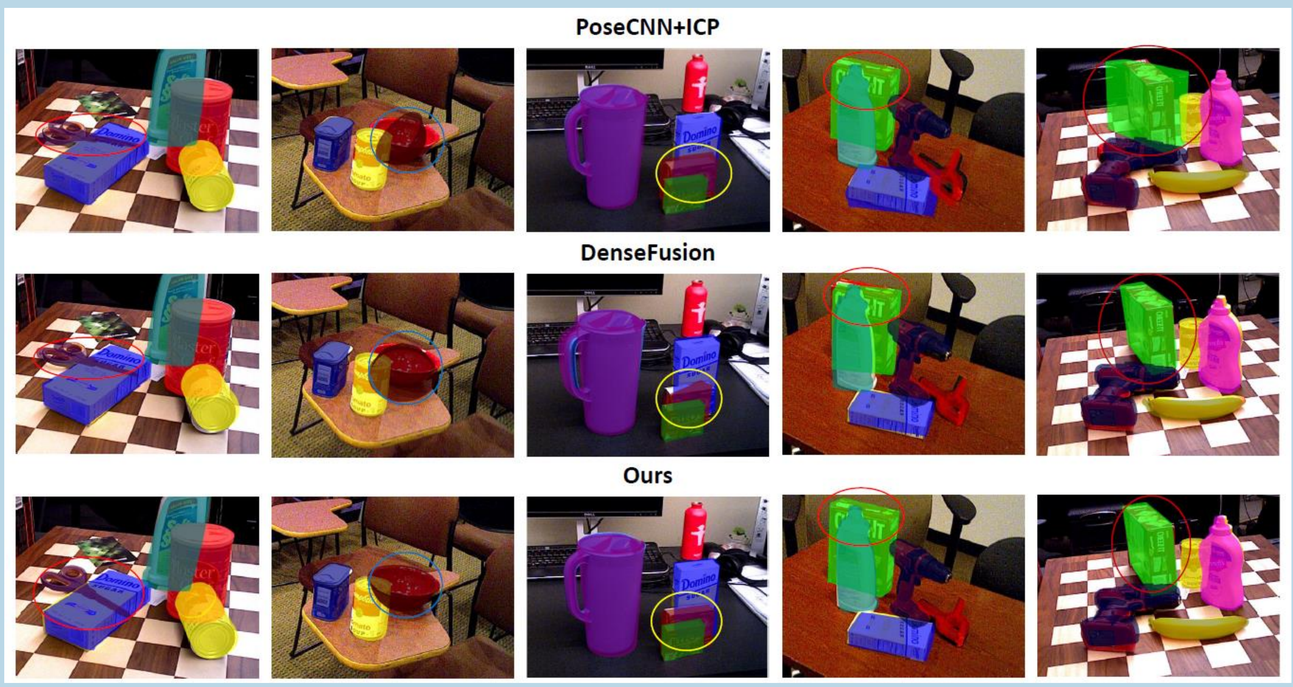
An Overview of Our Model

Overview of the proposed MixedFusion. Our method uses segmentation masks to get object information from RGB and depth images, and then gets the RGB features and the point cloud features obtained from the depth image. The obtained RGB features are matched with the point cloud features, in this process, the two features are automatically decoupled and reconnected. The posenet [2] outputs a 6D pose prediction of the object. We also have an iteratively refine network that is not shown here.



Experiments

Visualized results on the YCB-Video Dataset. All the methods are tested with the same segmentation masks, and we use twice optimization iterations for DenseFusion and ours. We get the poses of the point cloud based on the outputs by these networks, project them to an RGB image and visualize them, the differences have been highlighted with circles. From this figure, we can see that in most cases, the visualization of our model is better than DenseFusion and PoseCNN+ICP,



Results on the LineMod Dataset [3] and the YCB-Video Dataset [1]. Object with italic name are symmetirc. Best results in a given row are in bold.
Results on the LineMod Datset.

Object	SSD-6D +ICP	PointFusion	DenseFusion	Ours
ape	65.0	70.4	92.3	**96.9**
bench vi.	80.0	80.7	93.2	**97.7**
camera	78.0	60.8	94.4	**98.5**
can	86.0	61.1	93.1	**97.8**
cat	70.0	79.1	96.5	**98.2**
driller	73.0	47.3	87.0	**94.2**
duck	66.0	63.0	92.3	**96.8**
egg-box	100.0	99.9	99.8	100.0
glue	100.0	99.3	100.0	99.5
hole p.	49.0	71.8	92.1	**96.0**
iron	78.0	83.2	97.0	**98.6**
lamp	73.0	62.3	95.3	**98.4**
phone	79.0	78.8	92.8	**98.3**
Average	79.0	73.7	94.3	**97.8**

Results on the YCB-Video Datset.

Index	AUC			<1.0			<1.5			<2.0		
	P-ICP	DF	Ours	P-ICP	DF	Ours	P-ICP	DF	Ours	P-ICP	DF	Ours
002	95.77	**96.44**	95.98	99.50	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
003	92.72	95.78	**96.58**	84.79	97.00	**98.39**	90.09	98.36	**99.31**	91.59	99.31	**99.65**
004	**98.23**	97.59	97.87	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.0	100.00
005	94.45	94.44	**94.53**	96.80	96.73	**96.80**	96.87	96.94	96.94	96.94	96.94	96.94
006	**98.62**	97.35	97.71	98.88	98.88	**100.00**	99.72	100.00	100.00	100.00	100.00	100.00
007	**97.14**	97.07	96.77	97.62	**99.30**	96.65	98.24	**99.91**	99.82	99.74	100.00	100.00
008	**97.93**	95.93	95.75	**100.00**	99.07	98.60	100.00	99.53	100.00	100.00	99.53	100.00
009	**98.77**	97.96	98.33	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
010	**92.77**	90.82	91.40	83.44	**87.52**	84.36	90.54	**92.38**	91.33	**93.69**	92.90	92.77
011	**97.05**	96.34	96.95	94.99	98.68	**99.21**	99.47	99.74	**100.00**	99.74	100.00	100.00
019	**97.84**	97.51	97.68	99.65	**99.82**	99.65	100.00	100.00	100.00	100.00	100.00	100.00
021	**96.88**	95.90	96.20	95.23	99.12	**99.61**	96.59	99.81	**99.90**	99.42	99.81	**99.90**
024	81.15	**89.38**	87.91	43.39	**55.11**	44.64	48.13	**55.36**	44.64	55.61	94.51	**94.76**
025	94.93	96.69	**97.27**	97.61	97.61	**99.84**	99.84	99.68	**100.00**	99.84	99.84	**100.00**
035	**98.24**	95.99	96.60	**99.34**	97.07	98.49	**99.62**	98.58	99.43	**99.62**	98.86	99.43
036	87.51	92.77	**93.03**	74.58	**87.50**	87.50	79.17	**98.33**	97.92	80.00	**99.58**	98.75
037	91.67	92.06	**94.40**	67.96	73.48	**90.06**	86.19	95.58	**96.69**	95.58	**100.00**	99.45
040	97.21	97.58	**97.67**	97.07	100.00	100.00	99.38	100.00	100.00	99.69	100.00	100.00
051	**75.94**	73.22	72.96	**67.94**	34.04	26.95	71.91	71.77	**76.88**	75.60	**79.43**	79.15
052	64.24	69.79	**74.03**	38.52	17.33	**46.67**	45.33	24.59	**61.04**	48.89	74.37	**74.81**
061	**97.23**	91.96	95.26	99.65	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Average	93.01	93.20	**93.64**	89.69	89.08	**90.03**	91.98	92.57	**94.38**	93.32	96.70	**96.77**