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

Feng Hangtao

ICPR 2020

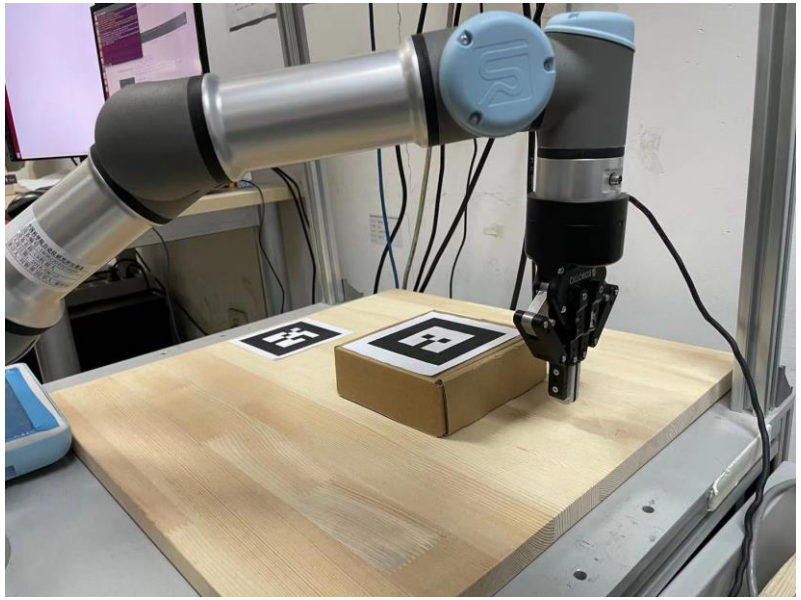
January 10-15, 2021

The plan

01 Introduction

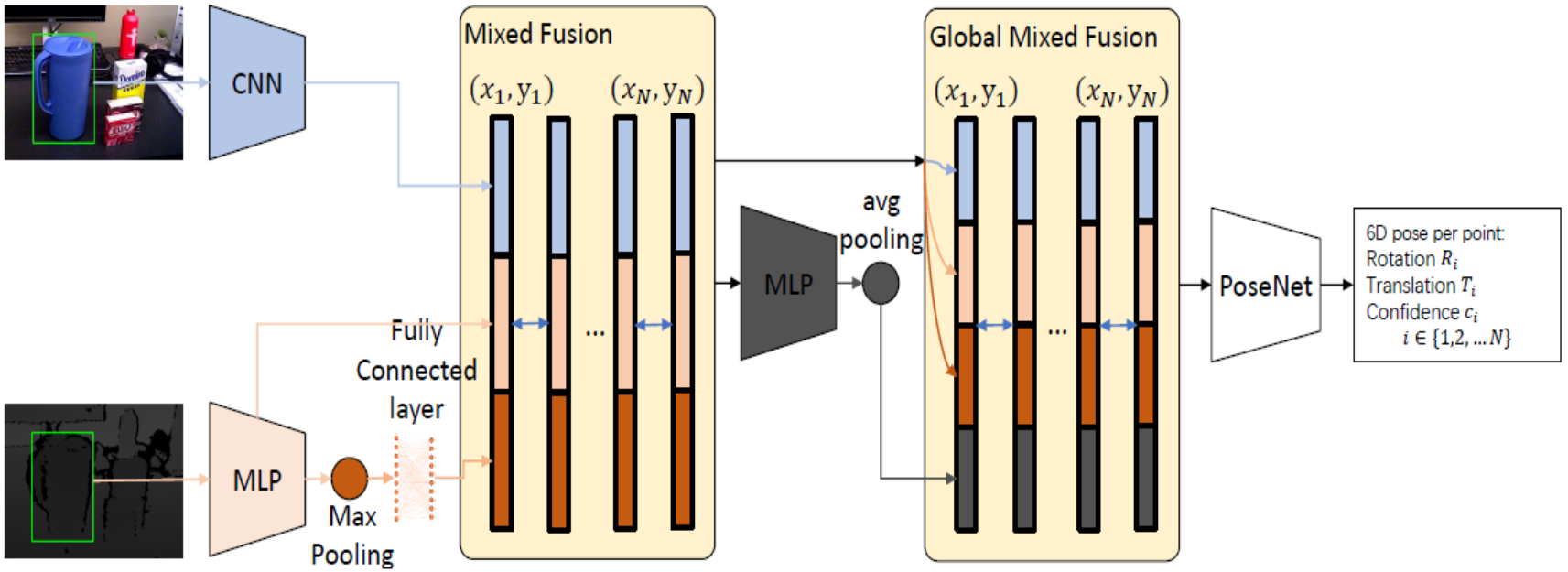
02 Contributions

03 Experiments



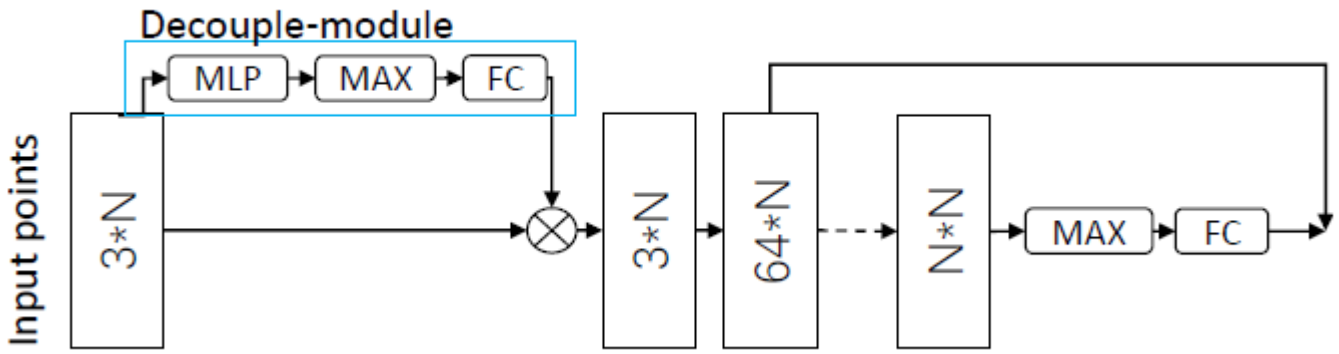
Estimating the 6D pose of objects is an important process for intelligent systems to achieve interaction with the real-world. As the RGB-D sensors become more accessible, the fusion-based methods have prevailed, since the point clouds provide complementary geometric information with RGB values. However, 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.

An overview of our model



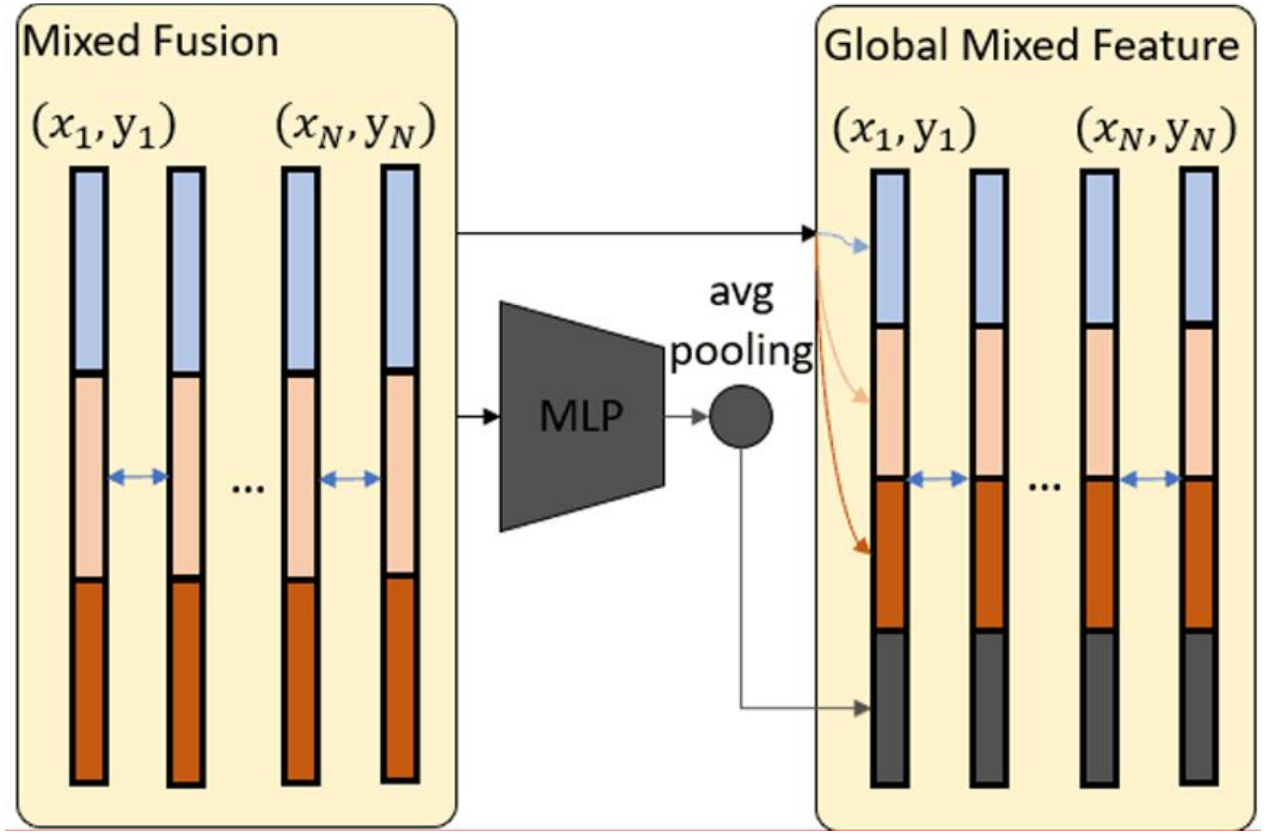
Overview of the proposed MixedFusion. Our method uses segmentation masks to get object information from RGB and depth images, and then gets the RGB features and the point cloud features obtained from the depth image. The obtained RGB features are matched with the point cloud features, in this process,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

Contribution 1:



Network architecture for processing point clouds. In this figure, decouple-module indicates a decouple-network, and a dotted line indicates a multilayer network structure. The core of this module is to use the fully connected operation 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:

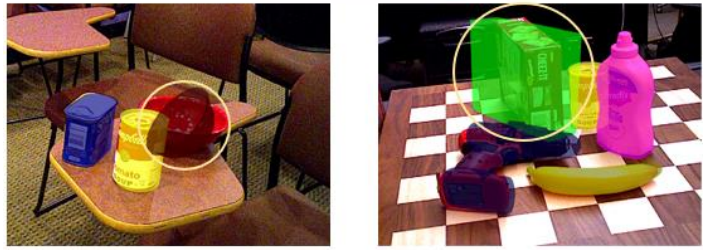


We design a mixed fusion module, and it concatenates the point embeddings and the color embeddings. We also concatenate the mixed fusion feature embeddings and the global feature embeddings as the input feature map of the posenet to fully mine the performance of the network.

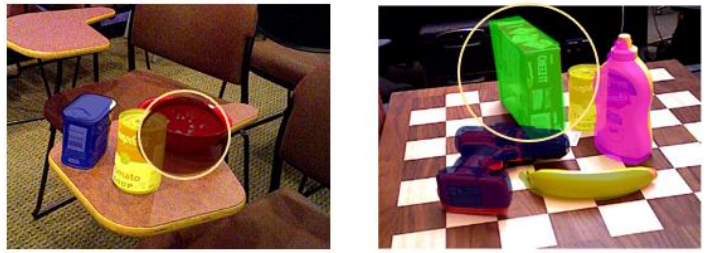
Experiment results on YCB-Video dataset:

Index	AUC			<1.0		
	P-ICP	DF	Ours	P-ICP	DF	Ours
002	95.77	**96.44**	95.98	99.50	100.00	100.00
003	92.72	95.78	**96.58**	84.79	97.00	**98.39**
004	**98.23**	97.59	97.87	100.00	100.00	100.00
005	94.45	94.44	**94.53**	96.80	96.73	**96.80**
006	**98.62**	97.35	97.71	98.88	98.88	**100.00**
007	**97.14**	97.07	96.77	97.62	**99.30**	96.65
008	**97.93**	95.93	95.75	**100.00**	99.07	98.60
009	**98.77**	97.96	98.33	100.00	100.00	100.00
010	**92.77**	90.82	91.40	83.44	**87.52**	84.36
011	**97.05**	96.34	96.95	94.99	98.68	**99.21**
019	**97.84**	97.51	97.68	99.65	**99.82**	99.65
021	**96.88**	95.90	96.20	95.23	99.12	**99.61**
024	81.15	**89.38**	87.91	43.39	**55.11**	44.64
025	94.93	96.69	**97.27**	97.61	97.61	**99.84**
035	**98.24**	95.99	96.60	**99.34**	97.07	98.49
036	87.51	92.77	**93.03**	74.58	**87.50**	87.50
037	91.67	92.06	**94.40**	67.96	73.48	**90.06**
040	97.21	97.58	**97.67**	97.07	100.00	100.00
051	**75.94**	73.22	72.96	**67.94**	34.04	26.95
052	64.24	69.79	**74.03**	38.52	17.33	**46.67**
061	**97.23**	91.96	95.26	99.65	100.00	100.00
Average	93.01	93.20	**93.64**	89.69	89.08	**90.03**

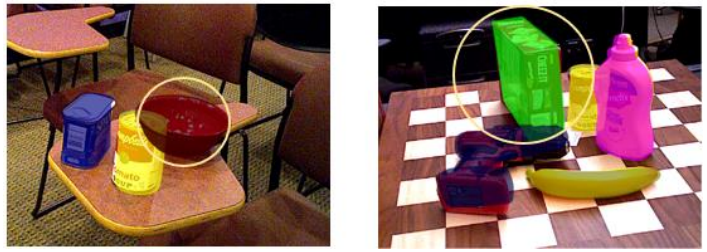
PoseCNN+ICP



DenseFusion



Ours



Experiment results on LineMod dataset:

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**

Thank you!

Ablation studies(results on LineMod Dataset)

Object	Ours (N/DM)	Ours (Avg)	Ours (N/MF)	Ours (N/M)	Ours (N/F)	Ours (ALL)
ape	89.1	84.6	94.1	92.9	94.5	**96.9**
bench vi.	93.5	90.8	93.7	93.6	95.9	**97.7**
camera	96.1	91.6	96.2	96.5	97.7	**98.5**
can	96.7	94.5	96.6	97.0	96.0	**97.8**
cat	96.0	95.4	97.4	97.6	97.1	**98.2**
driller	93.7	91.4	93.8	92.1	92.3	**94.2**
duck	90.6	89.7	94.8	93.4	95.7	**96.8**
egg-box	99.9	99.8	99.8	99.8	99.7	**100.0**
glue	99.6	99.6	99.5	99.5	**99.6**	99.5
hole p.	93.1	85.2	**96.8**	95.2	96.4	96.0
iron	97.1	95.8	**98.9**	96.6	97.8	98.6
lamp	97.5	95.4	97.2	97.9	96.0	**98.4**
phone	95.8	93.9	96.6	96.4	98.0	**98.3**
Average	95.3	92.9	96.6	96.1	96.7	**97.8**