Densely Regressed Pose Estimation

KIST
송명하

6D Pose Estimation with Correlation Fusion

Yi Cheng1，Hongyuan Zhu1，Cihan Acar1，Wei Jing2, 3, Y
an Wu1, 4, Liyuan Li1，Cheston Tan1，Joo-Hwee Lim

[2019 arxiv]

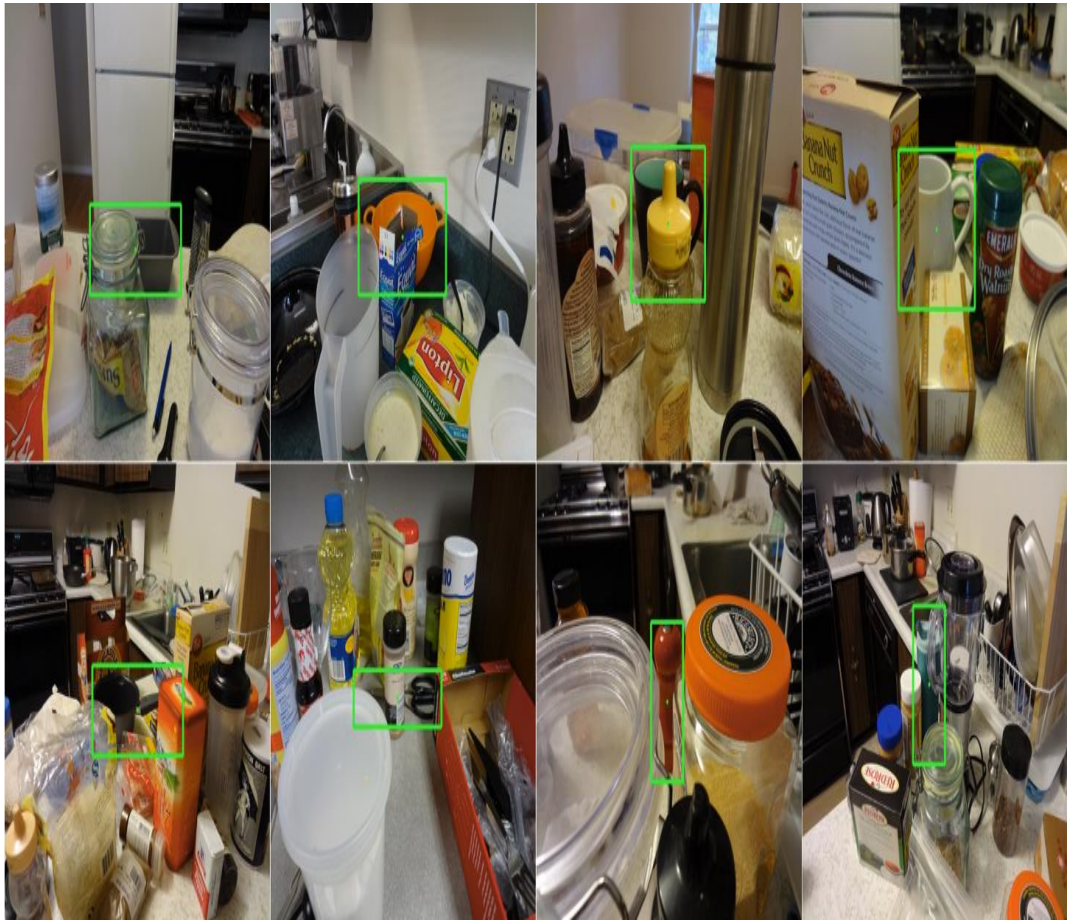
Content

1. Introduction

http://www.autodaily.co.kr/news/articleView.html?idxno=408344

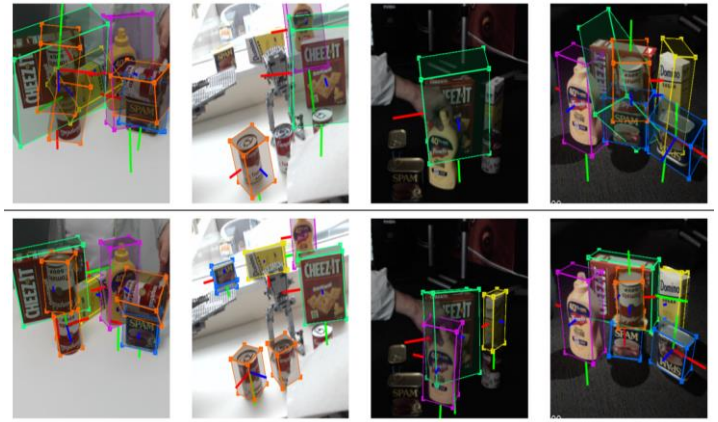
https://www.independent.co.uk/extras/indybest/gadgets-tech/video-games-consoles/augmented-reality-games-ar-android-ios-apple-google-play-tech-a8506831.html

1. Introduction



https://www.cs.cmu.edu/~hebert/occarbview.html

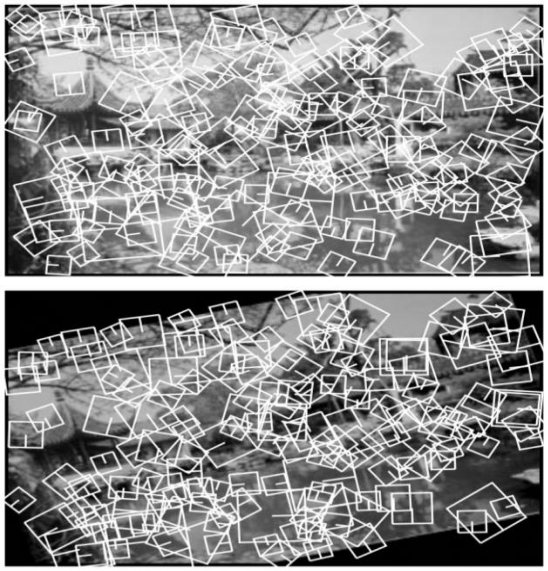
https://www.photoreview.com.au/tips/shooting/how-to-control-image-noise/



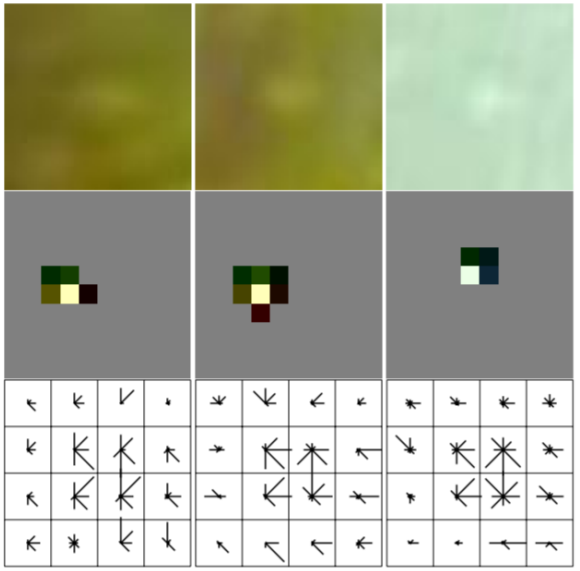
Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects

2. Related Work

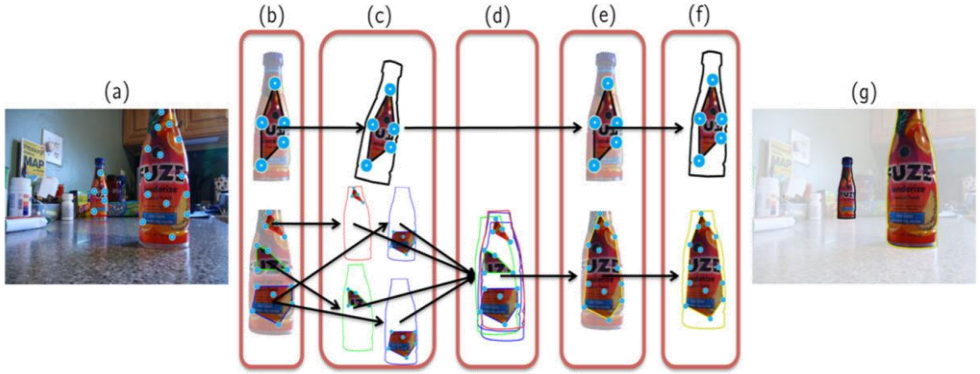
KeyPoints Matching.



Object Recognition from Local Scale-Invariant Features. 1999 ICCV



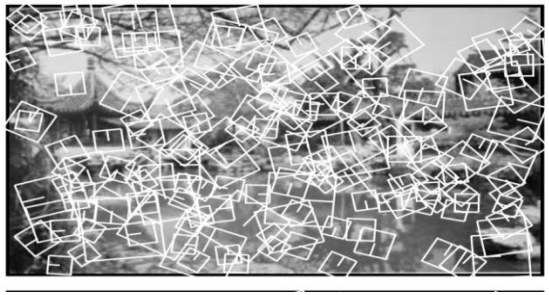
3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. IJCV 2006



The moped framework: Object recognition and pose estimation for manipulation 2011

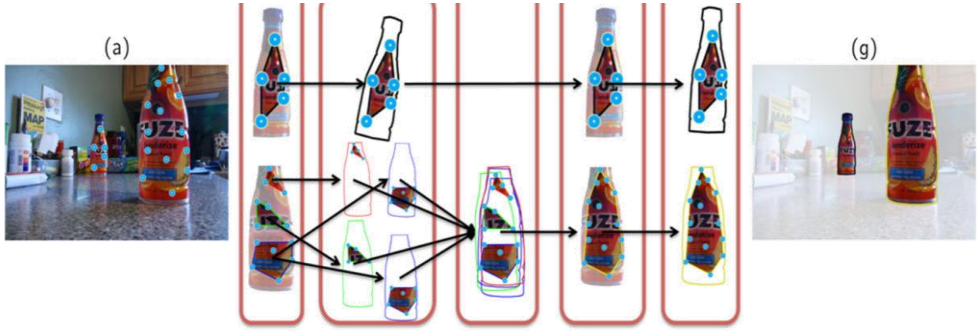
2. Related Work

KeyPoints Matching.



Occlusion Good

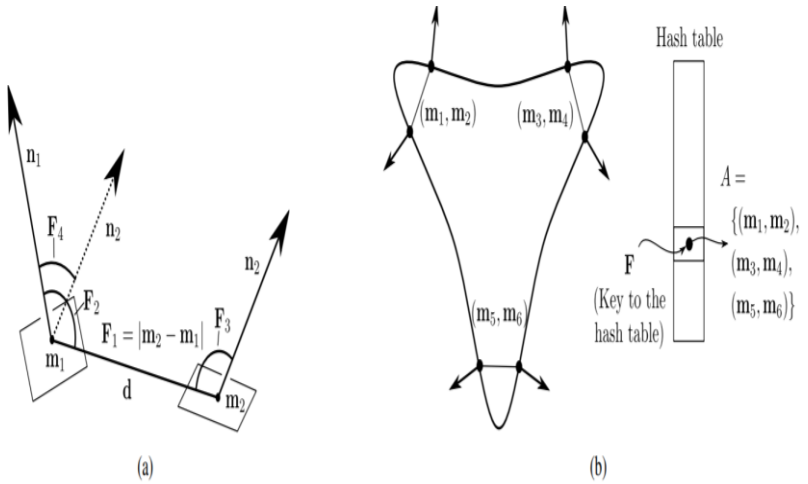
Texture Less Bad



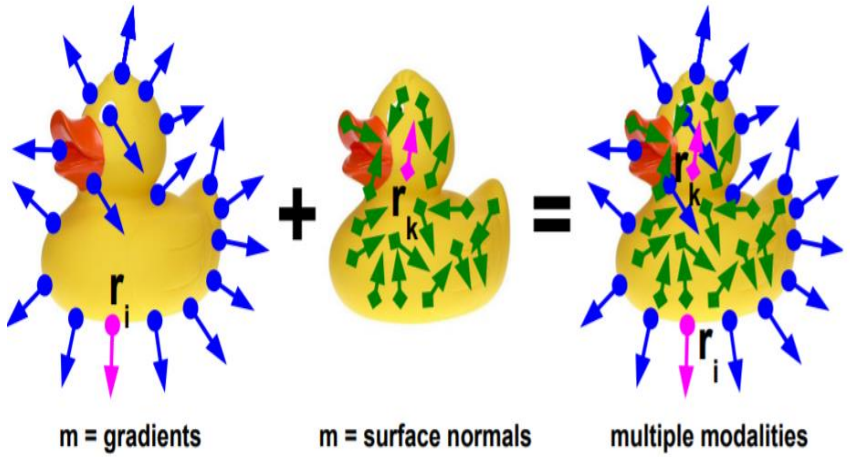
The moped framework: Object recognition and pose estimation for manipulation 2011

2. Related Work

Templated Based



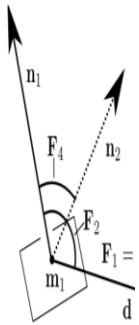
Match Locally: Efficient and Robust 3D Object Recognition 2010 CVPR

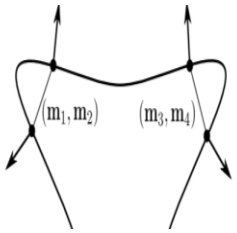


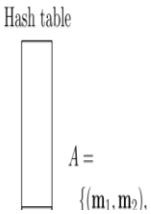
Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes,2011 ICCV

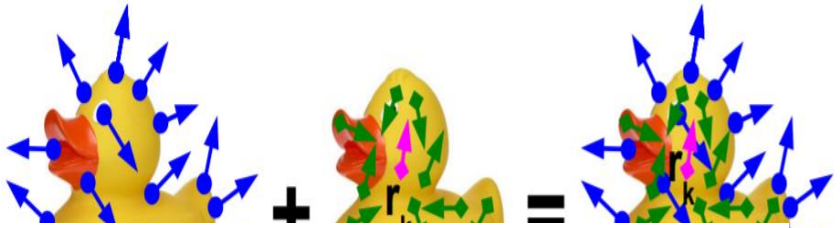
2. Related Work

Templated Based









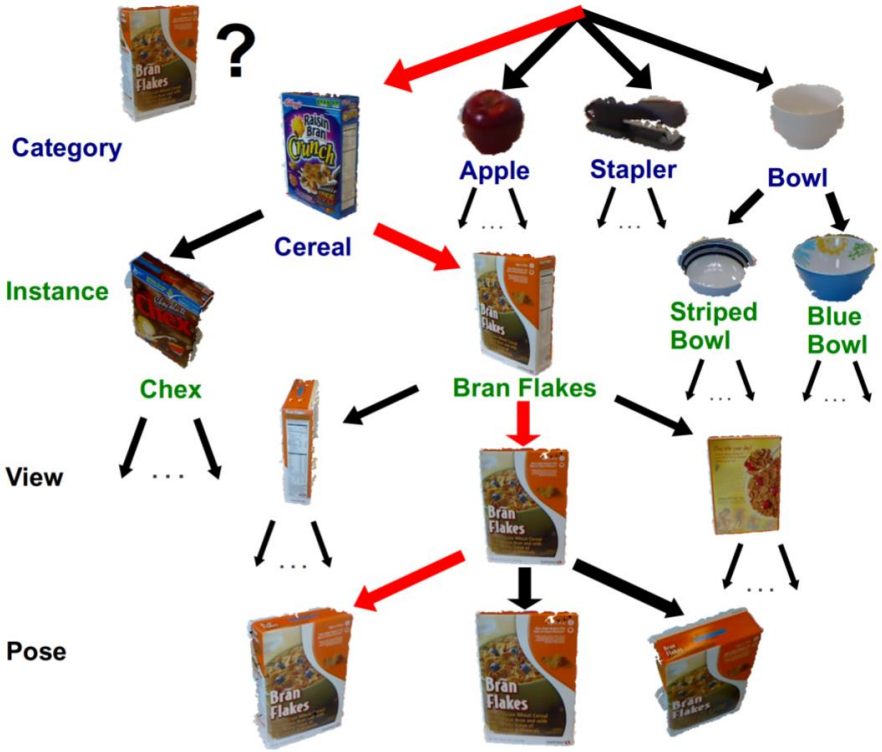
Texture Less Good

Occlusion Bad

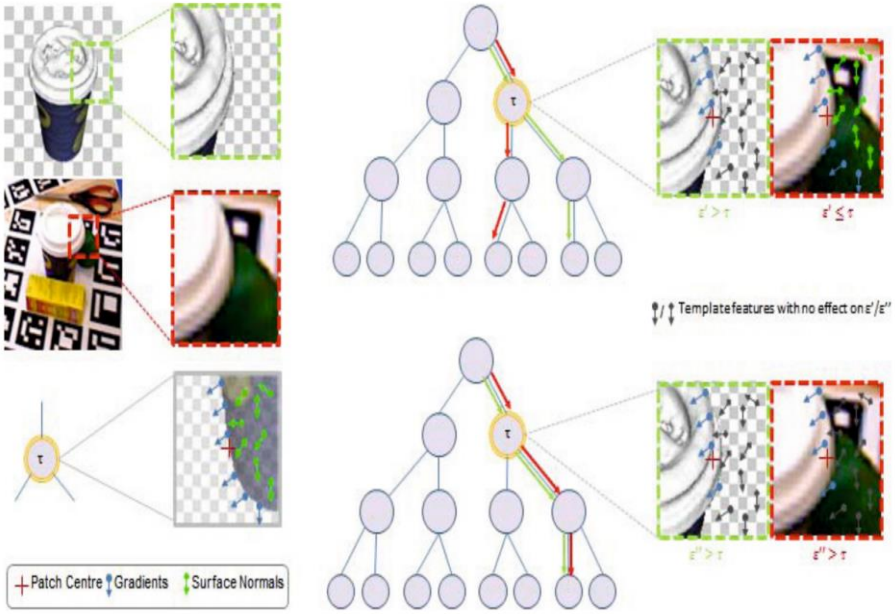
Match Locally

2. Related Work

Tree Based



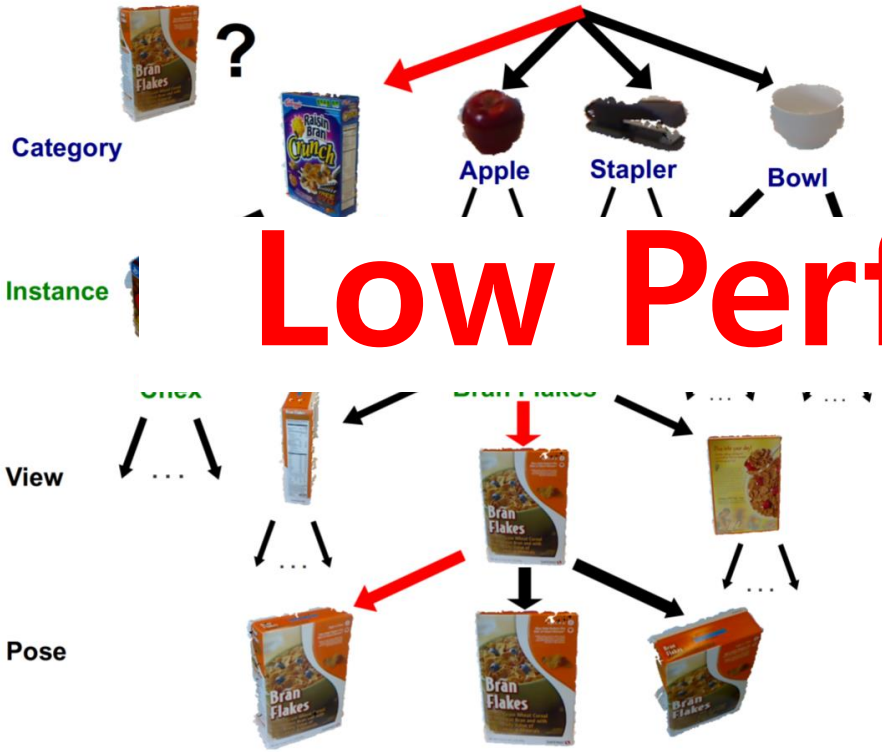
A Scalable Tree-based Approach for Joint Object and Pose Recognition 2011 AAAI



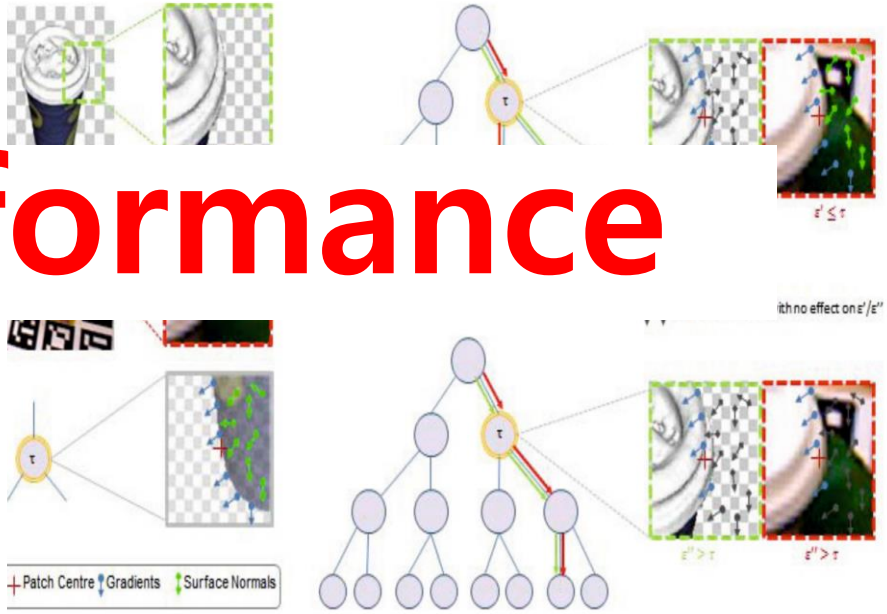
LatentClass Hough Forests for 3D Object Detection and Pose Estimation 2014 ECCV

2. Related Work

Tree Based



Low Performance

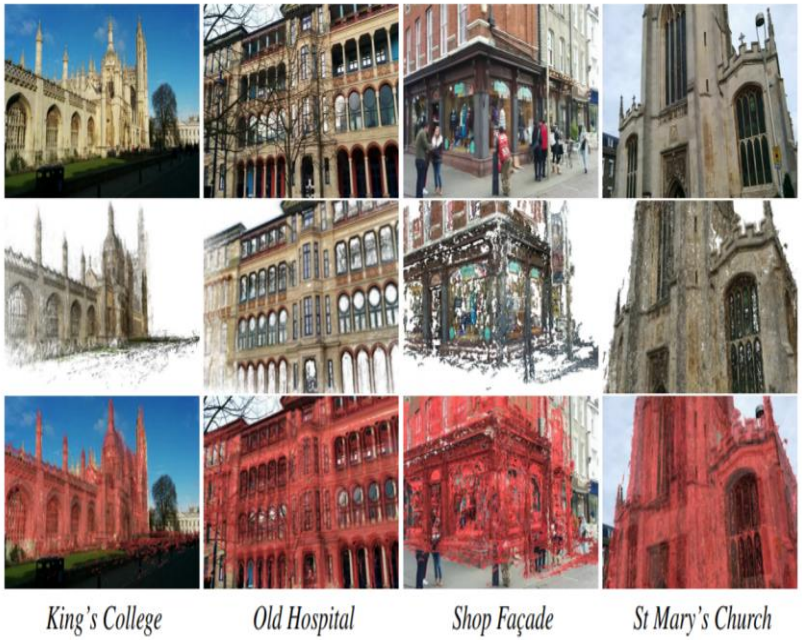


A Scalable Tree-based Approach for Joint Object and Pose Recognition 2011 AAAI

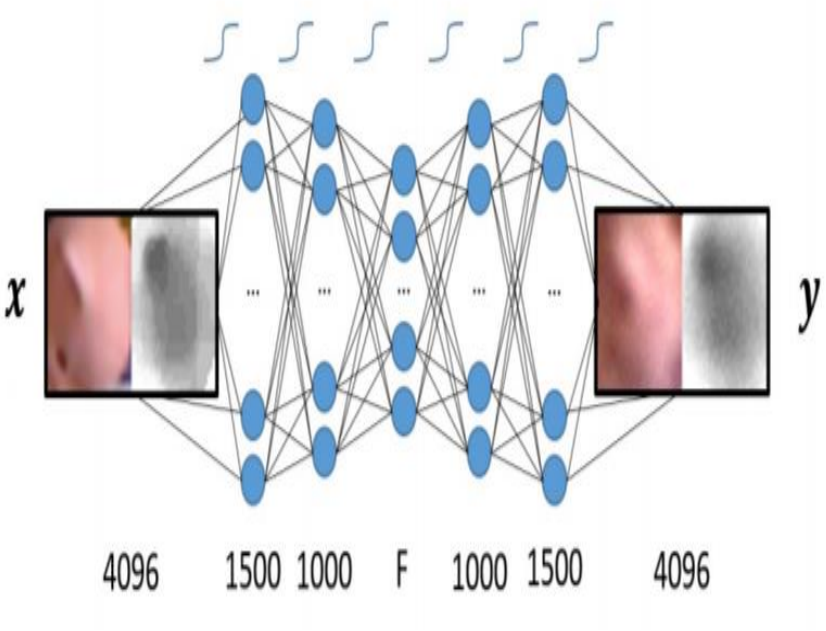
LatentClass Hough Forests for 3D Object Detection and Pose Estimation 2014 ECCV

2. Related Work

CNN



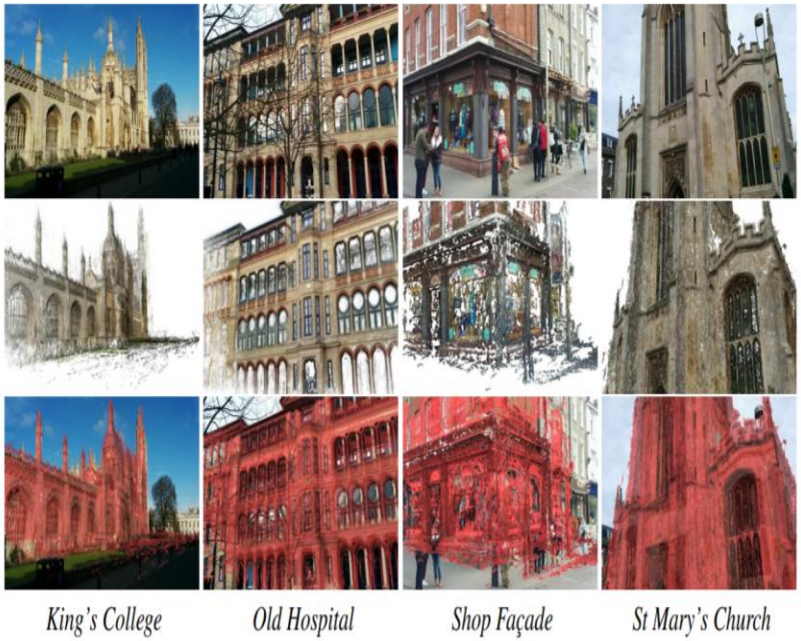
Geometric loss functions for camera pose regression with deep learning 2016 CVPR



Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation ECCV 2016

2. Related Work

CNN



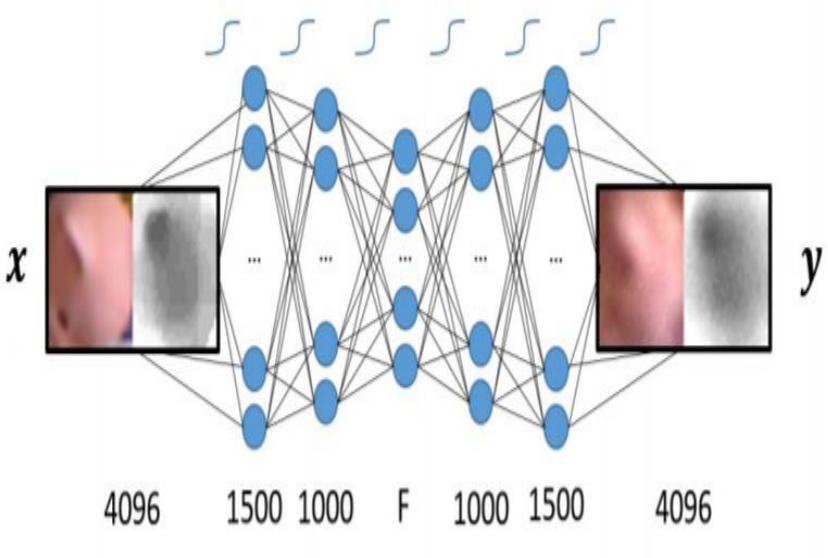
PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization 2015 ICCV

- **BackBone GoogLenet**
- **Direct Regression**
  3D Translation and 3D Rotation
- **카메라 포즈에 한정.**

$$loss(I) = \|\hat{\mathbf{x}} - \mathbf{x}\|_2 + \beta \left\| \hat{\mathbf{q}} - \frac{\mathbf{q}}{\|\mathbf{q}\|} \right\|_2 \qquad (2)$$

2. Related Work

CNN



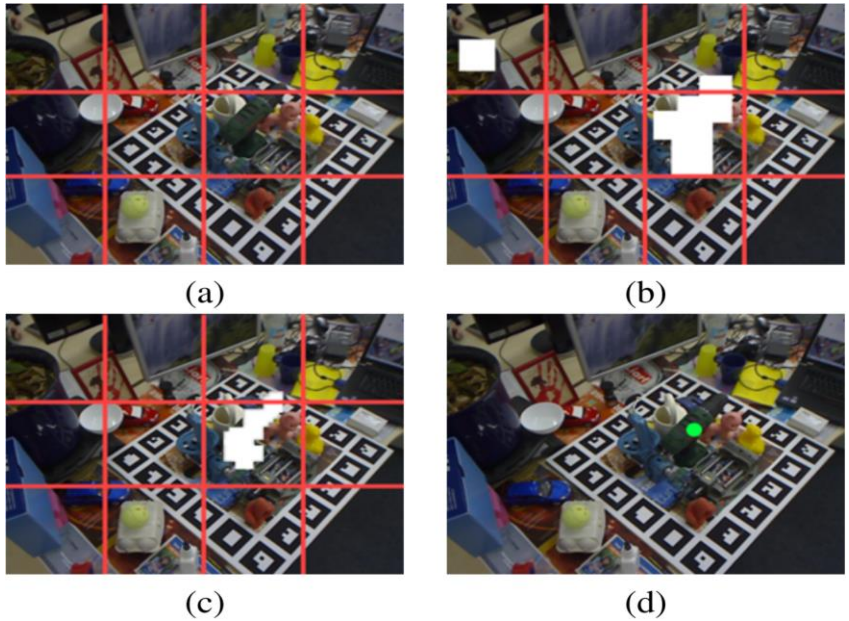
Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation ECCV 2016

- Convolutional Auto Encoder를 이용해서 모든 patch마다 Feature 생성 후 Codebook(DB)에 저장
- Voting 방식으로 pose를 찾음.
- Codebook을 이용해서 Search과정을 거쳐야하므로 시간 더 걸림.

2. Related Work

CNN

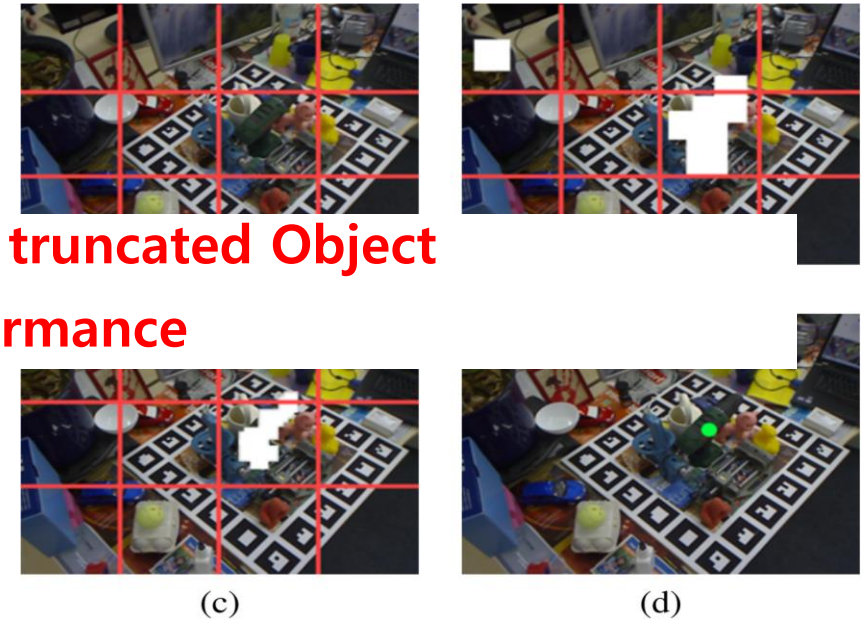
6-DoF Object Pose from Semantic Keypoints 2017 ICRA



BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth

2. Related Work

CNN



Occlusion or Partial truncated Object

Low performance

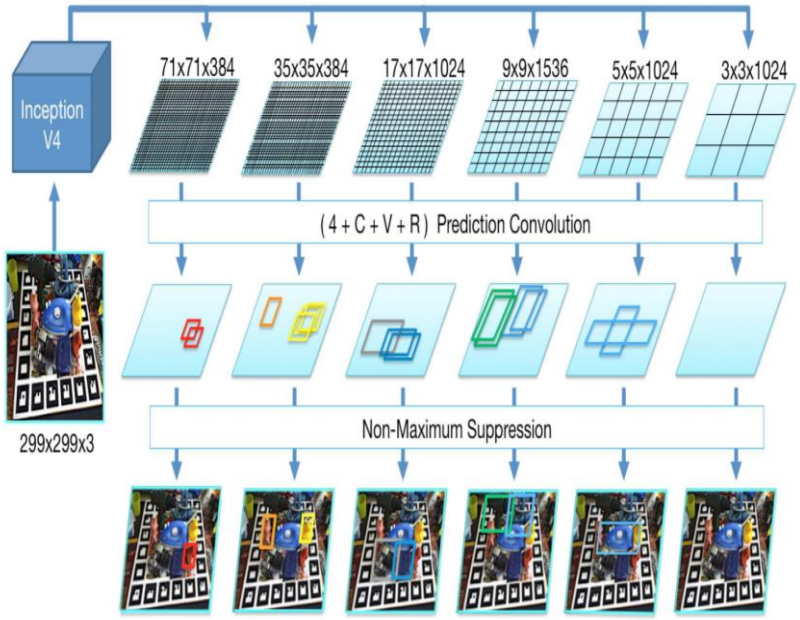
(c) (d)

6-DoF Object Pose from Semantic Keypoints 2017 ICRA

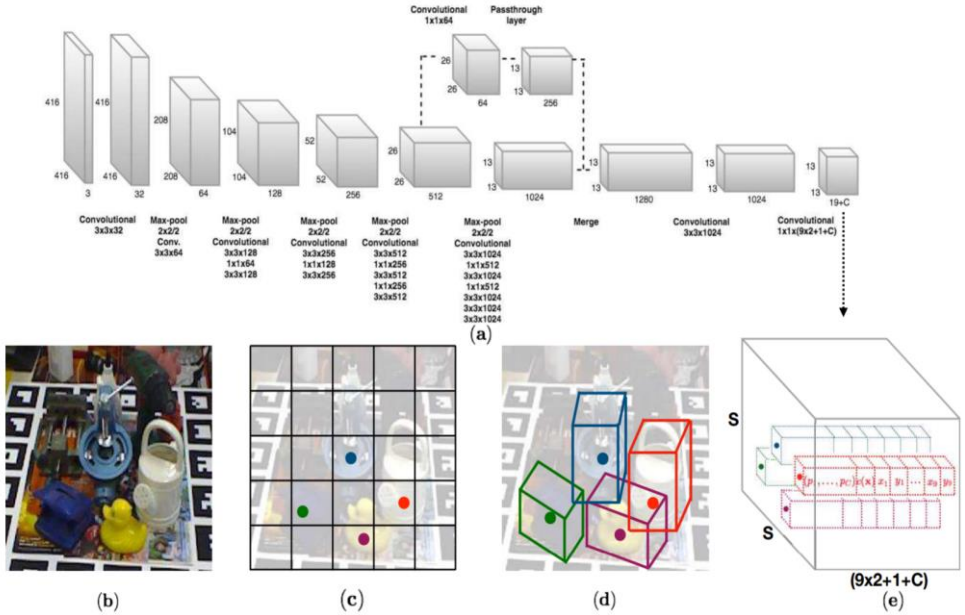
BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth

2. Related Work

CNN



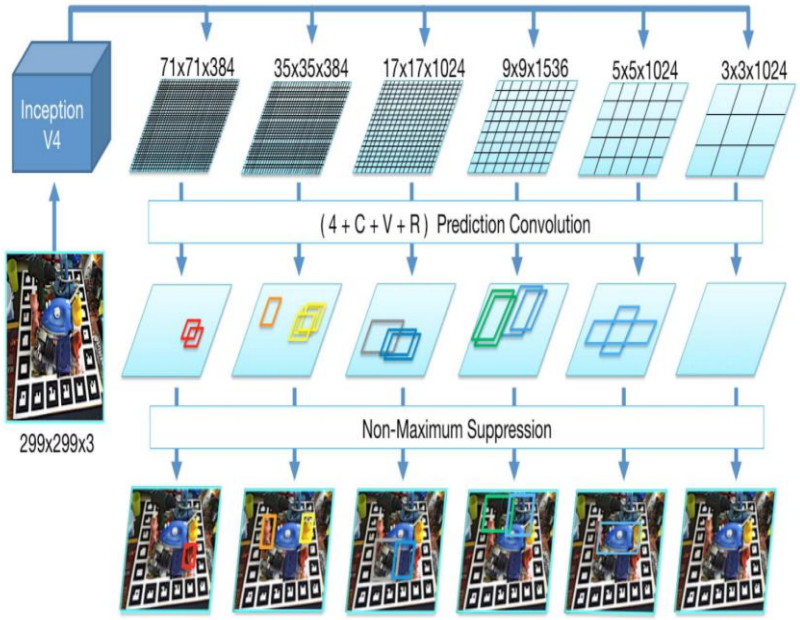
SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again 2017 ICCV



Real-Time Seamless Single Shot 6D Object Pose Prediction 2018 CVPR

2. Related Work

CNN



SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again 2017 ICCV

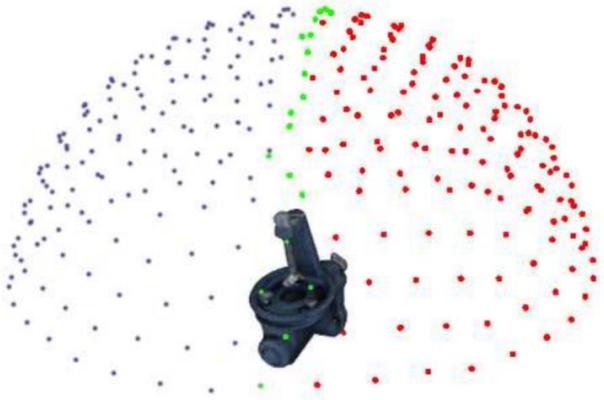
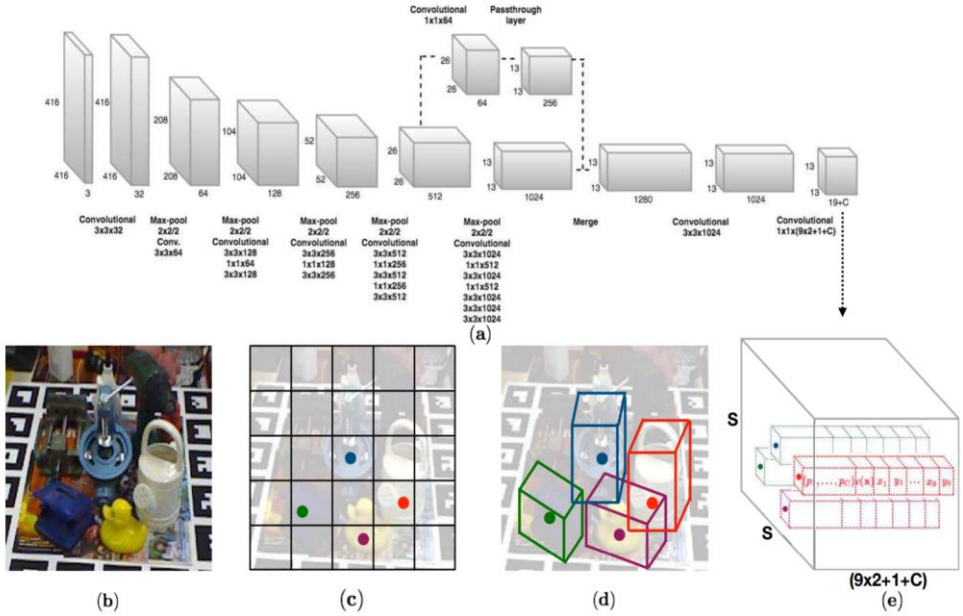


Figure 3: Discrete 6D pose space with each point representing a classifiable viewpoint. If symmetric, we use only the green points for view ID assignment during training whereas semi-symmetric objects use the red points as well.

- **2D Detection SSD 적용**
- **View Point 를 나눠 classification.**
- **매우빠름.**

2. Related Work

CNN

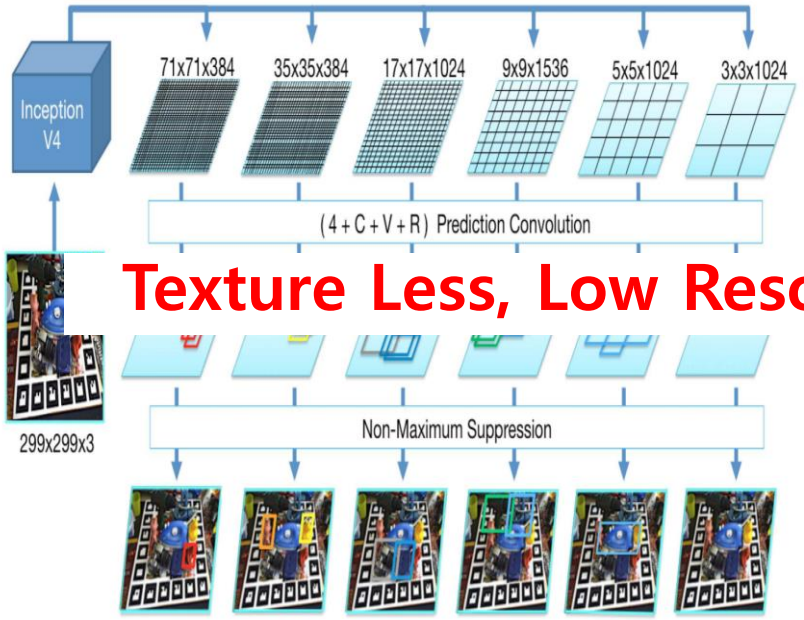


Real-Time Seamless Single Shot 6D Object Pose Prediction 2018 CVPR

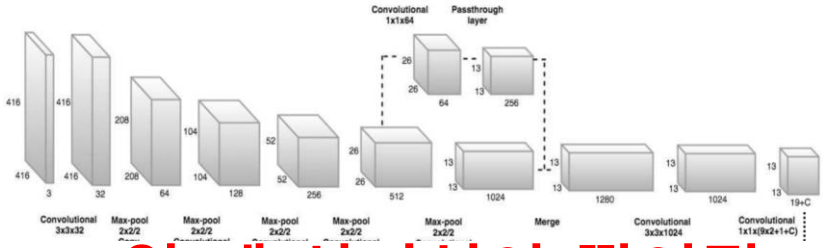
- YOLO9000
- 3D Bounding Box regression
- Solve PnP

2. Related Work

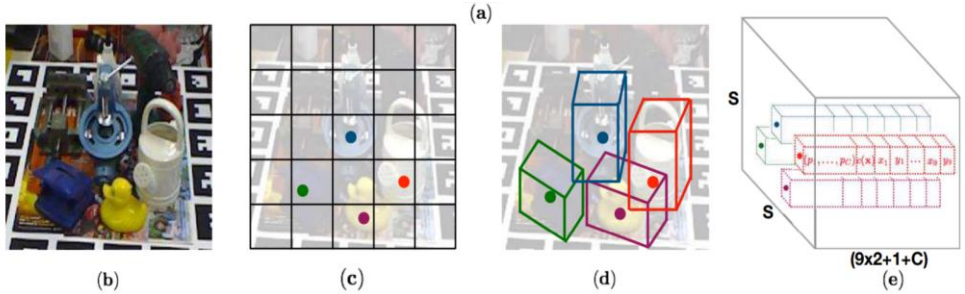
CNN



SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again 2017 ICCV



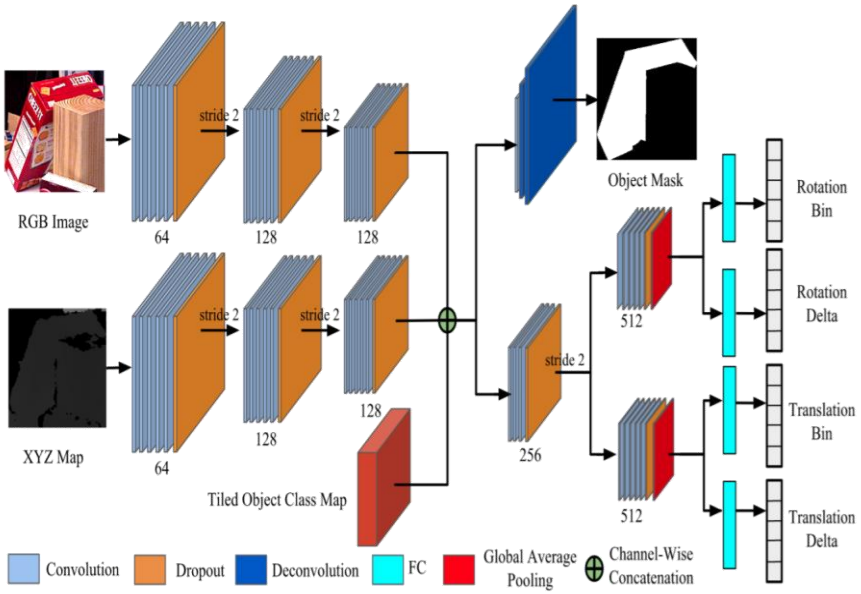
Texture Less, Low Resolution 일 때 신뢰성이 떨어짐



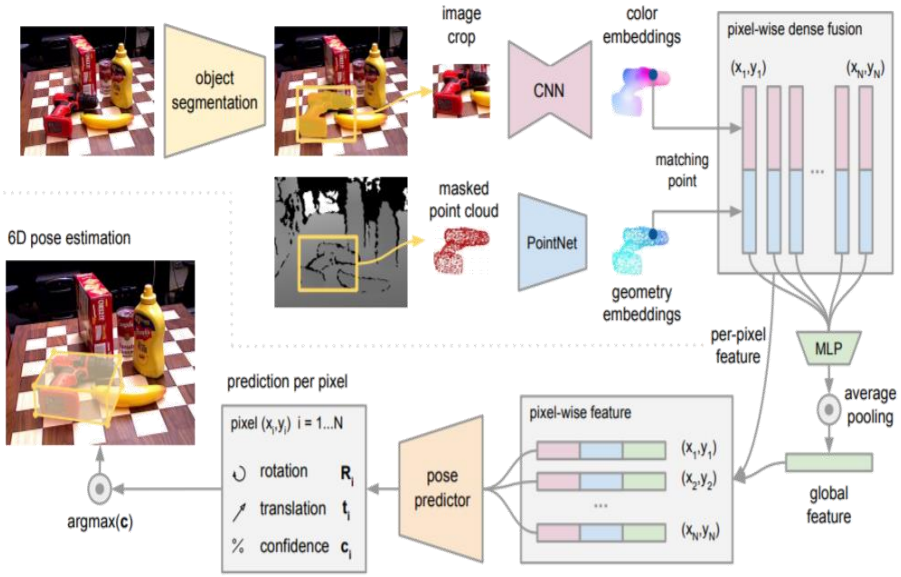
Real-Time Seamless Single Shot 6D Object Pose Prediction 2018 CVPR

2. Related Work

CNN



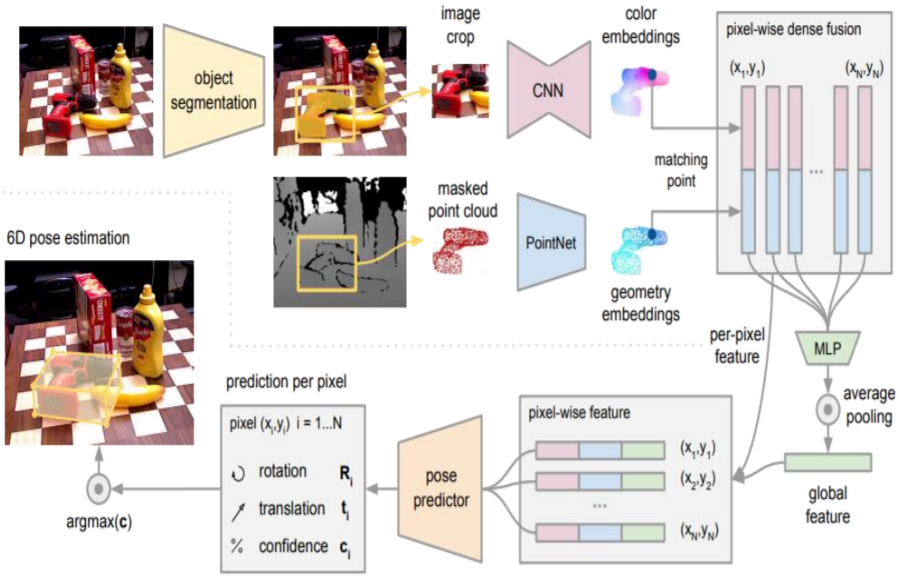
A Unified Framework for Multi-View Multi-Class Object Pose Estimation 2018 ECCV



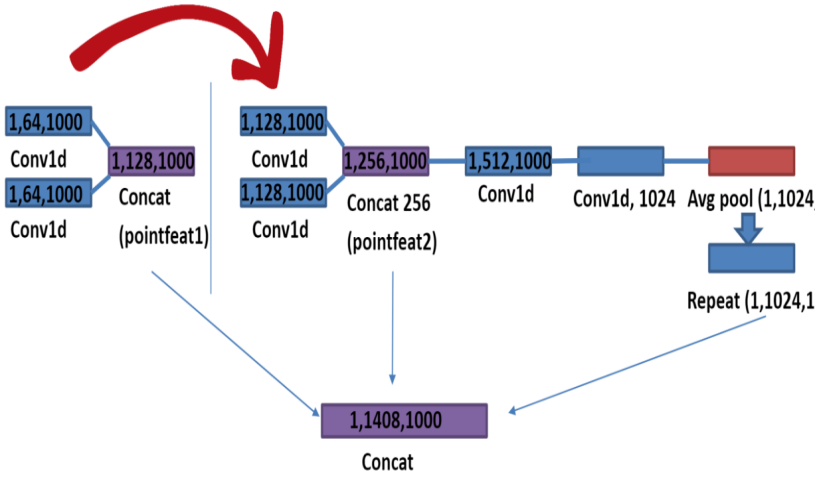
DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion 2019 CVPR

2. Related Work

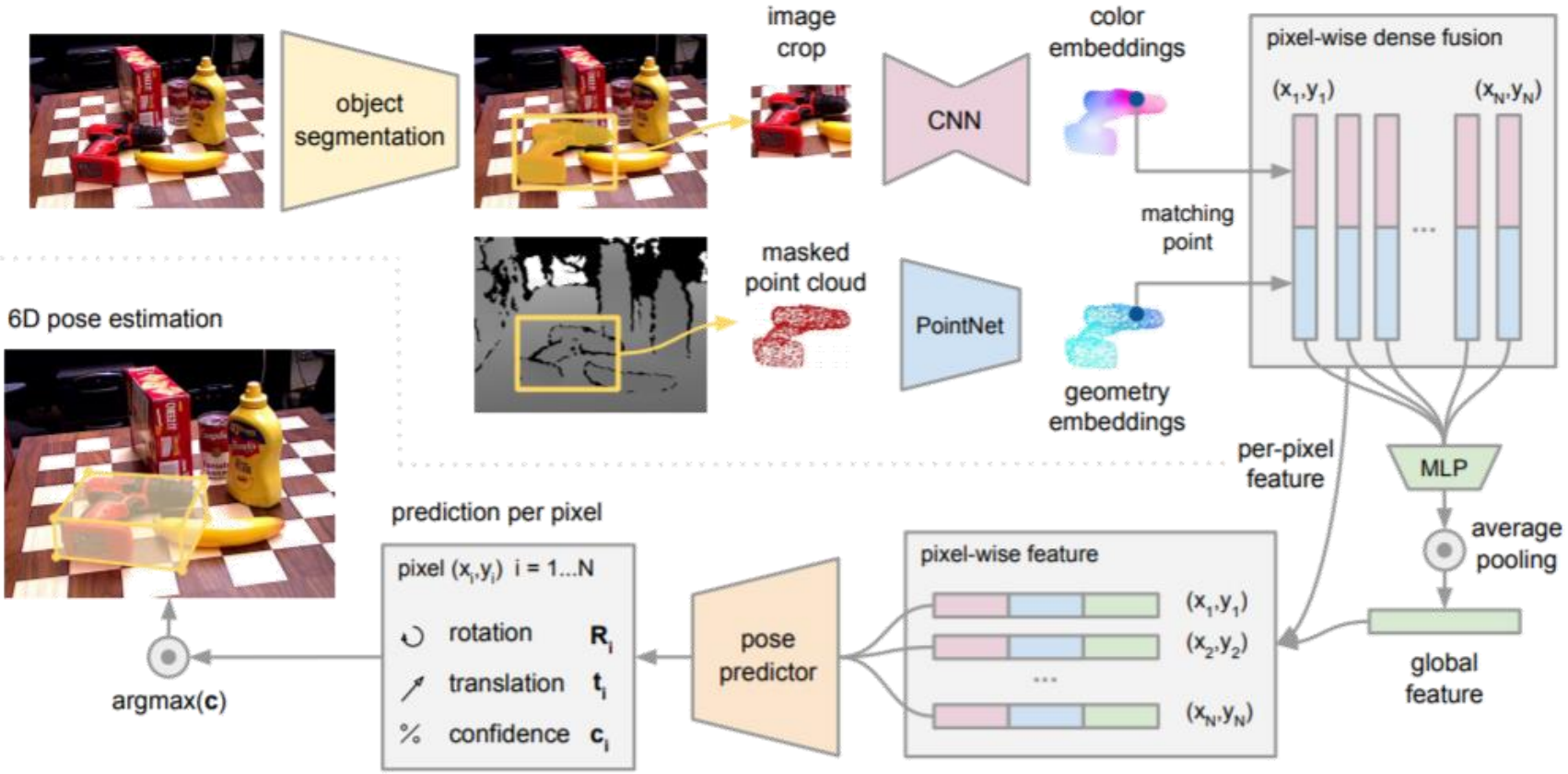
CNN



DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion 2019 CVPR



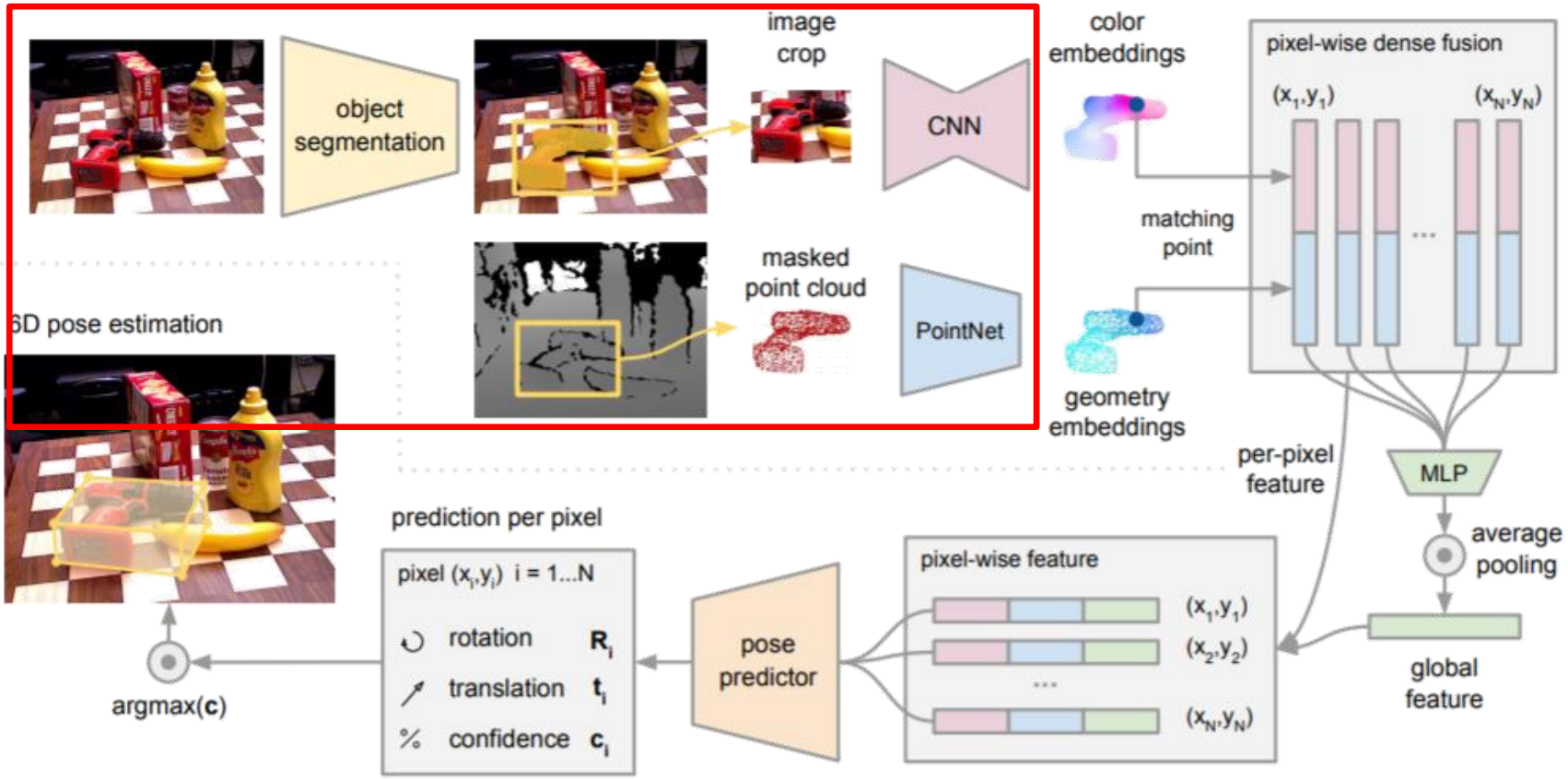
3. Methodology



3. Methodology

DenseFusion

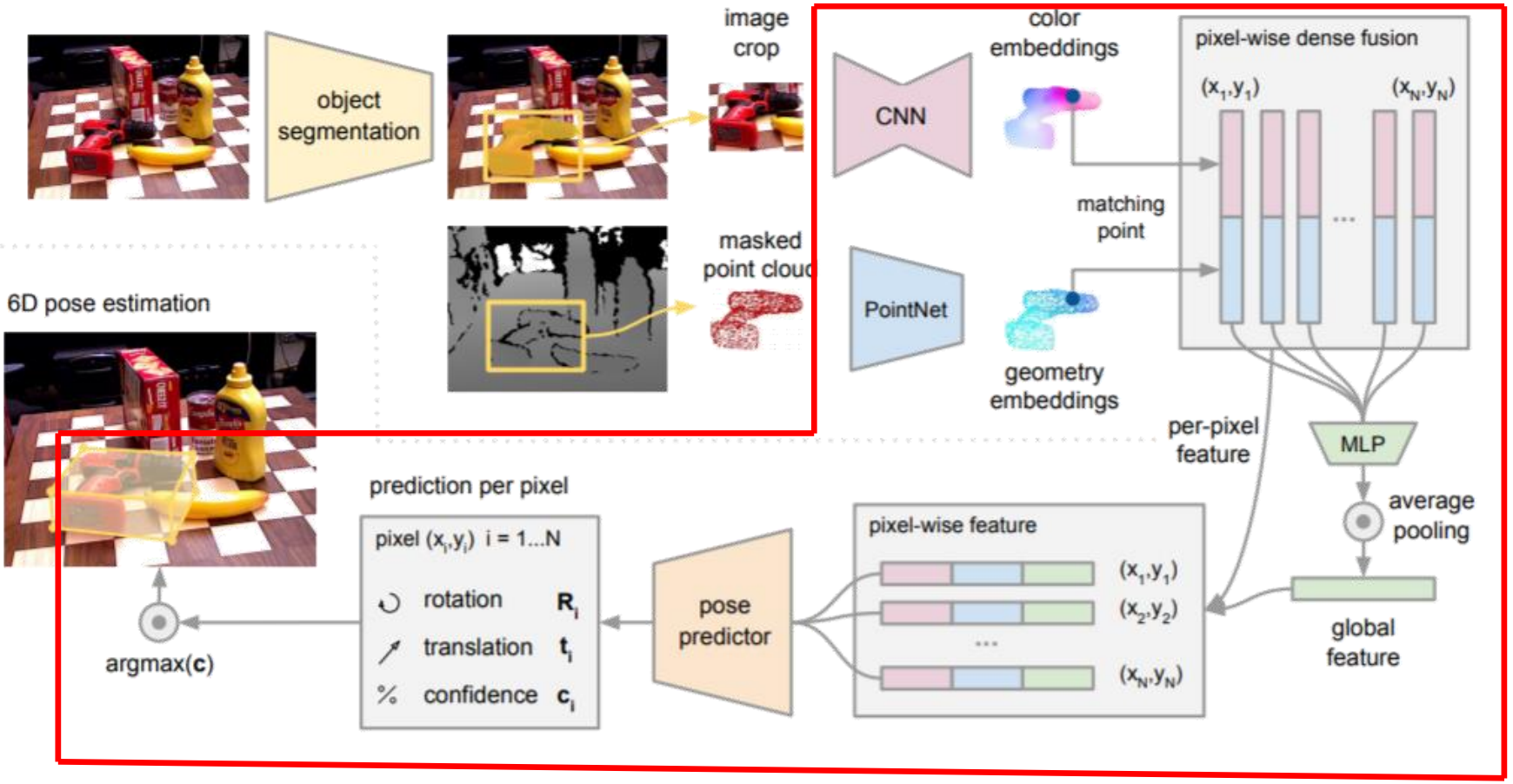
The first stage : take color image as input and performs semantic segmentation for each known object category.



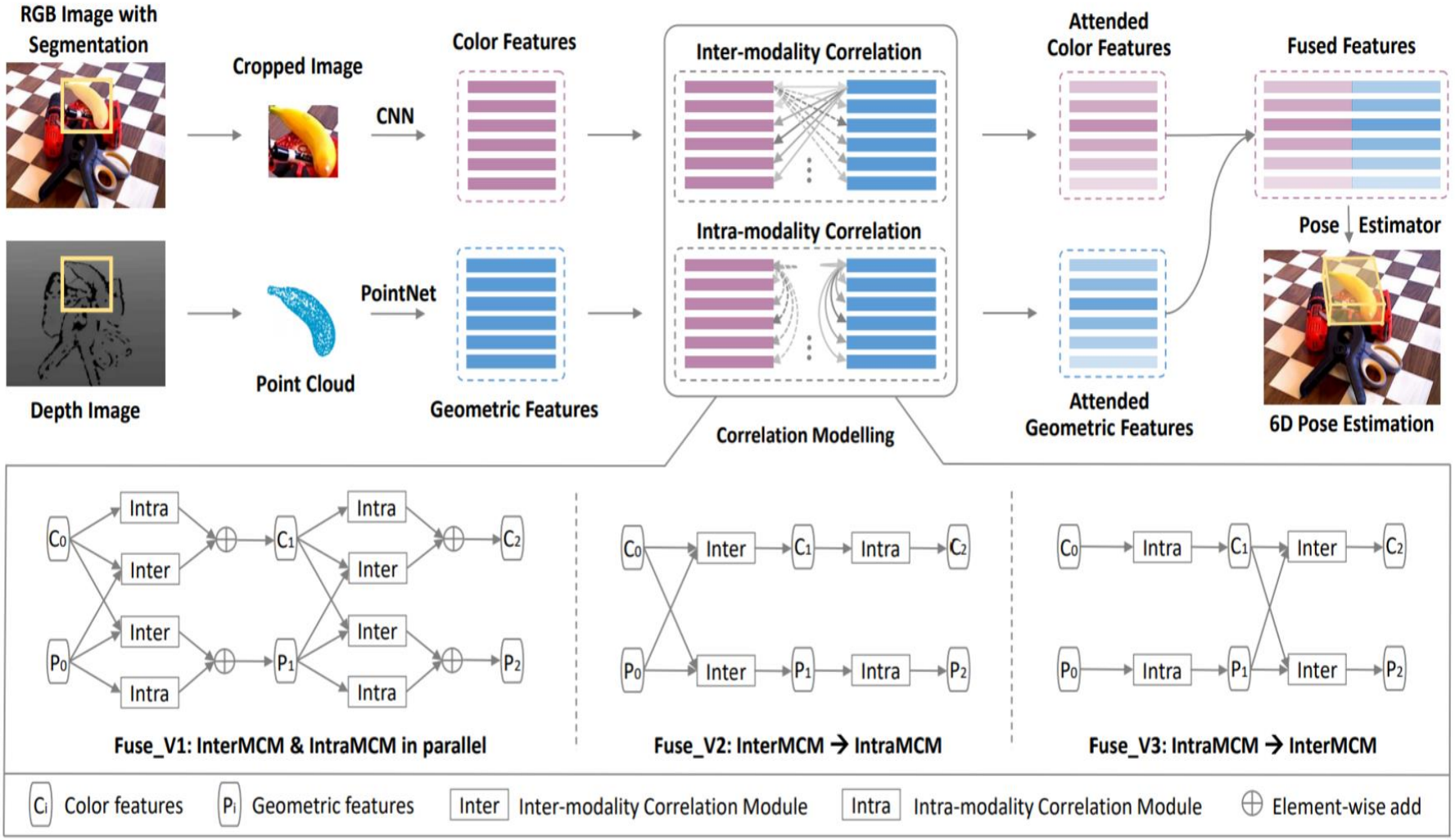
3. Methodology

DenseFusion

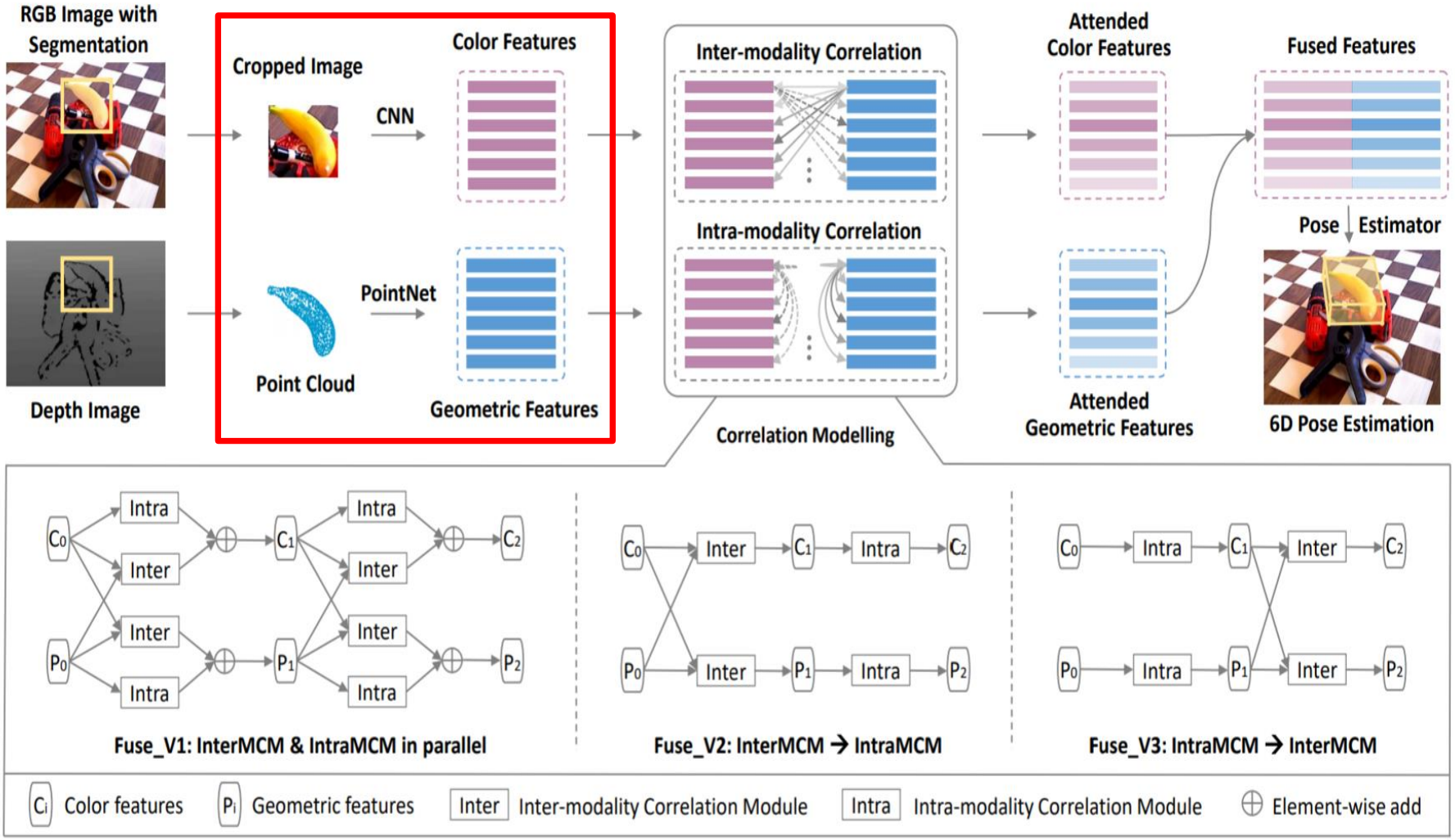
The second stage : The second stage processes the results of the segmentation and estimates the object's 6D pose.



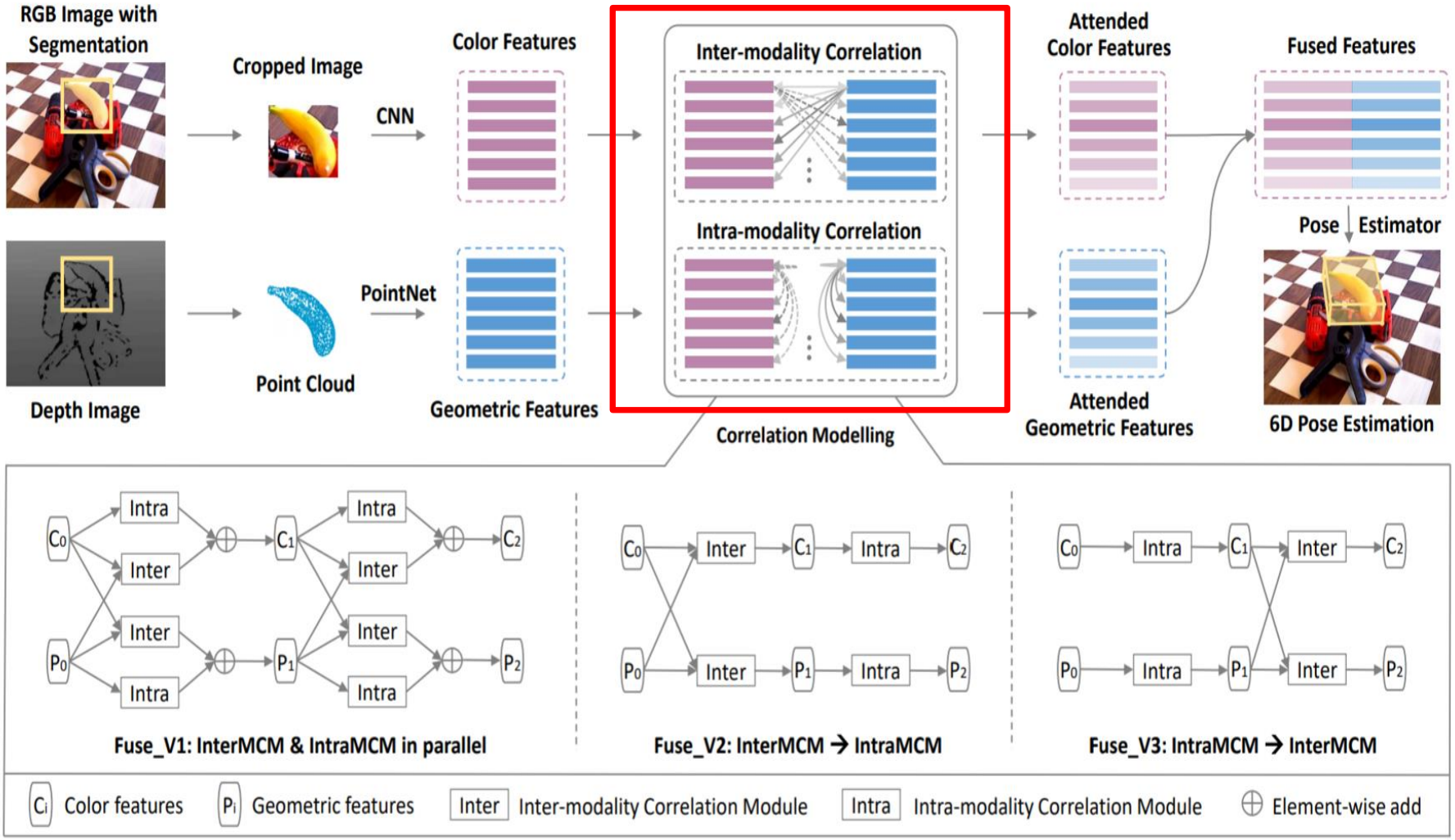
3. Methodology



3. Methodology

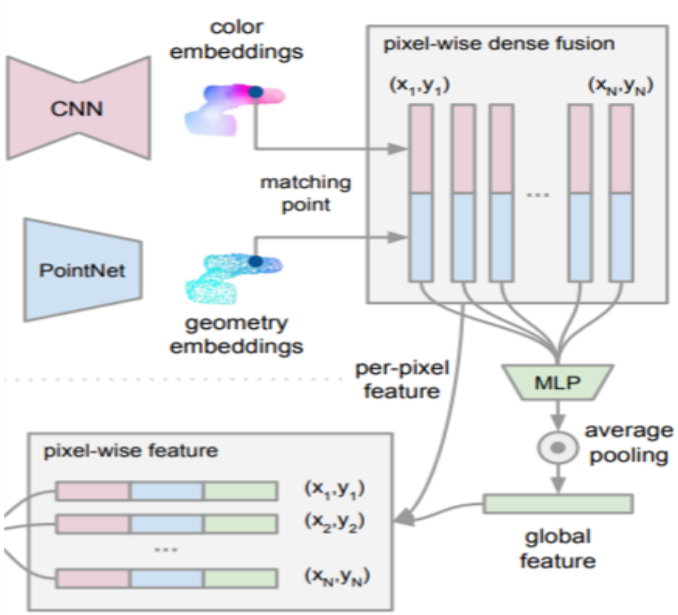


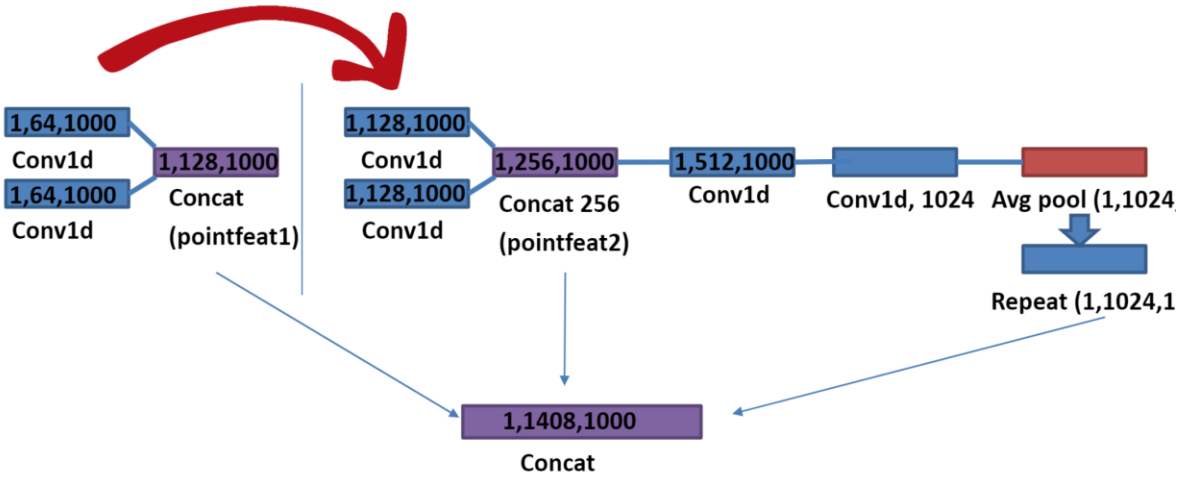
3. Methodology



3. Methodology

DenseFusion

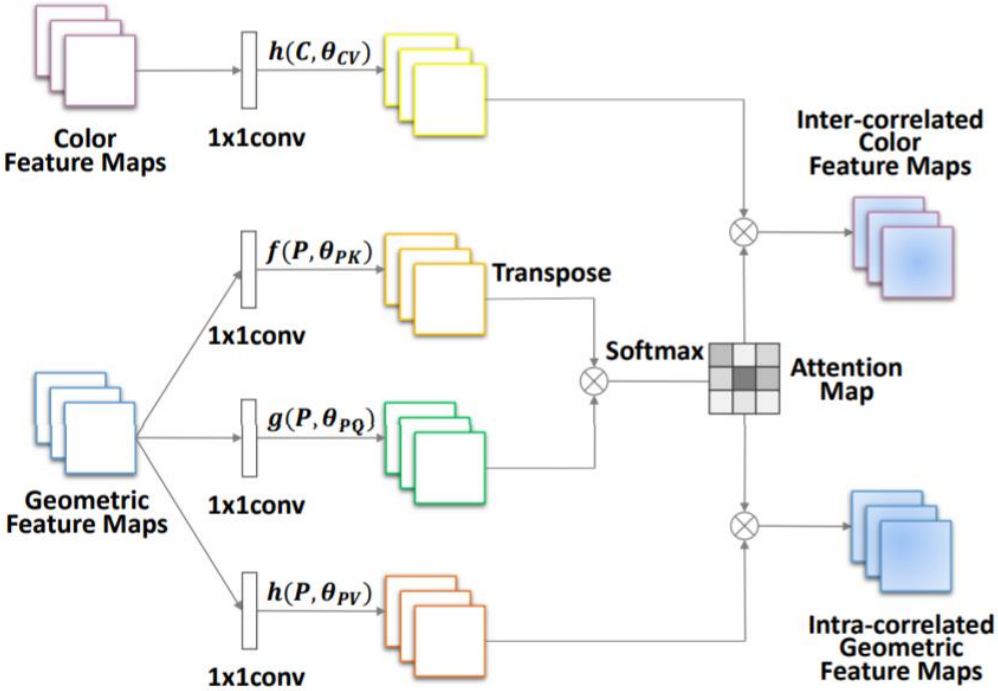




3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

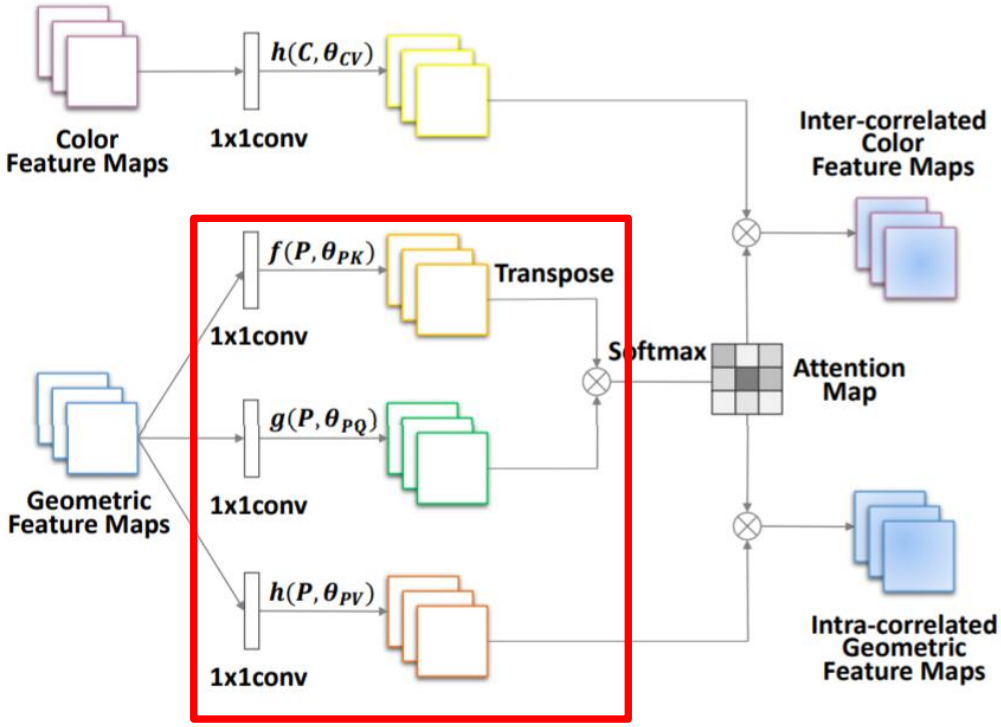
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

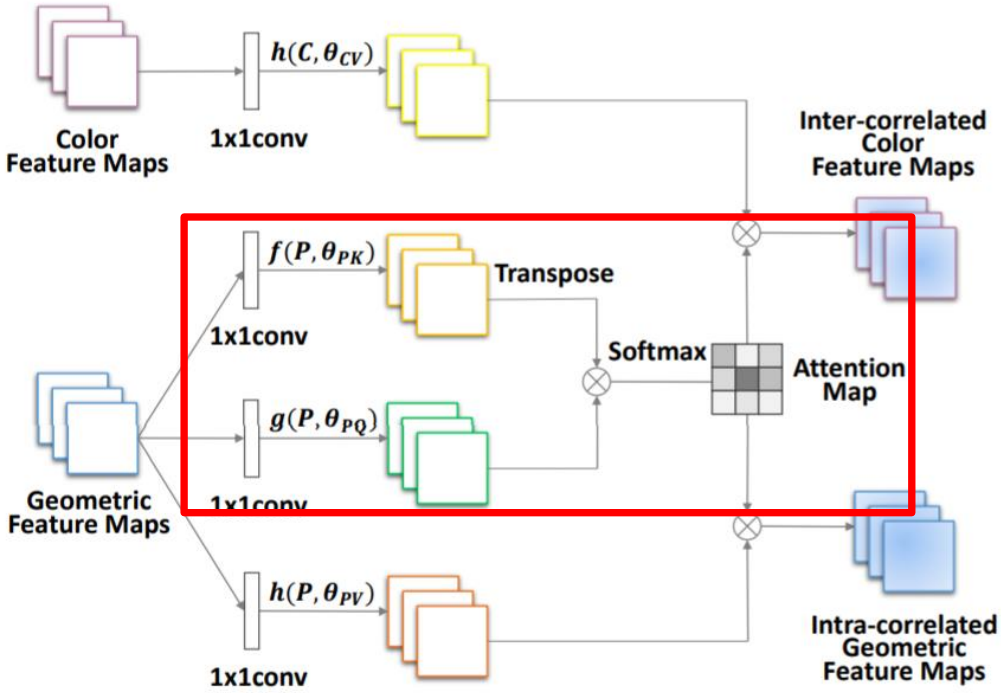
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

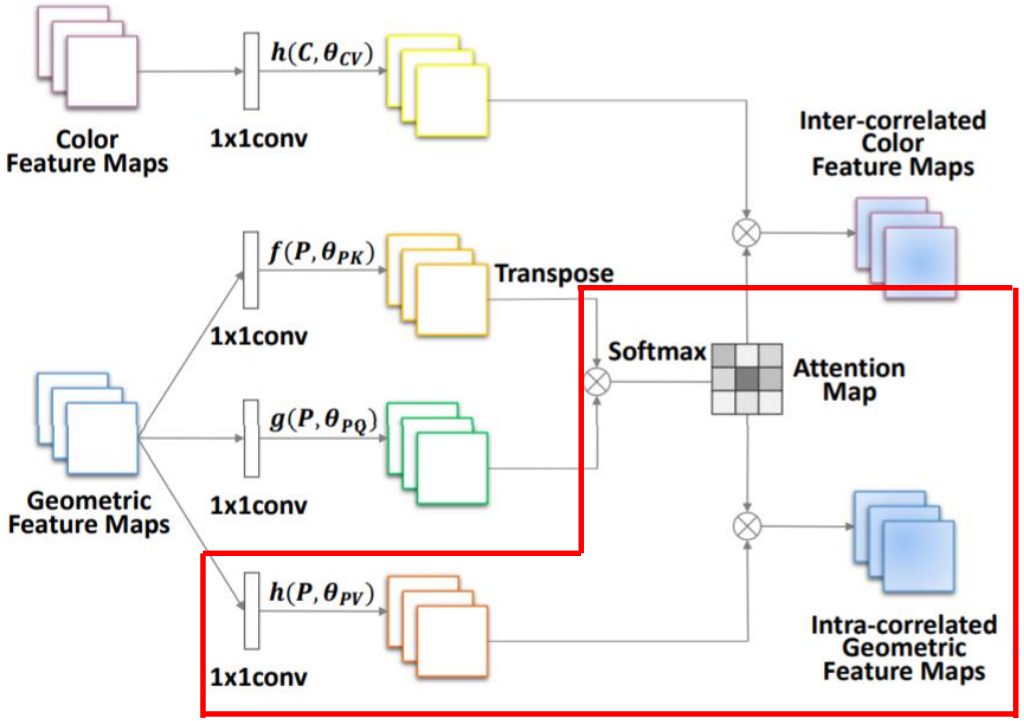
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

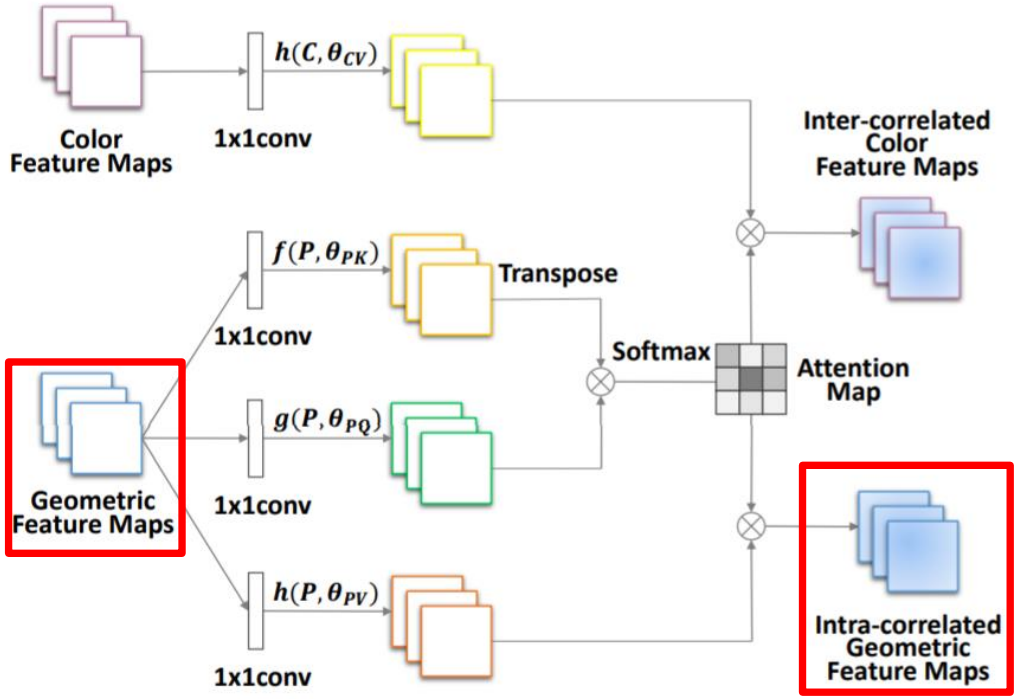
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

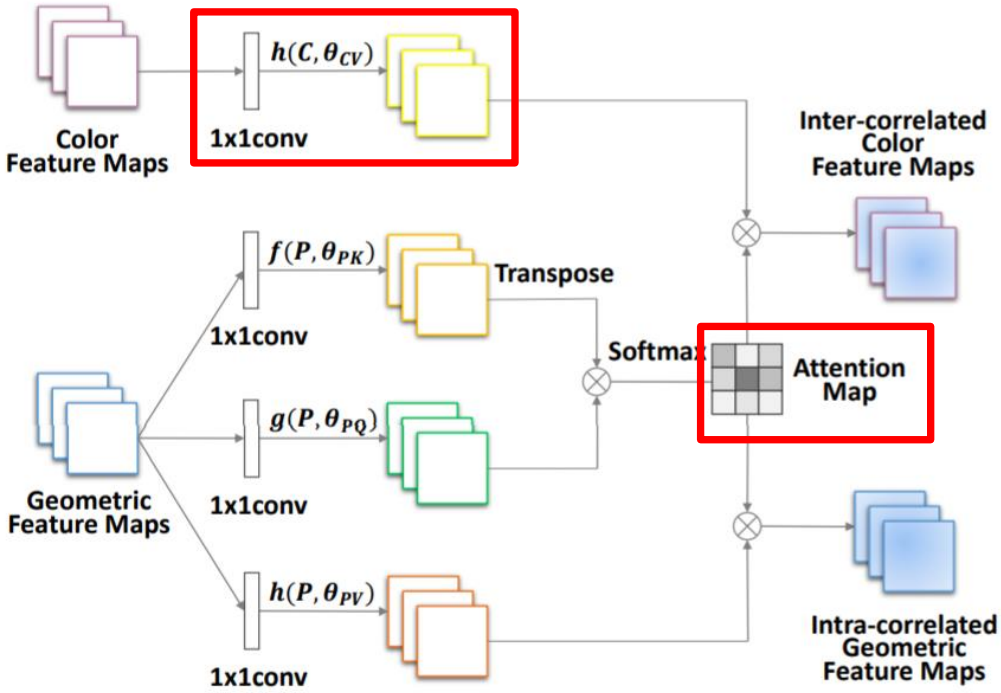
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

$$C_\triangle = C_A \times C_V, \tag{6}$$

$$P_\triangle = P_A \times P_V. \tag{7}$$

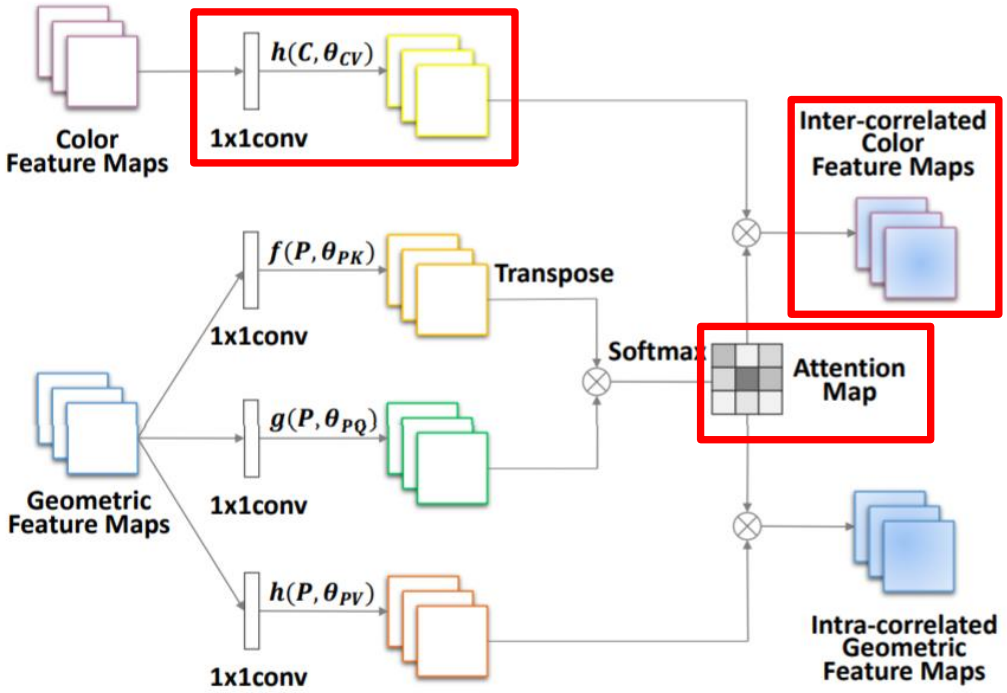
$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

Intra MCM(Intra-modality Correlation Modelling modules)

Inter MCM(Inter-modality Correlation Modelling modules)



$$C_K = \boldsymbol{f}(C;\theta_{CK}), \qquad P_K = \boldsymbol{f}(P;\theta_{PK}), \tag{1}$$

$$C_Q = \boldsymbol{g}(C;\theta_{CQ}), \qquad P_Q = \boldsymbol{g}(P;\theta_{PQ}), \tag{2}$$

$$C_V = \boldsymbol{h}(C;\theta_{CV}), \qquad P_V = \boldsymbol{h}(P;\theta_{PV}). \tag{3}$$

$$C_A = \text{softmax}(C_Q C_K^T), \tag{4}$$

$$P_A = \text{softmax}(P_Q P_K^T). \tag{5}$$

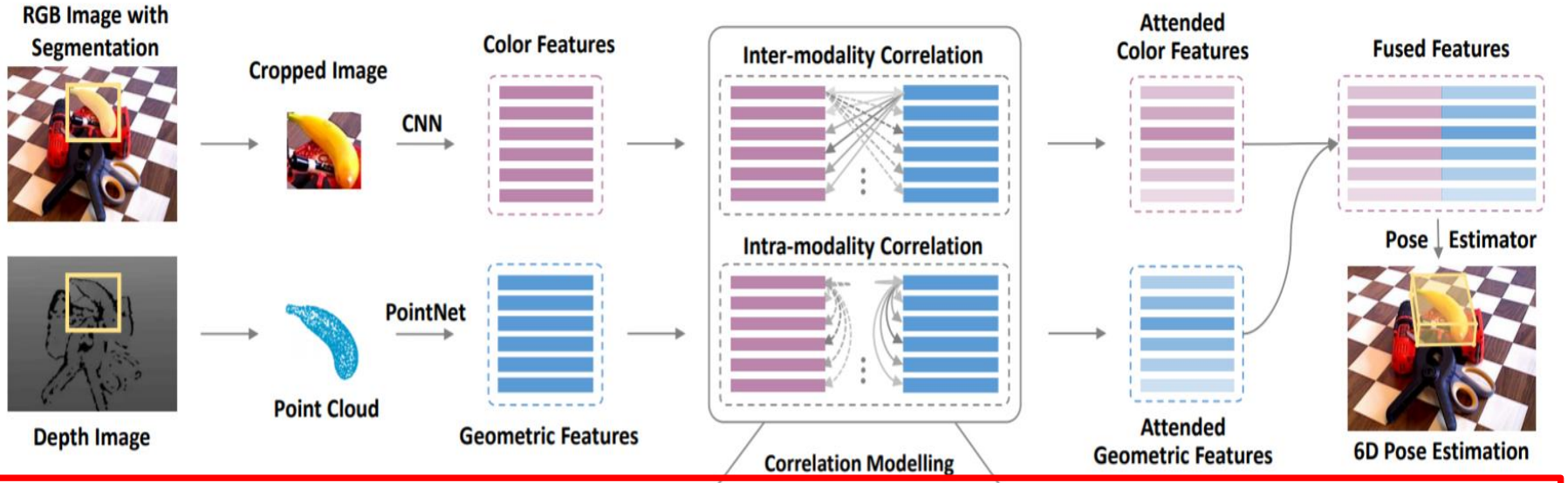
$$C_\triangle = C_A \times C_V, \tag{6}$$

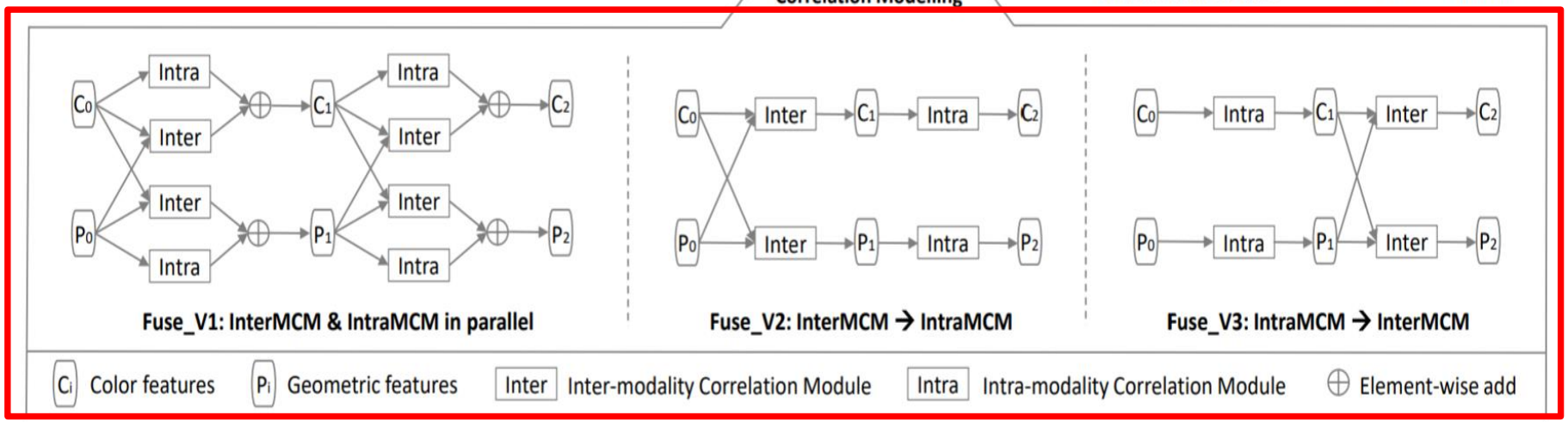
$$P_\triangle = P_A \times P_V. \tag{7}$$

$$C_\star = C + \lambda_{CC} \times C_\triangle, \tag{8}$$

$$P_\star = P + \lambda_{PP} \times P_\triangle. \tag{9}$$

3. Methodology

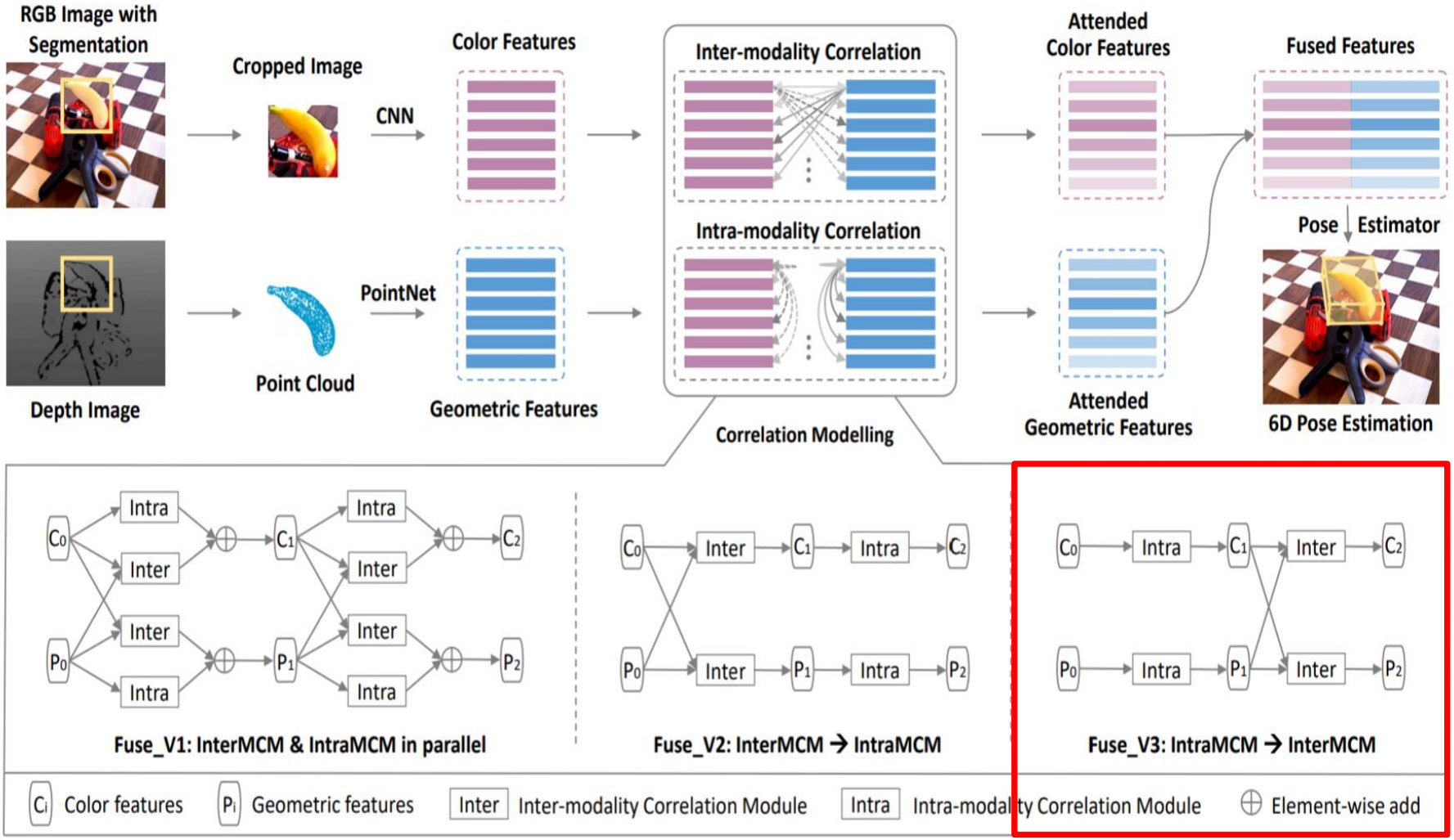




3. Methodology

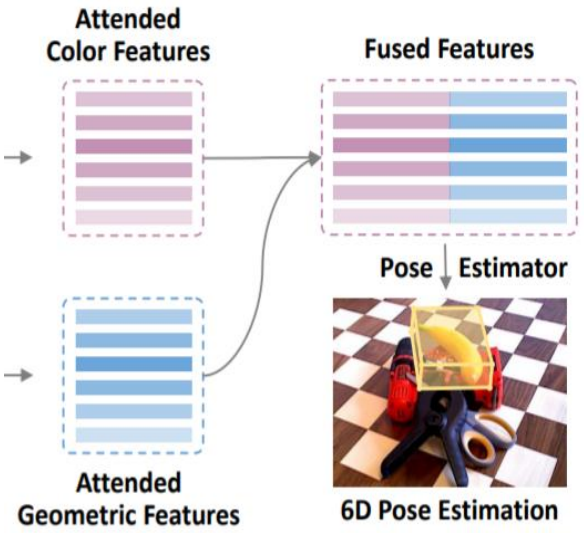
3. Methodology

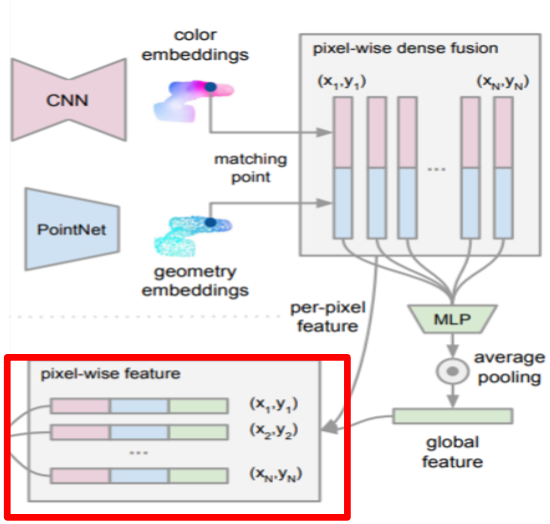
3. Methodology

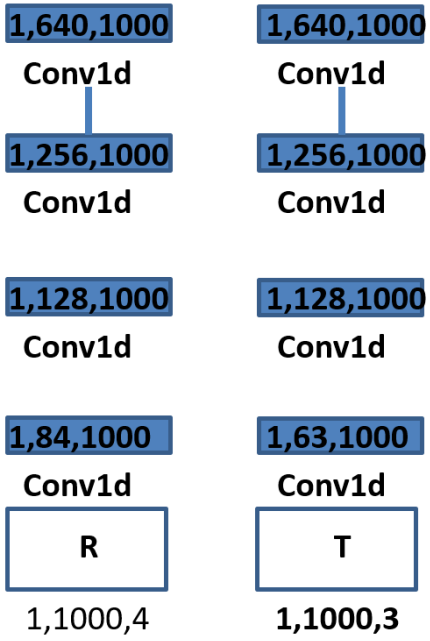


3. Methodology

DenseFusion

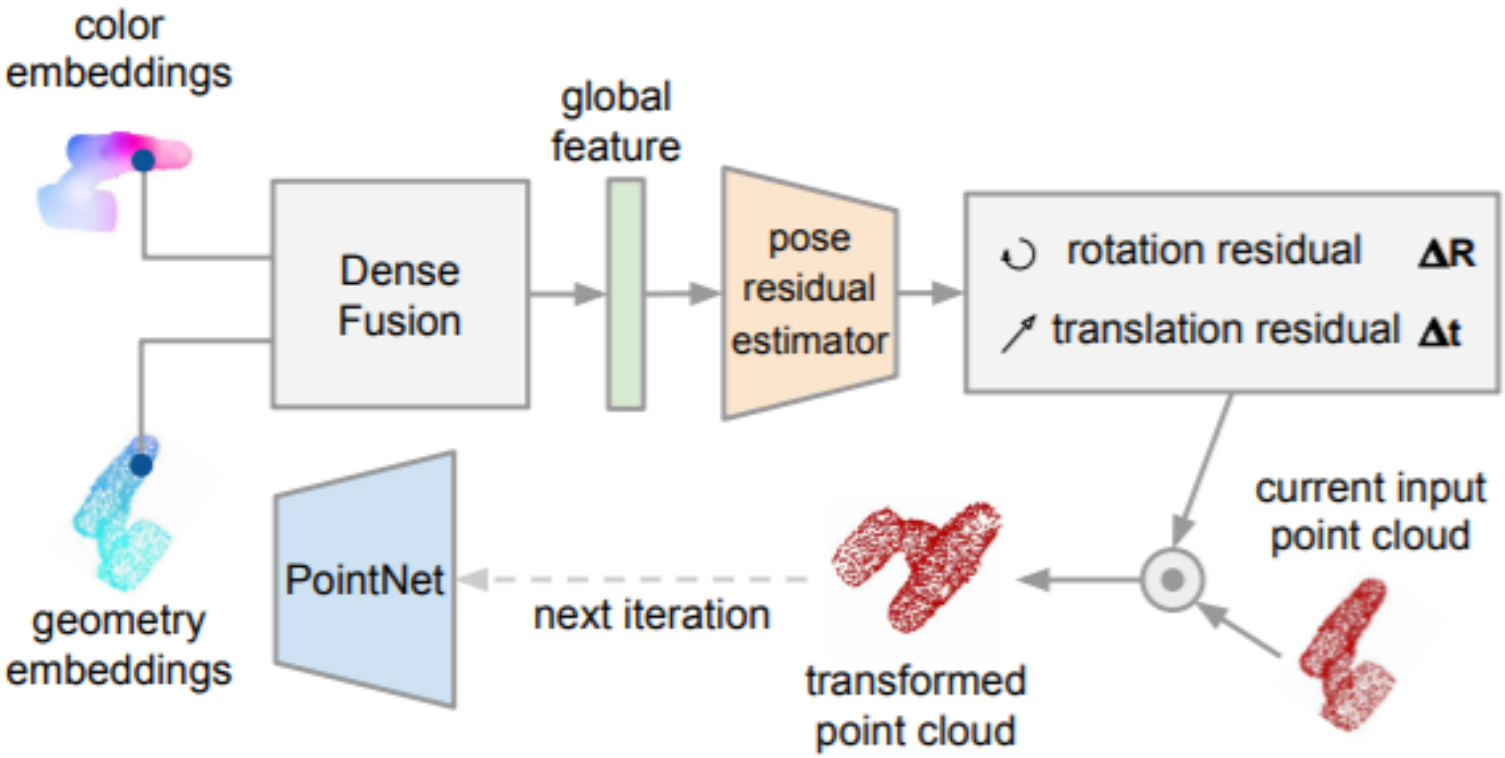






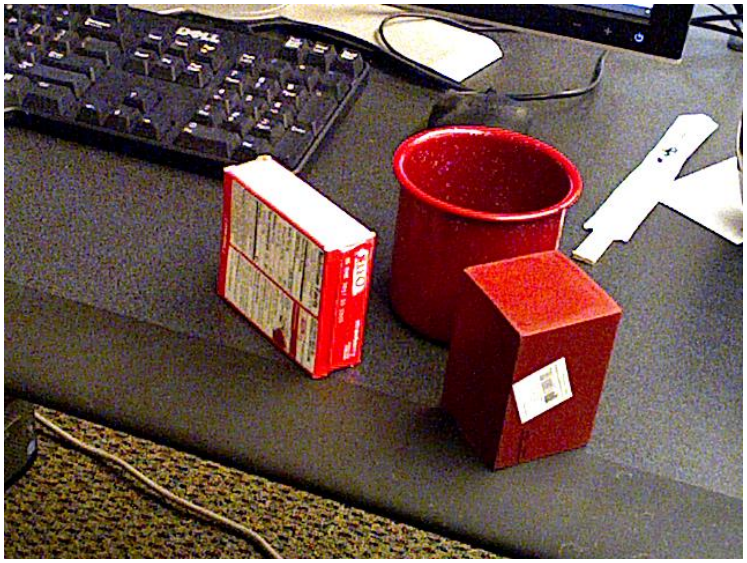
3. Methodology

DenseFusion



4. Experiments

YCB Video Dataset



LineMod

YCB Dataset 21개의 물체가 다양한 shape과 texture를 가지고 있음.

Linemod dataset : 13개의 low texture를 가지고 있는 물체

4. Experiments

Metrics

$$\text{ADD} = \frac{1}{m} \sum_{\mathbf{x} \in \mathcal{M}} \|(\mathbf{R}\mathbf{x} + \mathbf{T}) - (\tilde{\mathbf{R}}\mathbf{x} + \tilde{\mathbf{T}})\|, \qquad (15)$$

$$\text{ADD-S} = \frac{1}{m} \sum_{\mathbf{x}_1 \in \mathcal{M}} \min_{\mathbf{x}_2 \in \mathcal{M}} \|(\mathbf{R}\mathbf{x}_1 + \mathbf{T}) - (\tilde{\mathbf{R}}\mathbf{x}_2 + \tilde{\mathbf{T}})\|. \quad (16)$$

4. Experiments

TABLE I

THE 6D POSE ESTIMATION ACCURACY ON **YCB-VIDEO DATASET** IN TERMS OF THE **ADD(-S) <2CM AND THE AUC OF ADD(-S)**. THE OBJECTS WITH BOLD NAME ARE CONSIDERED AS SYMMETRIC. ALL THE METHODS USE RGB-D IMAGES AS INPUT.**(BEST ZOOM-IN AND VIEW IN PDF.)**

Methods	PoseCNN [5]		DenseFusion [21]		OURS (IntraMCM)		OURS (InterMCM)		OURS (Fuse_V1)		OURS (Fuse_V2)		OURS (Fuse_V3)	
Metrics	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm
002_master_chef_can	68.06	51.09	73.16	72.56	87.61	88.37	86.94	88.07	86.18	86.28	**92.47**	**98.71**	87.24	86.18
003_cracker_box	83.38	73.27	94.21	98.50	94.80	**99.54**	93.69	99.08	91.79	98.50	**95.45**	98.62	95.20	99.19
004_sugar_box	**97.15**	99.49	96.50	**100.00**	93.73	**100.00**	95.06	**100.00**	95.68	**100.00**	96.69	99.92	96.19	99.58
005_tomato_soup_can	81.77	76.60	85.42	82.99	91.50	95.42	90.23	93.19	**92.73**	95.56	92.02	**95.76**	91.51	95.56
006_mustard_bottle	**98.01**	98.60	94.61	96.36	92.27	98.04	93.10	98.60	89.66	91.04	94.82	97.48	95.29	**99.16**
007_tuna_fish_can	83.87	72.13	81.88	62.28	80.86	69.69	86.18	84.58	85.94	83.45	**88.85**	84.15	85.27	**86.31**
008_pudding_box	**96.62**	**100.00**	93.33	98.60	91.69	97.13	91.83	98.60	91.76	99.07	93.16	98.60	94.10	98.13
009_gelatin_box	**98.08**	100.00	96.68	100.00	95.35	100.00	95.06	100.00	95.92	100.00	95.68	100.00	97.28	100.00
010_potted_meat_can	83.47	77.94	83.54	79.90	85.01	83.55	83.77	80.81	84.07	82.90	**86.19**	83.94	86.03	**84.07**
011_banana	91.86	88.13	83.49	88.13	84.70	81.79	90.71	98.68	88.73	98.15	**92.57**	**98.94**	86.84	88.92
019_pitcher_base	**96.93**	97.72	96.78	99.47	95.76	98.02	96.55	**100.00**	96.07	**100.00**	95.43	98.42	95.97	99.65
021_bleach_cleanser	**92.54**	**92.71**	89.93	90.96	87.93	83.19	89.10	83.28	90.19	89.70	88.99	86.20	89.00	83.28
024_bowl	80.97	54.93	**89.50**	94.83	88.70	**97.78**	87.00	84.24	86.32	90.64	86.06	94.33	89.08	95.81
025_mug	81.08	55.19	88.92	89.62	91.84	92.77	92.00	94.97	91.06	91.98	**93.51**	94.81	93.44	**96.38**
035_power_drill	**97.66**	**99.24**	92.55	96.40	92.05	95.65	86.60	90.35	85.05	87.70	82.89	84.77	93.52	98.20
036_wood_block	87.56	80.17	**92.88**	**100.00**	91.44	98.35	90.16	**100.00**	91.46	99.59	92.32	99.59	92.35	98.76
037_scissors	78.36	49.17	77.89	51.38	**91.28**	86.37	78.98	67.40	79.25	64.70	90.15	**89.50**	88.38	86.74
040_large_marker	85.26	87.19	92.95	**100.00**	93.55	**100.00**	93.84	**100.00**	**94.10**	**100.00**	93.91	99.85	93.82	99.85
051_large_clamp	**75.19**	74.86	72.48	**78.65**	71.27	78.51	72.14	77.95	70.18	75.70	70.31	76.69	73.22	**78.65**
052_extra_large_clamp	64.38	48.83	69.94	75.07	70.11	**76.83**	**73.74**	75.51	69.71	75.22	69.53	74.49	70.80	76.25
061_foam_brick	**97.23**	100.00	91.95	100.00	94.36	100.00	94.15	100.00	93.08	100.00	94.62	100.00	94.89	100.00
MEAN	86.64	79.87	87.55	88.37	88.85	91.48	88.61	91.21	88.04	90.96	89.79	**93.08**	**89.97**	92.89

4. Experiments

TABLE II

THE 6D POSE ESTIMATION ACCURACY ON THE **LINEMOD DATASET** IN TERMS OF THE **ADD(-S)** METRIC. THE OBJECTS WITH BOLD NAME (GLUE AND EGGBOX) ARE CONSIDERED AS SYMMETRIC. ALL THE METHODS USE RGB-D IMAGES AS INPUT.

	SSD6D [16]	BB8 [13]	DenseFusion [21]	OURS (IntraMCM)	OURS (InterMCM)	OURS (Fuse_V1)	OURS (Fuse_V2)	OURS (Fuse_V3)
ape	65	40.4	92.3	94.9	95.2	94.8	**95.6**	95.4
bench	80	91.8	93.2	93.7	94.0	96.1	**96.9**	96.1
camera	78	55.7	94.4	97.5	95.6	96.0	**97.9**	97.5
can	86	64.1	93.1	95.4	95.7	92.2	**96.0**	95.0
cat	70	62.6	96.5	98.4	98.8	**99.2**	97.8	99.1
driller	73	74.4	87.0	92.2	92.7	91.4	**95.6**	94.7
duck	66	44.3	92.3	**96.2**	95.1	95.7	95.7	95.8
eggbox	**100**	57.8	99.8	**100.0**	99.6	**100.0**	99.9	99.9
glue	**100**	41.2	**100.0**	99.8	99.8	99.8	99.7	99.8
hole	49	67.2	92.1	95.2	95.6	95.8	96.7	**97.1**
iron	78	84.7	97.0	95.8	96.2	97.4	97.8	**98.4**
lamp	73	76.5	95.3	95.4	96.3	96.5	**97.0**	96.8
phone	79	54.0	92.8	97.3	**97.5**	95.6	97.0	97.4
MEAN	77	62.7	94.3	96.3	96.3	96.2	**97.2**	97.1

4. Experiment

TABLE III

SUCCESS RATE FOR THE GRASPING EXPERIMENTS WITH ROBOTIC ARM IN SIMULATION ENVIRONMENT OF GAZEBO.

Success Attempts (%)	tomato_soup_can	mustard_bottle	banana	bleach_cleanser
DenseFusion [21]	80.0	70.0	55.0	65.0
Ours	90.0	85.0	75.0	80.0

5. Conclusion

1. Intra-, inter-modality correlation을 이용해서 6D pose Estimation하는 Novel Correlation Fusion framework 제안.

2. IntraMCM은 modality-specific feature 를 학습, InterMCM은 modalit간 feature를 capture.

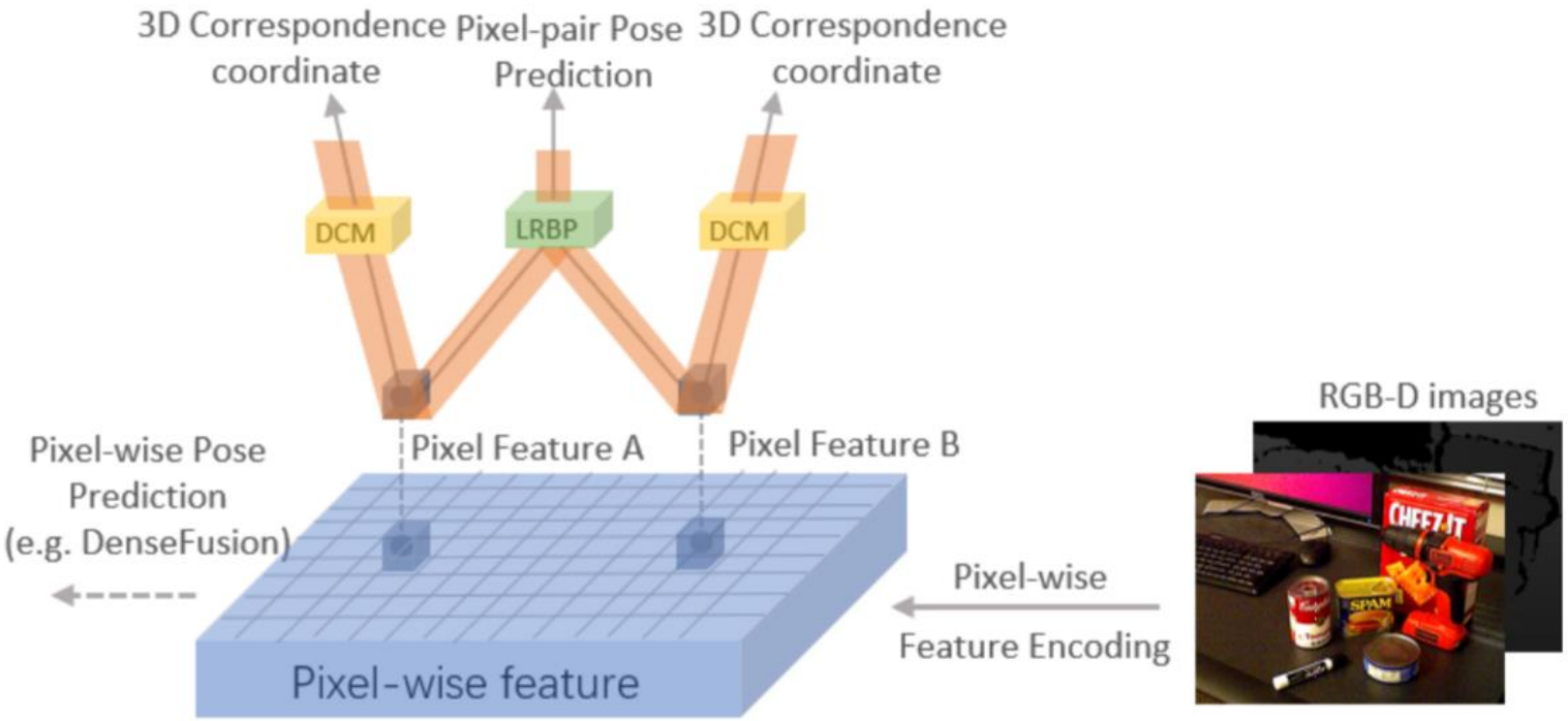
3. YCB, LINEMOD DB and robot grasping task demonstrate the superior performance

W-PoseNet: Dense Correspondence Regularized Pixel Pair Pose Regression

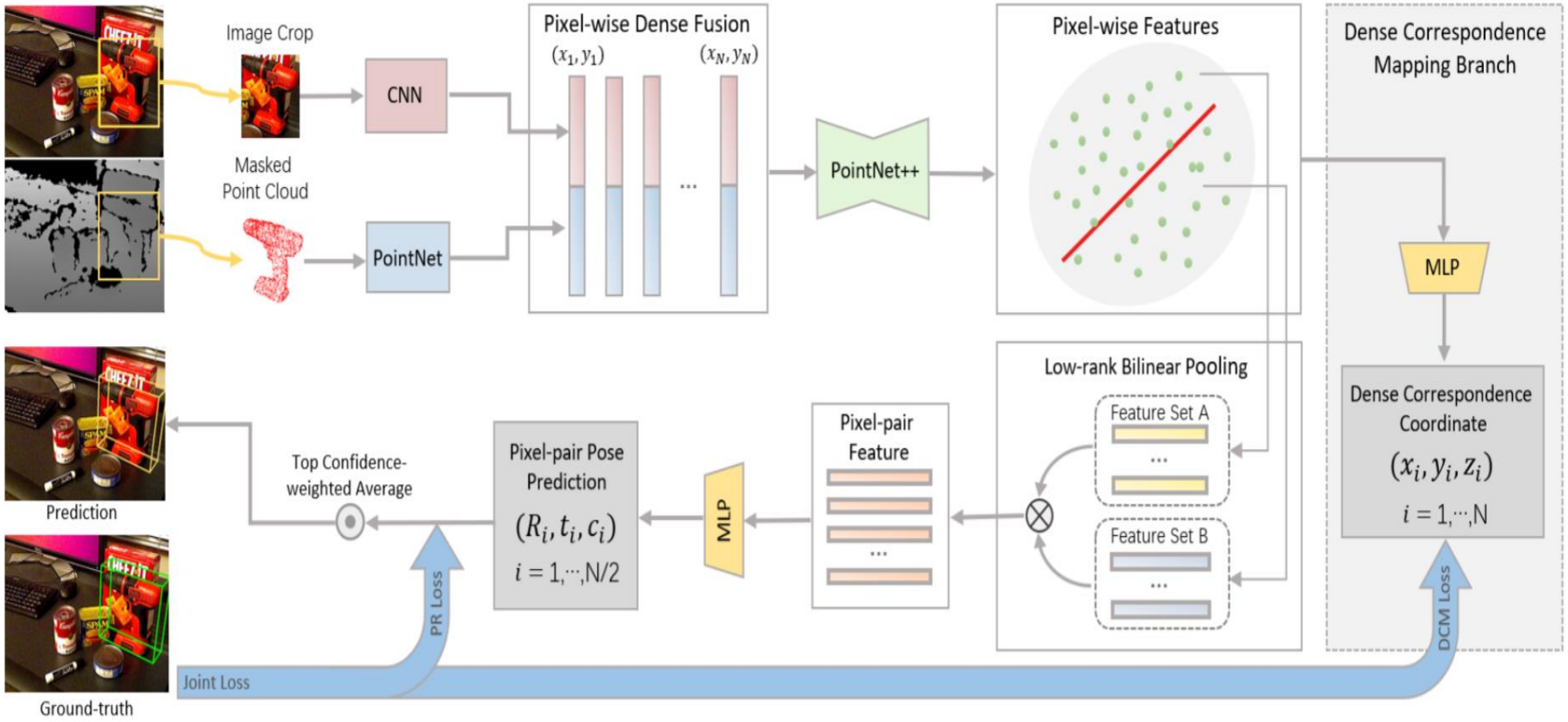
Zelin Xu* , Ke Chen* , and Kui Jia

[2019 CoRL]

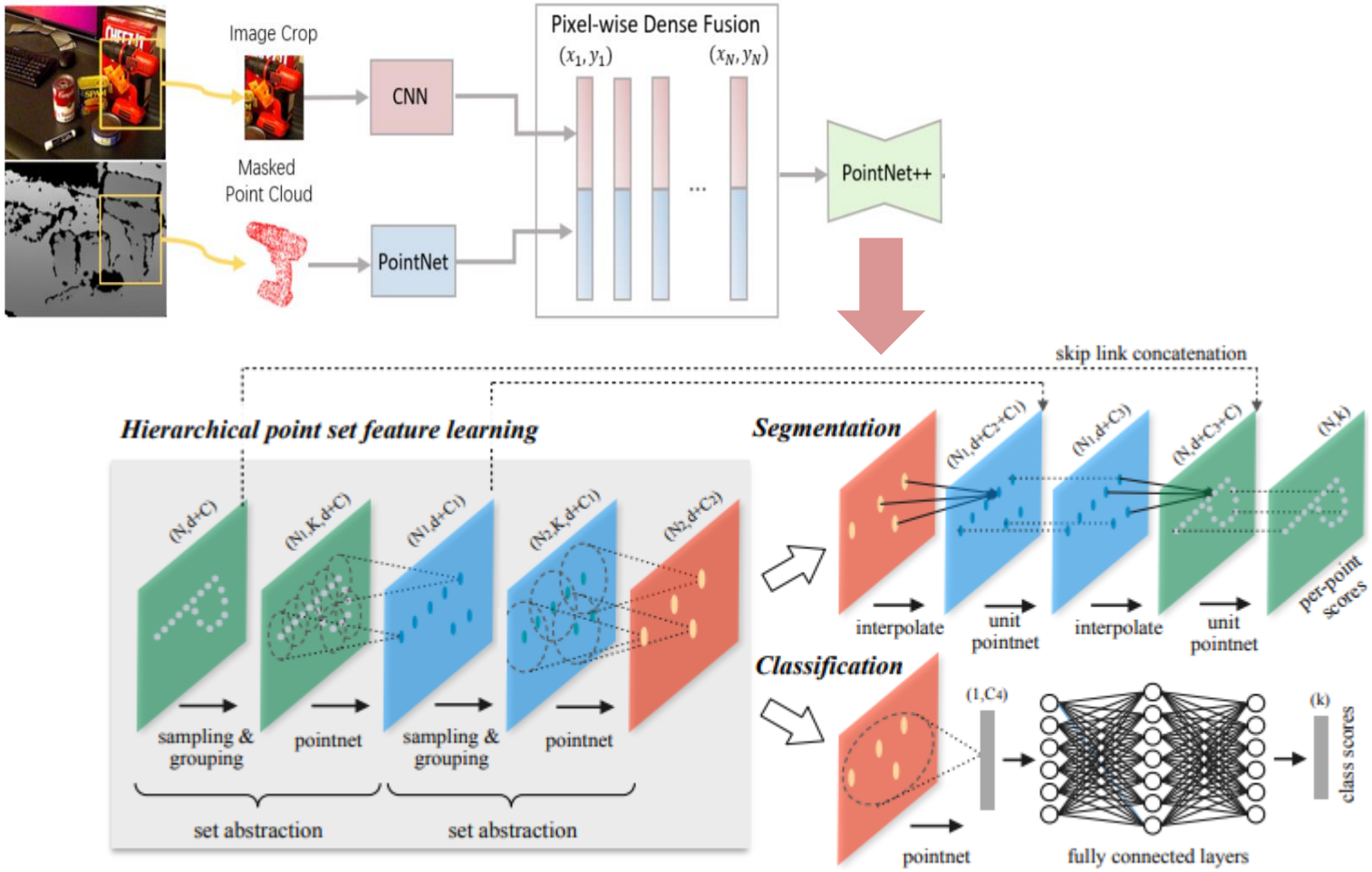
3. Methodology



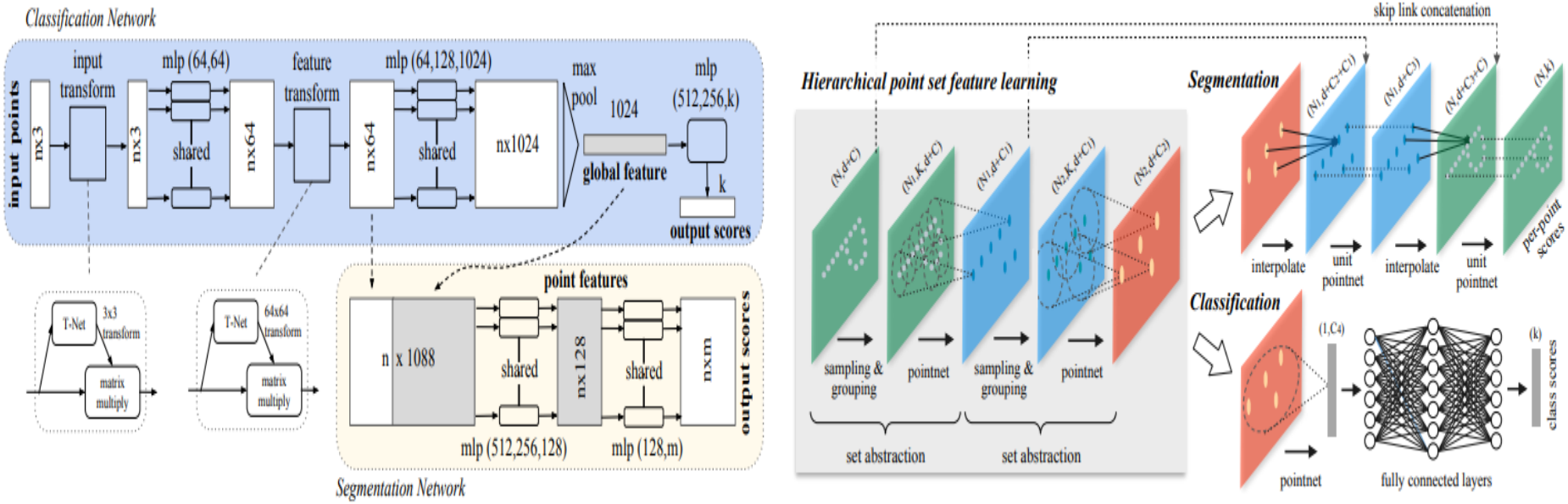
3. Methodology



3. Methodology



3. Methodology

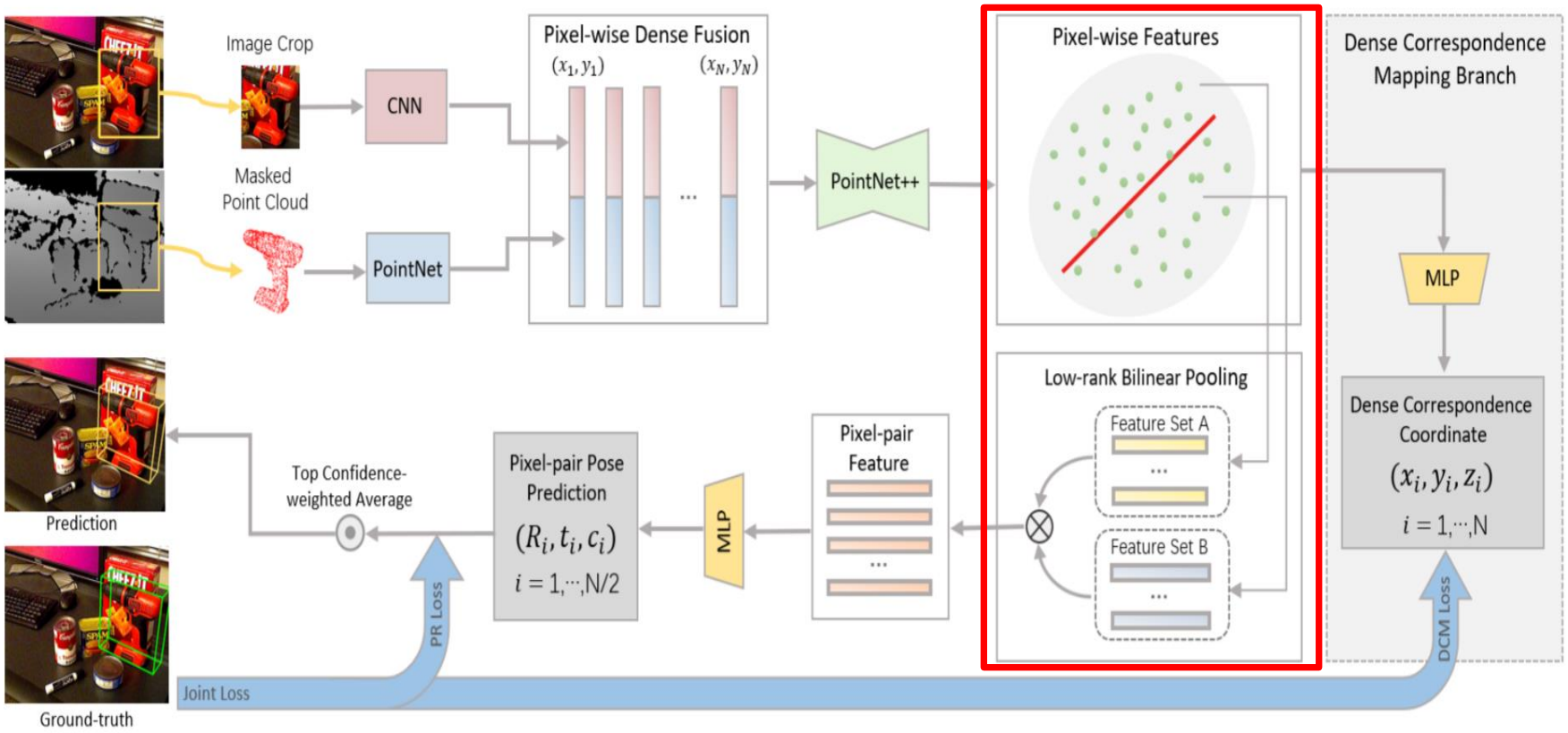


1. 입력 데이터를 Grouping 후, PointNet을 적용해서 Feature Vector 생성.
2. Feature vector들을 다시 Grouping 후, 다시 PointNet
3. 줄어든 Feature Vector들을 이용해서 Cls, Seg 진행.

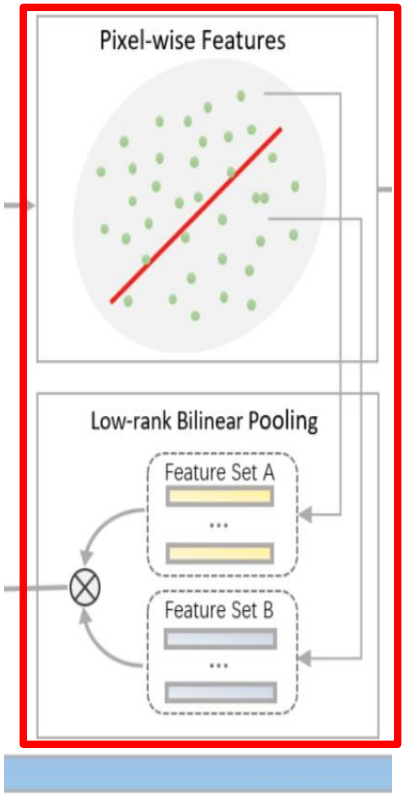
특징 :

Pointnet을 통과해서 나온 Global Feature를 local Feature가 전체 네트워크 구조에서는 Local Feature가 됨.

3. Methodology



3. Methodology



1. Pixel-Pair Pose Regression

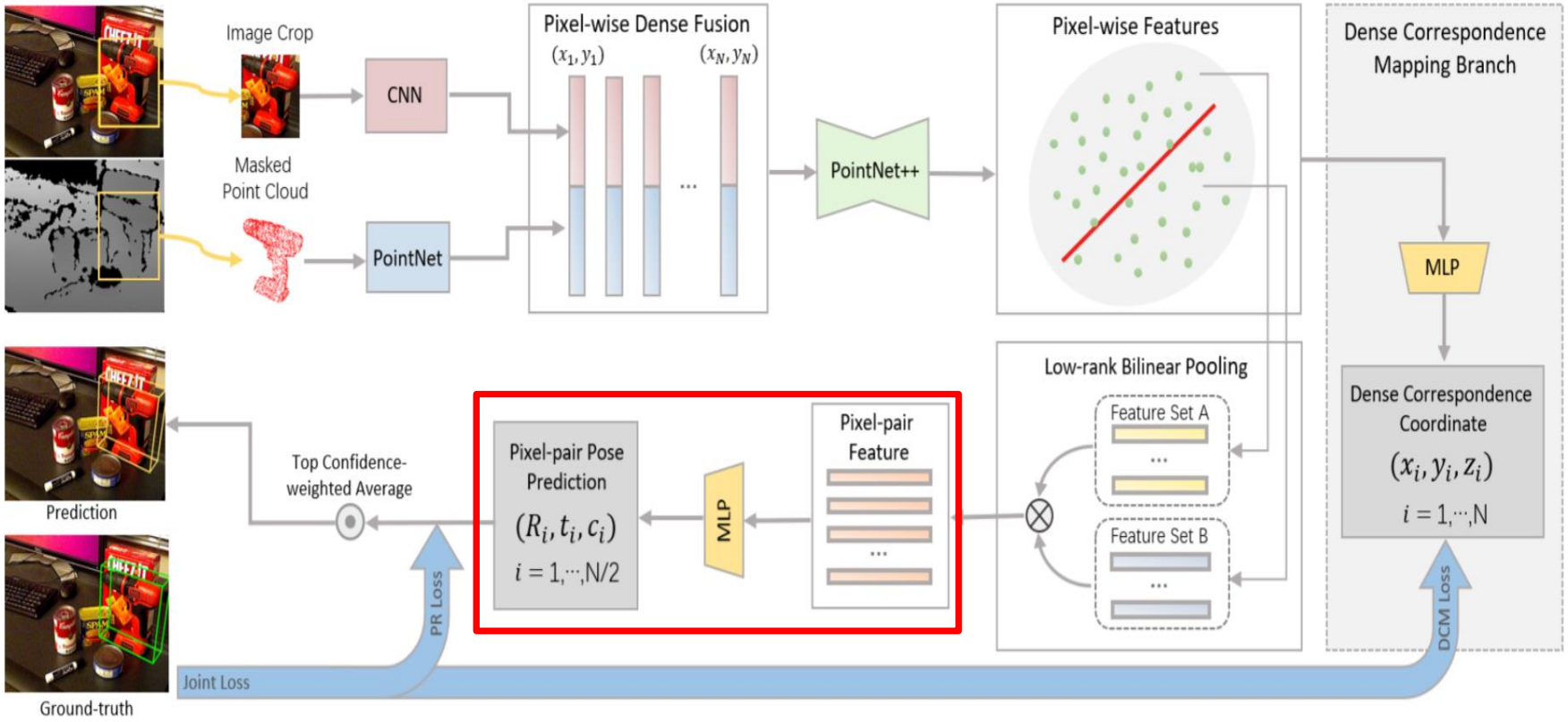
Pixel pair를 만들기 위해 각 pixel당 하나의 쌍을 만듦.

2. Pixel-Pair Feature (ppf) Encoding

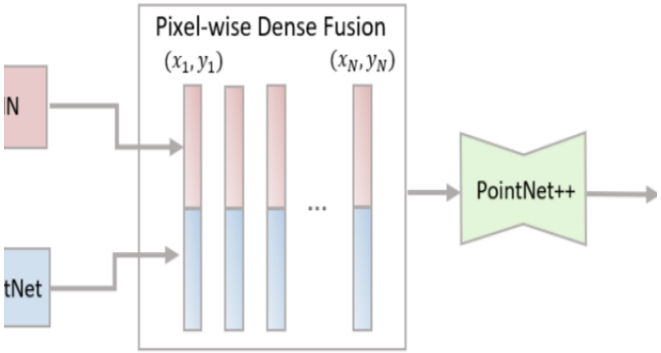
$$\mathcal{G}_{\text{ppf}}(\boldsymbol{a}, \boldsymbol{b}) = \text{vec}(\boldsymbol{a}\boldsymbol{b}^T) \in \mathbb{R}^{d^2_{\text{fusion}}}. \quad (1)$$

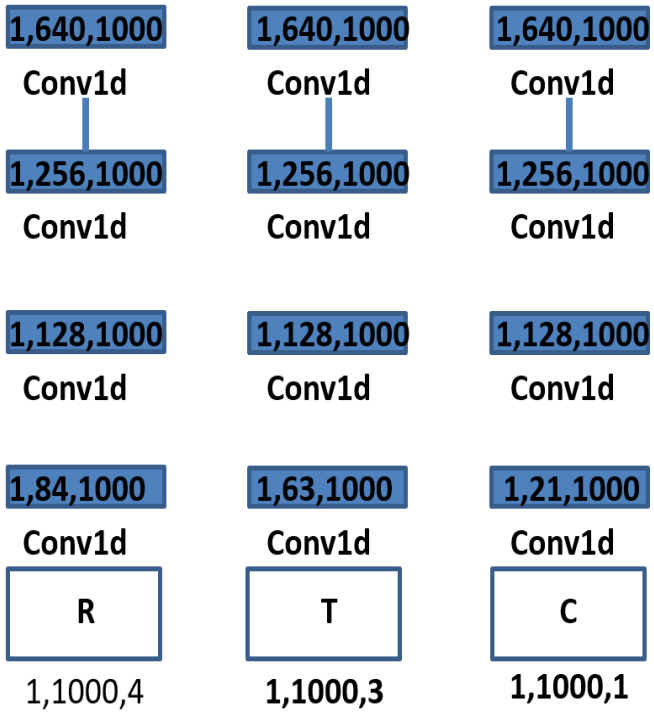
$$\mathcal{G}_{\text{ppf}}(\mathbf{a}, \mathbf{b}) = \boldsymbol{P}^T \sigma(\boldsymbol{U}^T \boldsymbol{a} \circ \boldsymbol{V}^T \boldsymbol{b}) \quad (2)$$

3. Methodology



3. Methodology



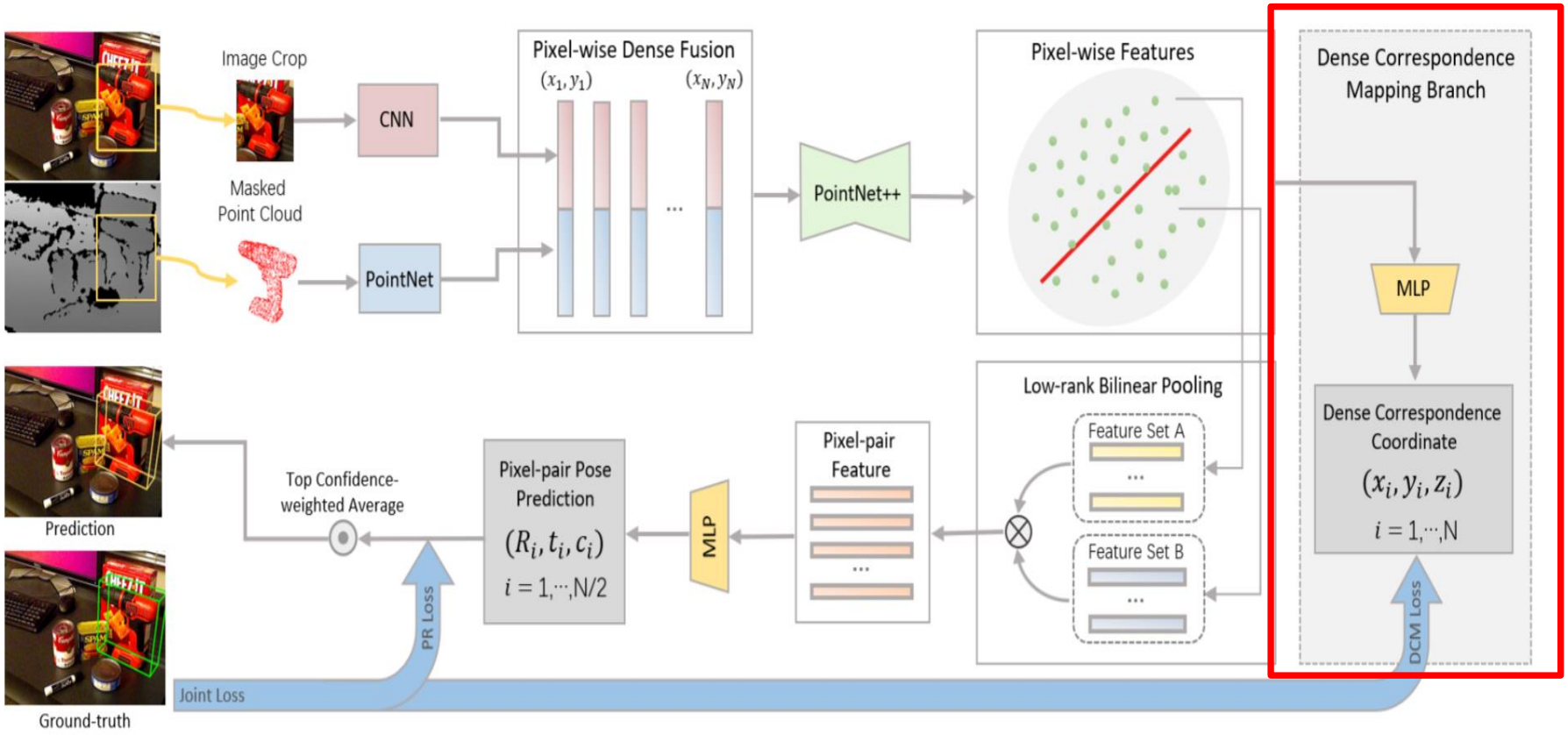


3. Methodology

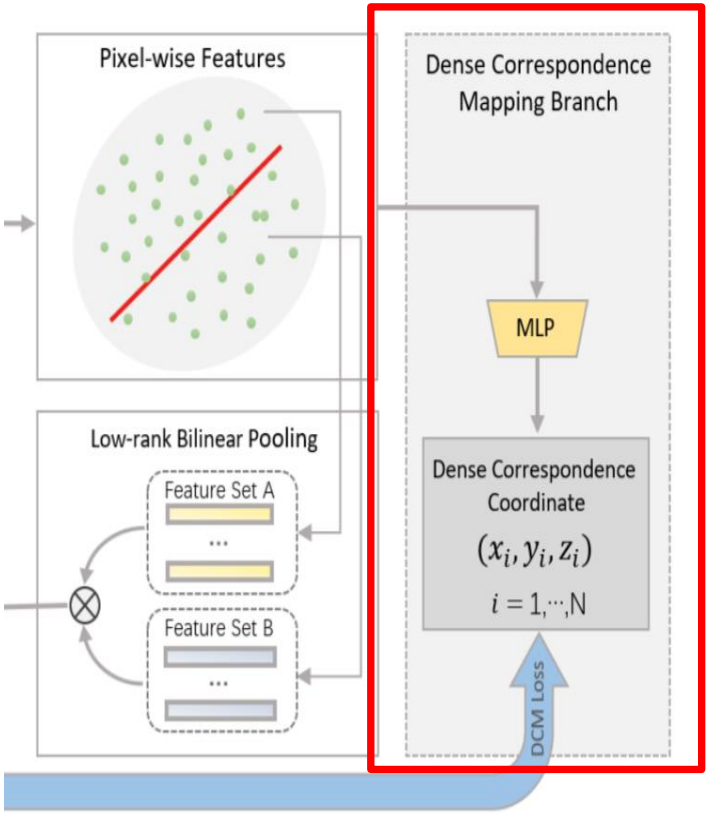
$$L_{\mathrm{PR}} = \frac{1}{N}\sum_{i}(L_i^{pair} c_i - w\log(c_i)), \tag{3}$$

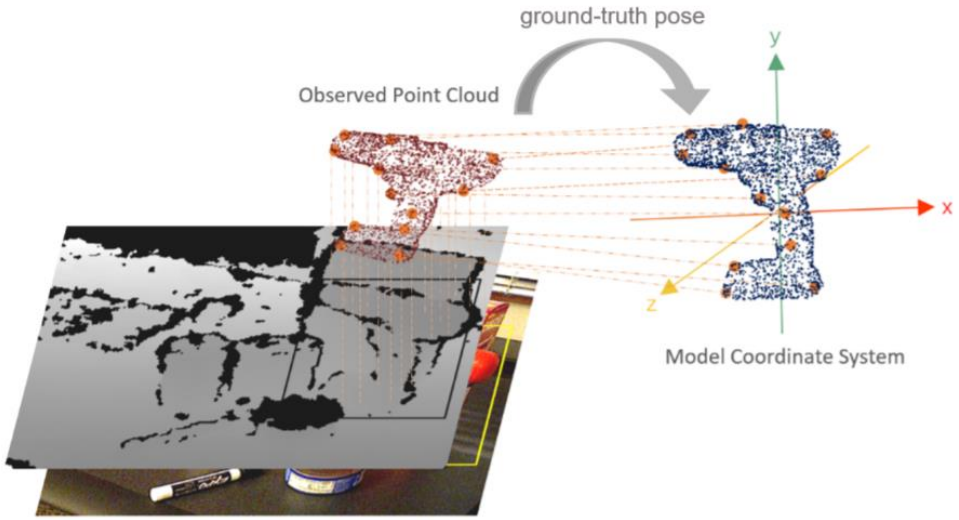
$$L_i^{pair} = \frac{1}{M}\sum_{j}\min_{0<k<M} ||(\boldsymbol{R}x_j + \boldsymbol{t}) - (\hat{\boldsymbol{R}}_i x_k + \hat{\boldsymbol{t}}_i)|| \tag{4}$$

3. Methodology



3. Methodology





$$L_{\text{DCM}} = \frac{1}{M} \sum_{i} ||p_i - \hat{p}_i||$$

for regularization….

4. Experiments

Table 1. Comparative evaluation of 6D pose estimation on the YCB-Video Dataset in terms of the ADD(-S)<2cm and the AUC of ADD(-S) metrics. We compare DenseFusion [39], PoseCNN+ICP [43] with the proposed $\mathcal{W}$-PoseNet with and without iterative pose refinement in Sec. 3.4. Objects in bold are symmetric. Results are report in units of %.

Method	without Refinement				with Refinement					
	DenseFusion		$\mathcal{W}$-PoseNet		PoseCNN+ICP		DenseFusion		$\mathcal{W}$-PoseNet	
	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm
002_master_chef_can	**70.7**	**70.7**	69.2	65.9	68.1	51.1	**73.3**	**72.3**	72.0	68.6
003_cracker_box	86.8	88.6	**87.8**	**90.0**	83.4	73.3	**94.2**	**98.2**	91.3	93.7
004_sugar_box	90.8	96.8	**91.5**	**98.5**	**97.5**	99.5	96.5	**100.0**	95.1	99.8
005_tomato_soup_can	84.7	82.8	**87.4**	**84.2**	81.8	76.6	85.5	83.0	**88.9**	**84.6**
006_mustard_bottle	90.9	94.1	**93.4**	**100.0**	**98.0**	98.6	94.7	96.1	96.5	**100.0**
007_tuna_fish_can	**79.5**	**58.5**	77.0	55.9	**83.9**	**72.1**	81.9	62.2	78.8	61.4
008_pudding_box	89.4	94.4	**91.8**	**98.1**	**96.6**	**100.0**	93.2	98.6	94.5	**100.0**
009_gelatin_box	**95.7**	**100.0**	94.6	**100.0**	**98.1**	**100.0**	96.7	**100.0**	96.0	**100.0**
010_potted_meat_can	**79.6**	76.9	79.0	**77.8**	83.5	77.9	**83.6**	79.9	82.6	**80.4**
011_banana	76.8	60.2	**87.9**	**87.3**	91.9	88.1	83.7	88.4	**92.8**	**98.9**
019_pitcher_base	87.1	87.2	**92.0**	**100.0**	**96.9**	97.7	**96.9**	**100.0**	95.0	**100.0**
021_bleach_cleanser	**87.5**	**85.4**	85.2	77.4	**92.5**	**92.7**	89.7	90.8	89.5	89.0
024_bowl	86.1	**61.3**	**86.2**	49.8	81.0	54.9	**89.5**	**95.1**	87.7	93.6
025_mug	83.9	**80.5**	**84.9**	79.7	81.1	55.2	**88.9**	88.8	88.2	**90.3**
035_power_drill	83.7	83.1	**91.1**	**98.8**	**97.7**	99.2	92.7	96.5	93.6	**99.5**
036_wood_block	**89.4**	**98.8**	86.3	96.3	87.6	80.2	**92.8**	**100.0**	87.0	97.5
037_scissors	77.1	50.8	**91.5**	**99.5**	78.4	49.2	77.5	48.6	**90.7**	**97.8**
040_large_marker	89.1	90.6	**90.9**	**96.3**	85.3	87.2	**93.0**	**100.0**	92.5	99.4
051_large_clamp	**71.5**	**78.0**	71.4	74.0	**75.2**	74.9	72.5	78.7	70.8	**79.2**
052_extra_large_clamp	**70.1**	**72.0**	68.0	60.4	64.4	48.8	**69.9**	**74.9**	69.6	72.7
061_foam_brick	92.2	**100.0**	**92.5**	**100.0**	**97.2**	**100.0**	91.9	**100.0**	92.9	**100.0**
MEAN	83.9	81.5	**85.7**	**85.2**	86.6	79.9	87.6	88.2	**87.9**	**90.8**

4. Experiments

Table 2. Comparative evaluation of 6D pose estimation in terms of ADD(-S) on the LineMOD dataset. Objects with bold name are symmetric. All methods use depth image. Results are report in units of %.

Method	without Refinement				with Refinement			
	DenseFusion [39]	*W*-PoseNet w/o DCM Loss	*W*-PoseNet w/o Pixel Pair Prediction	*W*-PoseNet	Implicit +ICP[33]	SSD-6D +ICP[12]	DenseFusion [39]	*W*-PoseNet
ape	79.5	80.1	82.9	**86.3**	20.6	65.0	92.3	**92.8**
bench vi.	84.2	91.7	95.8	**97.4**	64.3	80.0	93.2	**99.6**
camera	76.5	82.7	94.9	**98.3**	63.2	78.0	94.4	**99.0**
can	86.6	88.7	94.8	**97.5**	76.1	86.0	93.1	**99.3**
cat	88.8	87.1	94.2	**97.3**	72.0	70.0	96.5	**99.0**
driller	77.7	82.4	89.5	**95.9**	41.6	73.0	87.0	**97.8**
duck	76.3	81.3	83.2	**93.3**	32.4	66.0	92.3	**96.2**
eggbox	**99.9**	**99.9**	99.6	**99.9**	98.6	**100.0**	99.8	99.9
glue	99.4	99.7	99.5	**99.8**	96.4	**100.0**	**100.0**	99.9
hole p.	79.0	80.9	90.1	**95.7**	49.9	49.0	92.1	**97.3**
iron	92.1	88.1	**98.0**	96.6	63.1	78.0	97.0	**98.6**
lamp	92.3	93.5	96.3	**99.2**	91.7	73.0	95.3	**99.8**
phone	88.0	89.9	93.4	**96.4**	71.0	79.0	92.8	**98.3**
MEAN	86.2	88.2	93.2	**96.4**	64.7	77.0	94.3	**98.2**

4. Experiments

Table 3. Ablation studies in sparsity of pixel-pair features. Models are trained on the LineMOD dataset in terms of ADD(-S) metric.

Number of Pixel-pair	Accuracy
100	95.6%
250	96.4%
500	95.9%
750	95.4%
1000	94.9%

4. Experiments

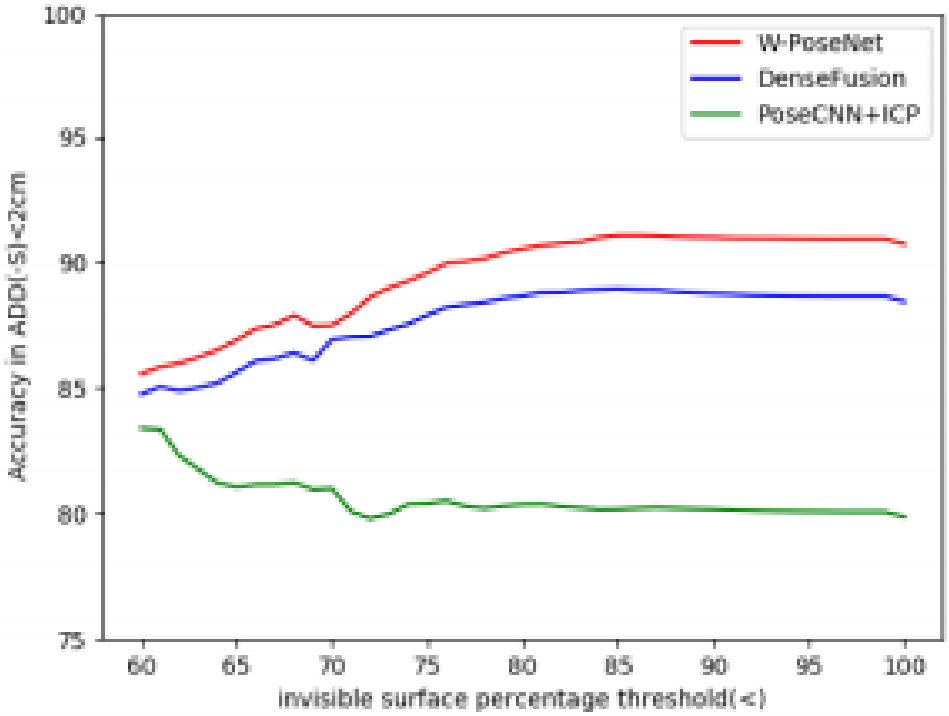


Figure 5. Performance comparison of our $\mathcal{W}$-PoseNet and two state-of-the-art methods under different degree of occlusion.

End