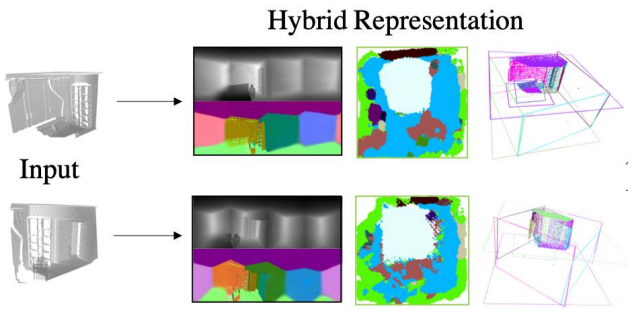
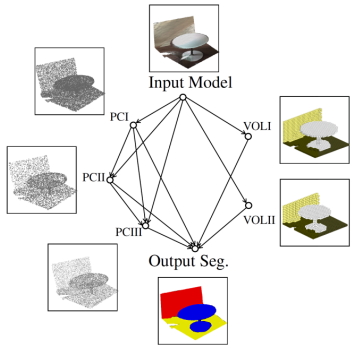
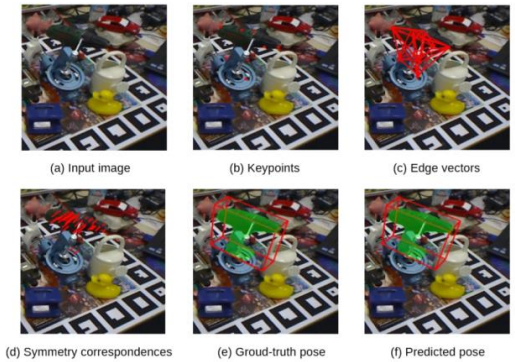
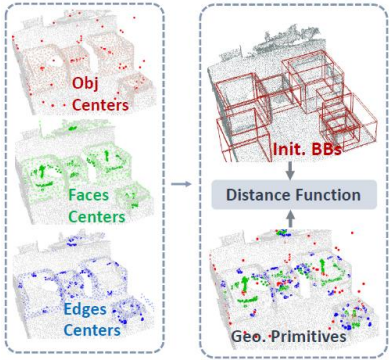
Hybrid 3D Representations for 3D Understanding



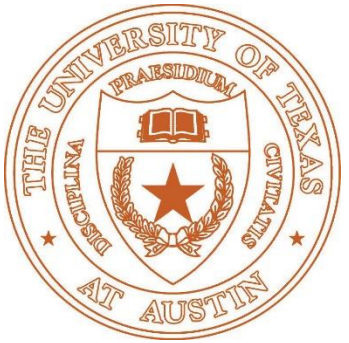






Qixing Huang

July 17th 2020

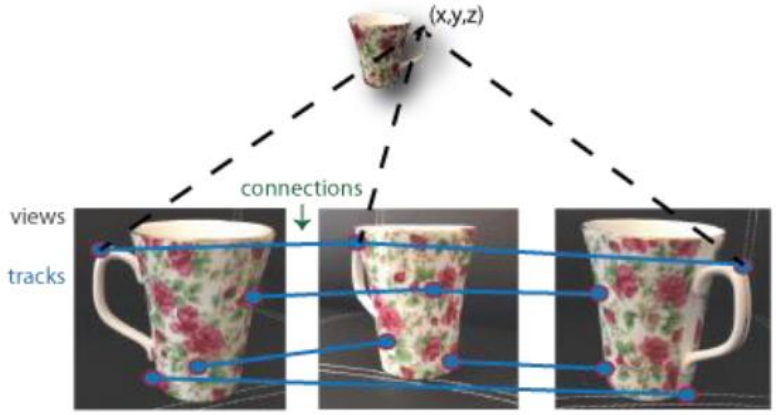


Ranking

The only top-10 program within 1000 miles

Live in Austin

3D Vision



Recovering the Underlying 3D structures from Images

Pose Estimation

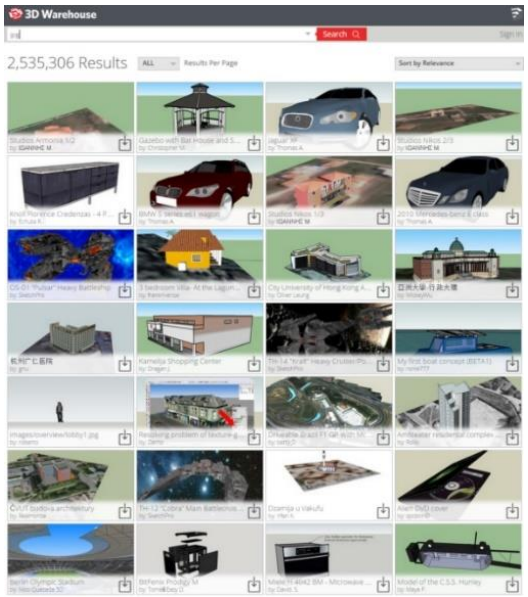
Structure-from-motion

Multi-view Stereo

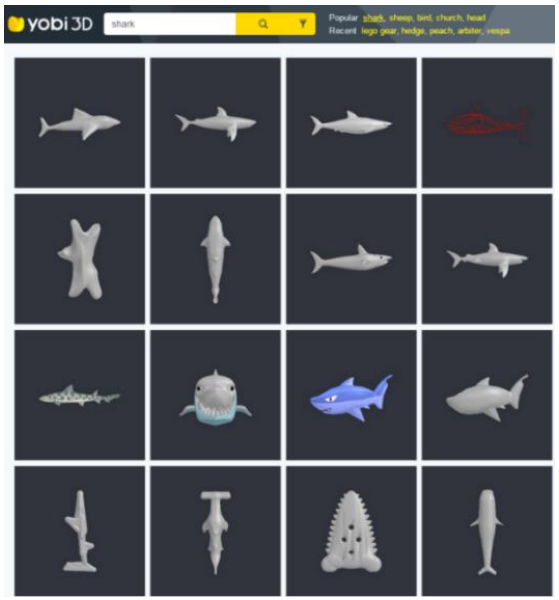
RGB images

RGB-D images

Large-scale 3D repositories



3D Warehouse



Yobi3D

3D-FRONT [Fu et al. 17]

70k Indoor Scenes

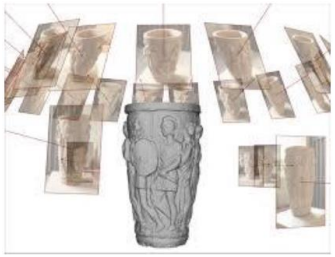
ScanNet [Dai et al. 17]

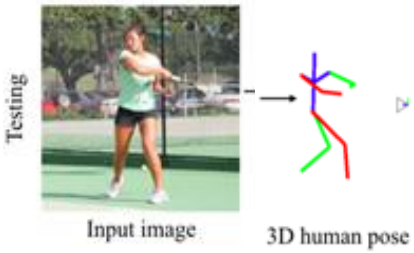
1k Indoor Scenes

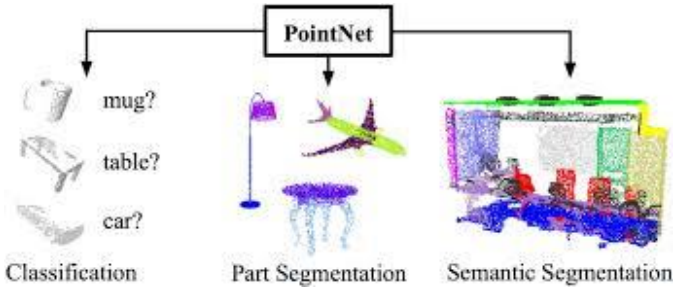
3M models in more than 4K categories

3D Vision (2015-)

RGB (RGBD) Images → 3D Structures → Understanding







Pose Estimation

Structure-from-motion

Multi-view Stereo

Classification

Segmentation

Detection

2D Vision versus 3D Vision

2D Vision

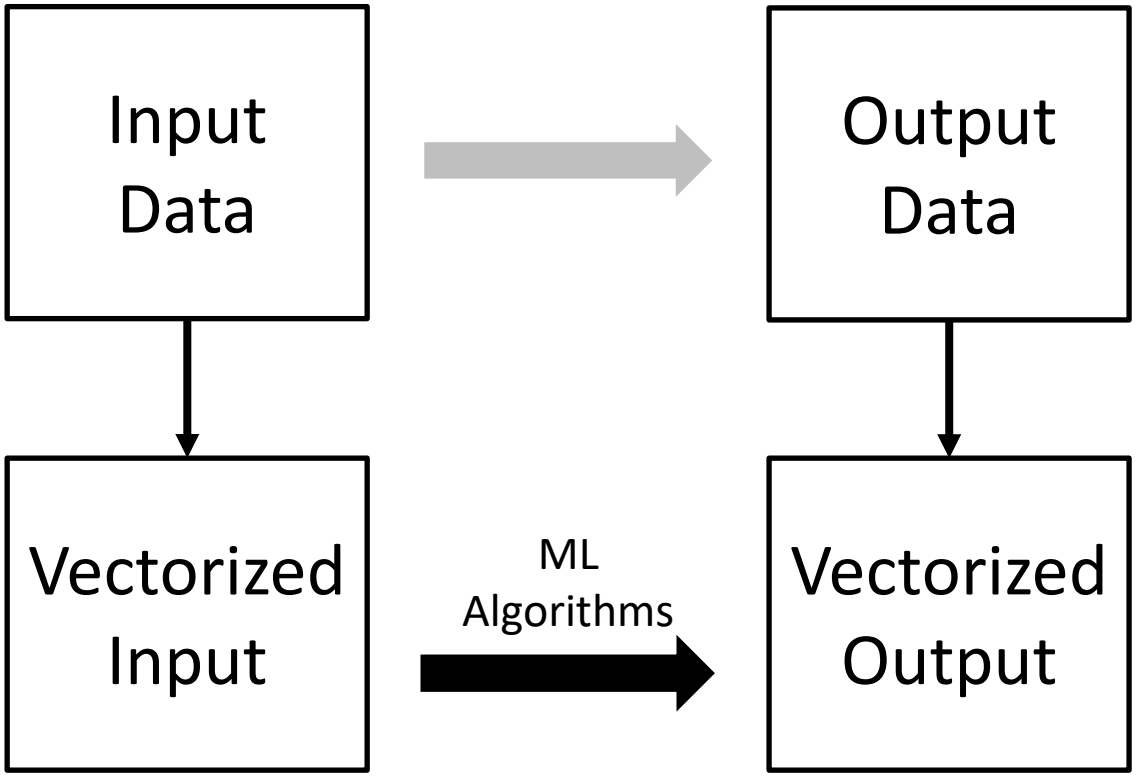
3D Vision

Classification

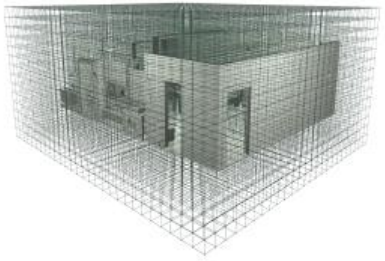
Segmentation

Detection

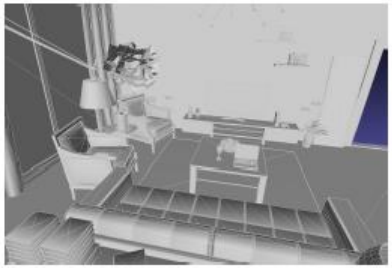
The notion of representation



Input and output 3D Representations

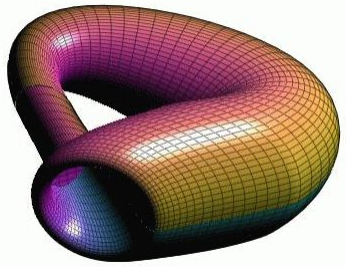


Volumetric



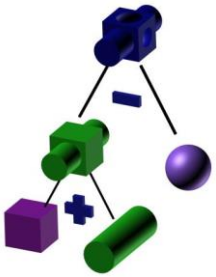
Triangular mesh

Point cloud



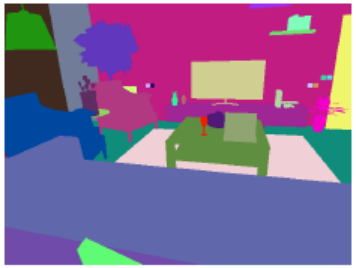
Parametric surfaces

Constructive solid geometry

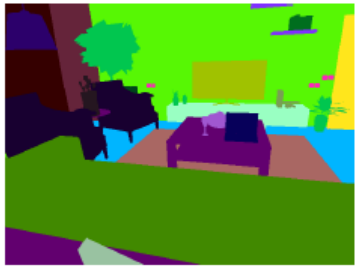


Multi-view

Scene-graph

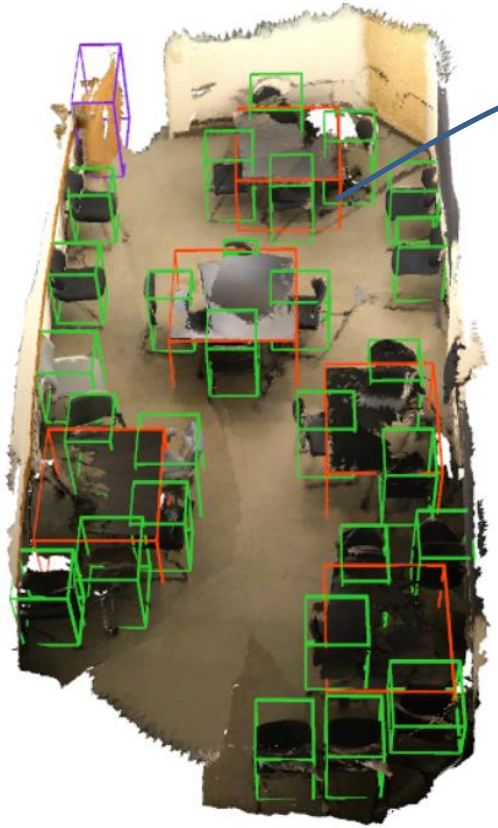


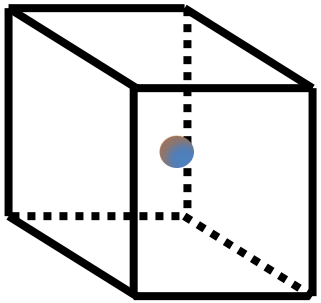
Semantic segments



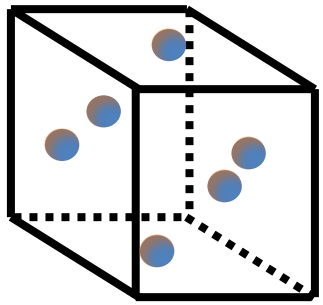
Input and output 3D Representations

[Zhang et al. 20]

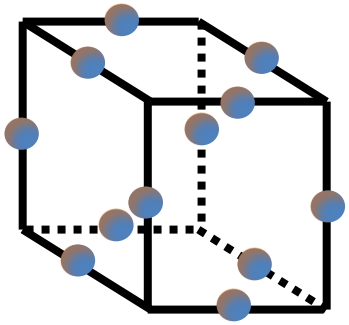




Box center

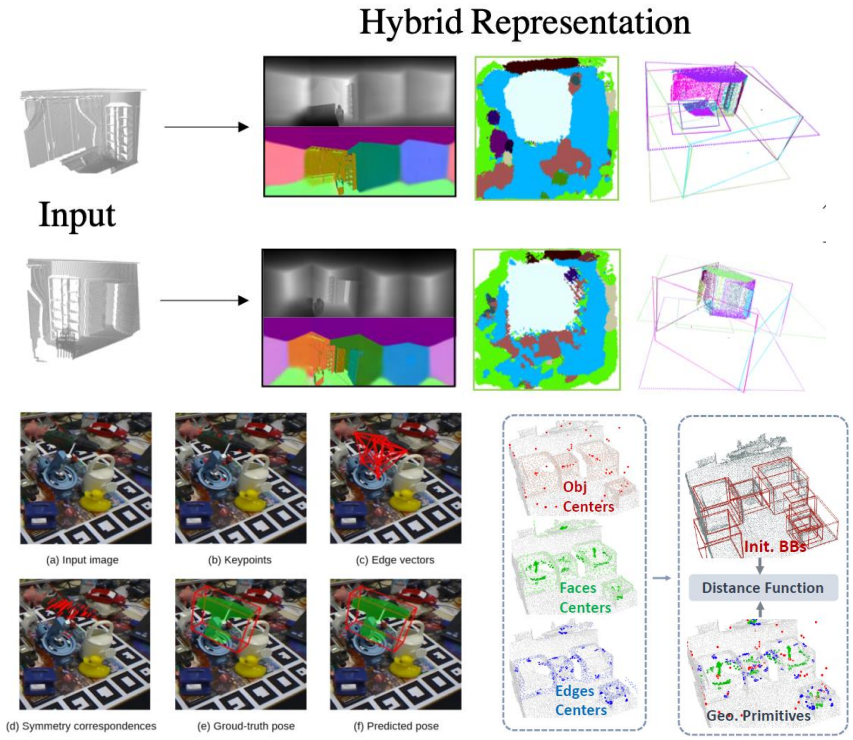


6 Face centers

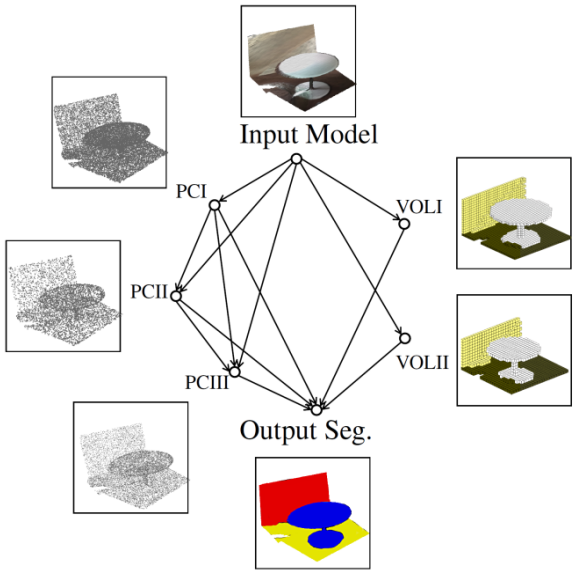


12 Edge centers

Theme: hybrid representations

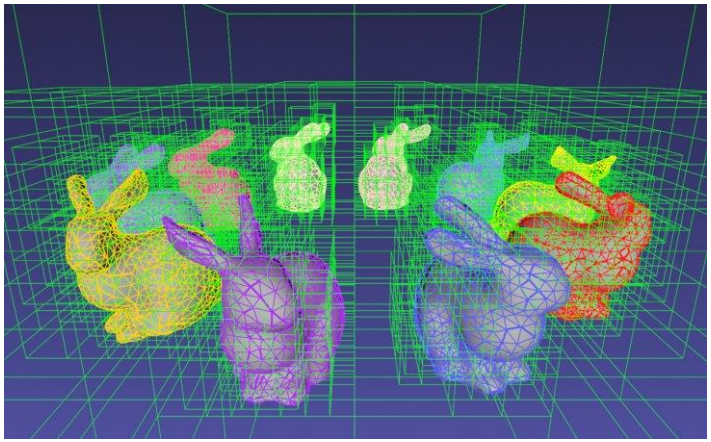


Supervised

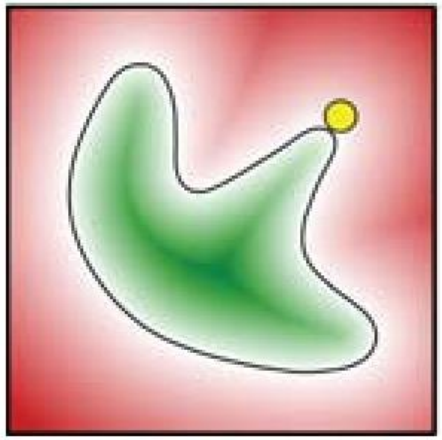


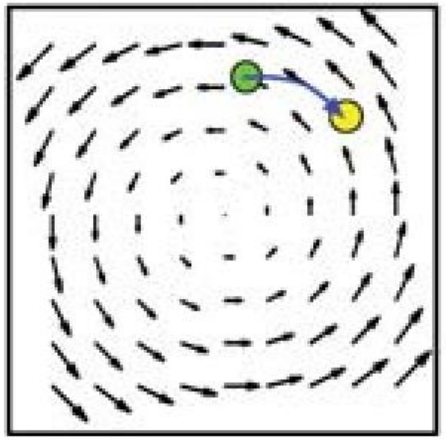
Semi-supervised

Hybrid Representations in Graphics



Mesh + Spatial Data Structure
for collision detection





Explicit + Implicit Reps.
for Fluid Simulation
[Bargteil et al. 06]

Figure credit: https://blog.kitware.com/octree-collision-imstk/

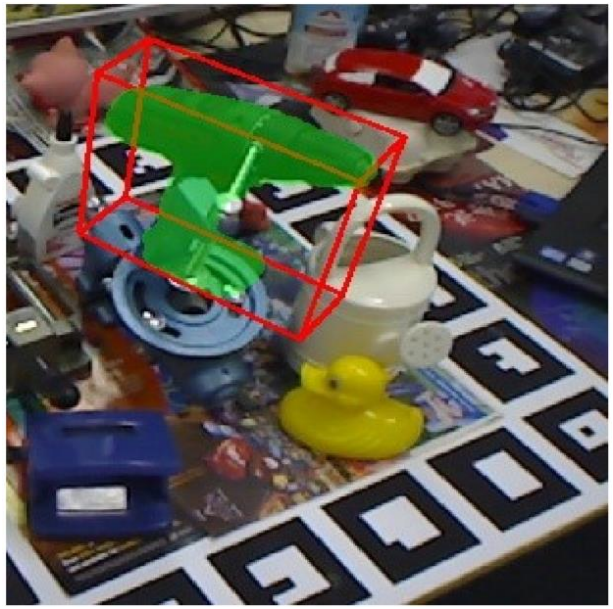
HybridPose: 6D Object Pose Estimation Under Hybrid Representations

[Song, Song, H., CVPR 2020]

Github: https://github.com/chensong1995/HybridPose

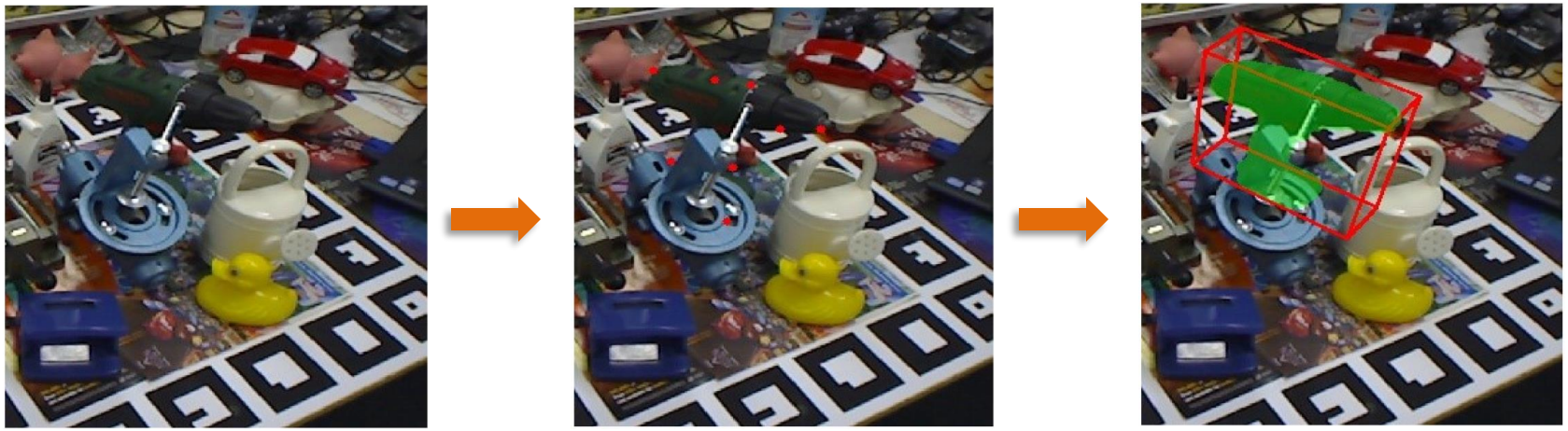
Task

Input image



Output pose

A popular framework

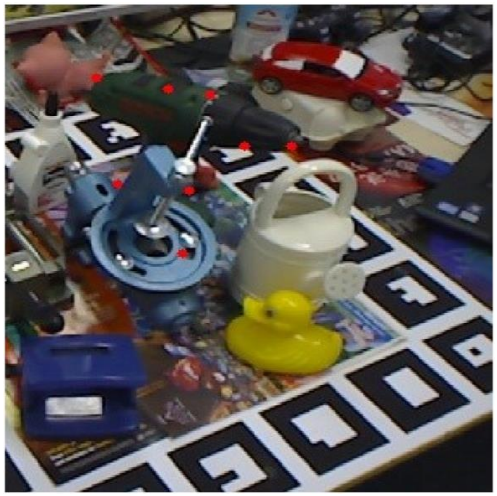


Keypoints

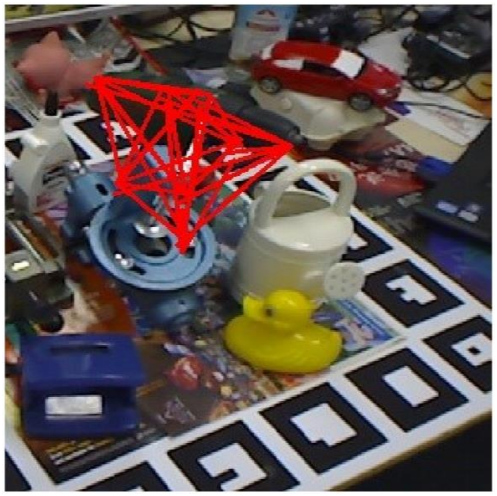
Pose regression

- ✓ Intermediate supervision
- ✓ Feature learning + geometric constraints

Hybrid intermediate representations



Keypoints

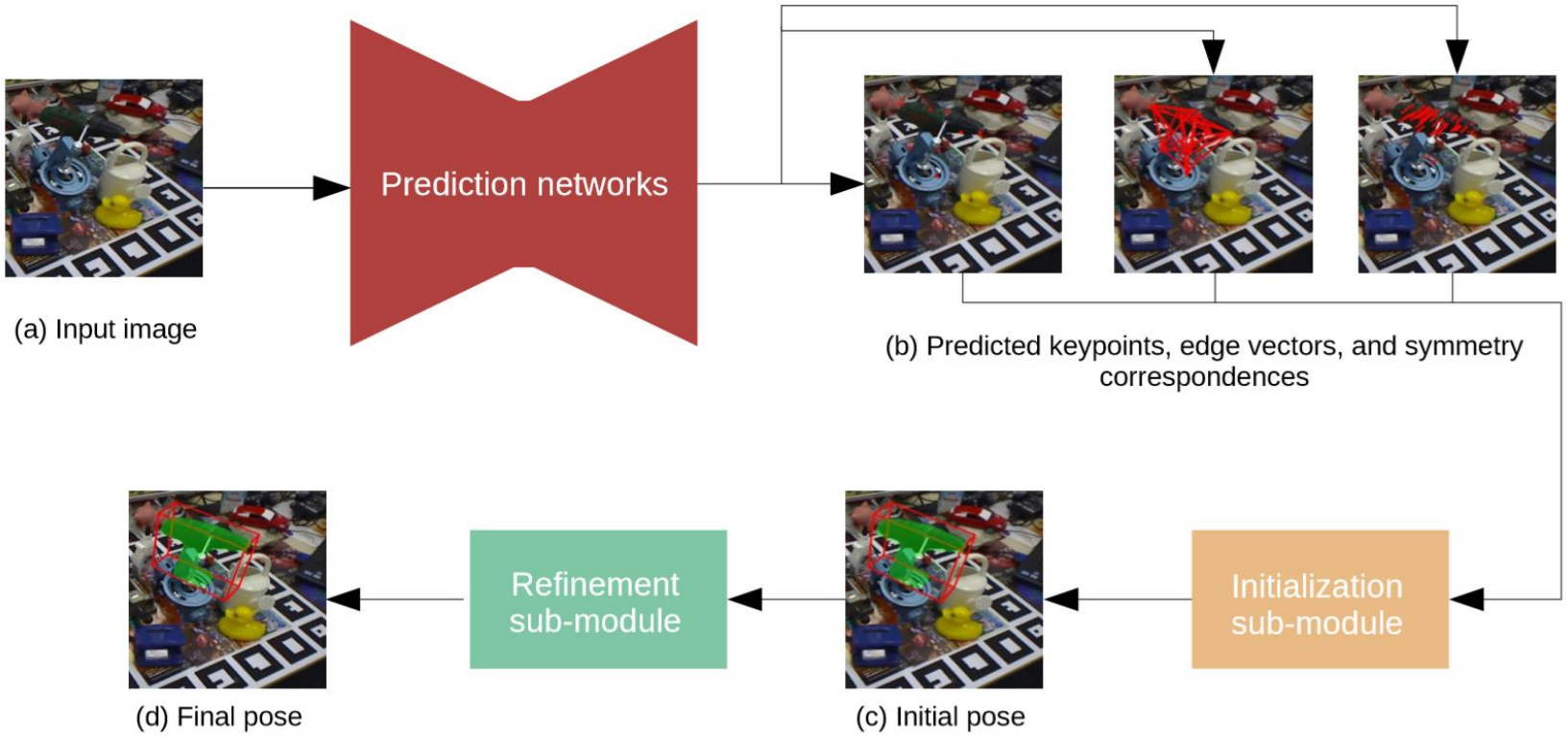


Edge vectors: edges between keypoints

Sym corres.: corres. of the underlying reflect. sym.

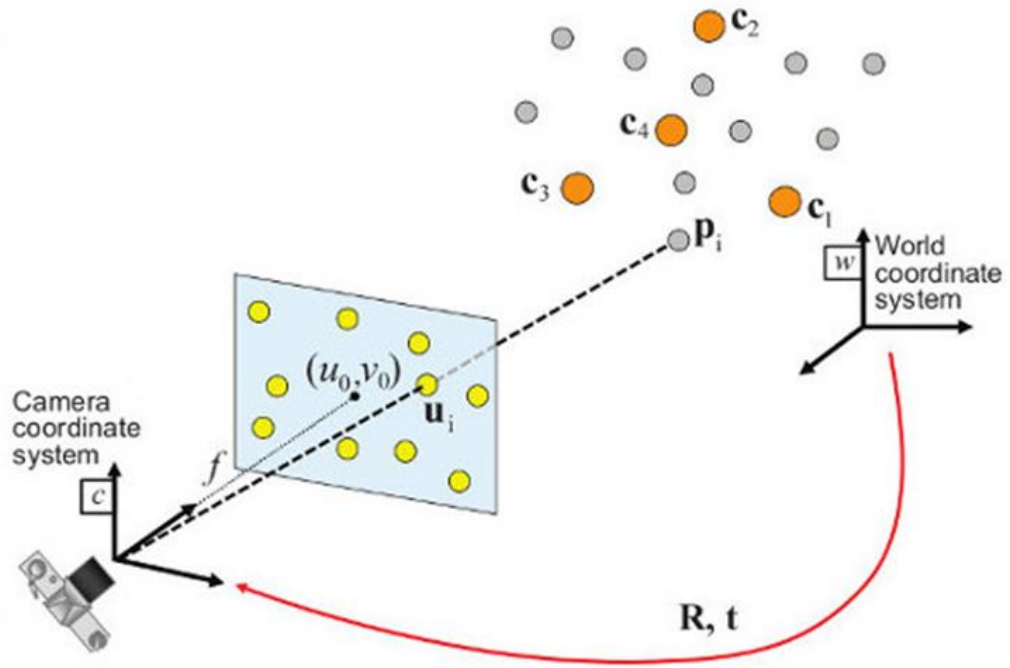
- Each intermediate representation tends to learn different features
- All provide meaningful constraints on the underlying pose

Neural architecture



PVNet [Peng et al. 19] as the back-bone network architecture

Point-based pose estimation



$$\overline{\boldsymbol{r}}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k) := \hat{\boldsymbol{p}}_k \times (R\overline{\boldsymbol{p}}_k + \boldsymbol{t}) = 0$$

Slide credit: https://docs.opencv.org/4.3.0/dc/d2c/tutorial_real_time_pose.html

Pose regression

Geometric constraints:

Keypoints: $\overline{\boldsymbol{r}}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k) := \hat{\boldsymbol{p}}_k \times (R\overline{\boldsymbol{p}}_k + \boldsymbol{t})$

Edge vectors: $\overline{\boldsymbol{r}}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e, \boldsymbol{p}_{e_s}) := \hat{\boldsymbol{v}}_e \times (R\overline{\boldsymbol{p}}_{e_t} + \boldsymbol{t}) + \hat{\boldsymbol{p}}_{e_s} \times (R\overline{\boldsymbol{v}}_e)$

Sym. corres.: $r^{\mathcal{S}}_{R,\boldsymbol{t}}(\boldsymbol{q}_{s,1}, \boldsymbol{q}_{s,2}) := (\hat{\boldsymbol{q}}_{s,1} \times \hat{\boldsymbol{q}}_{s,2})^T \boldsymbol{R}\overline{\boldsymbol{n}}_r$

Regression objective:

$$\min_{R,\boldsymbol{t}} \sum_{k=1}^{|\mathcal{K}|} \rho(\|\boldsymbol{r}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k)\|, \beta_{\mathcal{K}}) \|\boldsymbol{r}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k)\|^2_{\Sigma_k}$$
$$+ \frac{|\mathcal{K}|}{|\mathcal{E}|} \sum_{e=1}^{|\mathcal{E}|} \rho(\|\boldsymbol{r}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e)\|, \beta_{\mathcal{E}}) \|\boldsymbol{r}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e)\|^2_{\Sigma_e}$$
$$+ \frac{|\mathcal{K}|}{|\mathcal{S}|} \sum_{s=1}^{|\mathcal{S}|} \rho(r^{\mathcal{S}}_{R,\boldsymbol{t}}(\boldsymbol{q}_{s,1}, \boldsymbol{q}_{s,2}), \beta_{\mathcal{S}})$$

EPnP [Lepetit et al. 09] for initialization

+

Gauss-Newton for refinement

Robust norm: $\rho(x, \beta) := \beta_1^2/(\beta_2^2 + x^2)$

Experimental results

	keypoints		keypoints + symmetries		full model	
	Rot.	Trans.	Rot.	Trans.	Rot.	Trans.
ape	1.914°	0.107	1.809°	0.113	1.543°	0.092
can	1.472°	0.059	1.710°	0.073	0.912°	0.041
cat	1.039°	0.119	0.888°	0.117	0.751°	0.055
driller	1.180°	0.057	1.180°	0.057	0.803°	0.027
duck	1.773°	0.116	1.679°	0.115	1.439°	0.068
eggbox	1.675°	0.107	1.587°	0.105	1.052°	0.052
glue	1.796°	0.097	1.681°	0.099	1.224°	0.066
holepuncher	2.319°	0.141	2.192°	0.140	1.704°	0.051
mean	1.648°	0.100	1.590°	0.102	1.179°	0.057

Linmod-Occlusion: the median of absolute angular error in rotation, and the median of relative error in translation with respect to object diameter.

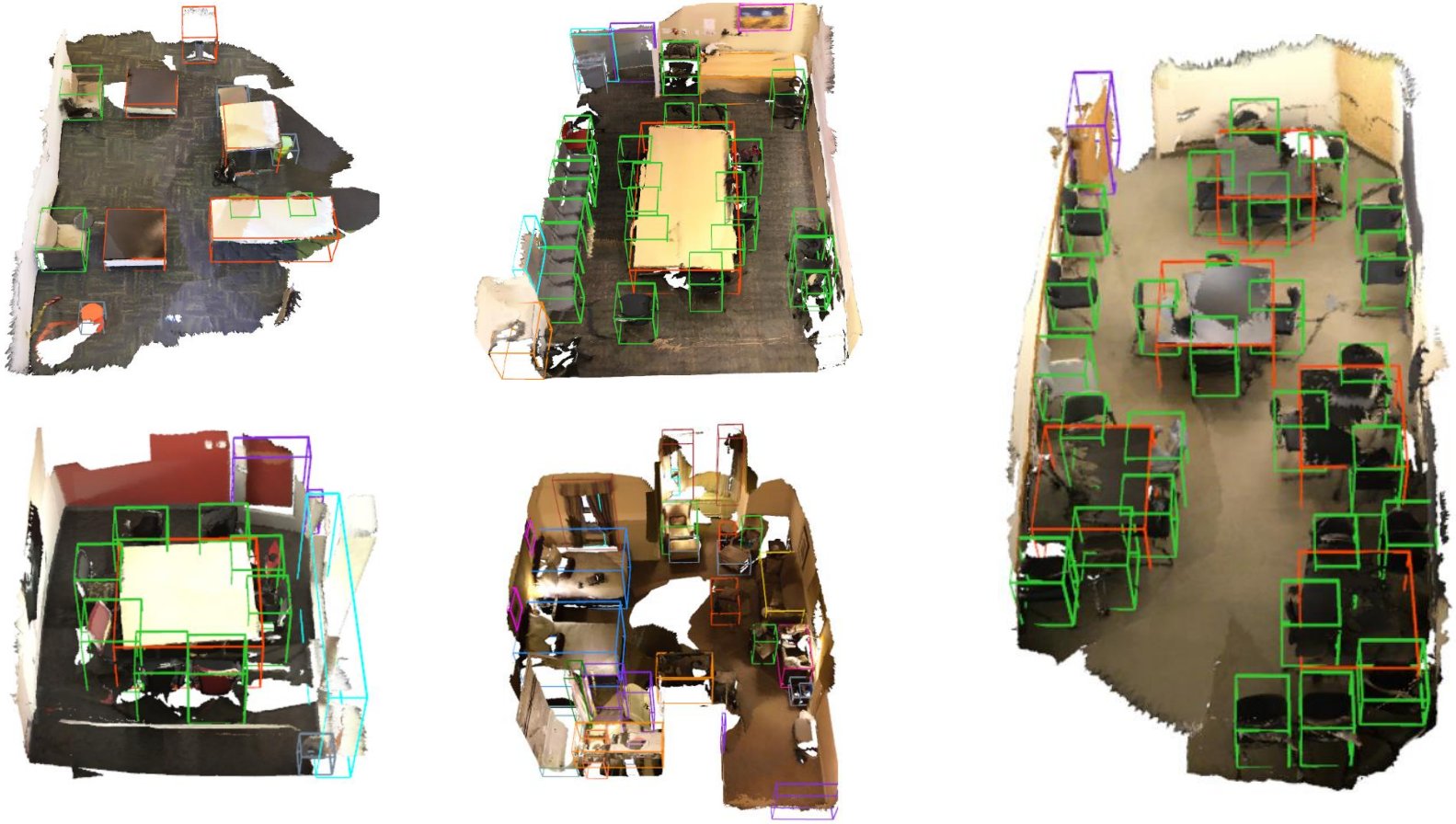
Hybrid

H3DNet: 3D Object Detection Using Hybrid Geometric Primitives

[Zhang, Su, Yang, H., ECCV 2020]

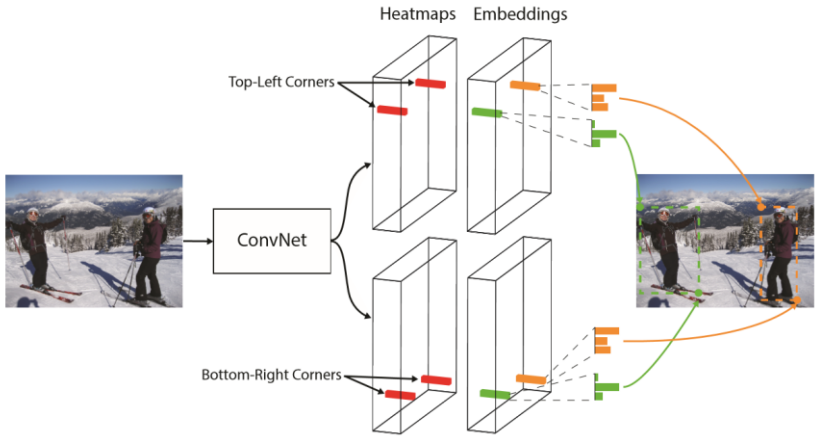
Github: https://github.com/zaiweizhang/H3DNet

Task

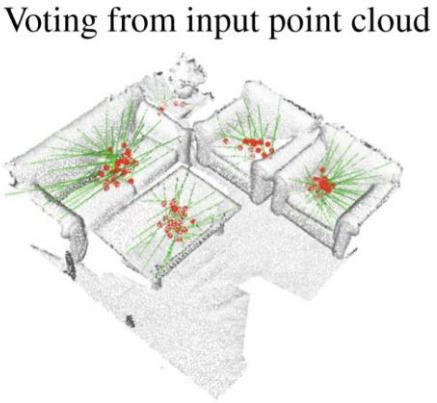


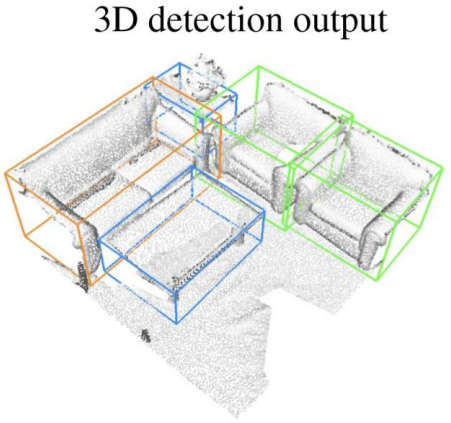
3D Scene -> Oriented object bounding boxes

Object detection as regression

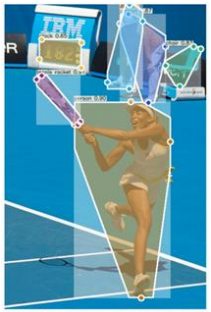


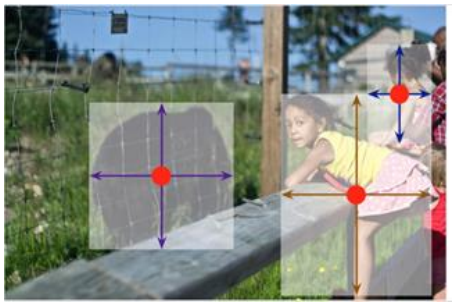
CornerNet [Law and Deng, 18]





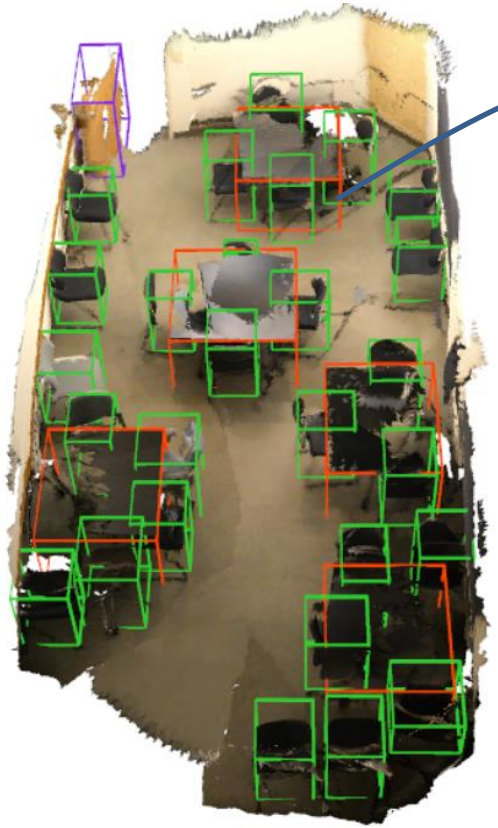
VoteNet [Qi et al. 19]

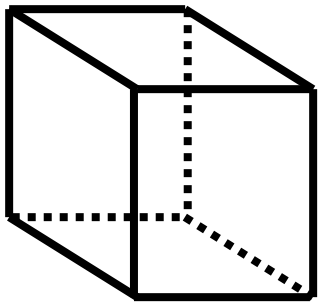


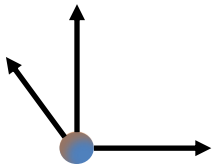


Extremal and/or center based
[Zhou et al. 19a, Zhou et al. 19b, Duan et al. 19]

Regression depends on bounding box representations



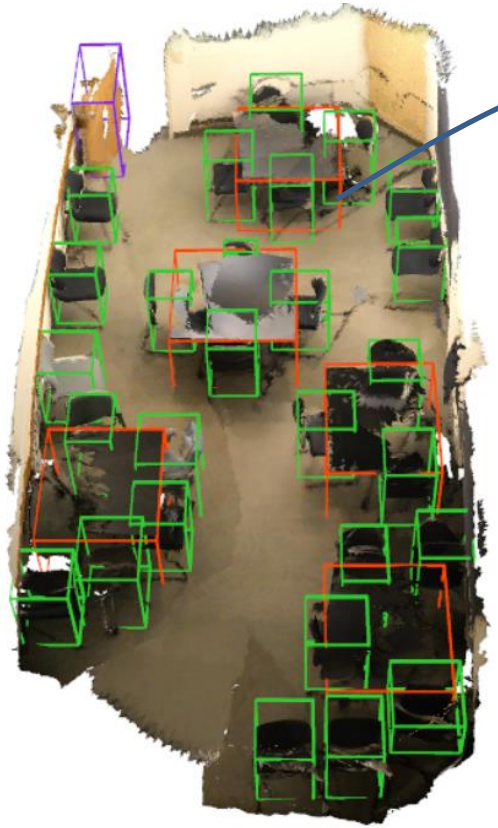


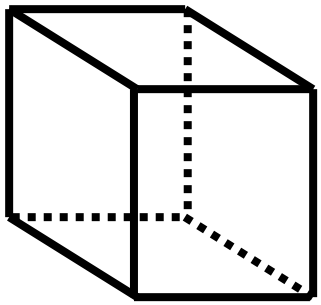


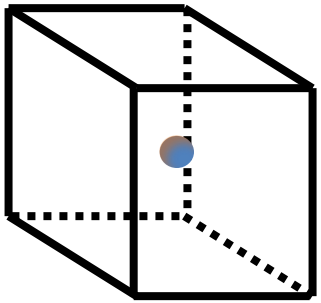
Center+Size [Qi et al. 19]

Our approach

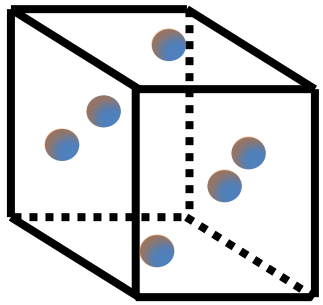
[Zhang et al. 20]



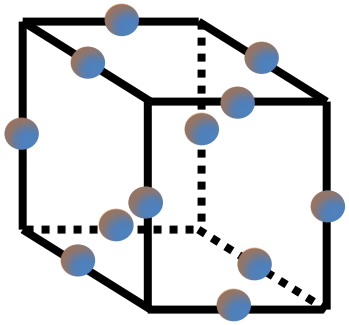




Box center

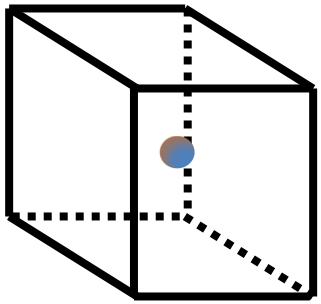


6 Face centers

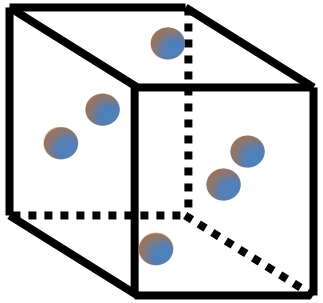


12 Edge centers

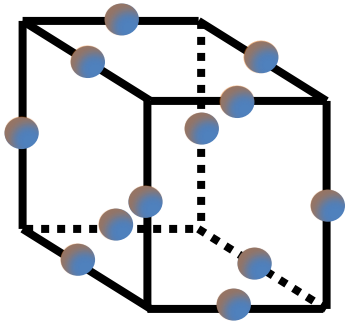
Different representations suitable for different object instances



Box
center

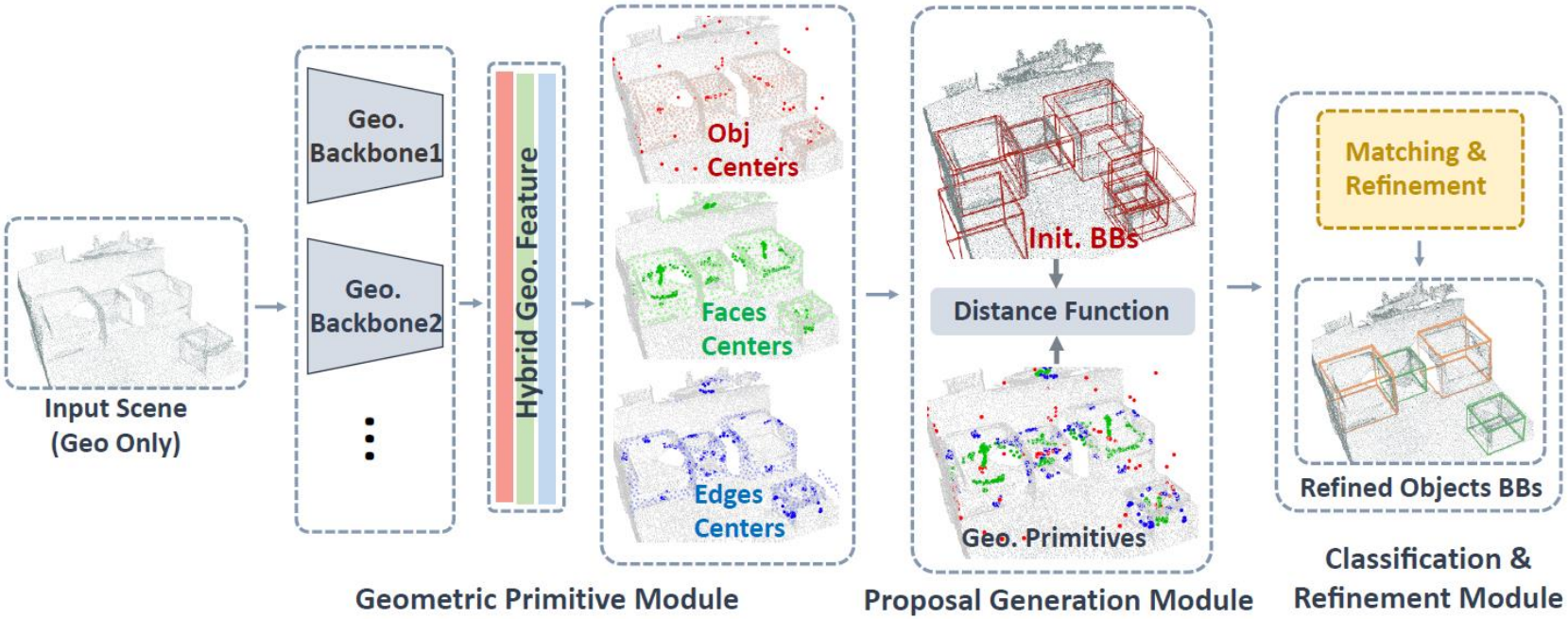


6 Face
centers

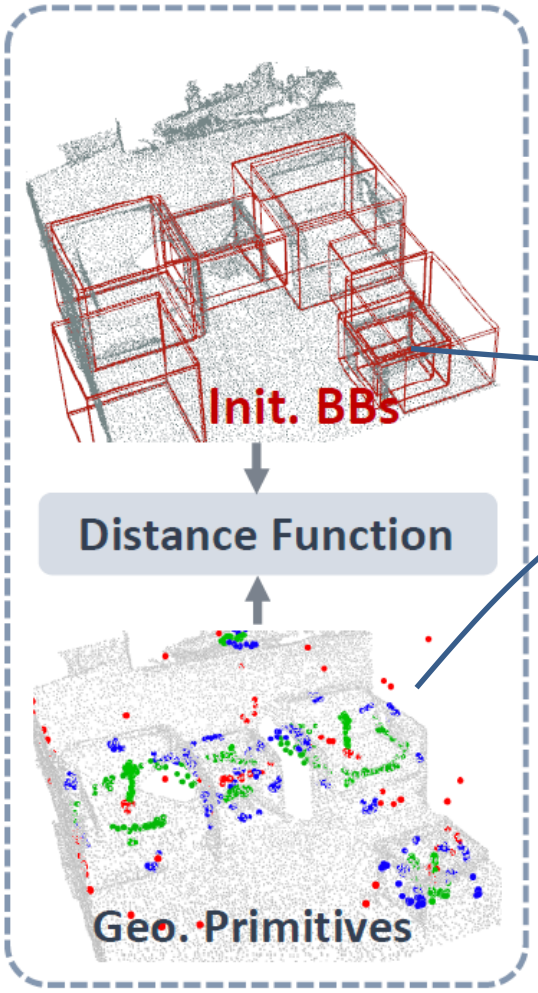


12 Edge
centers

Network architecture



Continuous optimization for object proposals

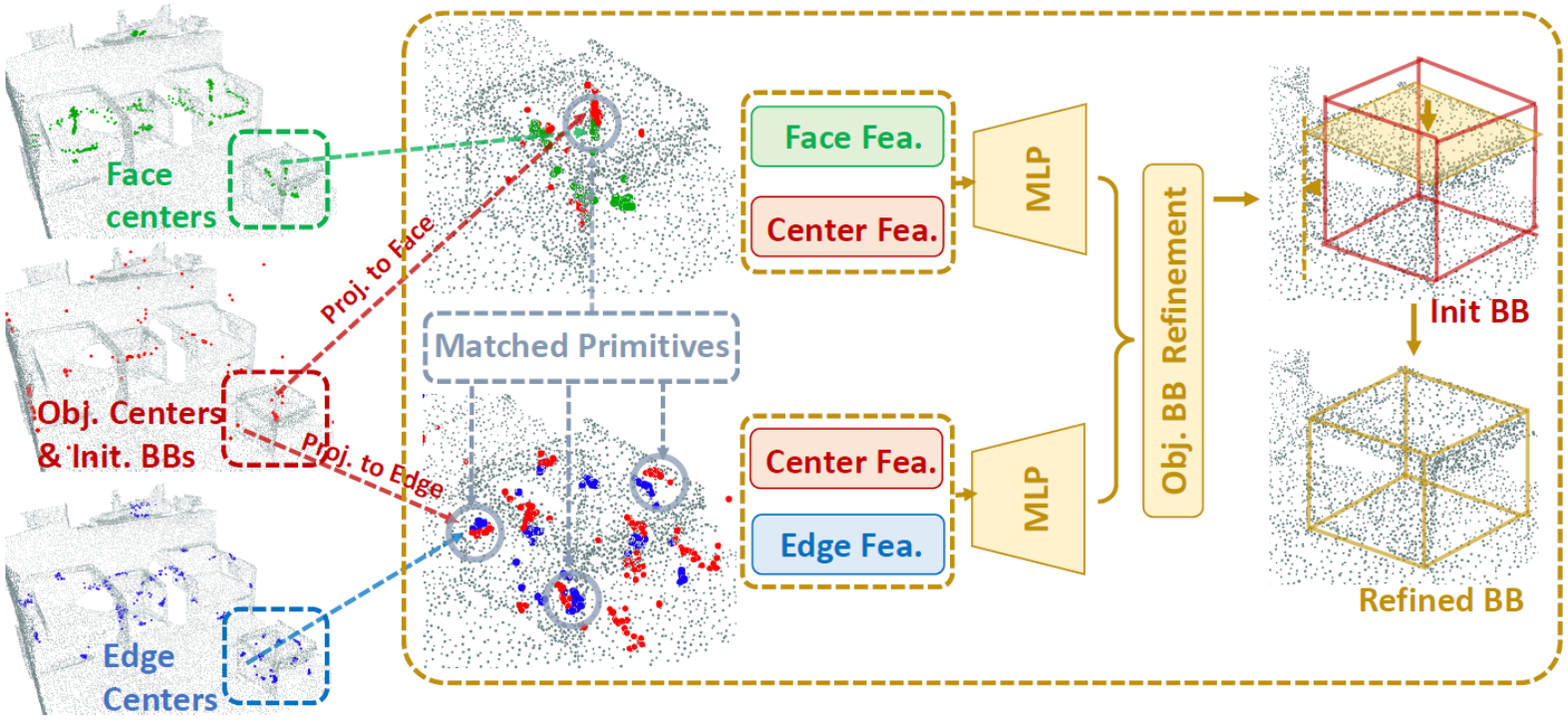


Objective function (Truncated L2):

$$F_S(o) := \sum_{t \in \mathcal{T}} \beta_t \sum_{\boldsymbol{c} \in \mathcal{P}_{t,S}} \min\left(\min_{1 \le i \le 19, t_i = t} \|\boldsymbol{c}_i - \boldsymbol{p}_i(o)\|^2 - \delta, 0 \right)$$

Back-propagate via
Implicit function theorem

Aggregate contextual information for proposal refinement



Experimental results

	Input	mAP@0.25	mAP@0.5
DSS	Geo + RGB	15.2	6.8
F-PointNet	Geo + RGB	19.8	10.8
GSPN	Geo + RGB	30.6	17.7
3D-SIS	Geo + 5 views	40.2	22.5
VoteNet	Geo only	58.7	33.5
Ours	Geo only	**67.2**	**48.1**
w\o refine	Geo only	60.2	37.3

3D object detection results on ScanNet V2 val dataset

Experimental results

	Input	mAP@0.25	mAP@0.5
DSS	Geo + RGB	42.1	-
COG	Geo + RGB	47.6	-
2D-driven	Geo + RGB	45.1	-
F-PointNet	Geo + RGB	54.0	-
VoteNet	Geo only	57.7	32.9
Ours	Geo only	**60.1**	**39.0**
w\o refine	Geo only	58.5	34.2

3D object detection results on SUN RGB-D V1 val dataset

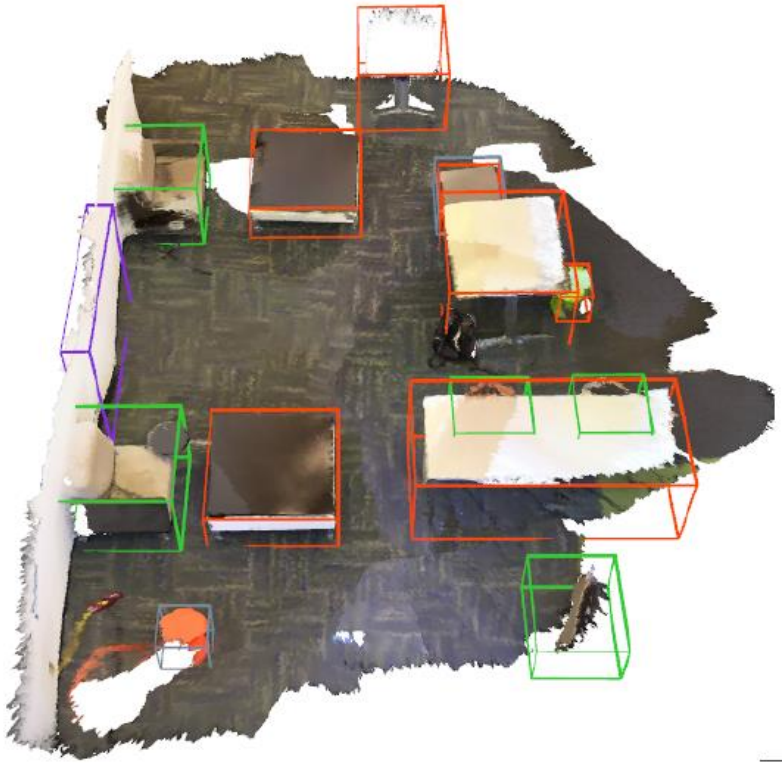
Experimental results

	RGB	cab	bed	chair	sofa	tabl	door	wind	bkshf	pic
3DSIS-5	✓	19.8	69.7	66.2	71.8	36.1	30.6	10.9	27.3	0.0
3DSIS	✗	12.8	63.1	66.0	46.3	26.9	8.0	2.8	2.3	0.0
Votenet	✗	36.3	87.9	88.7	89.6	58.8	47.3	38.1	44.6	7.8
Ours	✗	**49.4**	**88.6**	**91.8**	**90.2**	**64.9**	**61.0**	**51.9**	**54.9**	**18.6**
w\o refine	✗	37.2	89.3	88.4	88.5	64.4	53.0	44.2	42.2	11.1

	RGB	cntr	desk	curt	fridg	showr	toil	sink	bath	ofurn
3DSIS-5	✓	10.0	46.9	14.1	53.8	36.0	87.6	43.0	84.3	16.2
3DSIS	✗	6.9	33.3	2.5	10.4	12.2	74.5	22.9	58.7	7.1
Votenet	✗	56.1	71.7	47.2	45.4	57.1	94.9	54.7	92.1	37.2
Ours	✗	**62.0**	**75.9**	**57.3**	**57.2**	**75.3**	**97.9**	**67.4**	**92.5**	**53.6**
w\o refine	✗	51.2	59.8	47.0	54.3	74.3	93.1	57.0	85.6	43.5

3D object detection results on ScanNet V2 val dataset

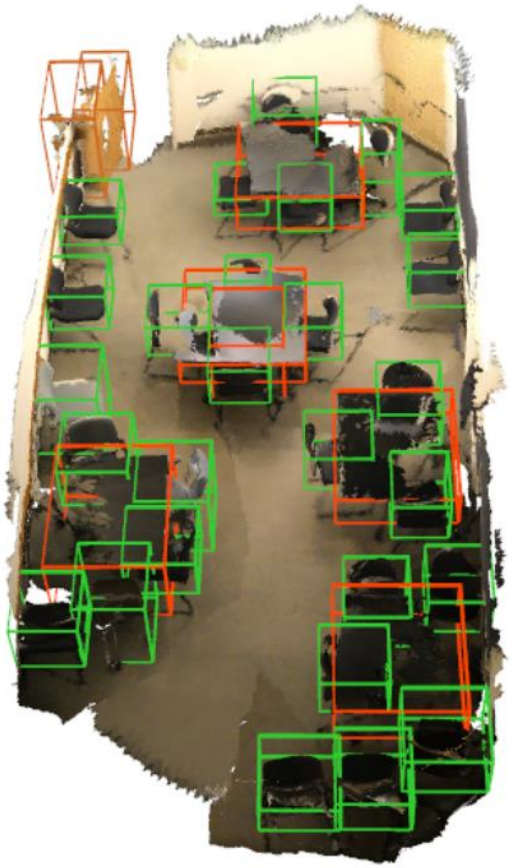
Qualitative results on ScanNet v2



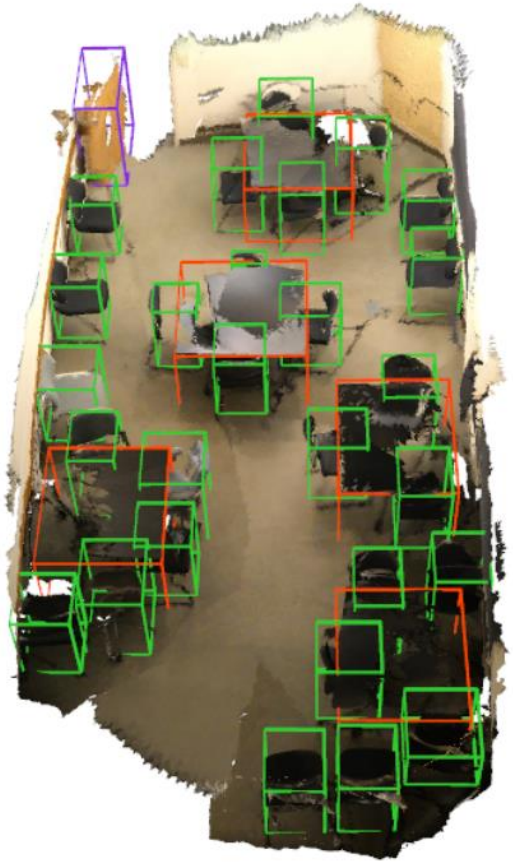
Ours

GT

Qualitative results on ScanNet v2

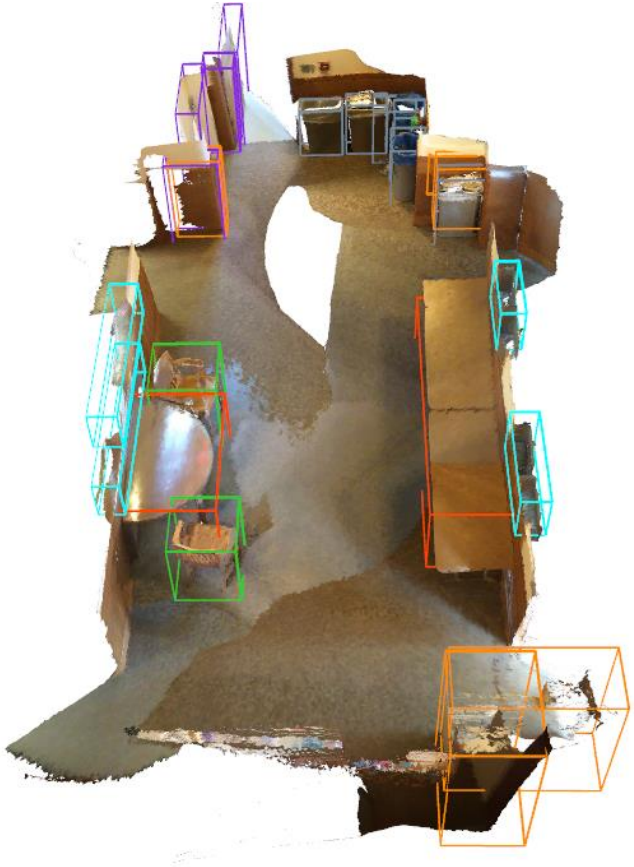


Ours

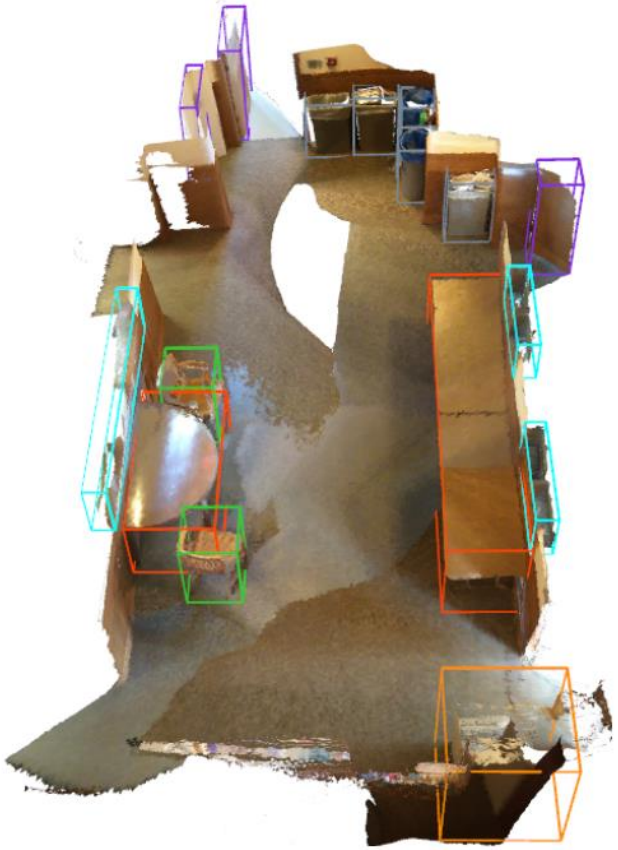


GT

Qualitative results on ScanNet v2

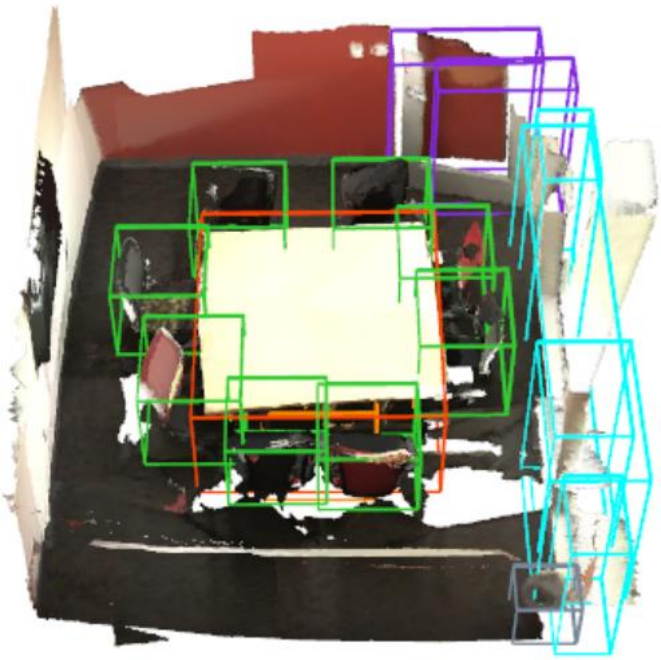


Ours

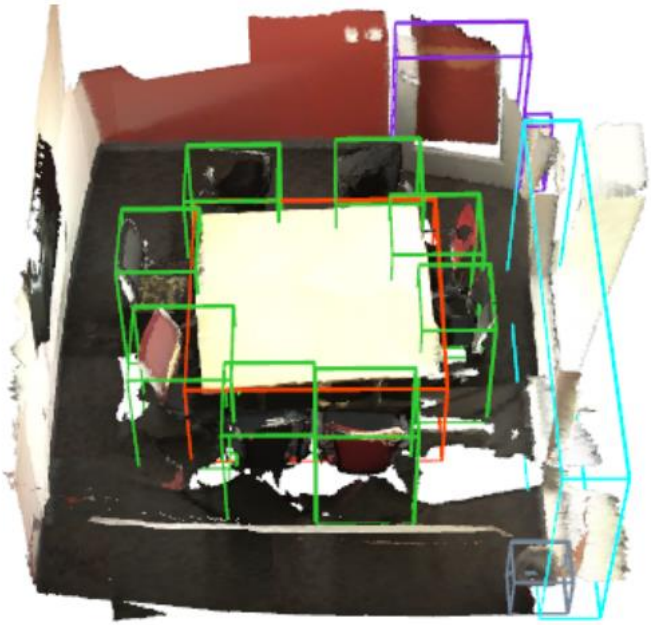


GT

Qualitative results on ScanNet v2

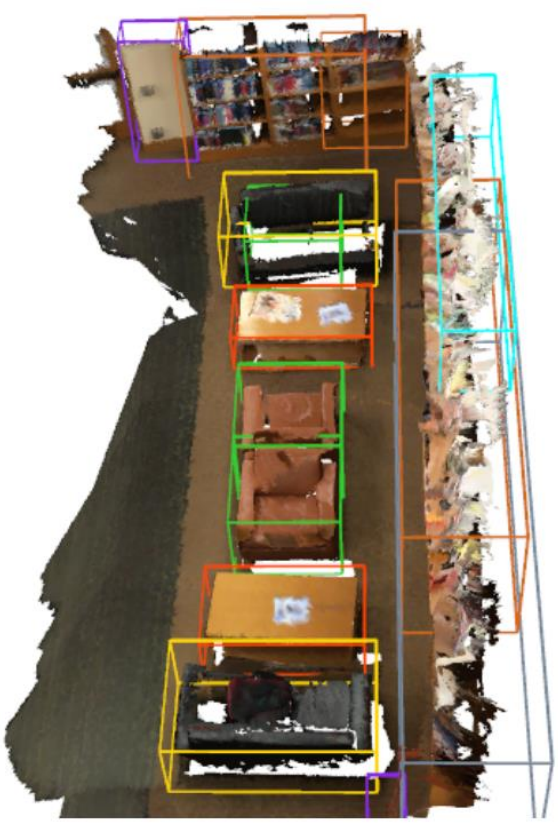


Ours

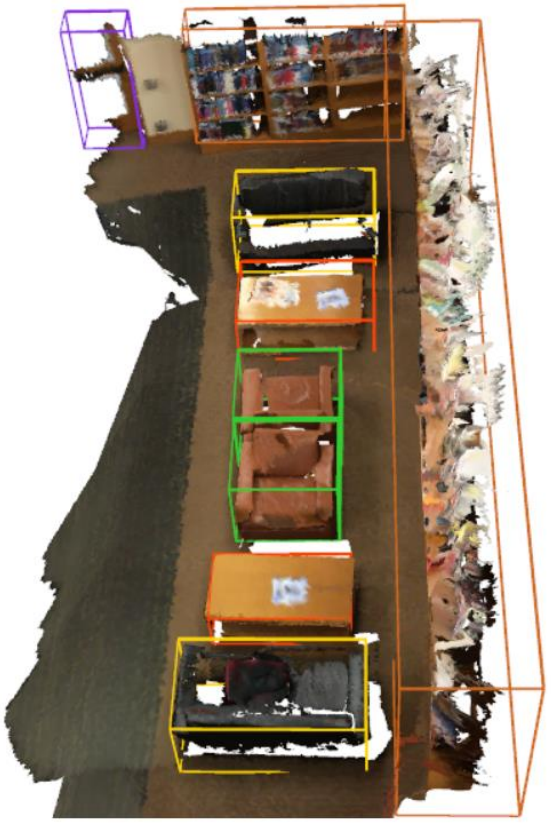


GT

Qualitative results on ScanNet v2

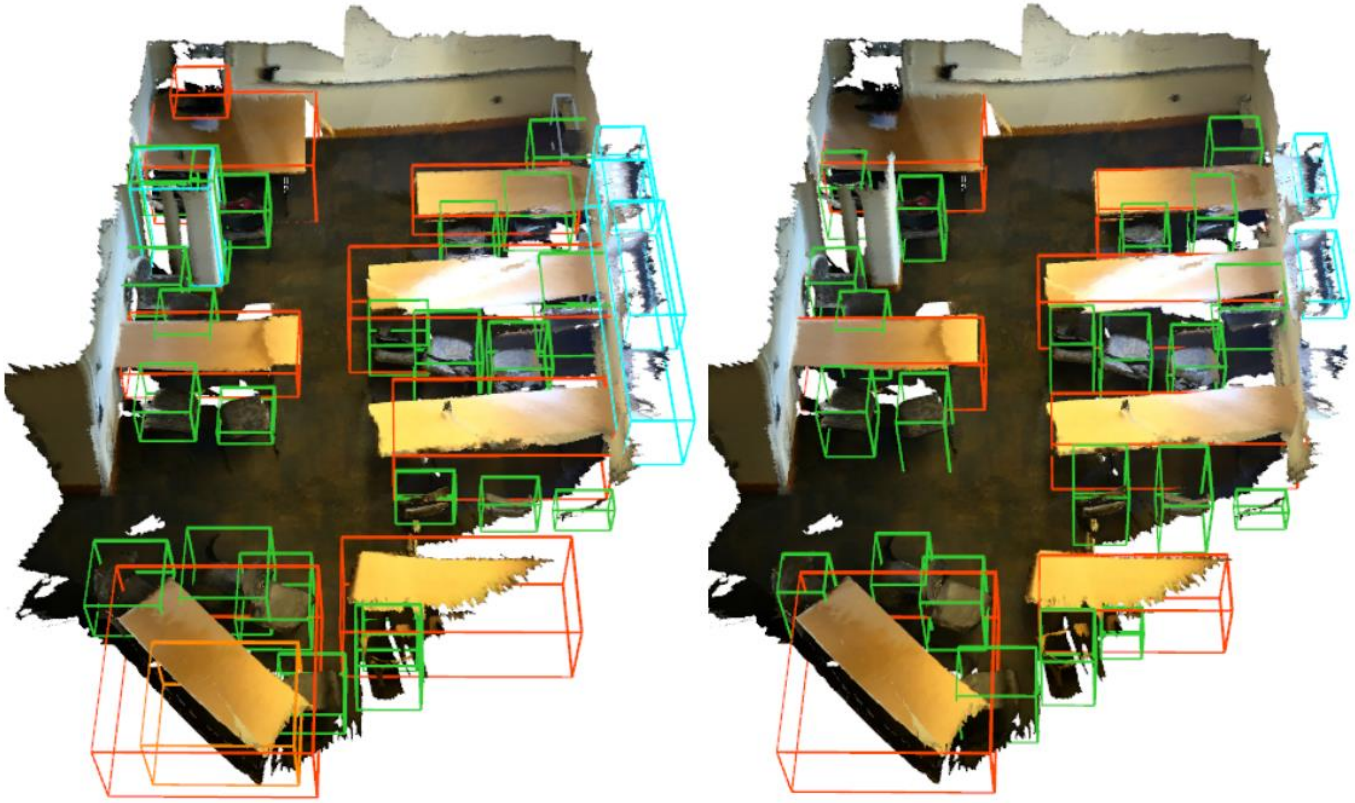


Ours



GT

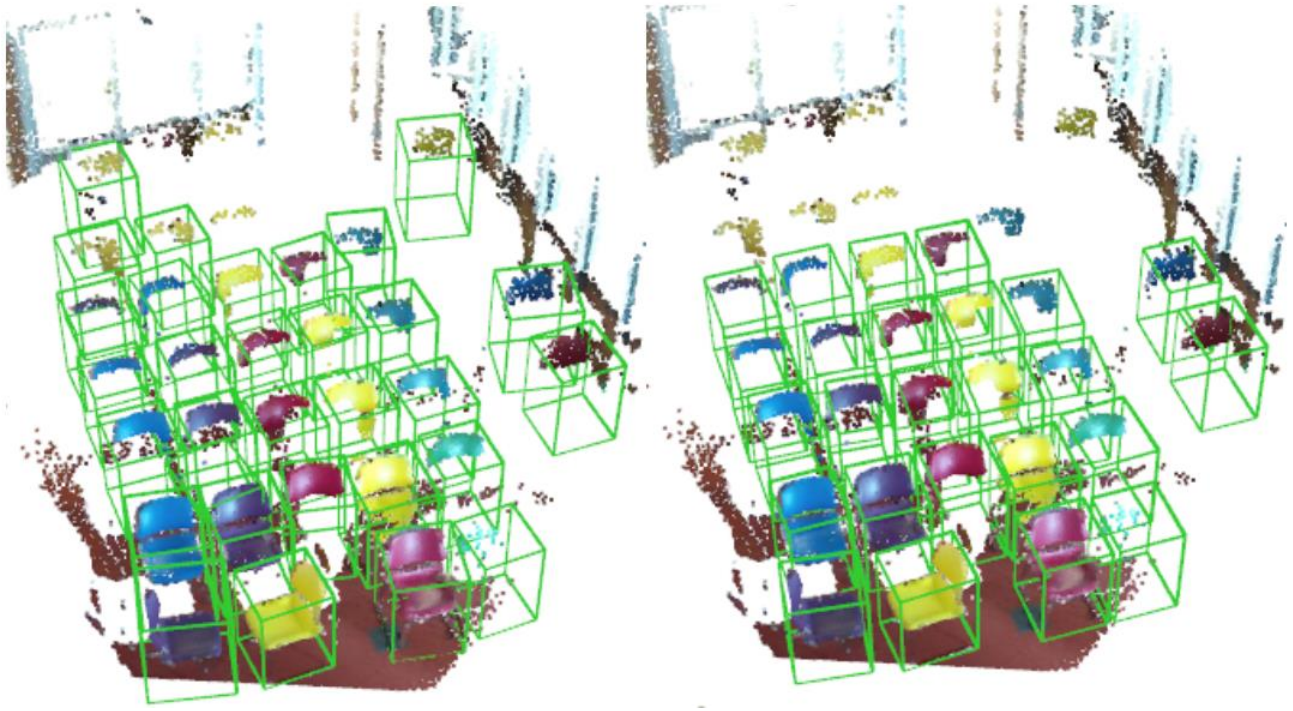
Qualitative results on ScanNet v2



Ours

GT

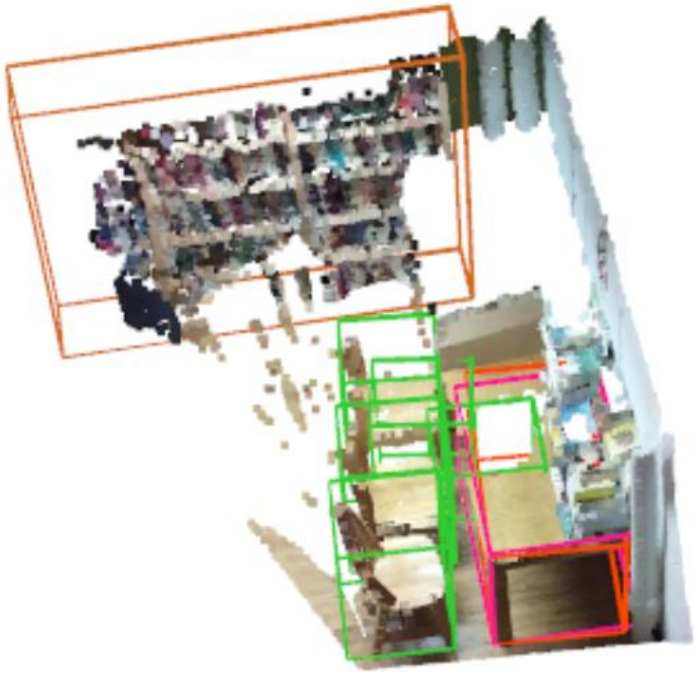
Qualitative results on SUNRBGD v1



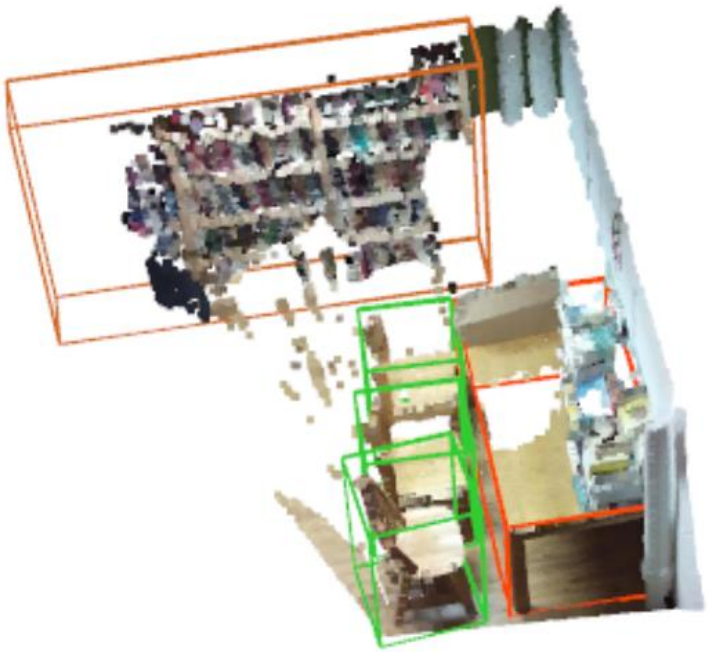
Ours

GT

Qualitative results on SUNRBGD v1

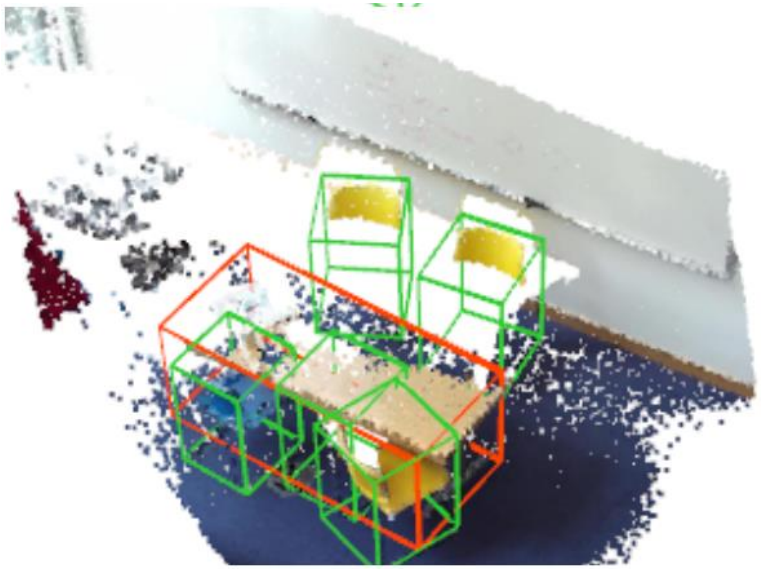


Ours

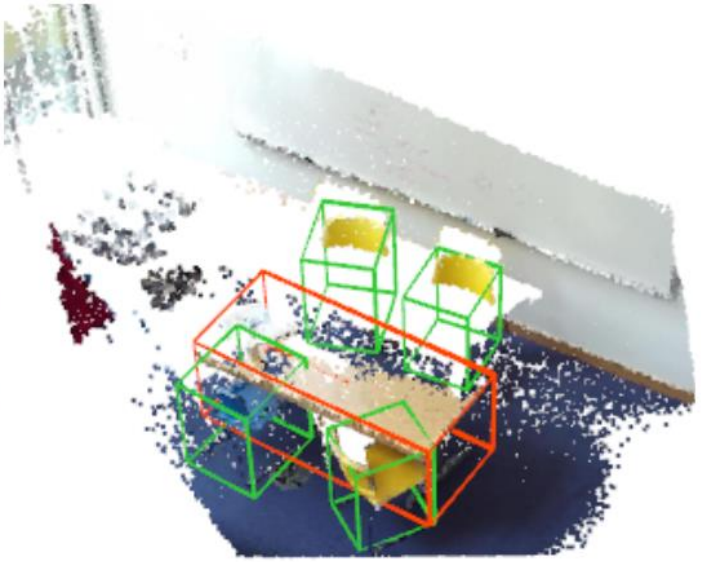


GT

Qualitative results on SUNRBGD v1



Ours

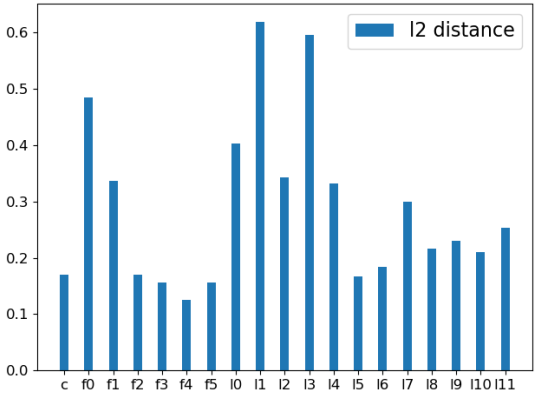


GT

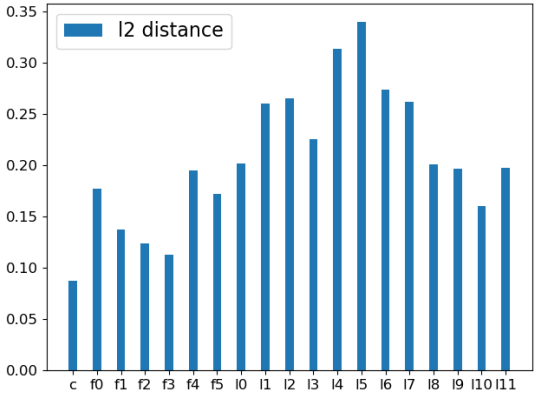
Analysis of Hybrid 3D Representations

Argument I: Different reps. learn different features and have different gen. behavior

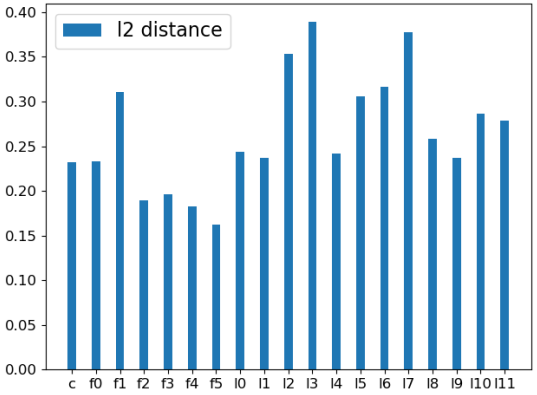
Door



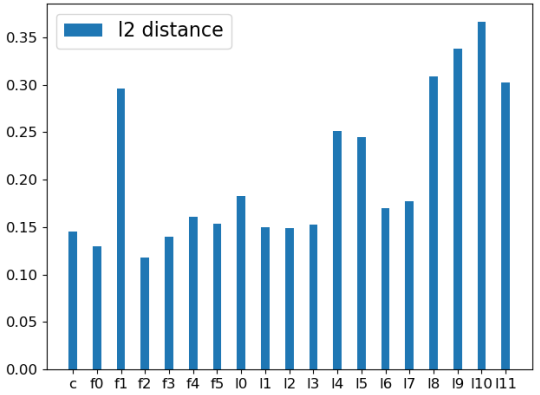
Chair



Cabinet



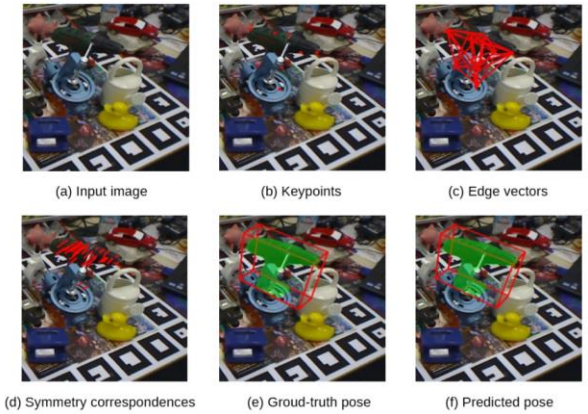
Table



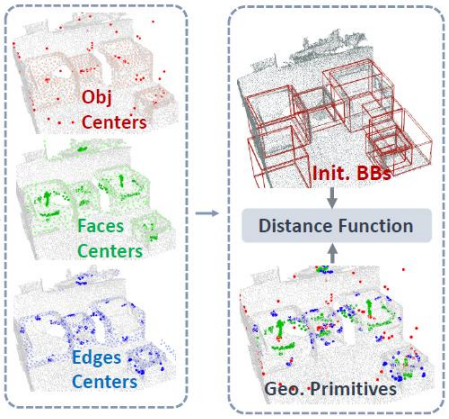
Prediction errors of geometric primitives of four categories from the ScanNet dataset

Adaptative feature selection via robust optimization and geometric constraints

- 6D Pose estimation
  - Geman Mcclure loss



- 3D object detection
  - Truncated L2 loss



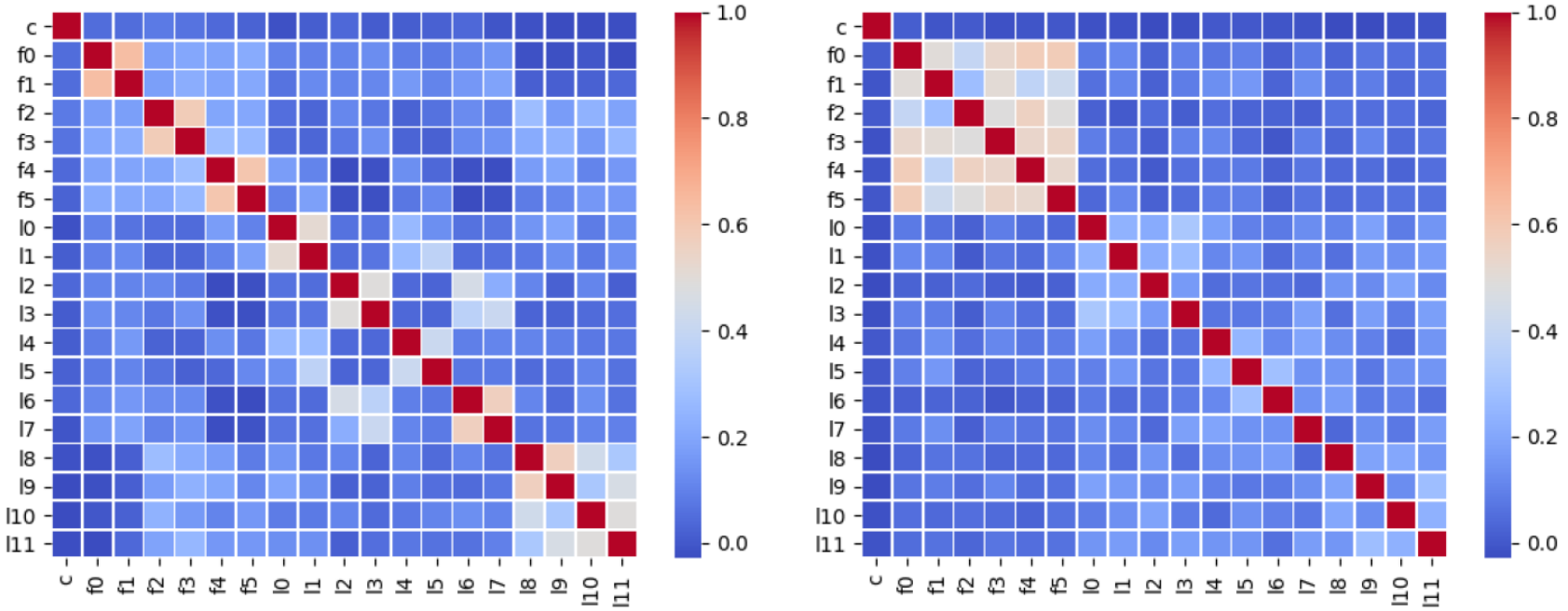
Argument II: Predictions under different representations are not strongly correlated

- In the over-parameterized regime, the optimal network parameters are close to the initial network parameters [Du et al. 19…]

- Therefore, if the network parameters under different representations are initialized independently, then the resulting network parameters are not strongly correlated, so do the predictions.

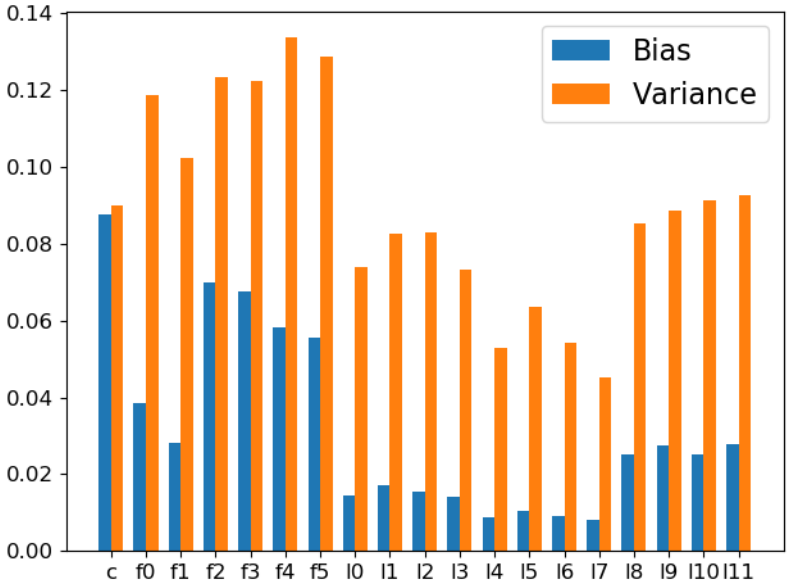
Argument II: Predictions under different representations are not strongly correlated

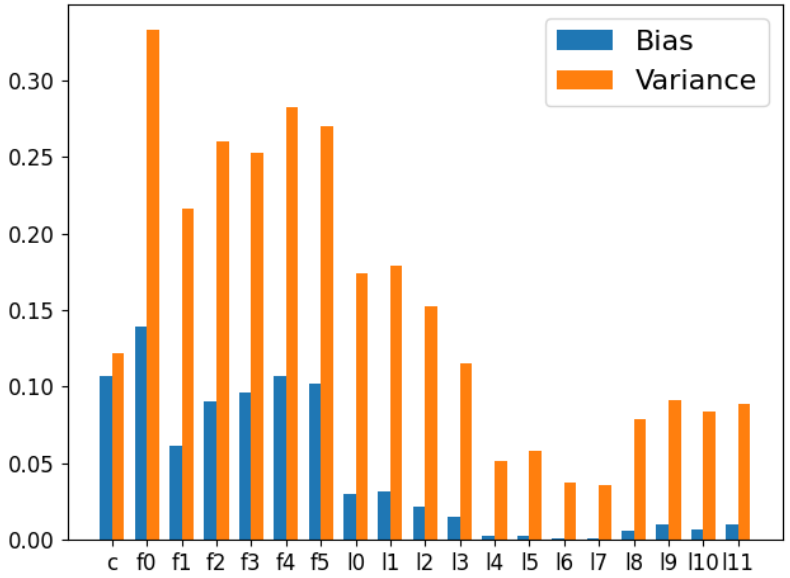
[Zhang et al. 20]



Left: covariance matrix of ScanNet. Right: covariance matrix of SUN RGB-D.
c represents object center, f0-f6 represent 6 BB face centers, and l0-l11 represent 12 BB edge centers.

Argument III: Bias is smaller than the variance (square-root)





Left: magnitudes of bias and variance (square-root) of geometric primitive predictions on ScanNet. Right: magnitudes of bias and variance (square-root) of geometric primitive predictions on SUNRBGD

Statistical analysis --- simple setting

Simple regression problem: $$x^\star := \arg\min_x \sum_{i=1}^n \|x - y_i\|^2$$

Independent random variables y_i : $$E[y_i] = \mu_i, \qquad V[y_i] = \sigma_i^2.$$

Assumption: $$|\mu_i - \mu| \leq \delta, \qquad \sigma_i >> \delta.$$

Optimal solution: $$x^\star = \frac{1}{n} \sum_{i=1}^n y_i.$$

Properties: $$E[x^\star] = \frac{1}{n} \sum_{i=1}^n \mu_i, \qquad V[x^\star] \approx \frac{1}{n^2} \sum_{i=1}^n \sigma_i,$$

$$|E[y_i| - \mu| \leq \delta, \qquad \sqrt{|V[x^\star]|} \approx \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\sum_{i=1}^n \sigma_i^2}{n}}.$$

Statistical analysis --- general setting

Generalized regression problem: $\boldsymbol{x}^\star := \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} f_i^2(\boldsymbol{x}, \boldsymbol{y}_i)$

Approximated solution (assume zero residuals for the ground-truth, i.e., $f_i(\boldsymbol{x}^{gt}, \boldsymbol{y}_i) = 0$):

$$\boldsymbol{x}^\star \approx \Big(\sum_{i=1}^{n} \frac{\partial f_i}{\partial \boldsymbol{x}} \cdot \frac{\partial f_i}{\partial \boldsymbol{x}}^T\Big)^{-1} \Big(\frac{\partial f_i}{\partial \boldsymbol{x}} \frac{\partial f_i}{\partial \boldsymbol{y}_i}^T \cdot \boldsymbol{y}_i\Big), \qquad \frac{\partial f_i}{\partial \boldsymbol{x}} := \frac{\partial f_i}{\partial \boldsymbol{x}}(\boldsymbol{x}^{gt}, \boldsymbol{0}),\ \frac{\partial f_i}{\partial \boldsymbol{y}_i} := \frac{\partial f_i}{\partial \boldsymbol{y}_i}(\boldsymbol{x}^{gt}, \boldsymbol{0}).$$

Variance:

$$V[\boldsymbol{x}^\star] \approx \Big(\sum_{i=1}^{n} \frac{\partial f_i}{\partial \boldsymbol{x}} \cdot \frac{\partial f_i}{\partial \boldsymbol{x}}^T\Big)^{-1} \Big(\sum_{i=1}^{n} \frac{\partial f_i}{\partial \boldsymbol{x}} \frac{\partial f_i}{\partial \boldsymbol{y}_i}^T \cdot V[\boldsymbol{y}_i] \cdot \frac{\partial f_i}{\partial \boldsymbol{y}_i} \frac{\partial f_i}{\partial \boldsymbol{x}}^T\Big) \cdot \Big(\sum_{i=1}^{n} \frac{\partial f_i}{\partial \boldsymbol{x}} \cdot \frac{\partial f_i}{\partial \boldsymbol{x}}^T\Big)^{-1}$$

Characteristics of variance reduction depend on the configuration among different representations

Variance reduction -- HybridPose

Geometric constraints:

Keypoints: $\overline{\boldsymbol{r}}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k) := \hat{\boldsymbol{p}}_k \times (R\overline{\boldsymbol{p}}_k + \boldsymbol{t})$

Edge vectors: $\overline{\boldsymbol{r}}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e, \boldsymbol{p}_{e_s}) := \hat{\boldsymbol{v}}_e \times (R\overline{\boldsymbol{p}}_{e_t} + \boldsymbol{t}) + \hat{\boldsymbol{p}}_{e_s} \times (R\overline{\boldsymbol{v}}_e)$

Sym. corres.: $r^{\mathcal{S}}_{R,\boldsymbol{t}}(\boldsymbol{q}_{s,1}, \boldsymbol{q}_{s,2}) := (\hat{\boldsymbol{q}}_{s,1} \times \hat{\boldsymbol{q}}_{s,2})^T \boldsymbol{R}\overline{\boldsymbol{n}}_r$

Regression objective:

$$\min_{R,\boldsymbol{t}} \sum_{k=1}^{|\mathcal{K}|} \rho(\|\boldsymbol{r}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k)\|, \beta_{\mathcal{K}}) \|\boldsymbol{r}^{\mathcal{K}}_{R,\boldsymbol{t}}(\boldsymbol{p}_k)\|^2_{\Sigma_k}$$

$$+ \frac{|\mathcal{K}|}{|\mathcal{E}|} \sum_{e=1}^{|\mathcal{E}|} \rho(\|\boldsymbol{r}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e)\|, \beta_{\mathcal{E}}) \|\boldsymbol{r}^{\mathcal{E}}_{R,\boldsymbol{t}}(\boldsymbol{v}_e)\|^2_{\Sigma_e}$$

$$+ \frac{|\mathcal{K}|}{|\mathcal{S}|} \sum_{s=1}^{|\mathcal{S}|} \rho(r^{\mathcal{S}}_{R,\boldsymbol{t}}(\boldsymbol{q}_{s,1}, \boldsymbol{q}_{s,2}), \beta_{\mathcal{S}})$$

EPnP [Lepetit et al. 09] for initialization

+

Gauss-Newton for refinement

Robust norm: $\rho(x, \beta) := \beta_1^2/(\beta_2^2 + x^2)$

Variance reduction -- HybridPose

[Song et al. 20]

Tradeoff weights

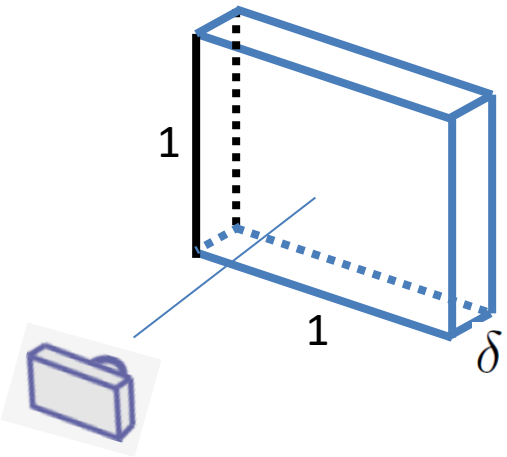
Covariances of each type

$$\mathrm{Var}([\boldsymbol{c}, \overline{\boldsymbol{c}}]) \approx (H_{\mathcal{K}} + \lambda H_{\mathcal{E}} + \mu H_{\mathcal{S}})^{-1} (\sigma_{\mathcal{K}}^2 H_{\mathcal{K}} + \lambda^2 \sigma_{\mathcal{E}}^2 H_{\mathcal{E}} + \mu^2 \sigma_{\mathcal{S}}^2 H_{\mathcal{S}}) (H_{\mathcal{K}} + \lambda H_{\mathcal{E}} + \mu H_{\mathcal{S}})^{-1}$$

$$H_{\mathcal{K}} := \begin{pmatrix} c_1(\delta) & 0 & 0 & 0 & -c_2(\delta) & 0 \\ 0 & c_1(\delta) & 0 & c_2(\delta) & 0 & 0 \\ 0 & 0 & \frac{12\delta^2}{8} & 0 & 0 & 0 \\ 0 & c_2(\delta) & 0 & 1 & 0 & 0 \\ -c_2(\delta) & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{12\delta^2}{8} \end{pmatrix}$$

$$H_{\mathcal{E}} := \begin{pmatrix} \frac{11}{7}\delta^4 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{11}{7}\delta^4 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{24}{7}\delta^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{24}{7}\delta^2 \end{pmatrix}$$

$$H_{\mathcal{S}} := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{4}{3}\delta^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{4}{9}\delta^4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



Edge vectors and sym. corres. help reduce the variance along the viewing direction (for both translation and rotation)

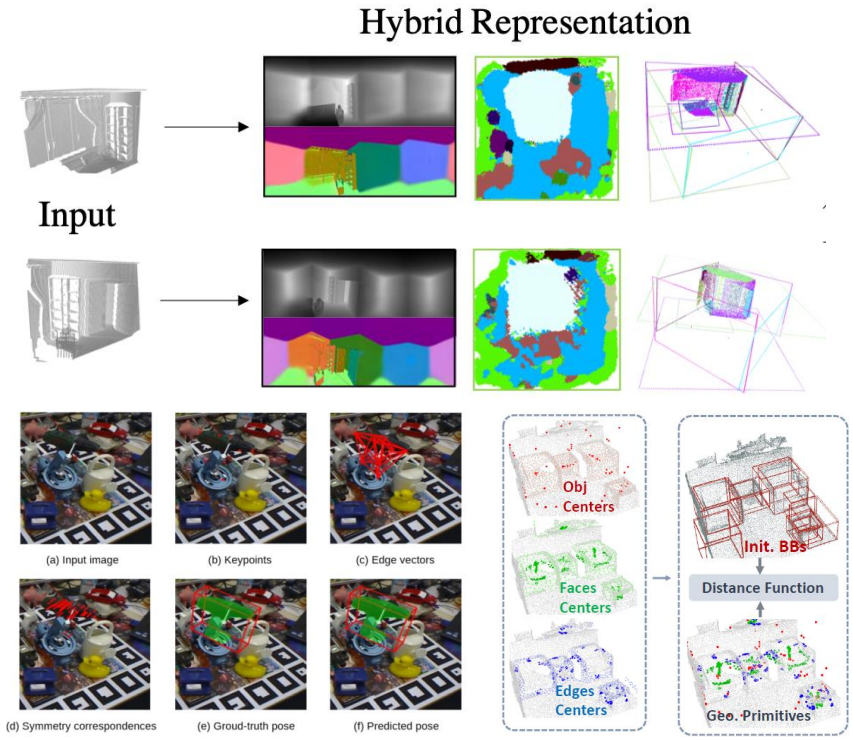
Further Discussion

- Intermediate predictions are mostly uncorrelated
- Bias is smaller than the variance

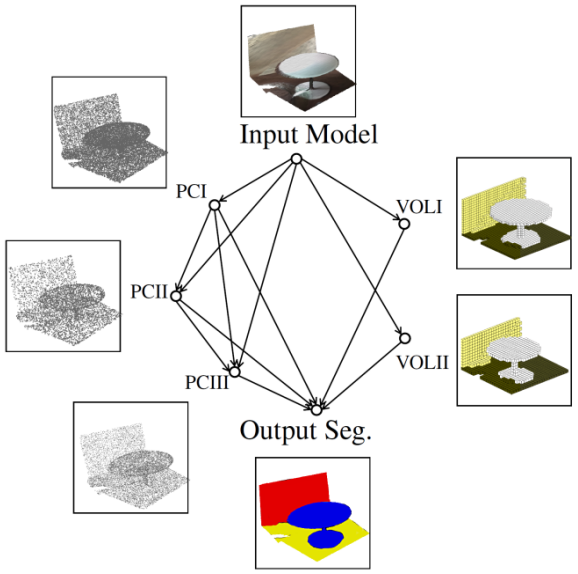
- Single rep. + multiple networks
  - Variance reduction but bias reduction is less effective

- Multiple reps. + single network per rep.
  - Potentially salient variance reduction and bias reduction

Theme: hybrid representations

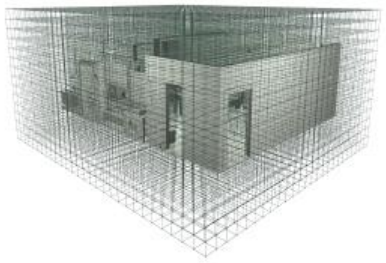


Supervised

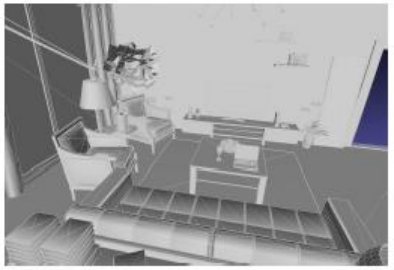


Semi-supervised

A network of 3D representations

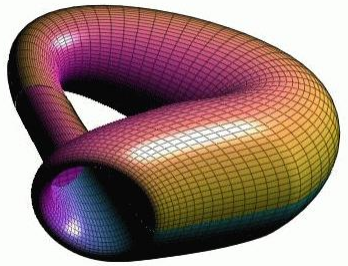


Volumetric

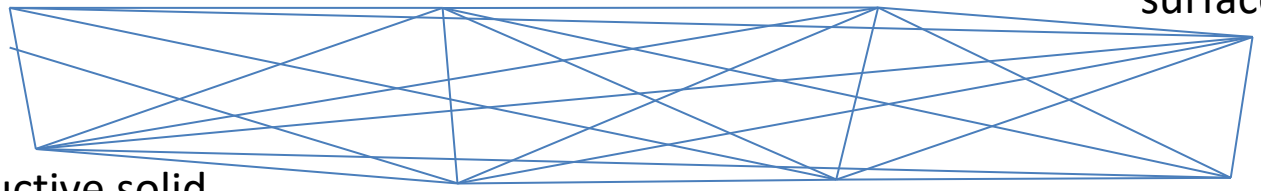


Triangular mesh

Point cloud



Parametric surfaces

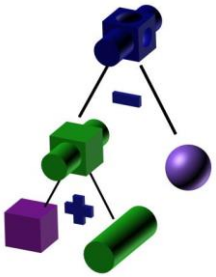


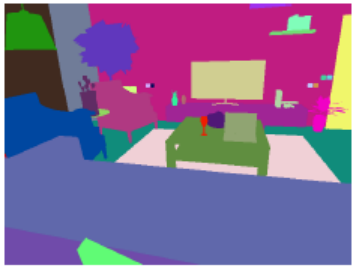
Constructive solid geometry

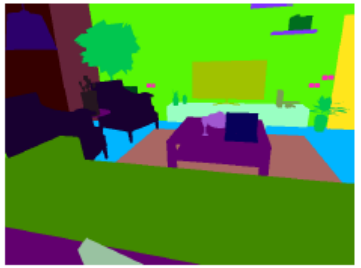
Multi-view

Scene-graph

Semantic segments



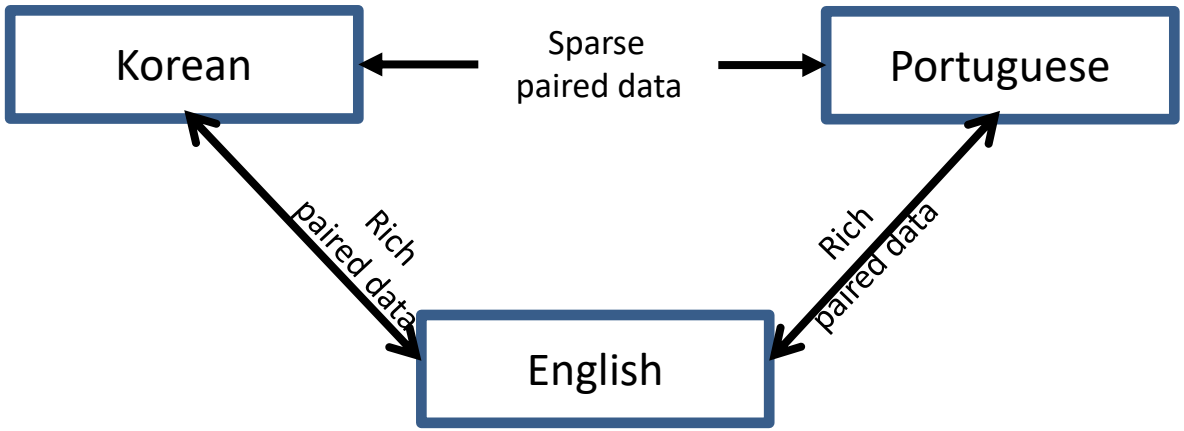




Advantage I: Leverage more training data

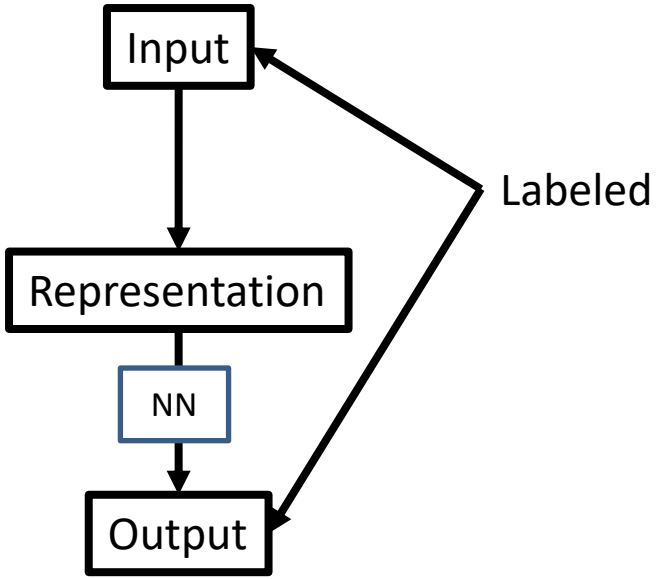
A toy example

[Johnson et al. 16]

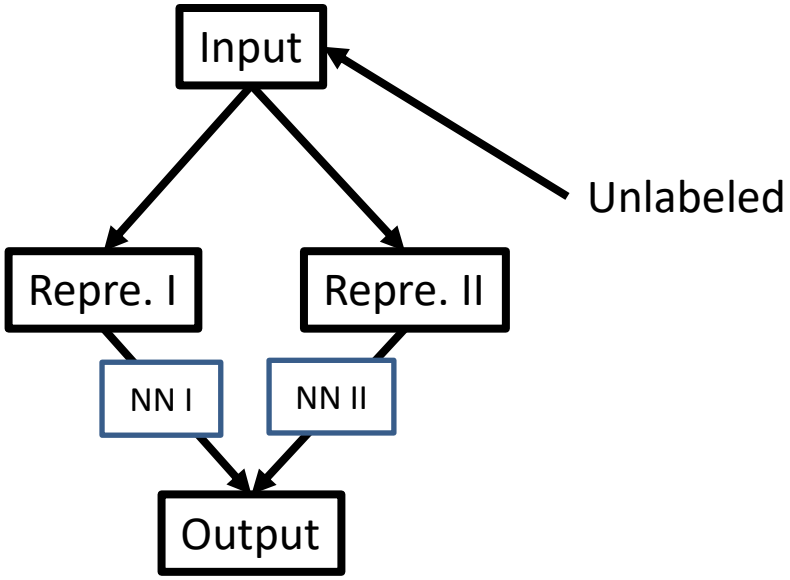


Advantage II: Leverage Unlabeled Data

A toy example



Standard setting:



Joint setting:

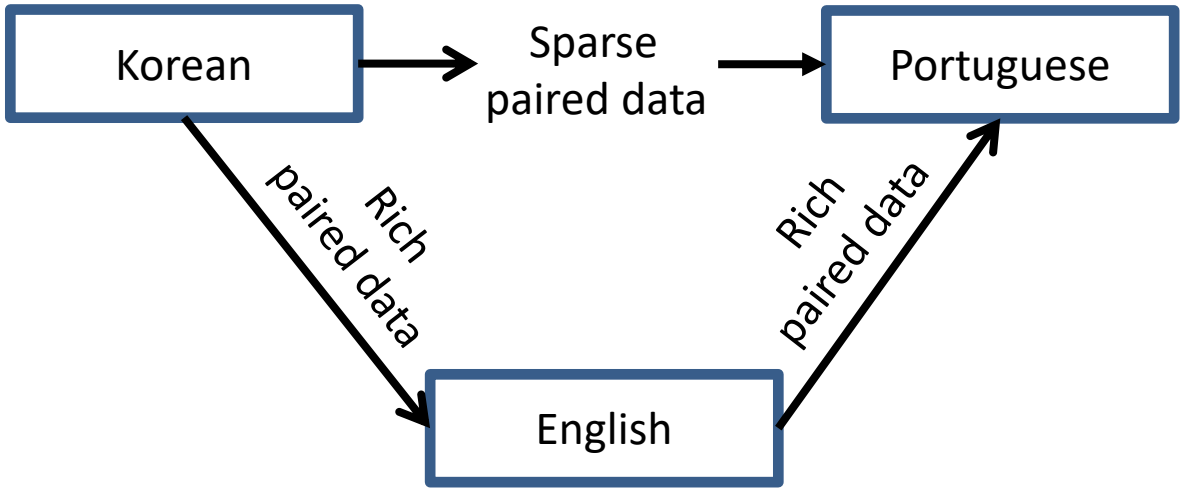
Note that the supervised setting and the unsupervised setting use different characteristics of the hybrid representations

Path-invariant map networks

[Zhang, Liang, Wu, Zhou, H, CVPR' 2019]

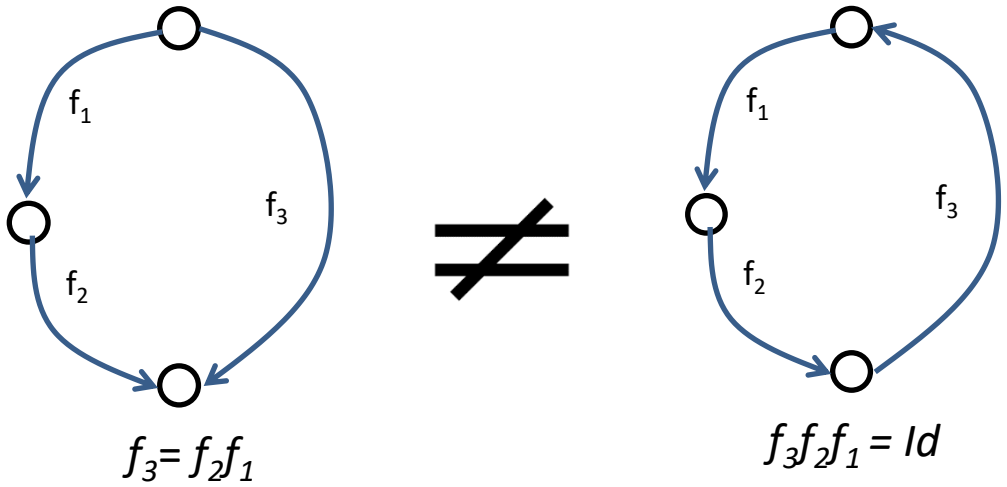
Multi-lingual translation

[Johnson et al. 16]



Abstraction

[Zhang et al. CVPR 19]



$f_3 = f_2 f_1$

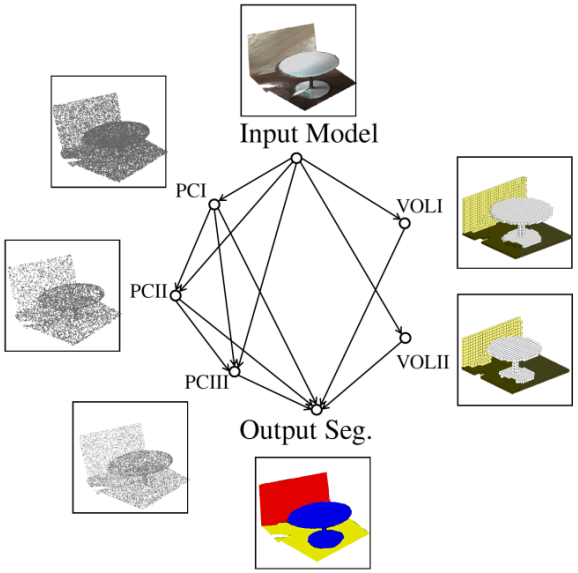
Path-invariance

$f_3 f_2 f_1 = Id$

Cycle-consistency

Path-invariance

[Zhang et al. CVPR 19]



Definition 3. *Let $\mathcal{G}_{\text{path}}(u, v)$ collect all paths in $\mathcal{G}$ that connect u to v. We define the set of all possible path pairs of $\mathcal{G}$ as*

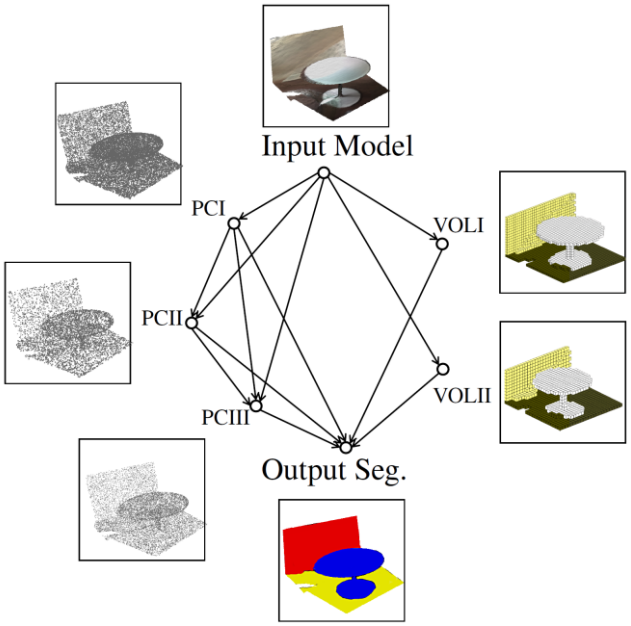
$$\mathcal{G}_{\text{pair}} = \bigcup_{u,v \in \mathcal{V}} \{(p, q) | p, q \in \mathcal{G}_{\text{path}}(u, v)\}.$$

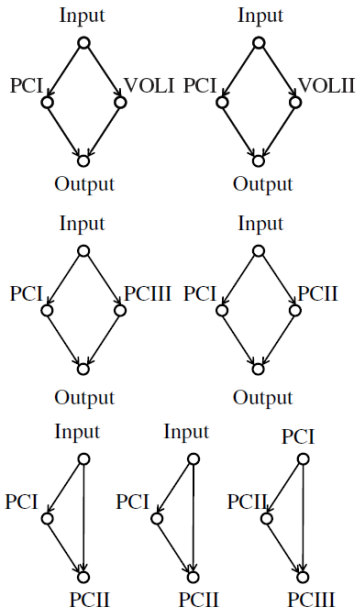
We say $\mathcal{F}$ is path-invariant if

$$f_p = f_q, \qquad \forall (p, q) \in \mathcal{G}_{\text{pair}}.$$

Path-invariance basis

[Zhang et al. CVPR 19]

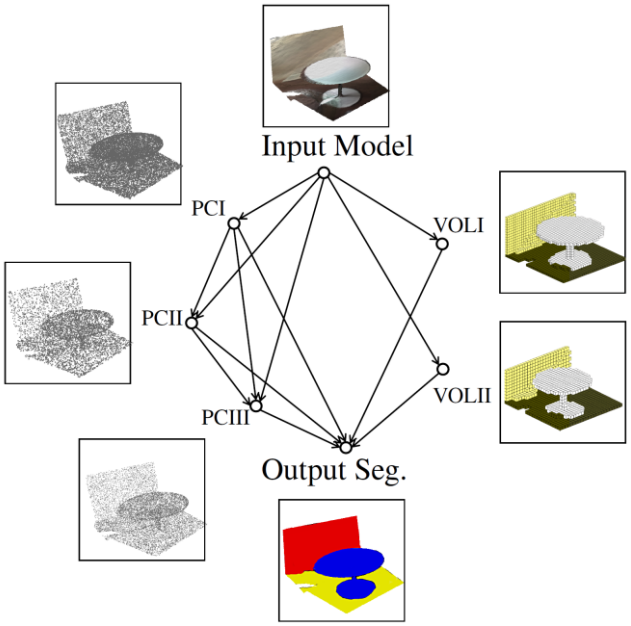


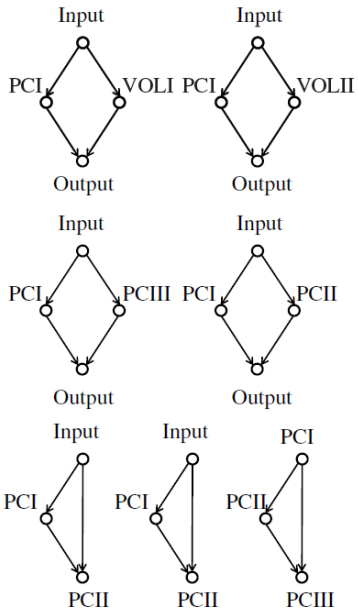


Can induce the path-invariance property of the entire graph

Path-invariance provides a regularization for training neural networks

[Zhang et al. CVPR 19]



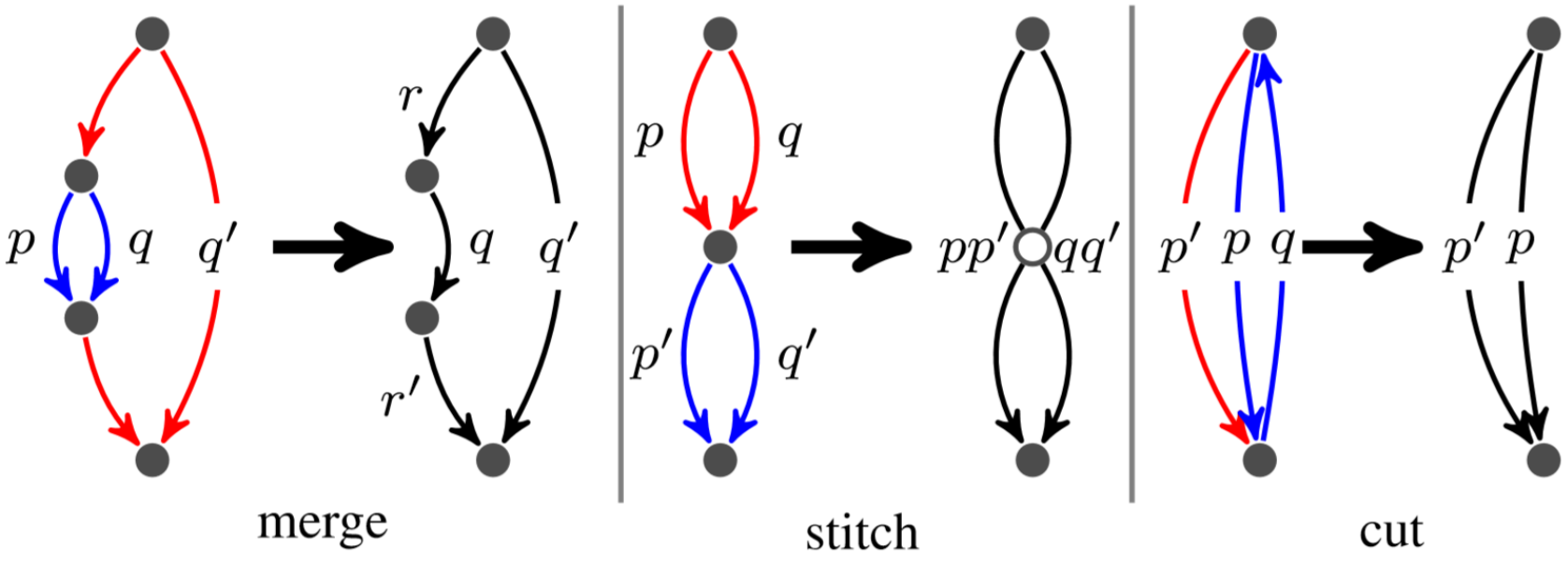


$$\min_{\Theta} \sum_{(i,j)\in\overline{\mathcal{E}}} l_{ij}(f^{\Theta}_{\widehat{v_i v_j}}) + \lambda \sum_{(p,q)\in\mathcal{B}} \mathop{E}_{v\sim P_{p_t}} d_{\mathcal{D}_{p_t}}(f^{\Theta}_{p}(v), f^{\Theta}_{q}(v))$$

Supervised loss

Unsupervised loss

Induction operations



Primitive operations that preserve the path-invariance property

Main result

[Zhang et al. CVPR 19]

- Theorem: *Given a directed graph with n vertices and m edges, there exists a path-invariance basis with size at most O(nm)*

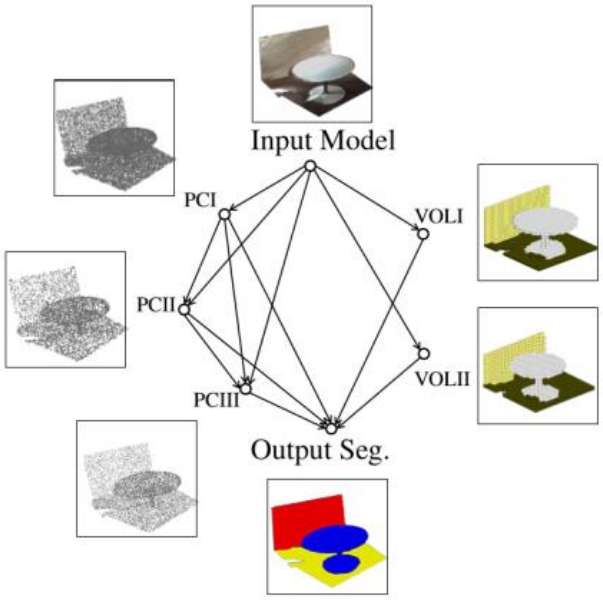
- Main idea for the proof
  - A directed graph is a directed acyclic graph (DAG) of strongly connected components
  - Use a vertex order to construct a path-invariance basis for DAG

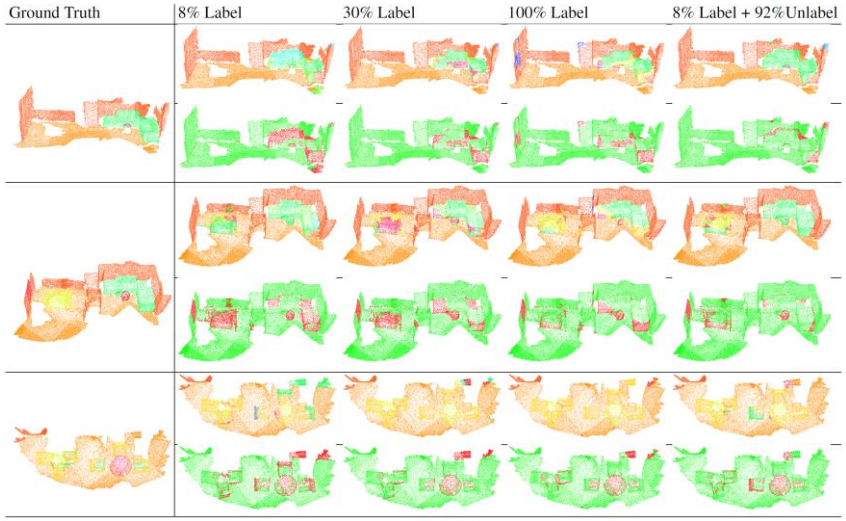
Three advantages over randomly sampling path-pairs

[Zhang et al. CVPR 19]

- One may need to sample many (exponentially number of) path pairs to ensure the path-invariance property
  - Many long path pairs
- There is a cost of implementing one path pair
- Convergence of stochastic algorithms

Semantic segmentation on ScanNet

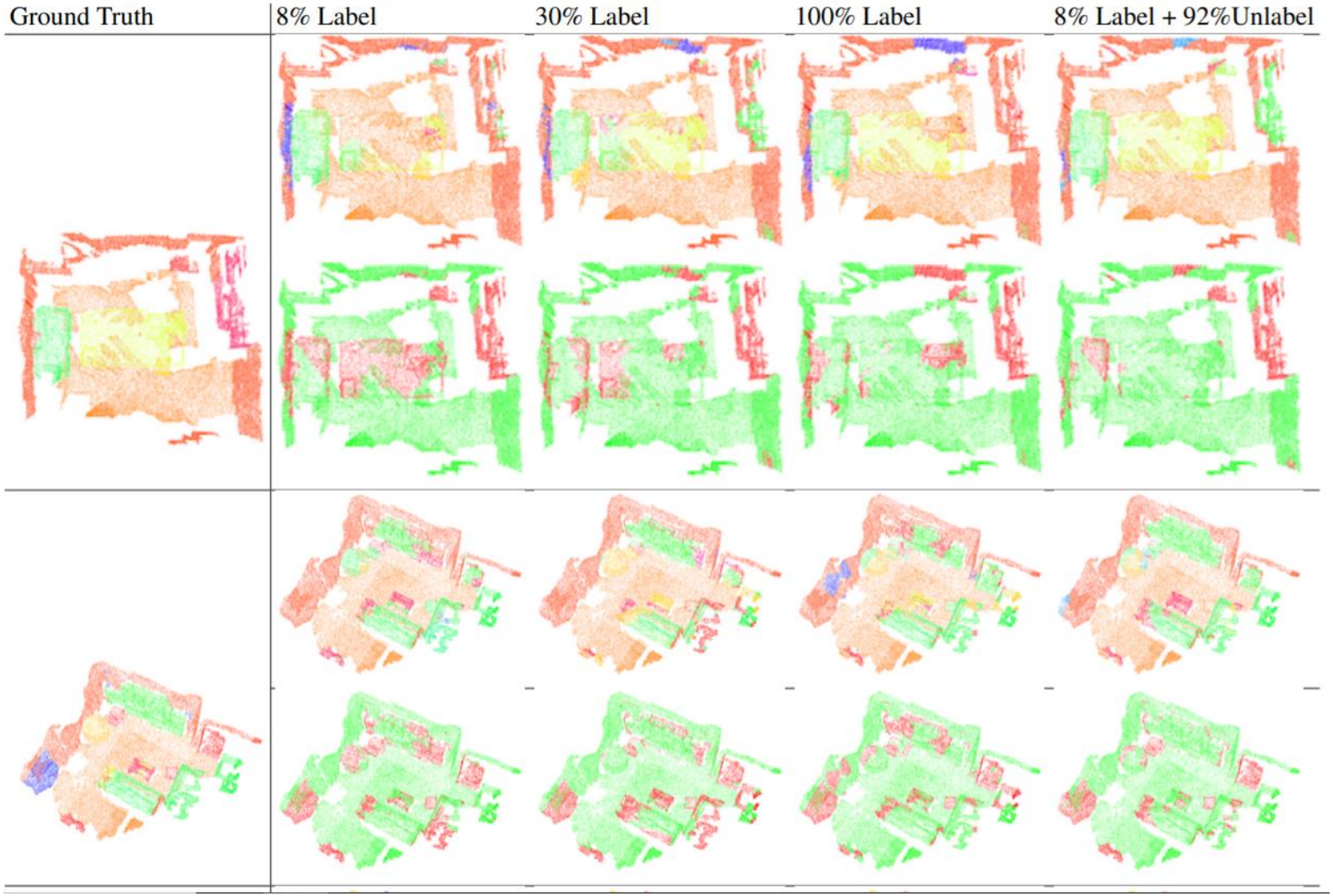




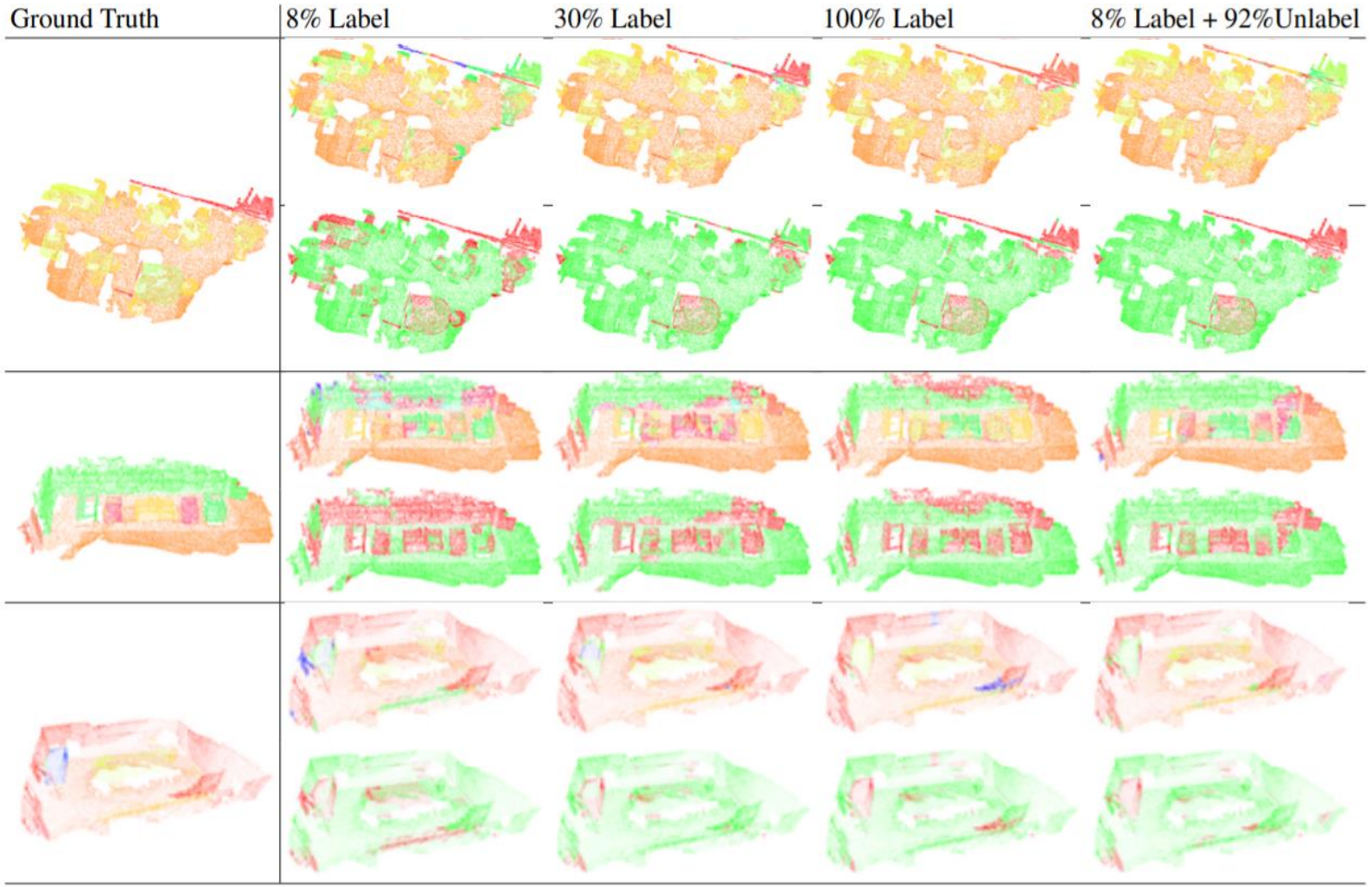
	PCI	PCII	PCIII	VOLI	VOLII
100% Label (Isolated)	84.2	83.3	83.4	81.9	81.5
8% Label (Isolated)	79.2	78.3	78.4	78.7	77.4
8% Label + 92%Unlabel (Joint)	81.7	81.7	81.4	81.1	78.7
30% Label (Isolated)	80.8	81.9	81.2	80.3	79.5

8% labeled + 92% unlabeled $\approx$ 30% labeled

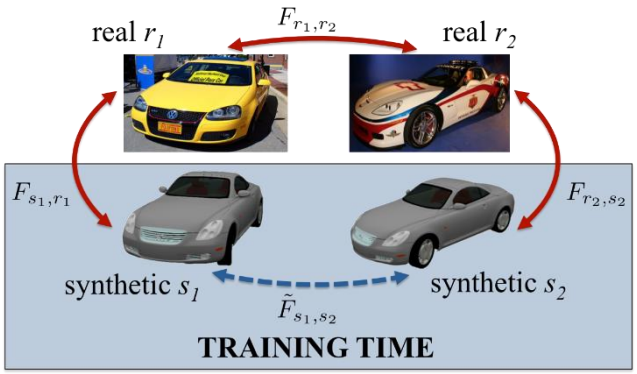
Qualitative results



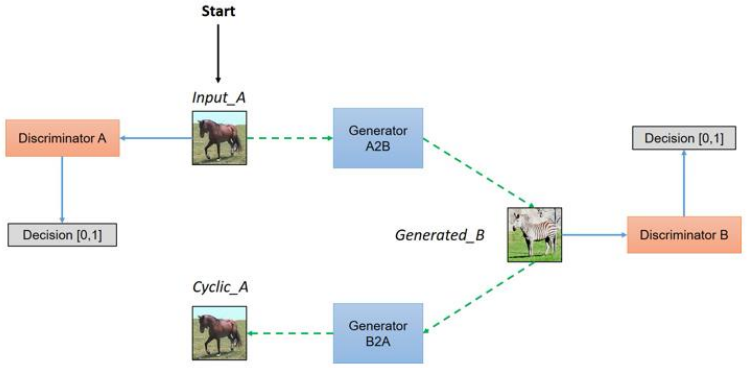
Qualitative results



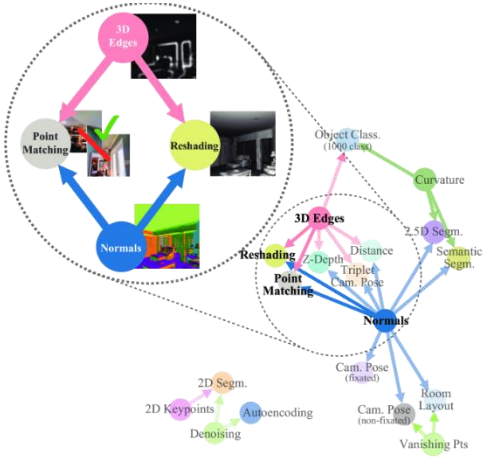
Other relevant works



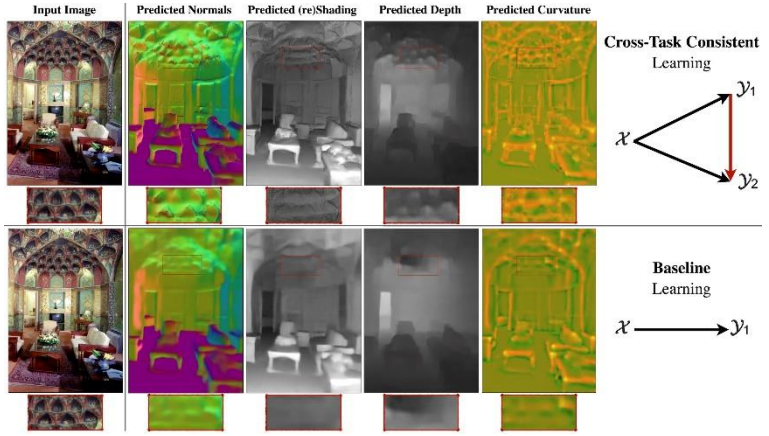
[Zhou et al. 16]



[Zhu et al. 17]

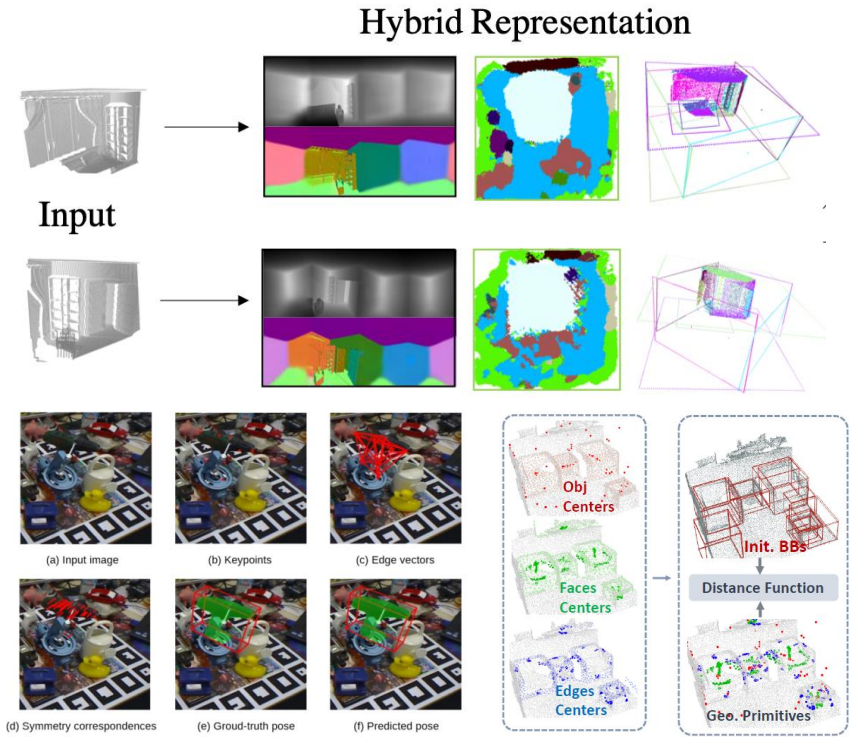


[Zamir et al. 18]

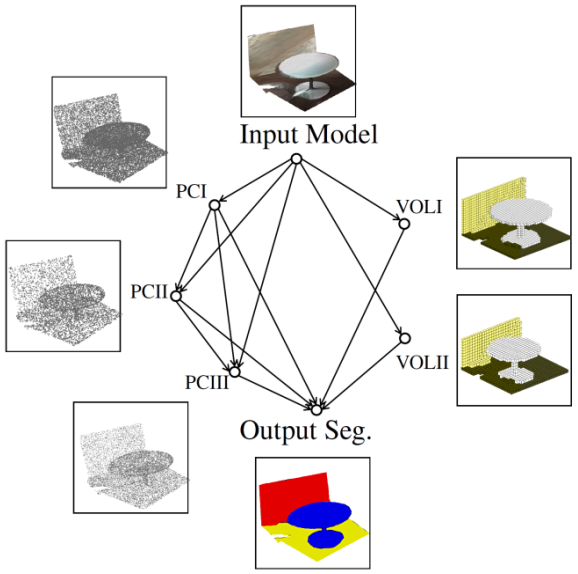


[Zamir et al. 20]

Theme: hybrid representations



Supervised



Semi-supervised