Deep Hough Voting for 3D Object Detection in Point Clouds

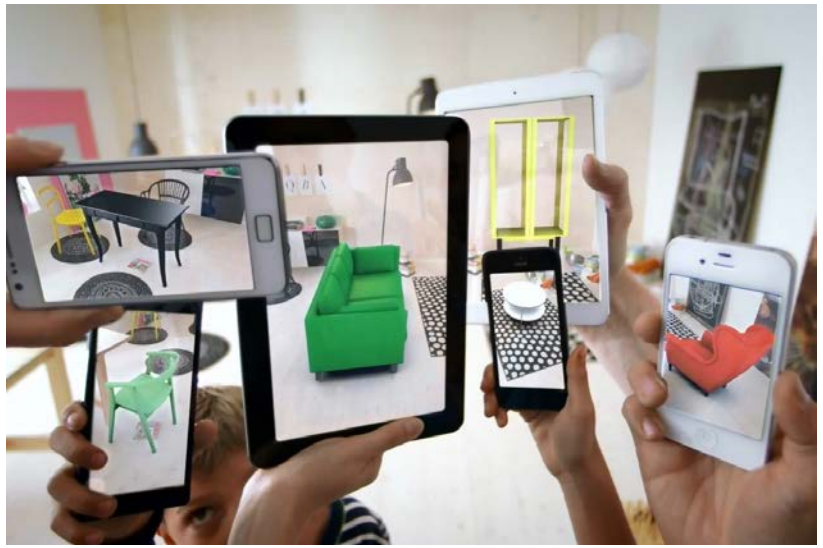
Charles Qi (祁芮中台)

GAMES Webinar December 5th, 2019

Joint work with Or Litany, Kaiming He, Leonidas Guibas. ICCV 2019.

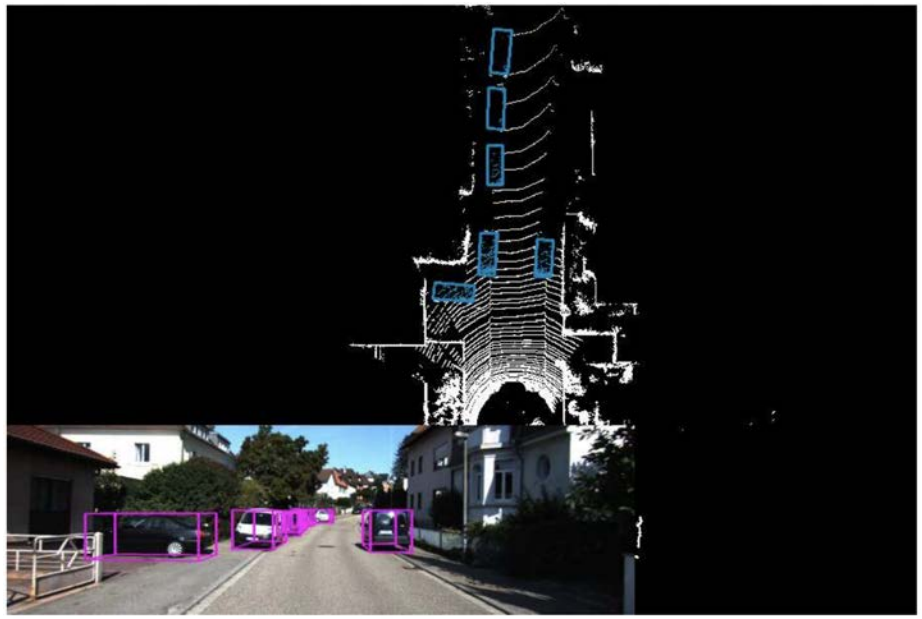
3D object detection

Estimate **oriented 3D bounding boxes** and **semantic classes** from sensor data.



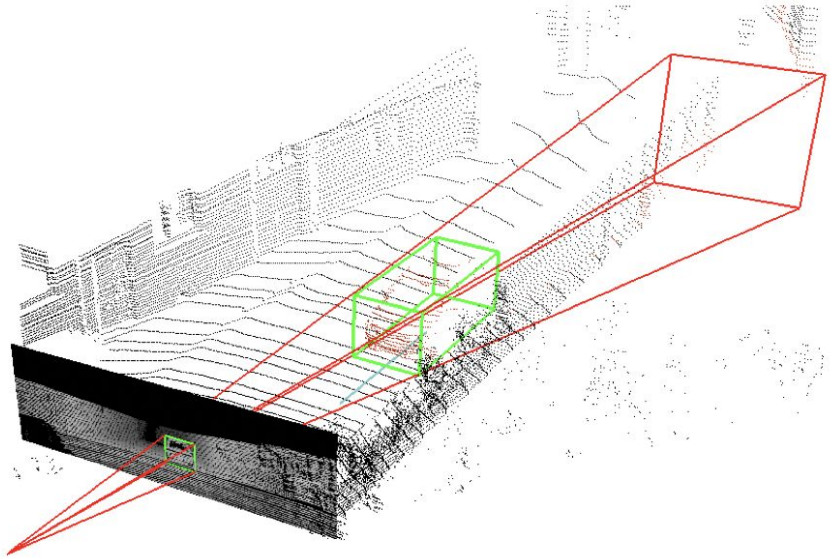
Prior work relies on 2D object detection

Bird's eye view detector



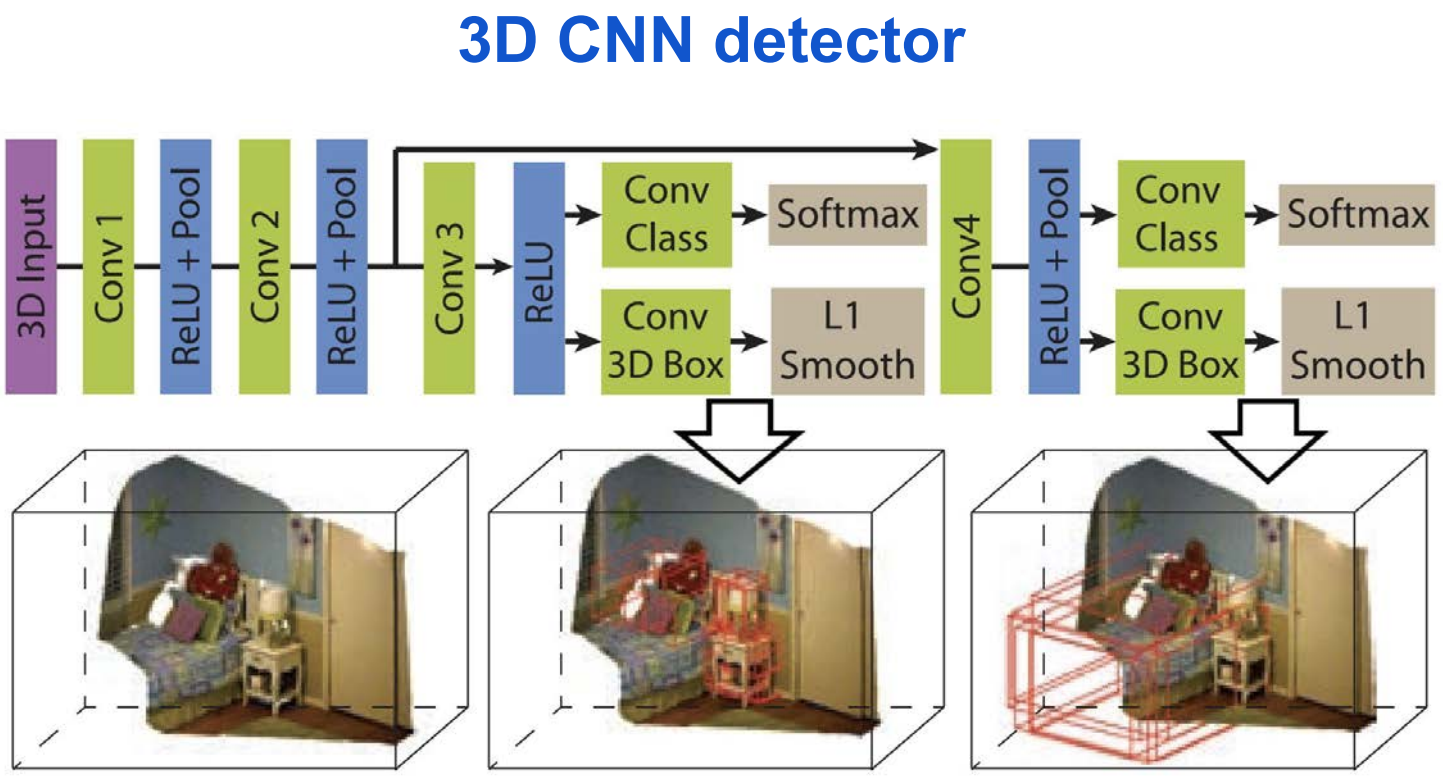
[MV3D by Chen et al. CVPR 2017]

Frustum-based detector



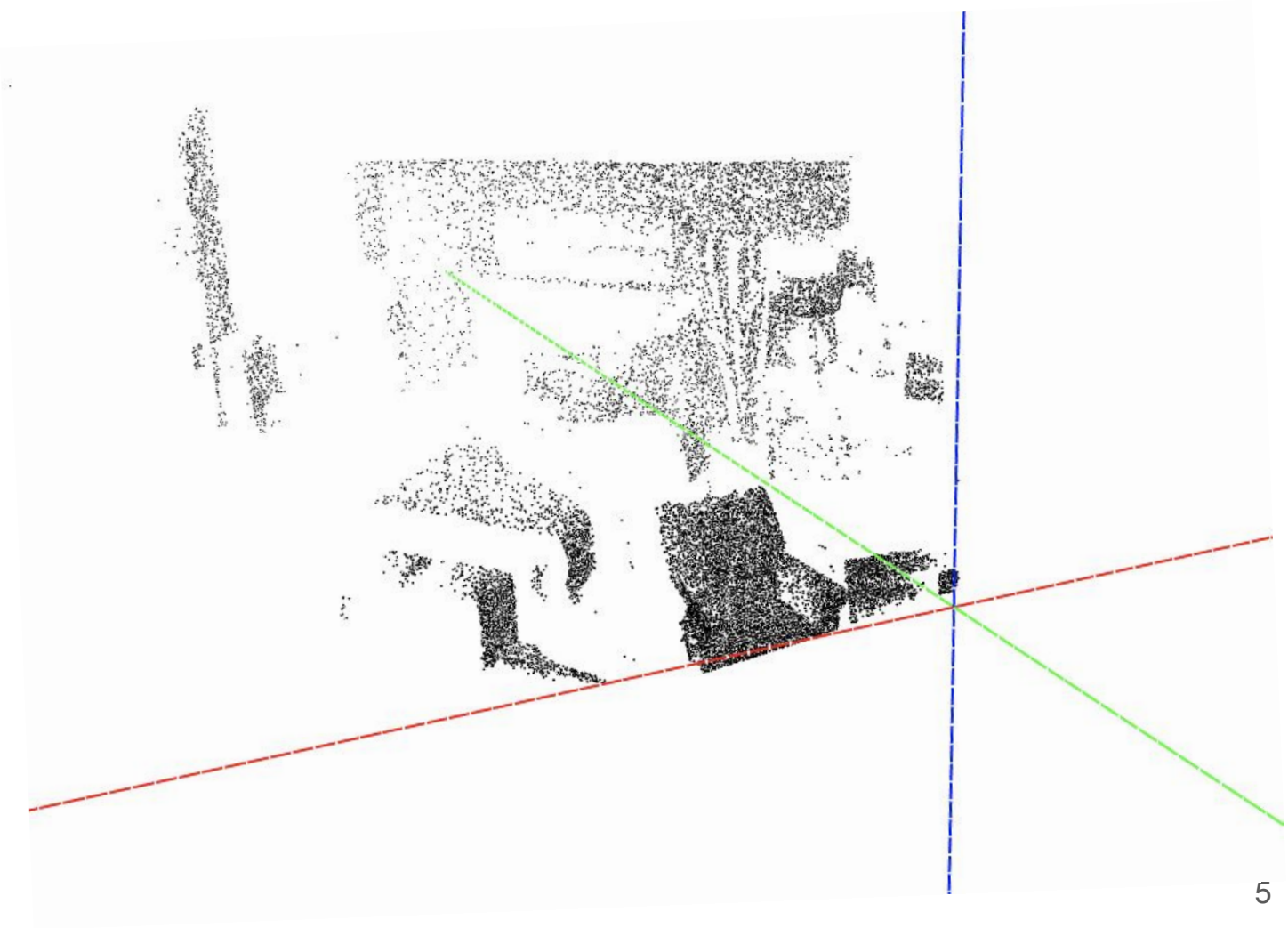
[F-PointNet by Qi et al. CVPR 2018]

Prior work relies on 2D object detection

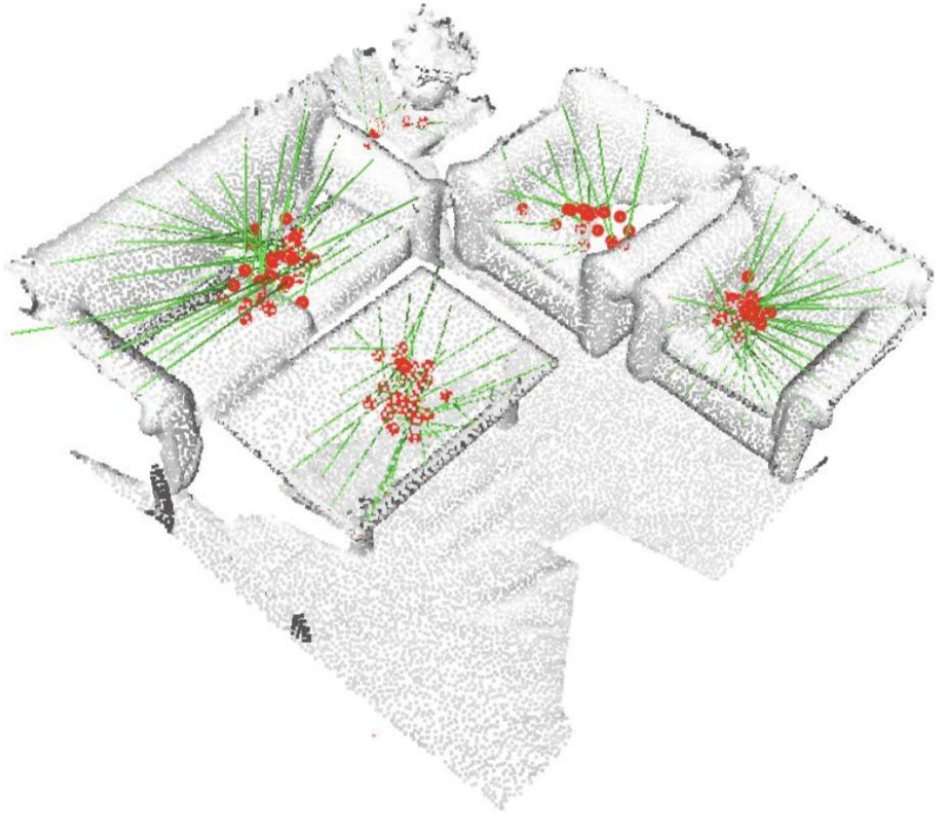


[Deep Sliding Shapes by Song et al. CVPR 2016]

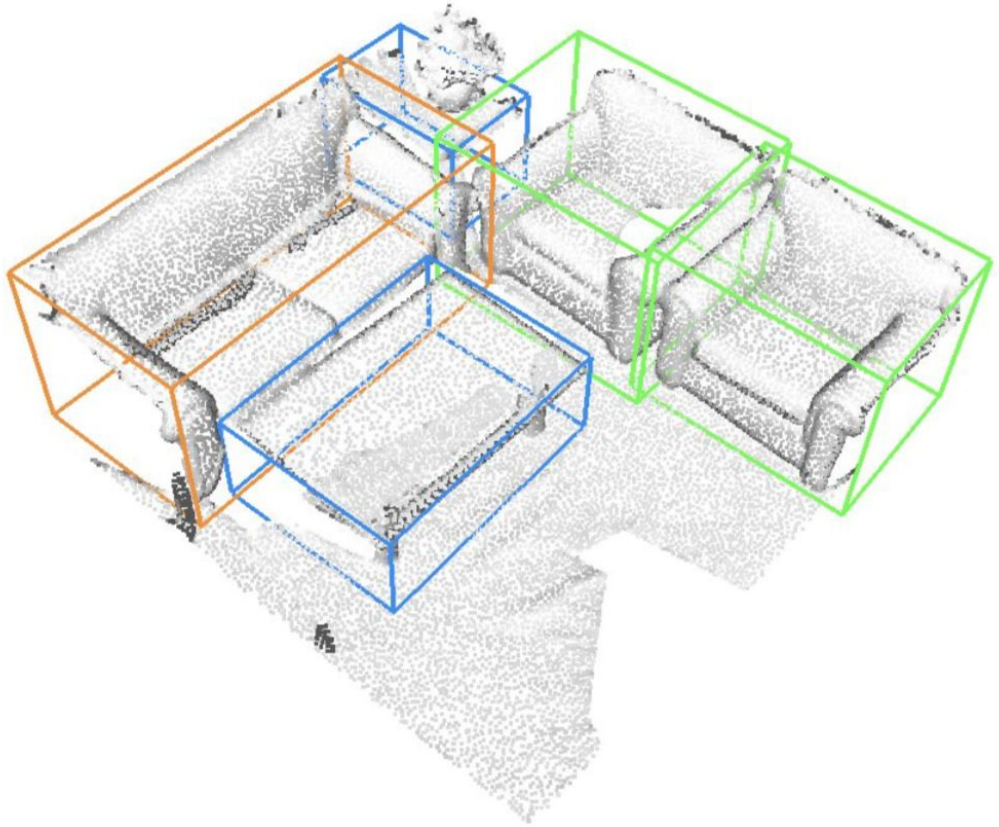
Observation: 2D v.s. 3D



Our idea: “ask” the points to *vote* for object centers



Voting from surface points



Detected 3D bounding boxes

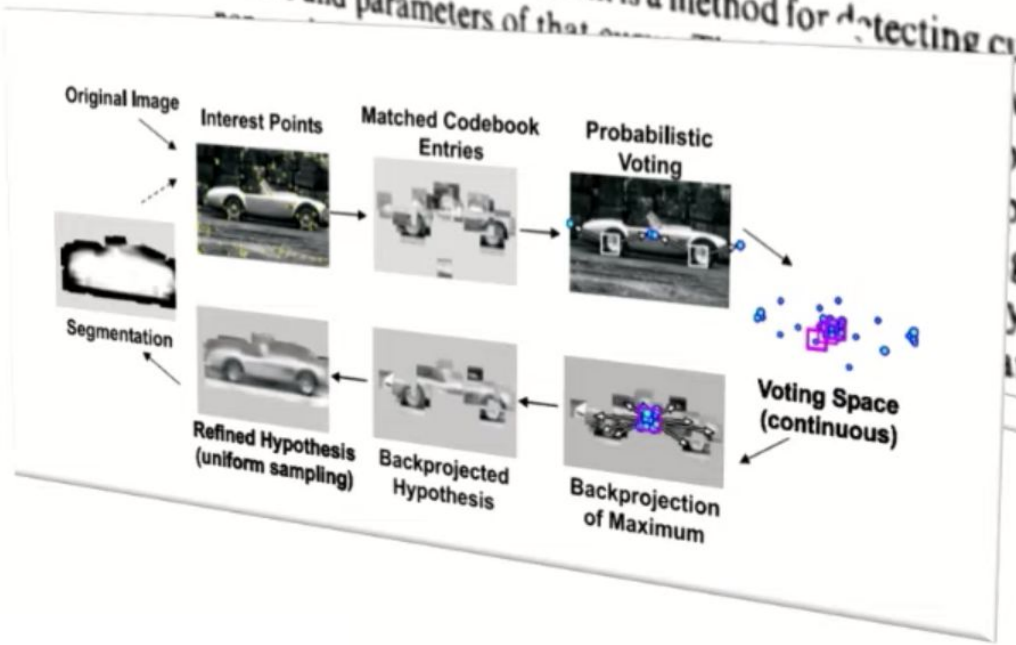
GENERALIZING THE HOUGH TRANSFORM TO DETECT ARBITRARY SHAPES*

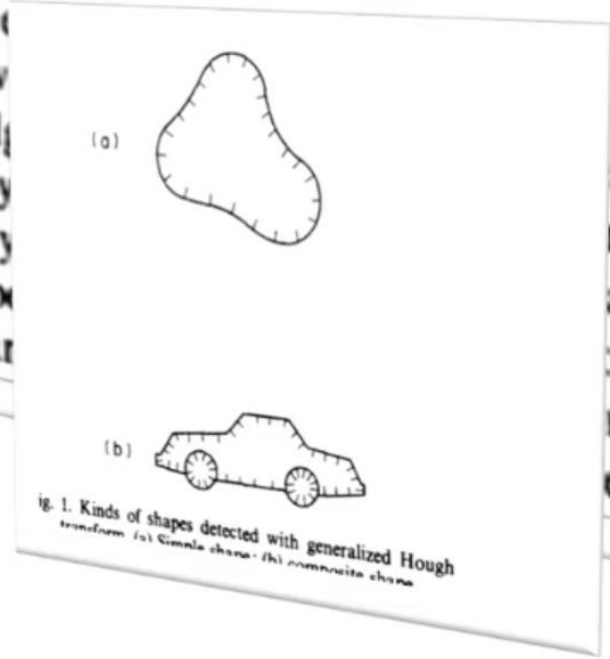
D. H. BALLARD

Computer Science Department, University of Rochester, Rochester, NY 14627, U.S.A.

(Received 10 October 1979; in revised form 9 September 1980; received for publication 23 September 1980)

Abstract—The Hough transform is a method for detecting curves by
a curve and parameters of that





ig. 1. Kinds of shapes detected with generalized Hough

points on
s[1,2] and
ralized to
as.[6] The
ons.[7,8,9]
mapping
es of that

Hough voting detector recap

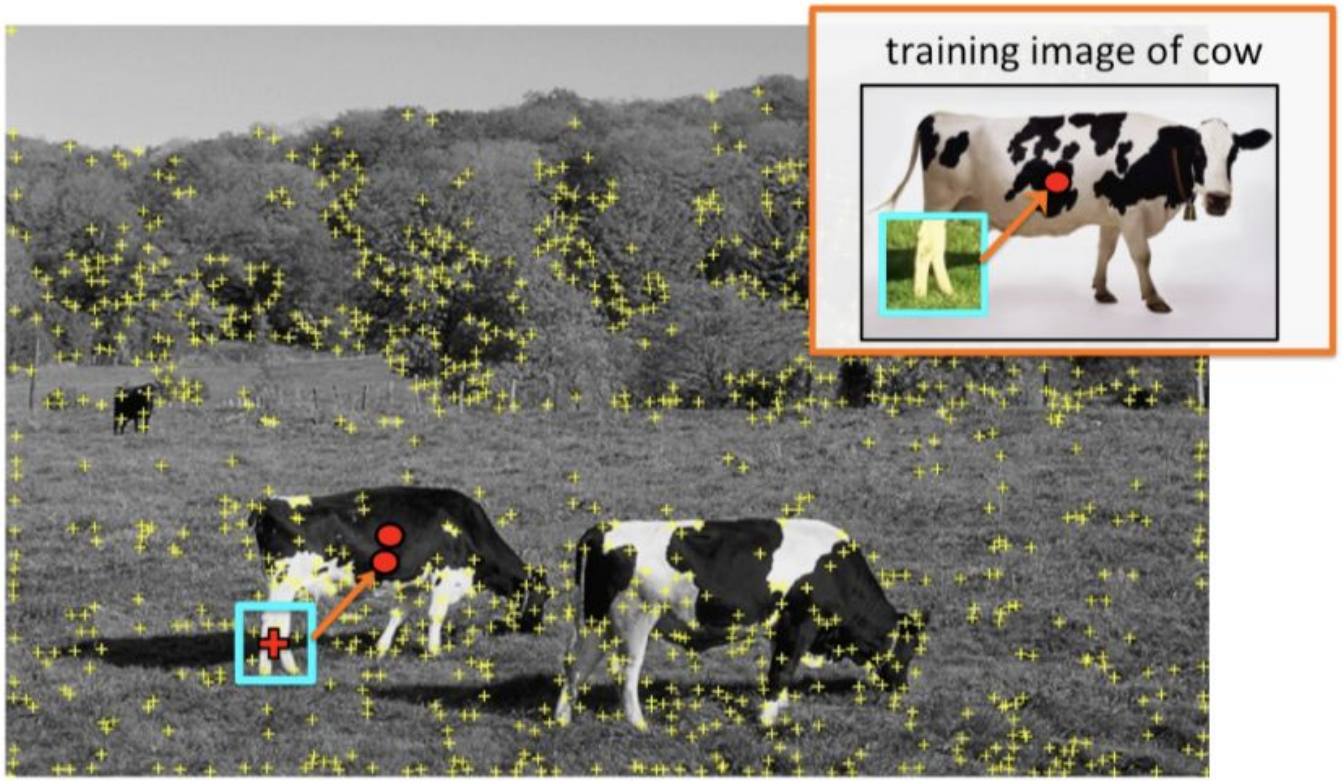
vote for center of object

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance

From U. Toronto CSC420

Hough voting detector recap



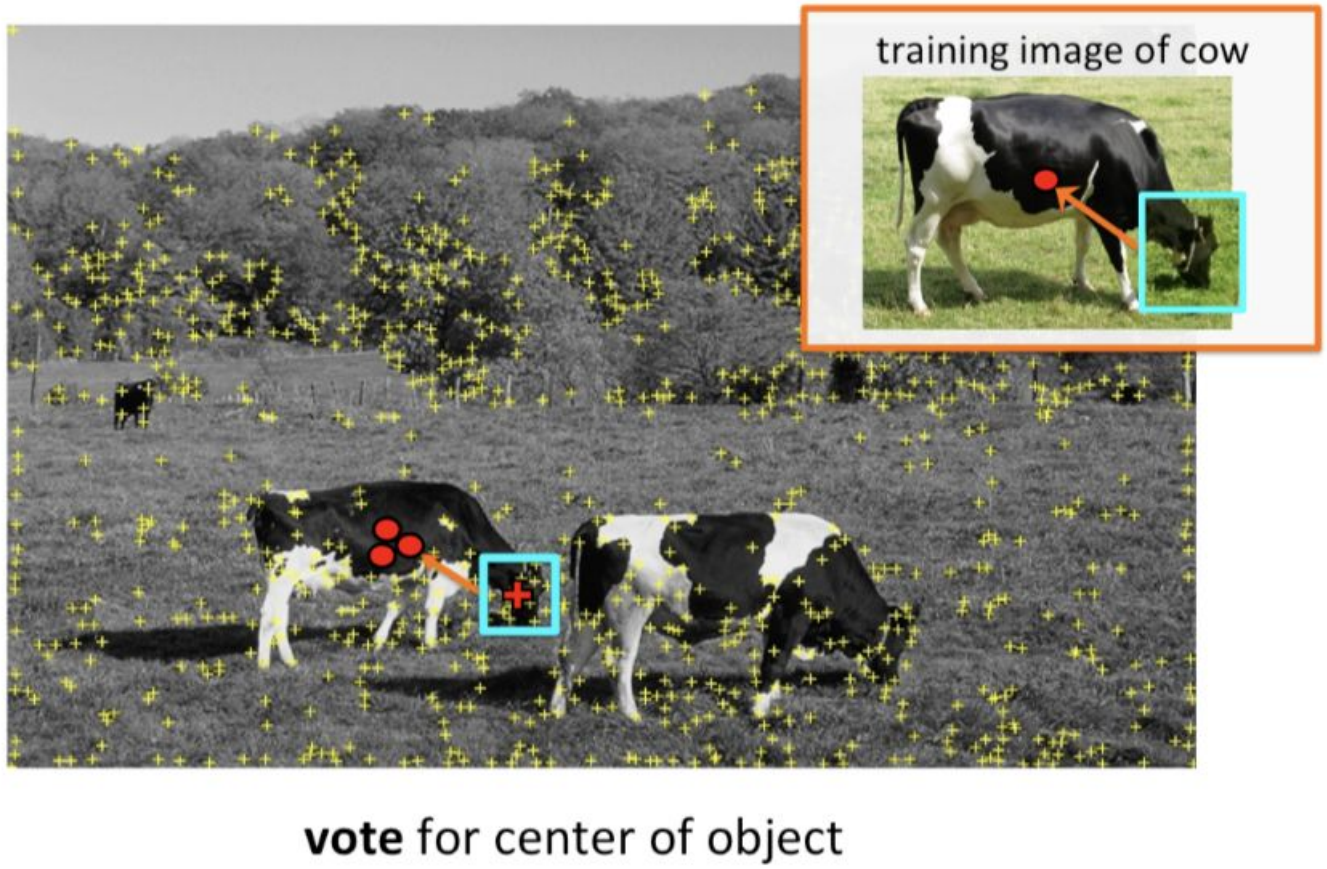
vote for center of object

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance

From U. Toronto CSC420

Hough voting detector recap



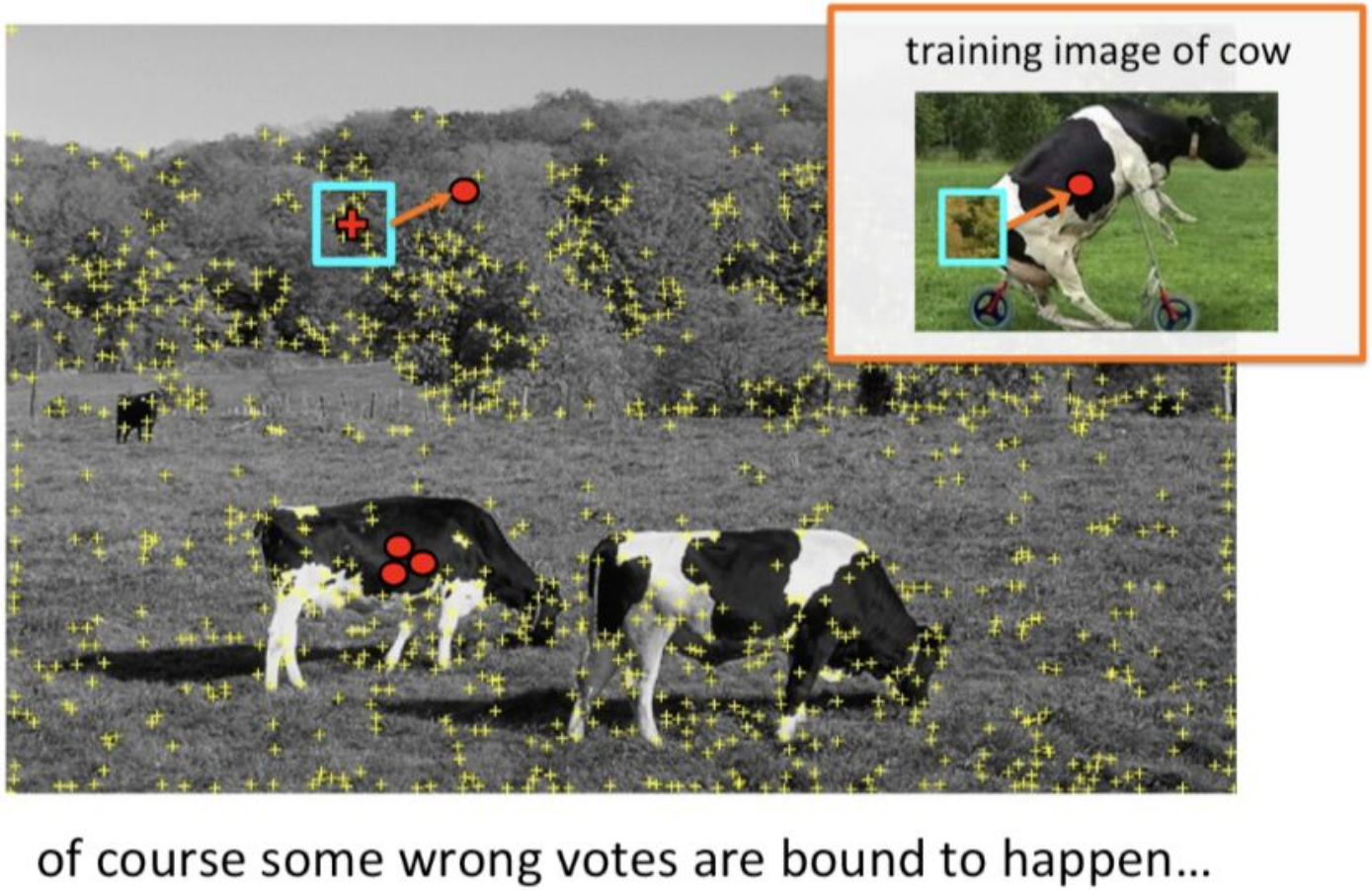
vote for center of object

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance

From U. Toronto CSC420

Hough voting detector recap

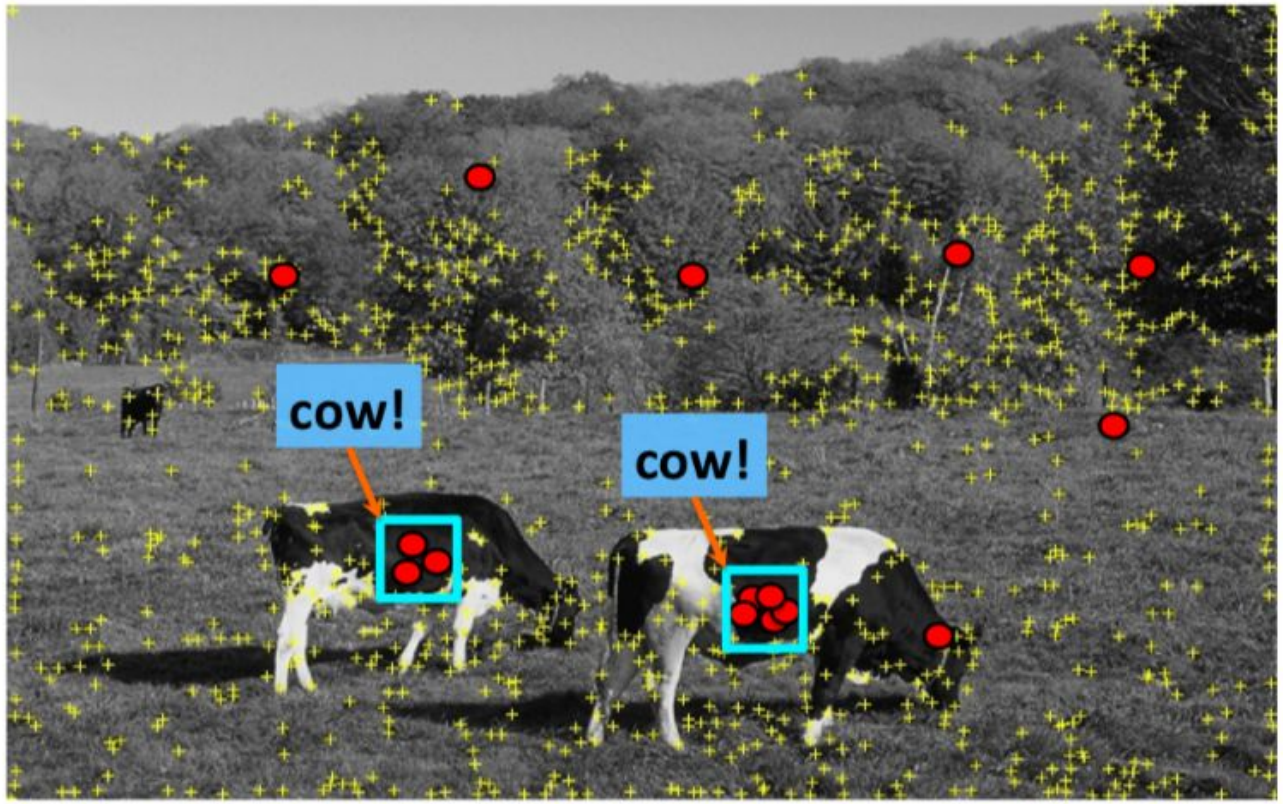


Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance

From U. Toronto CSC420

Hough voting detector recap



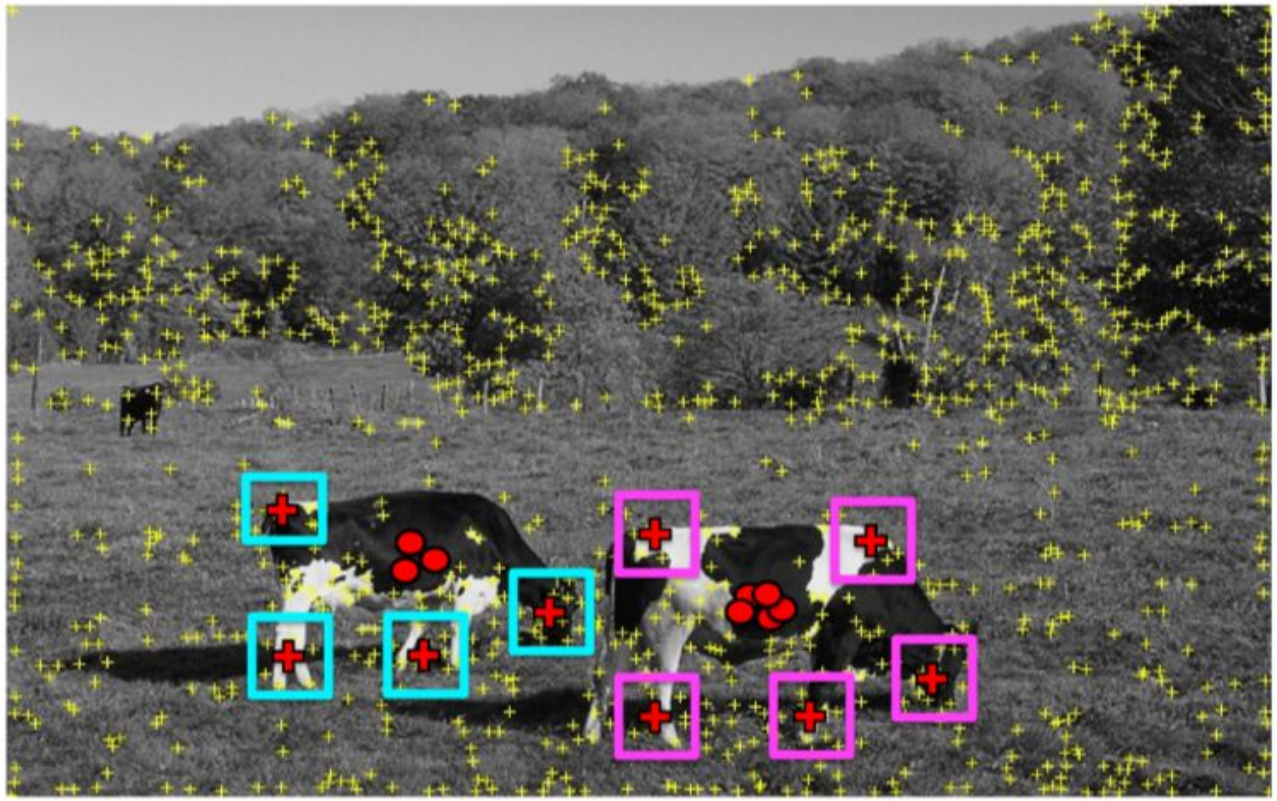
But that's ok. We want only **peaks** in voting space.

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance
- **Votes clustering to find peaks**

From U. Toronto CSC420

Hough voting detector recap



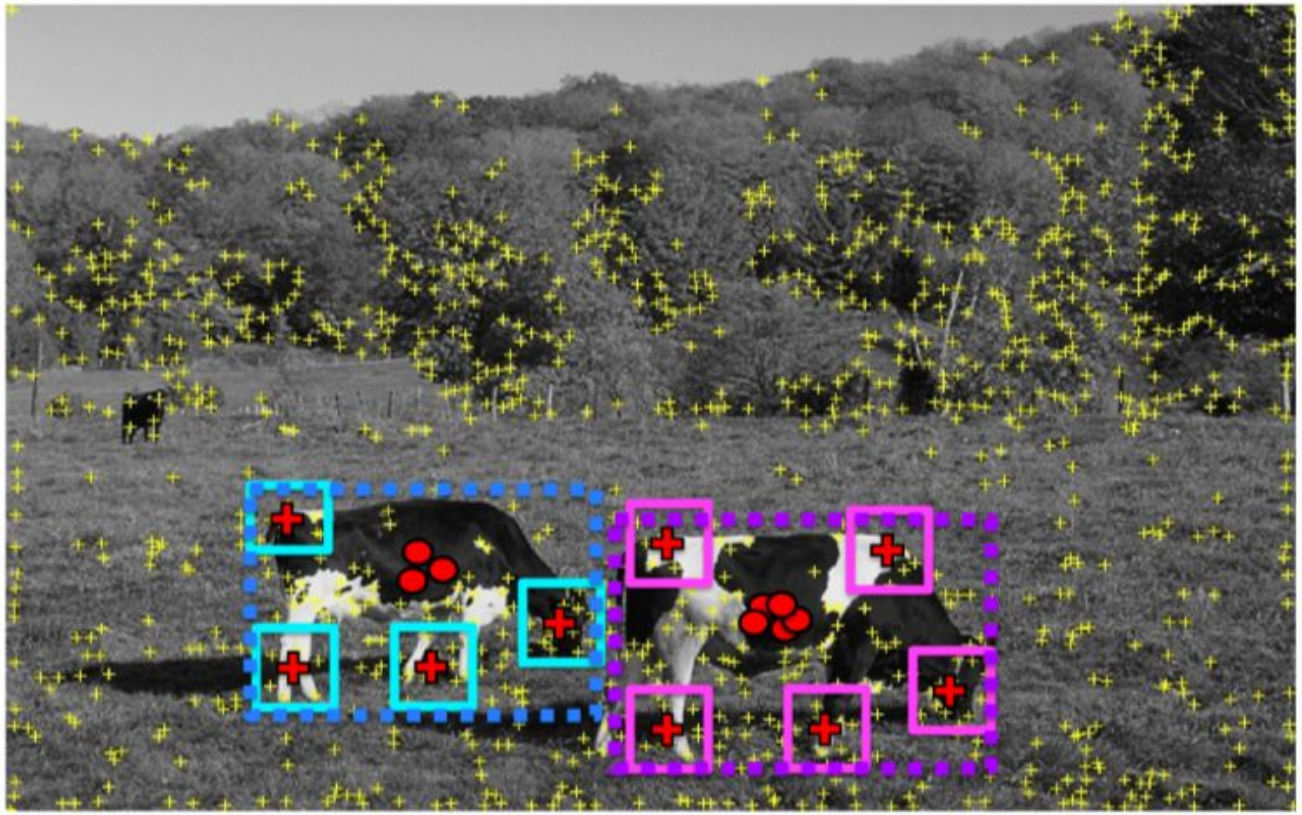
Find patches that voted for the peaks (back-projection).

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance
- Votes clustering to find peaks
- **Find patches that voted for the peaks by back-projection**

From U. Toronto CSC420

Hough voting detector recap



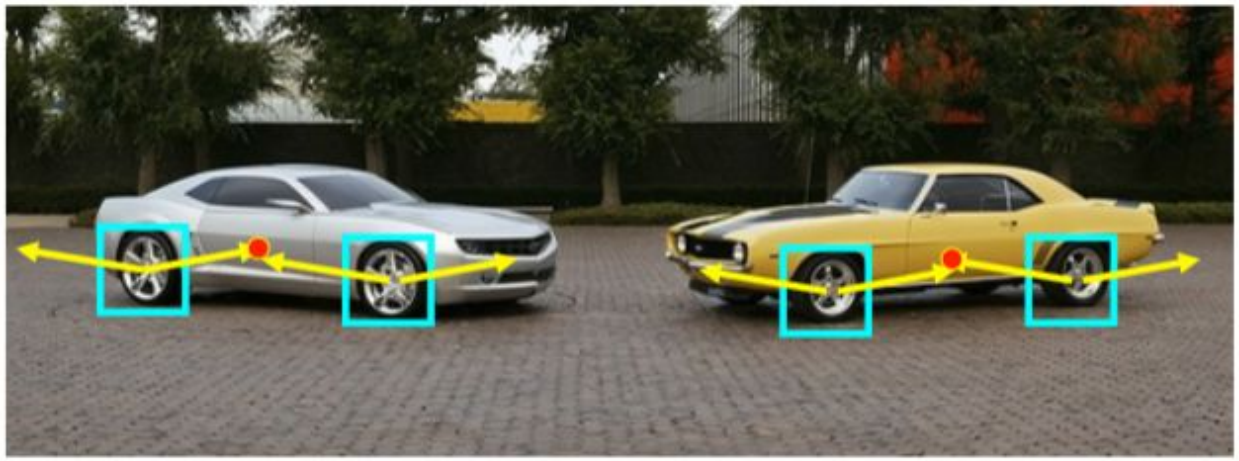
Find full objects based on the back-projected patches.

Hough voting pipeline (on 2D images):

- Select interest points
- Match patch around each interest point to a training patch (codebook)
- Vote for object center given that training instance
- Votes clustering to find peaks
- Find patches that voted for the peaks by back-projection
- **Find full objects based on back-projected patches**

From U. Toronto CSC420

Hough voting detector recap



+ Computation is only on “interest” points instead of on all pixels/voxels.
+ Support “templates” (used in 6DoF pose estimation)
- **Not end-to-end optimizable**

From U. Toronto CSC420

3D object proposal: A return of hough voting!

Deep hough voting with PointNet++

Interest points → seed points sampled from the point clouds

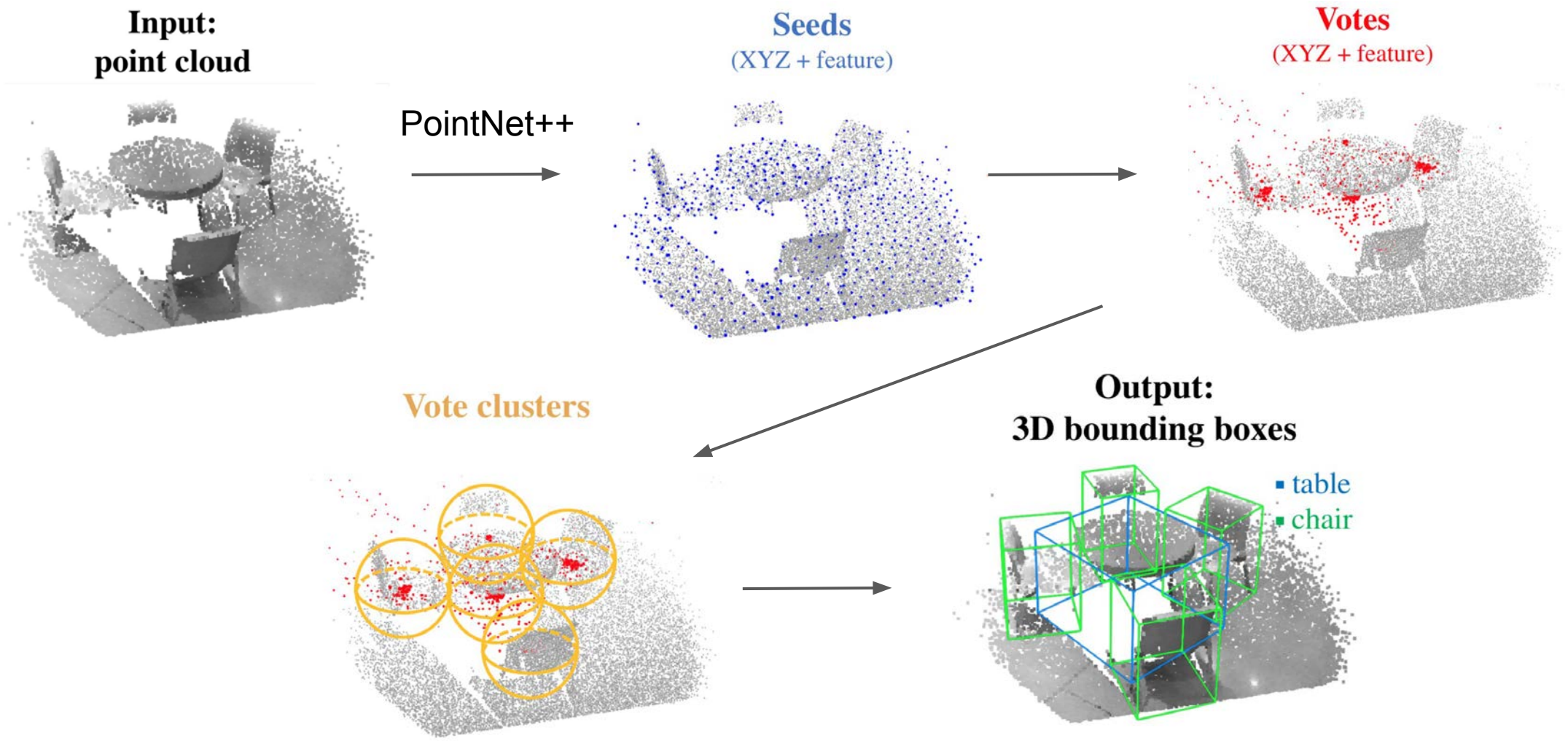
Votes → learned mapping from point features to votes

Clustering → local pointnet layers to group and aggregate local votes

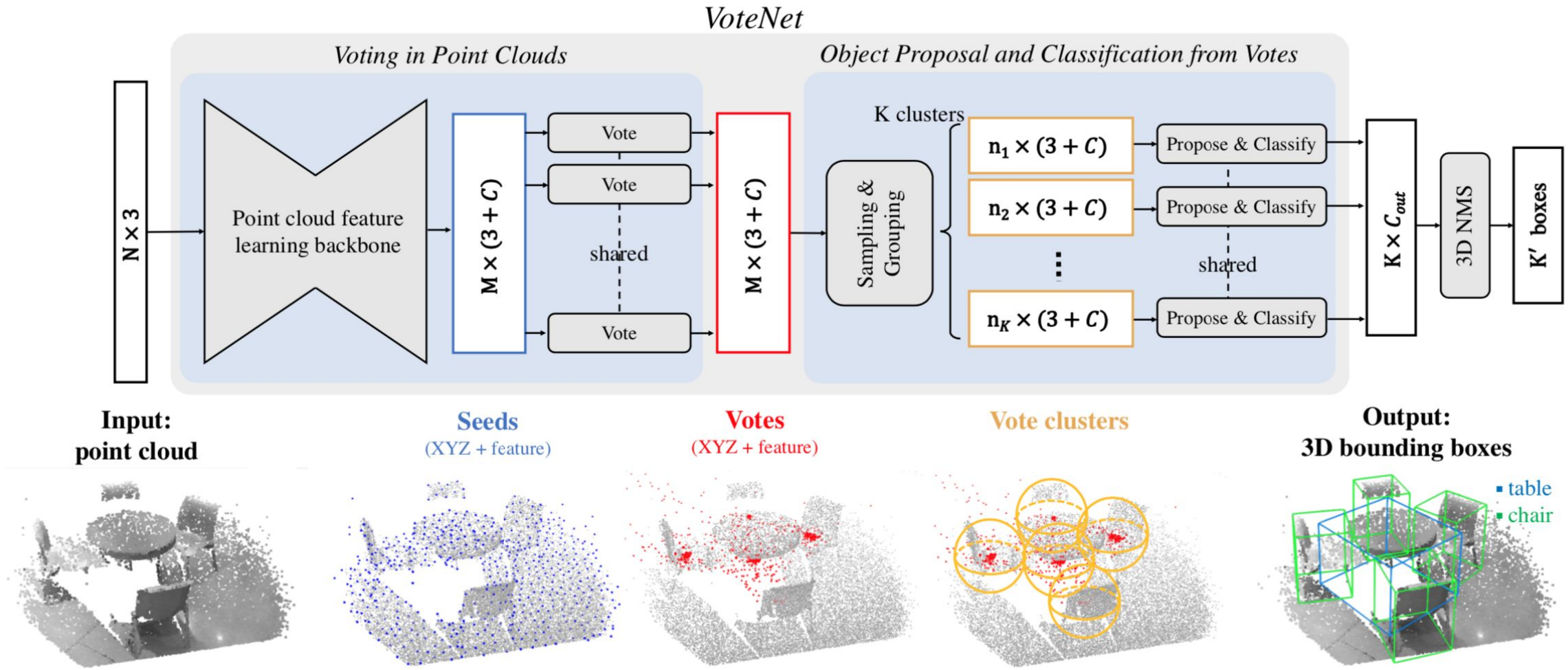
Object recovery → learned bounding box predictor

End-to-end optimizable!

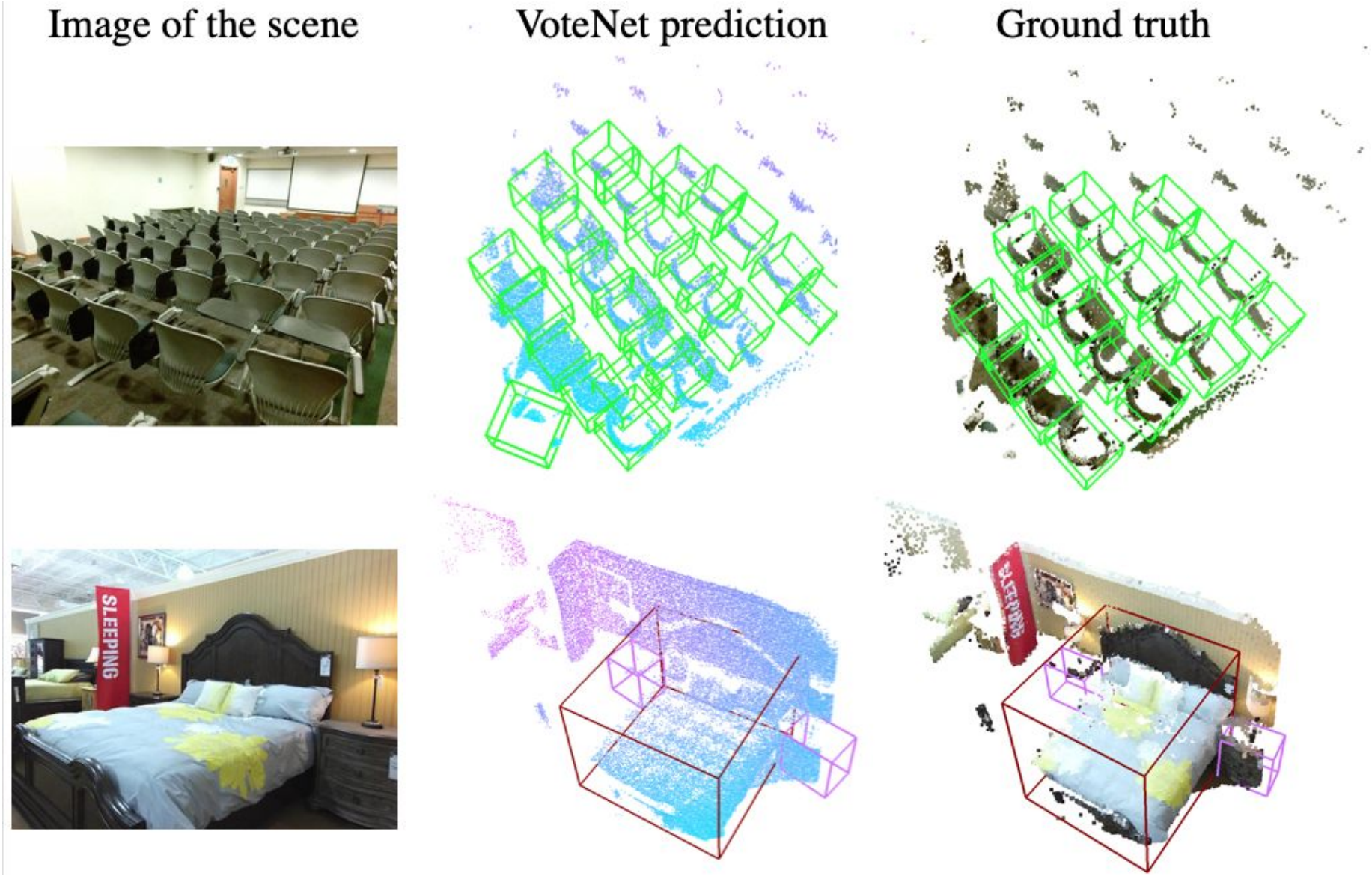
Deep Hough voting: Detection pipeline



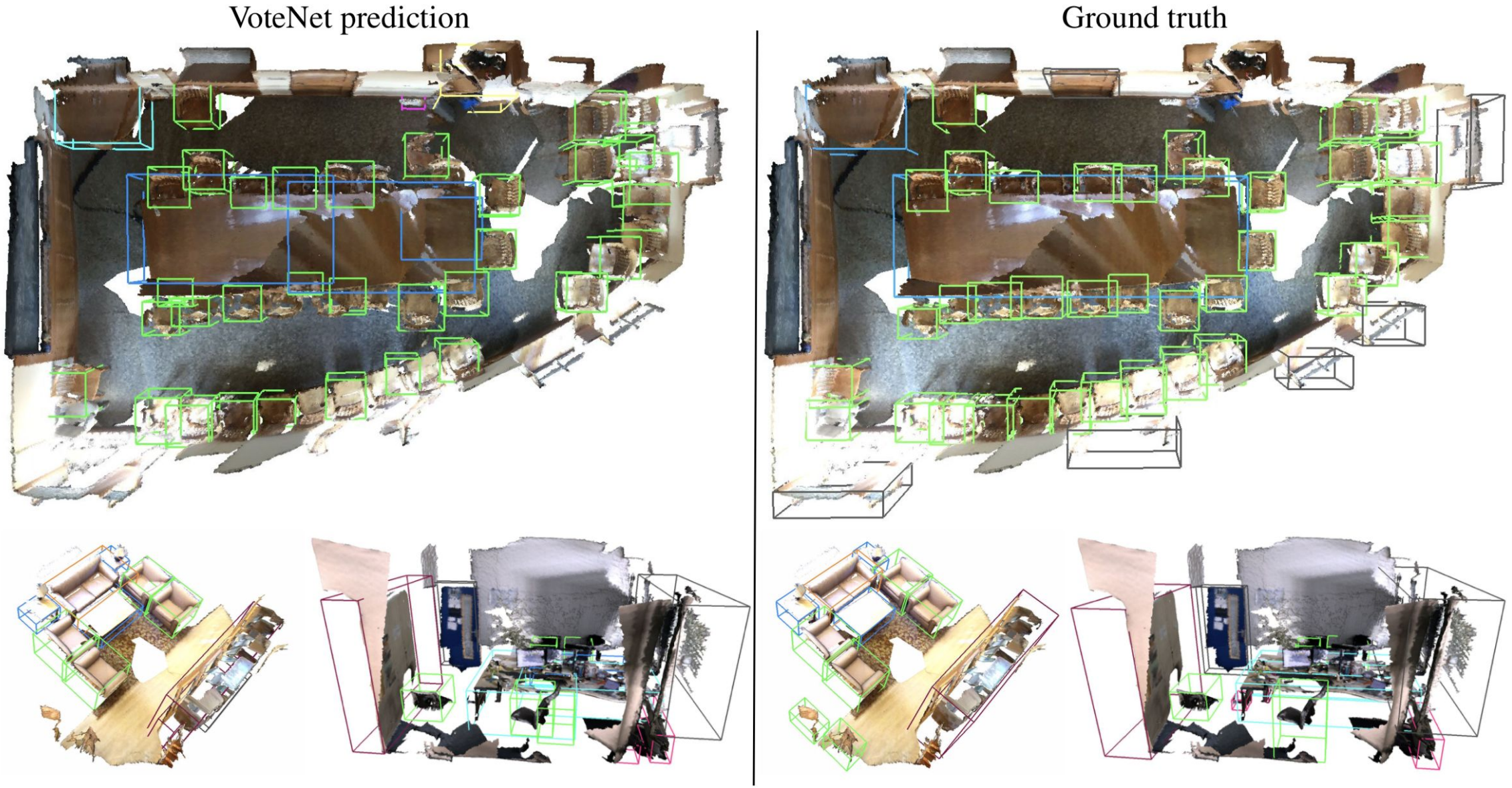
Deep Hough voting: Detection pipeline



Results: SUN RGB-D (single depth images)



Results: ScanNet (3D reconstructions)



Comparing with previous methods

	Input	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	table	toilet	mAP
DSS [42]	Geo + RGB	44.2	78.8	11.9	61.2	20.5	6.4	15.4	53.5	50.3	78.9	42.1
COG [38]	Geo + RGB	58.3	63.7	31.8	62.2	**45.2**	15.5	27.4	51.0	**51.3**	70.1	47.6
2D-driven [20]	Geo + RGB	43.5	64.5	31.4	48.3	27.9	25.9	41.9	50.4	37.0	80.4	45.1
F-PointNet [34]	Geo + RGB	43.3	81.1	**33.3**	64.2	24.7	**32.0**	58.1	61.1	51.1	**90.9**	54.0
VoteNet (ours)	**Geo only**	**74.4**	**83.0**	28.8	**75.3**	22.0	29.8	**62.2**	**64.0**	47.3	90.1	**57.7**

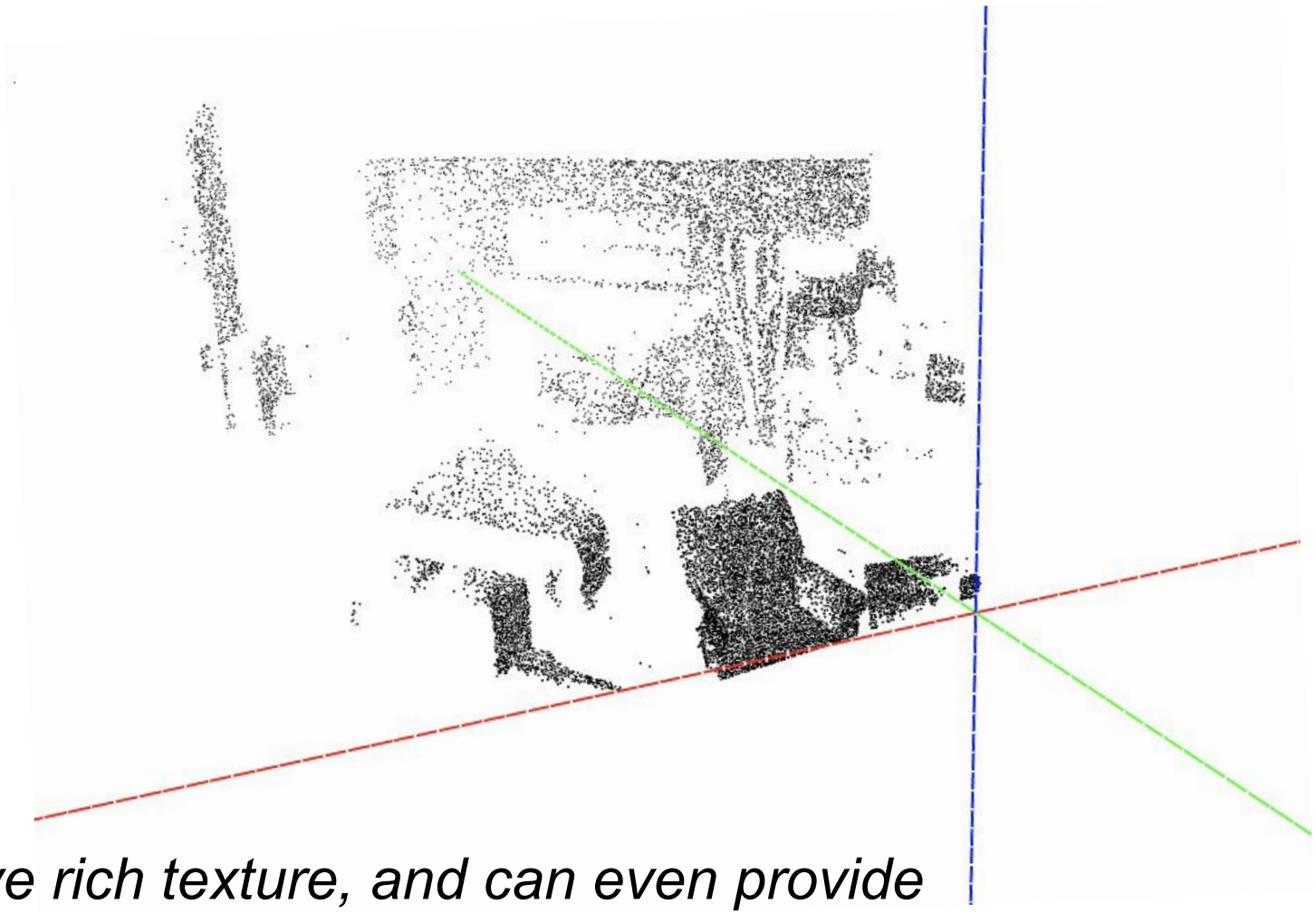
SUN RGB-D: *+3.7mAP* with just 3D geometry data as input.

Comparing with previous methods

	Input	mAP@0.25	mAP@0.5
DSS [42, 12]	Geo + RGB	15.2	6.8
MRCNN 2D-3D [11, 12]	Geo + RGB	17.3	10.5
F-PointNet [34, 12]	Geo + RGB	19.8	10.8
GSPN [54]	Geo + RGB	30.6	17.7
3D-SIS [12]	Geo + 1 view	35.1	18.7
3D-SIS [12]	Geo + 3 views	36.6	19.0
3D-SIS [12]	Geo + 5 views	40.2	22.5
3D-SIS [12]	Geo only	25.4	14.6
VoteNet (ours)	Geo only	**58.6**	**33.5**

ScanNet: *+18.3 mAP* compared with prior art (3D CNN based method) with 3D & multi-view images.

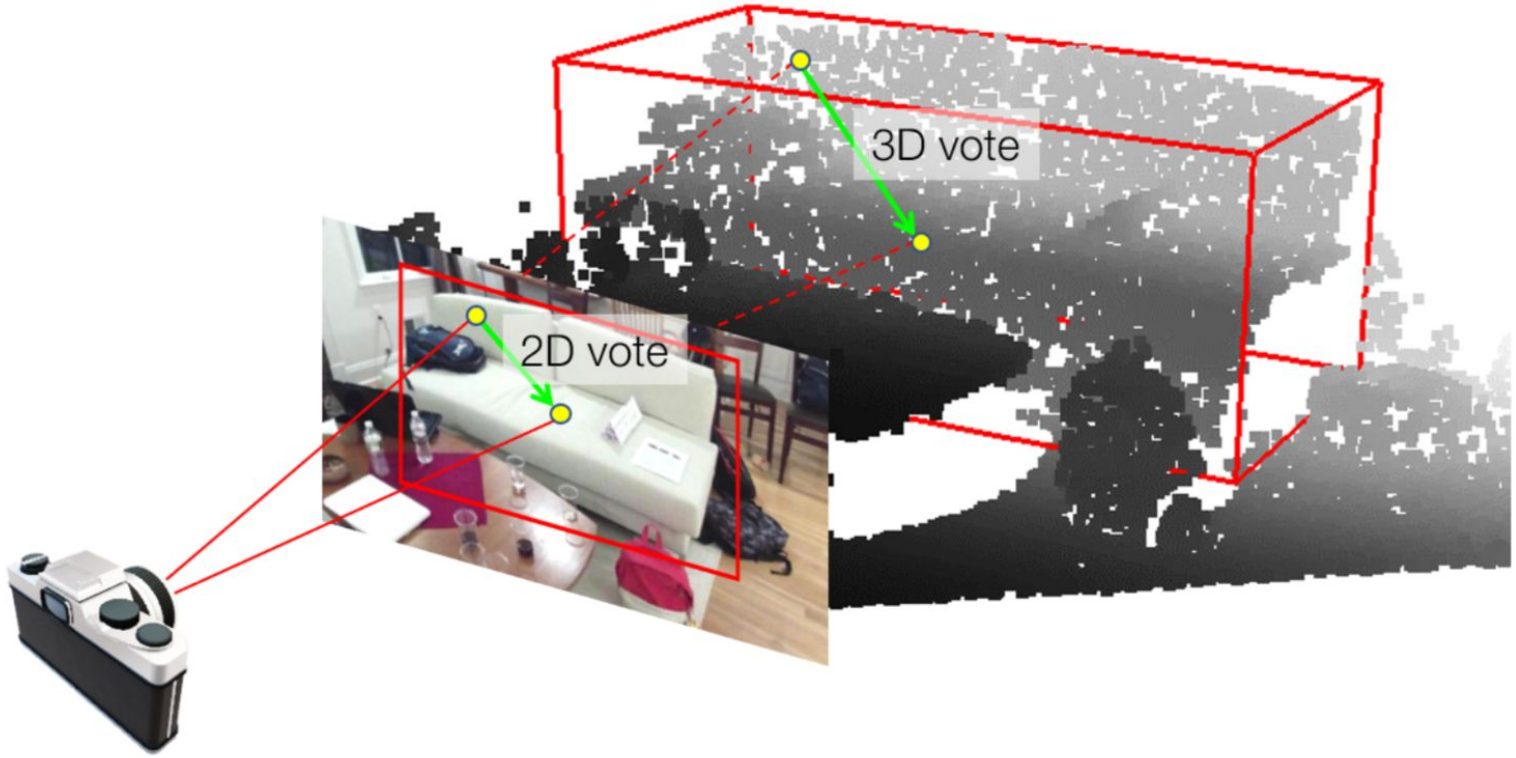
Can images help the VoteNet detection?



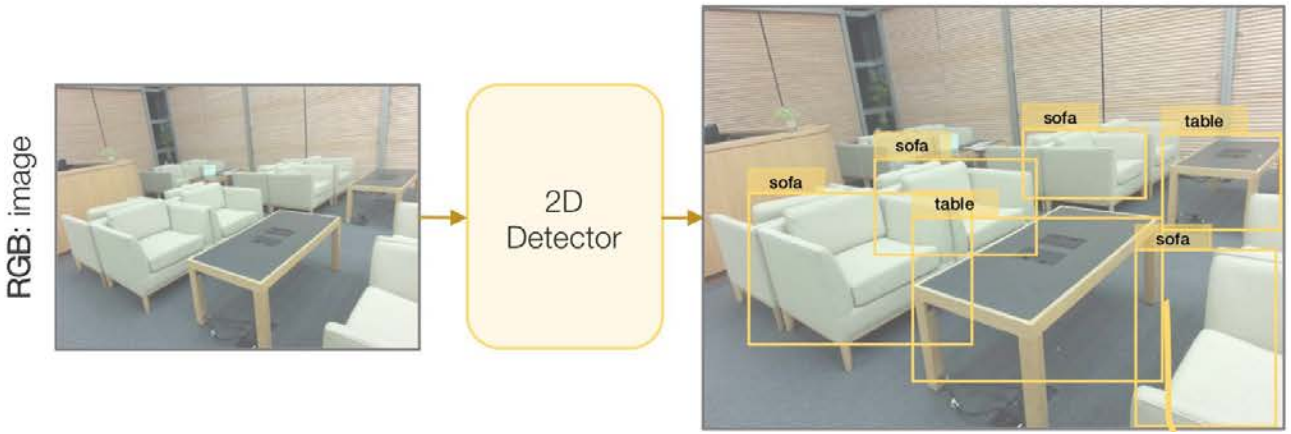
Images are in high resolution, have rich texture, and can even provide useful geometric cues for object localization & shape/pose estimation.

ImVoteNet: Boosting 3D Object Detection in Point Clouds with Image Votes

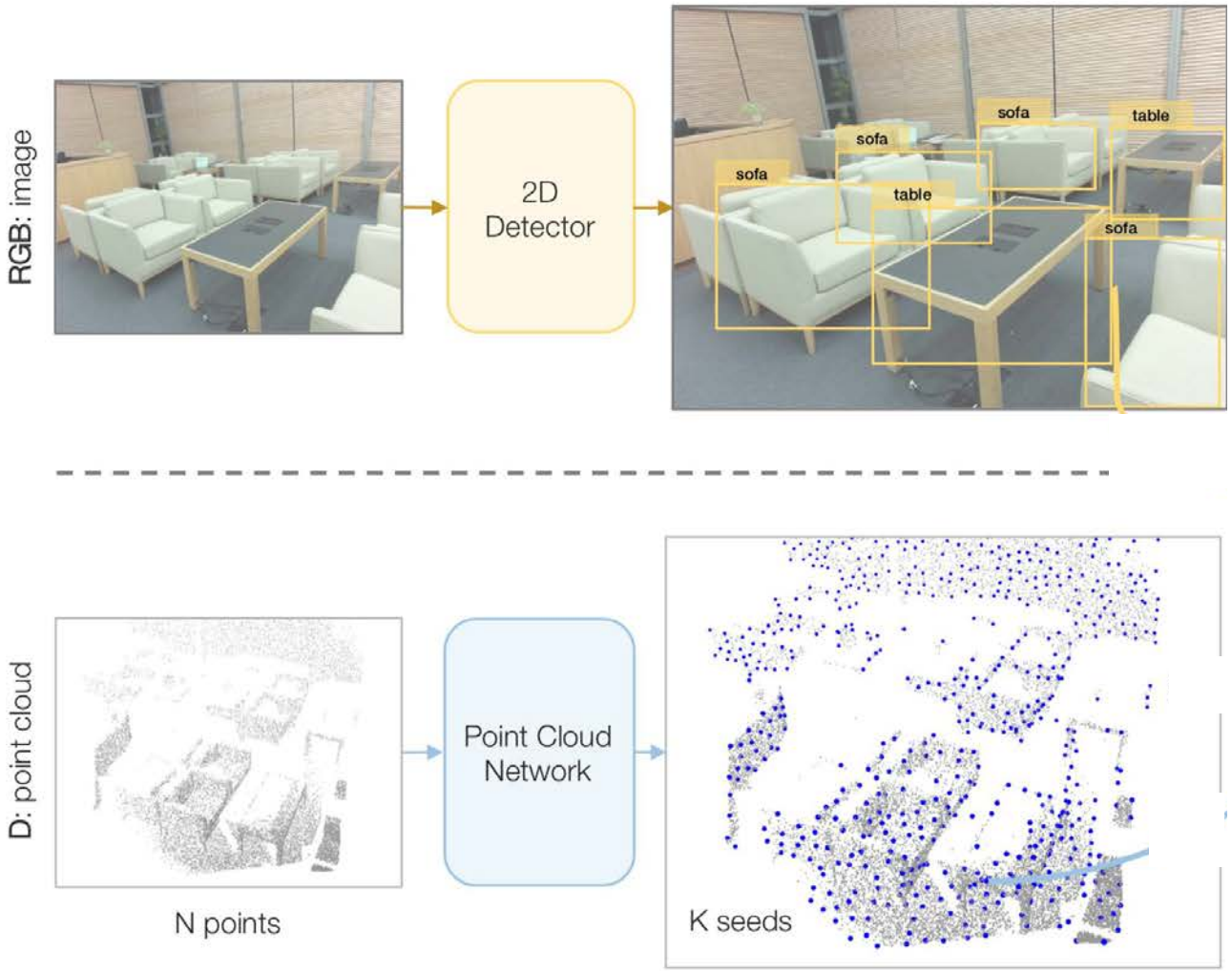
On-going work with Xinlei Chen, Or Litany and Leonidas Guibas



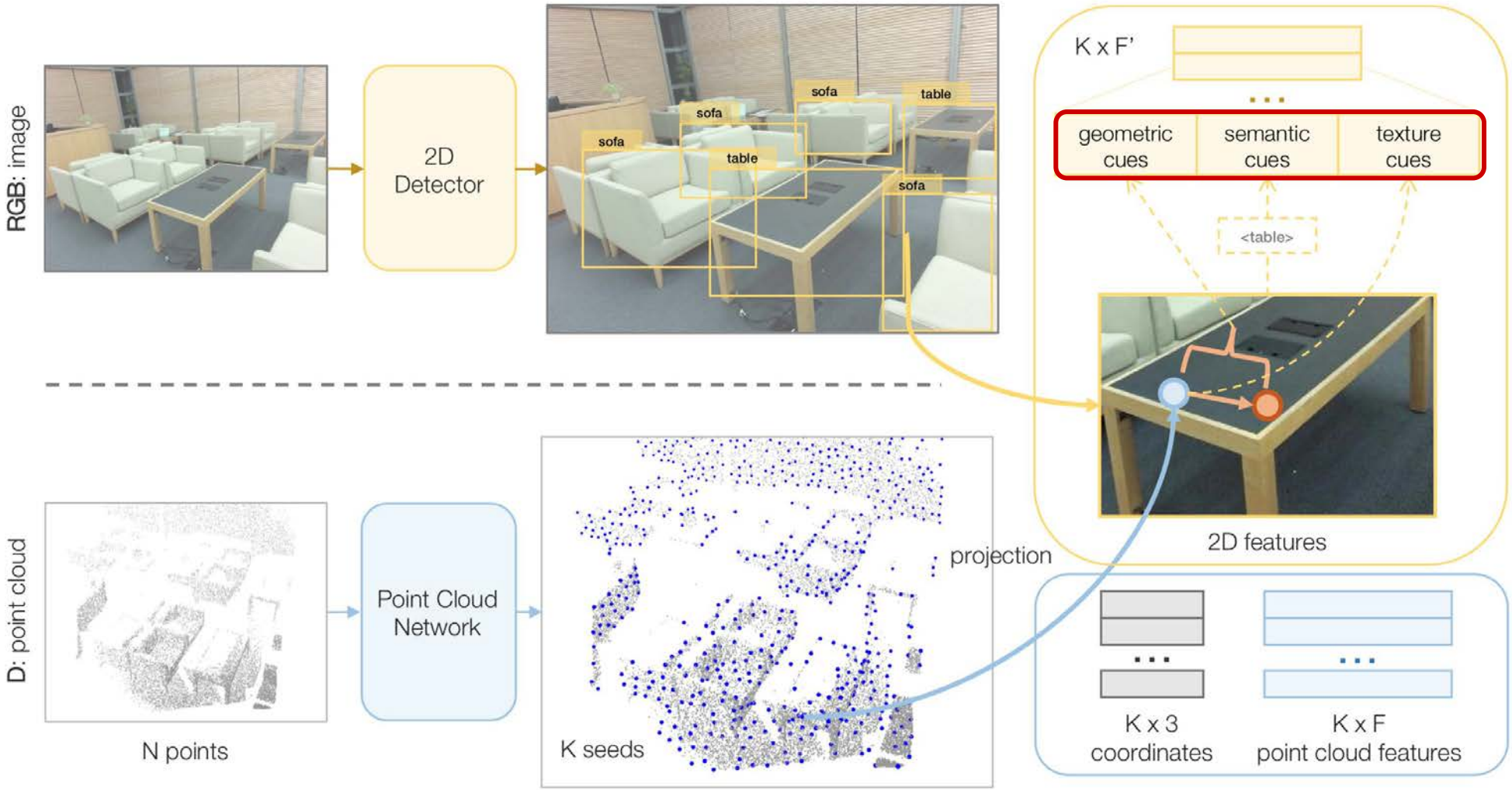
ImVoteNet detection pipeline



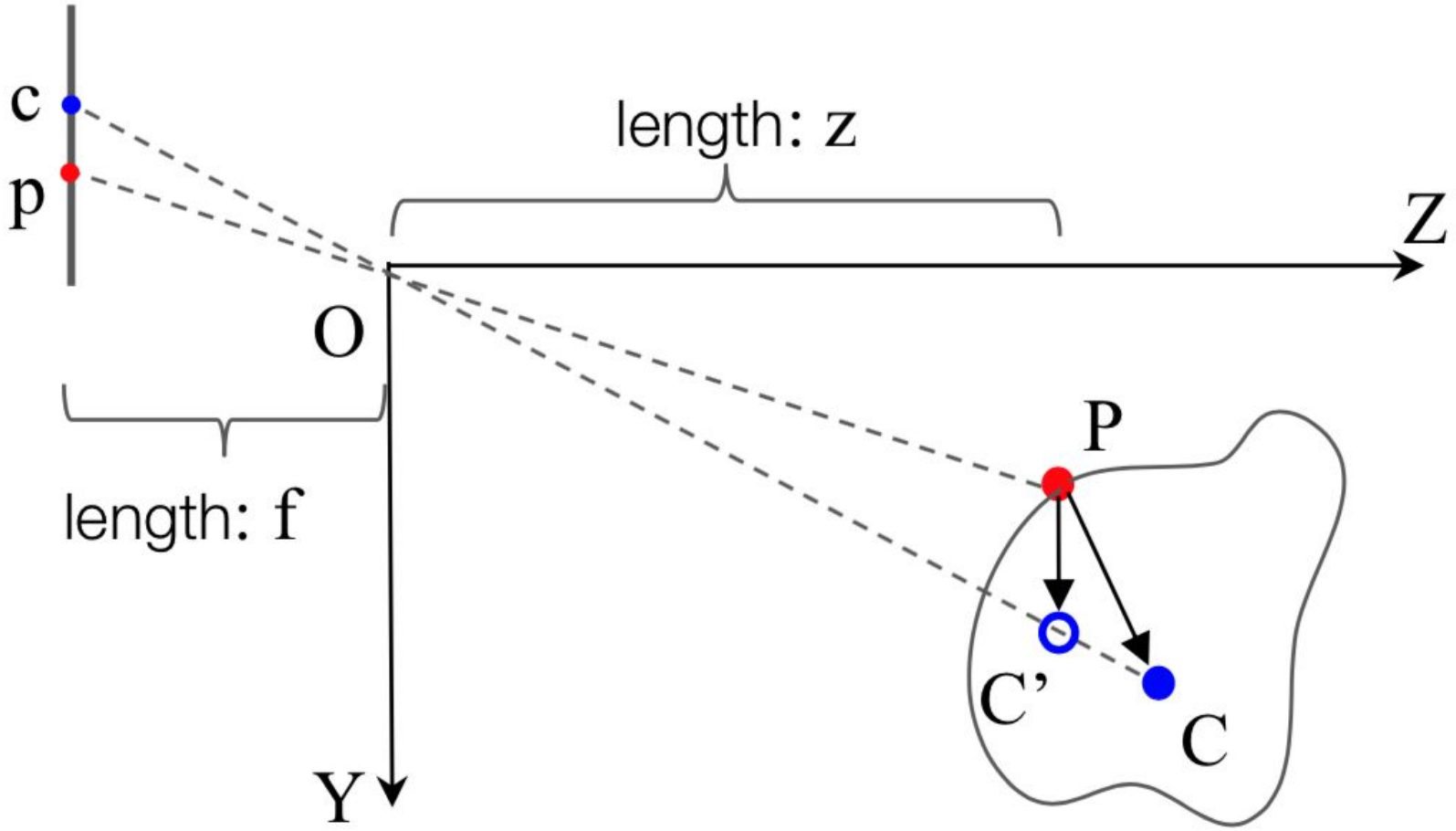
ImVoteNet detection pipeline



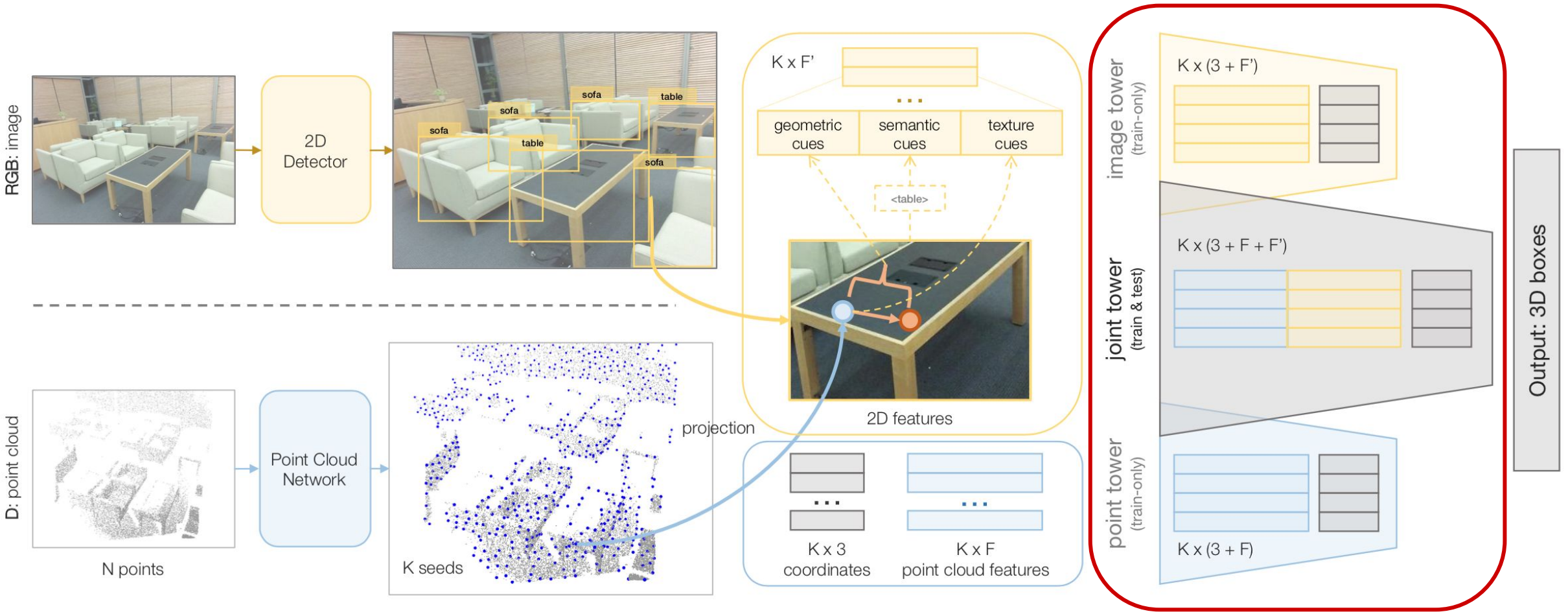
ImVoteNet detection pipeline



Geometric cues from images: Lifted image votes

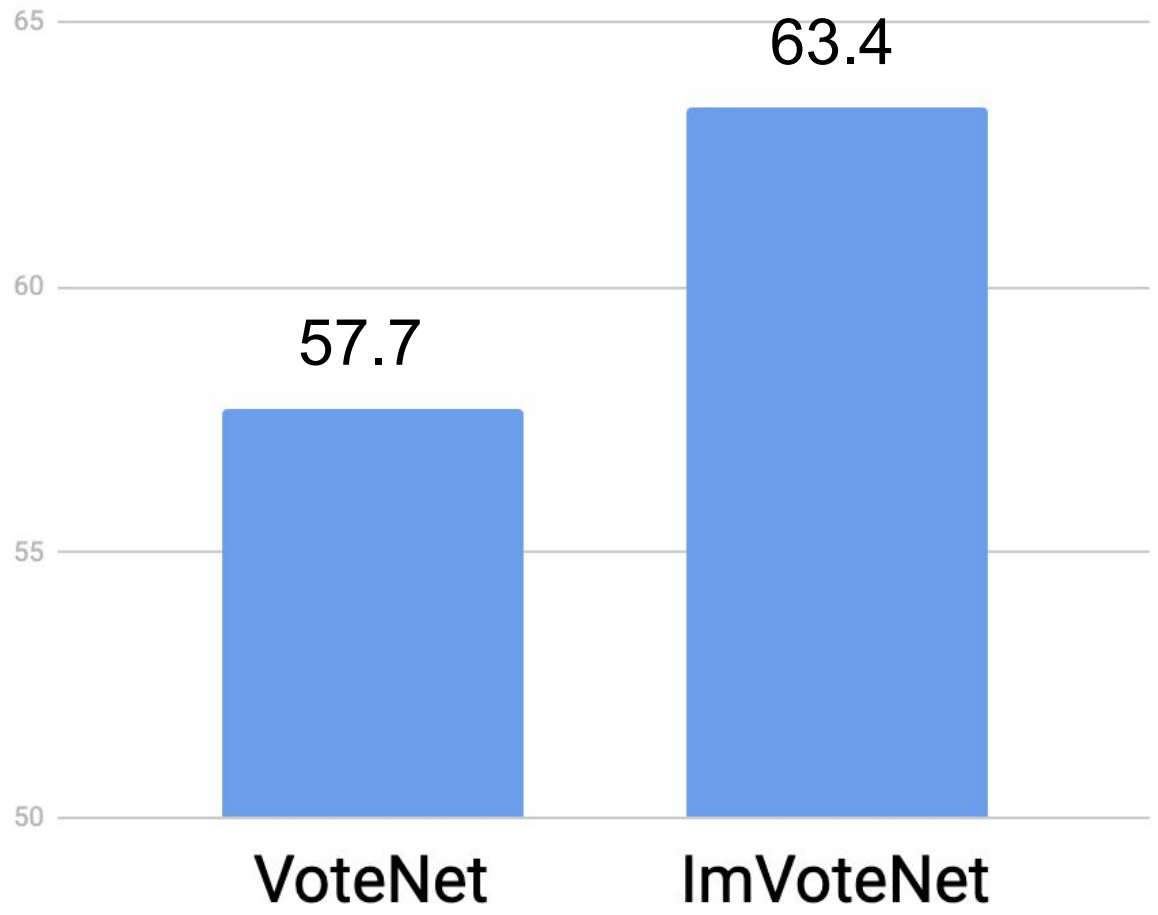


ImVoteNet detection pipeline



Results on SUN RGB-D

+5.7mAP with lifted image cues for voting

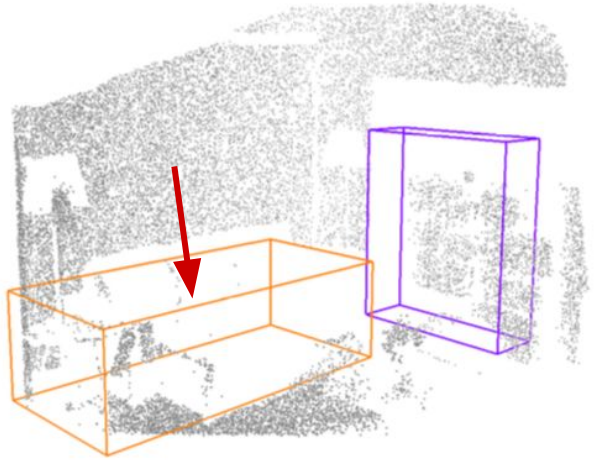


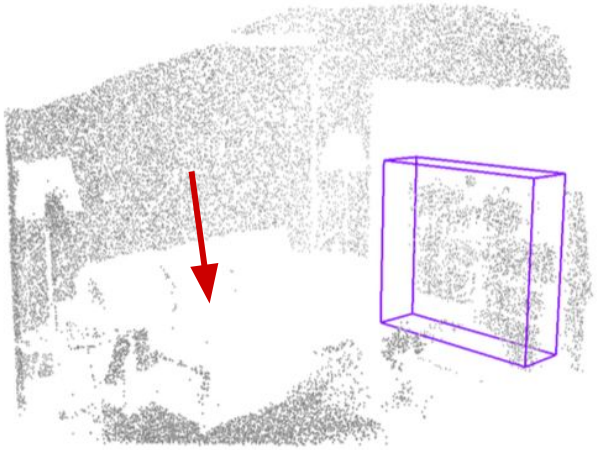
Results on SUN RGB-D

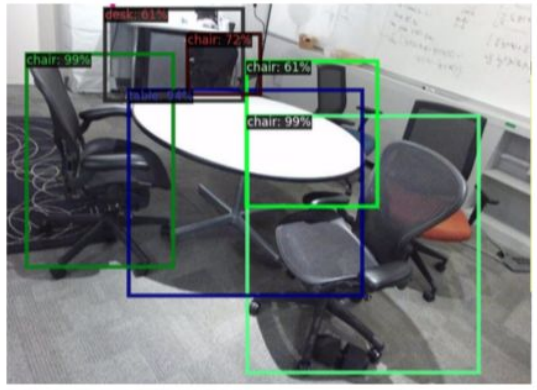
Ours 2D detection

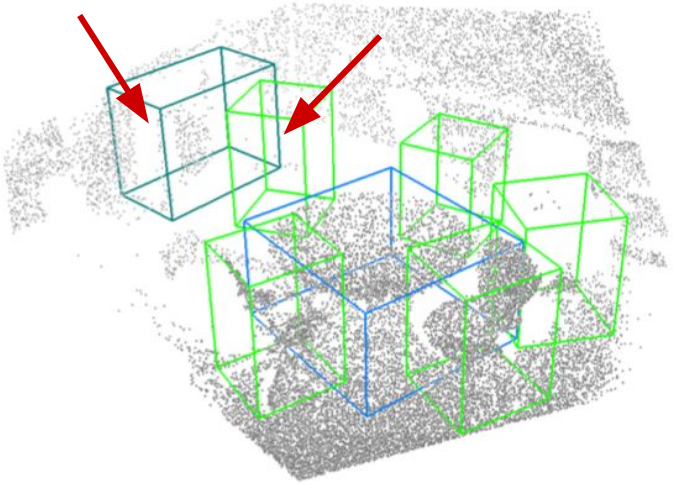
Ours 3D detection

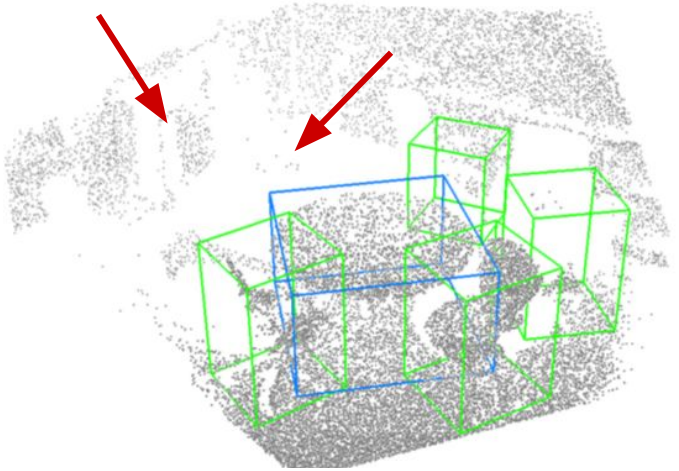
VoteNet











sofa bookshelf chair table desk

Summary

VoteNet: a revival of Hough voting with 3D deep learning.

- End-to-end optimizable hough voting with point cloud deep nets.
- A new detection model with a simple design shows state-of-the-art results on SUN RGB-D and ScanNet with geometry data only.

Code: https://github.com/facebookresearch/votenet

ImVoteNet: boosting 3D detection with lifted image votes.

Many open possibilities to extend the pipeline (e.g. 6D pose estimation, template based detection).