物体检测中的深度学习方法探索

Wanli Ouyang （欧阳万里）

wanli.ouyang@sydney.edu.au

The Chinese University of Hong Kong

The University of Sydney

Advertisement

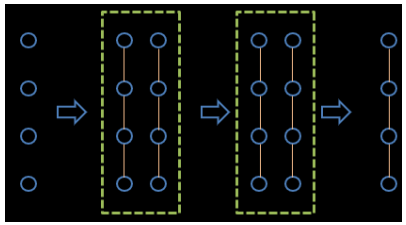
- Deep learning for generic object detection: a survey

Outline

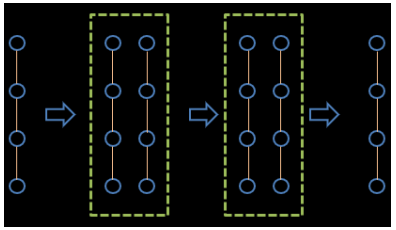
Introduction

Structured deep learning

Structured output and features



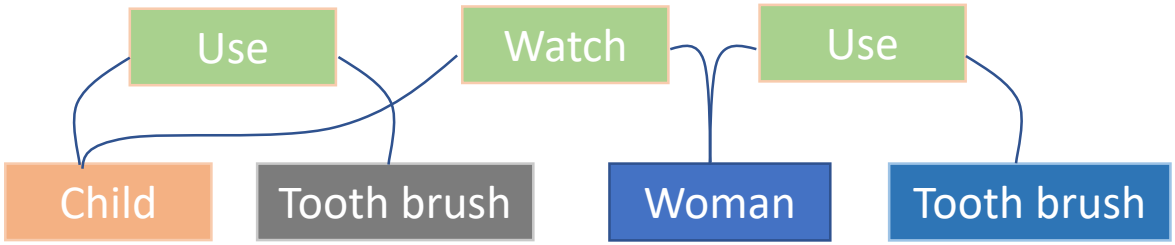
Structured input



Conclusion

Outline

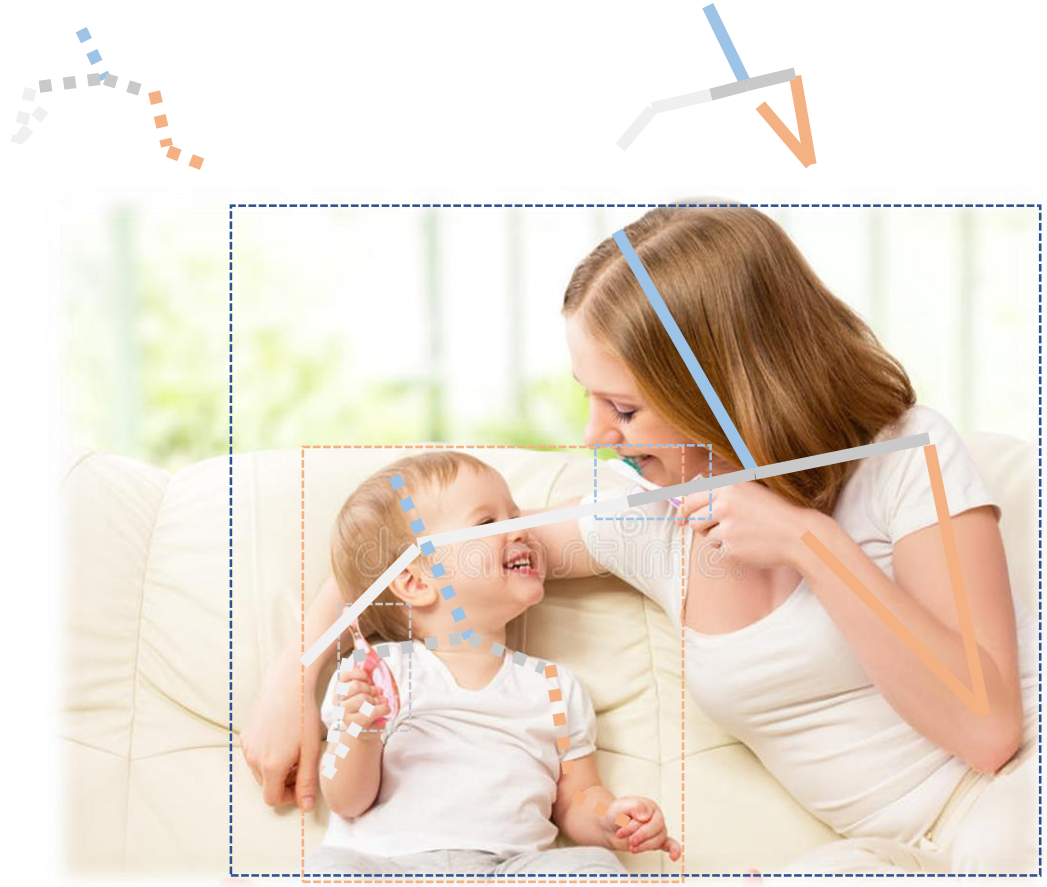
Introduction



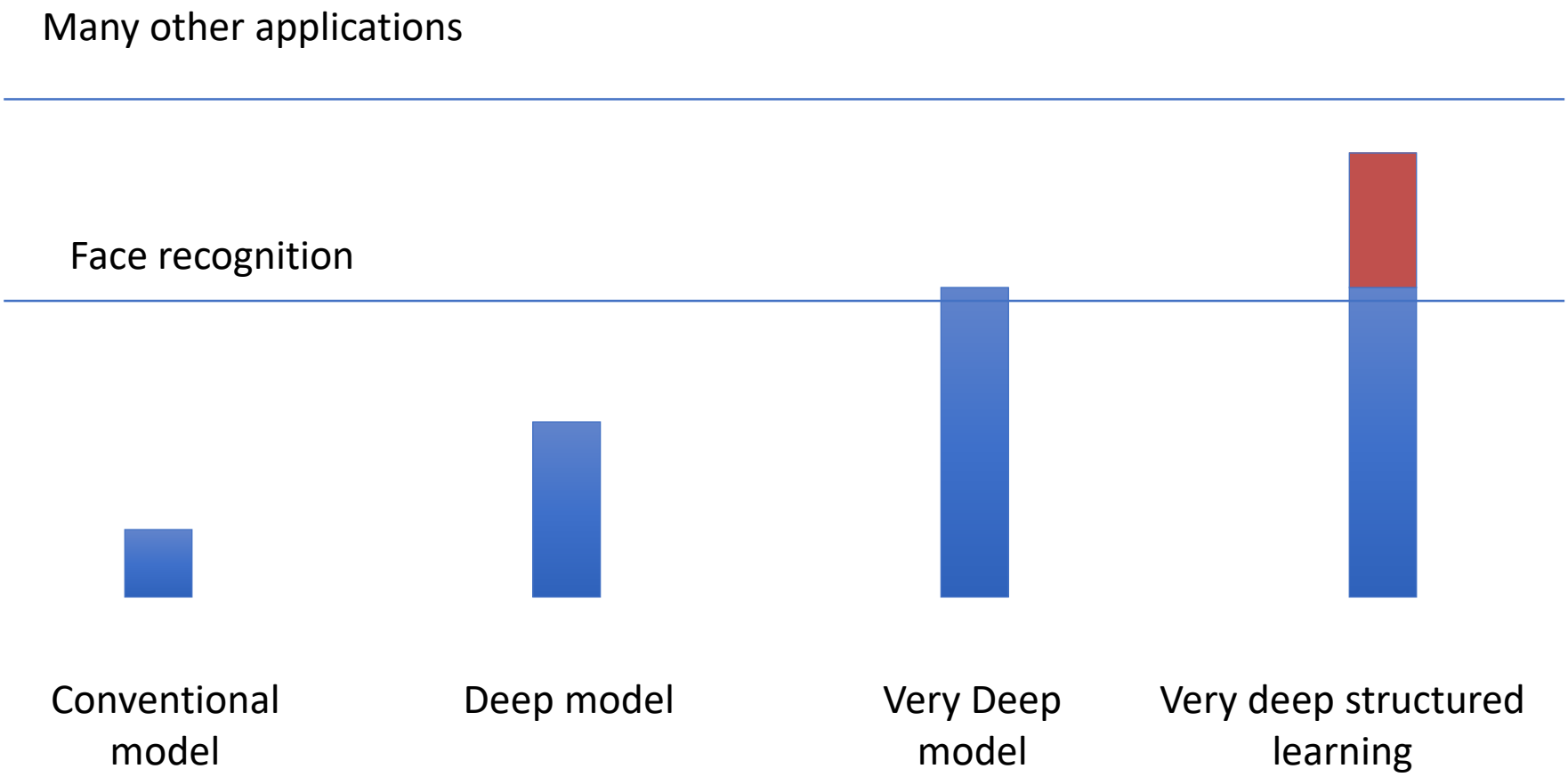
Relationship detection

Object detection

Human pose estimation



Performance vs practical need

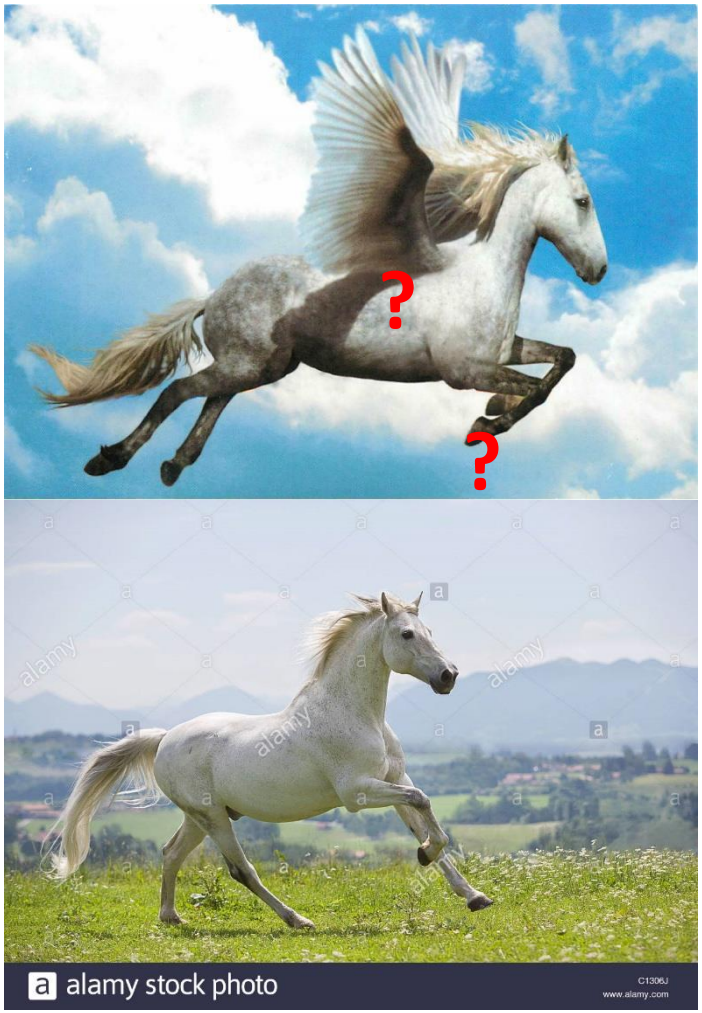


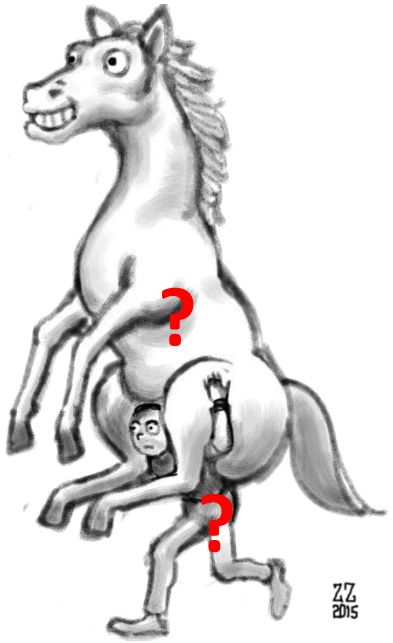
Structure

- Structure: the arrangement （布局） of and relations （关系）between the parts or elements of something complex.

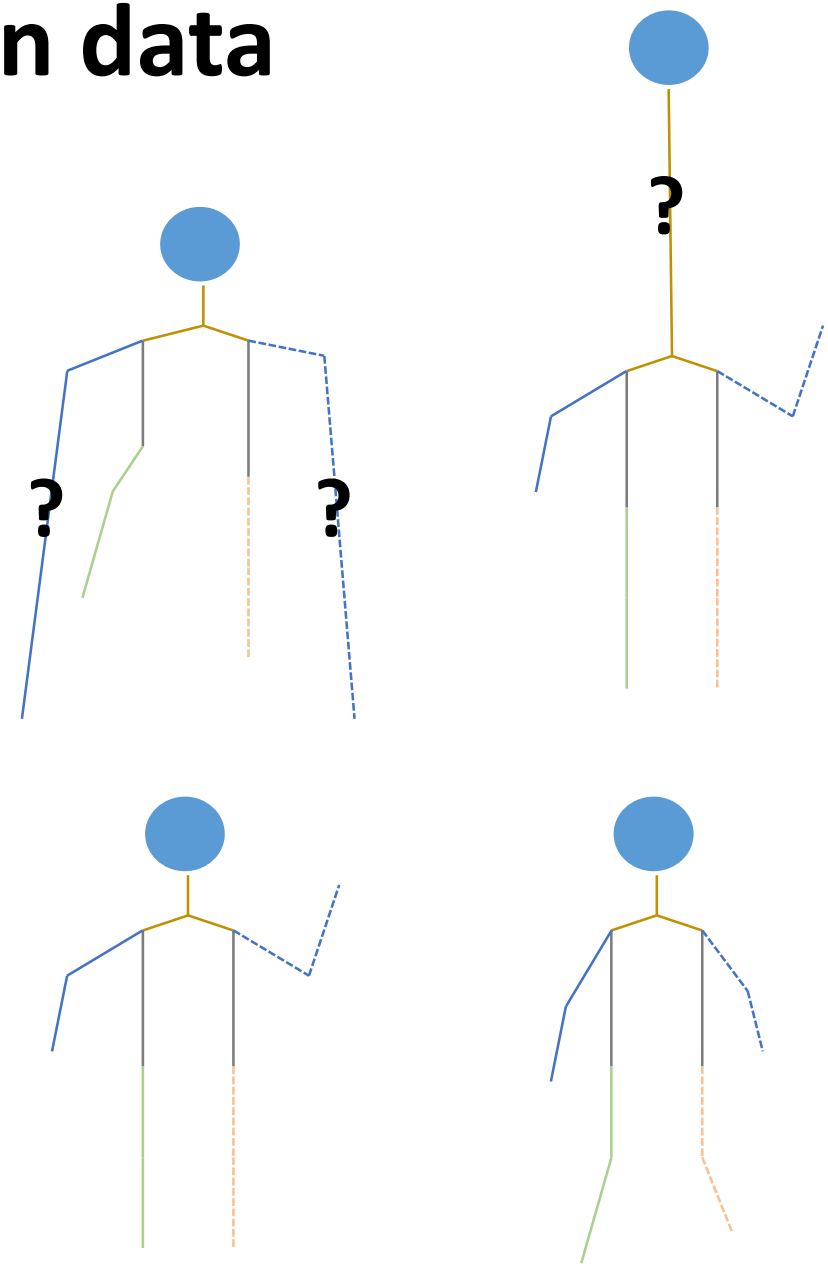
- Elements are correlated.

Structure **in data**

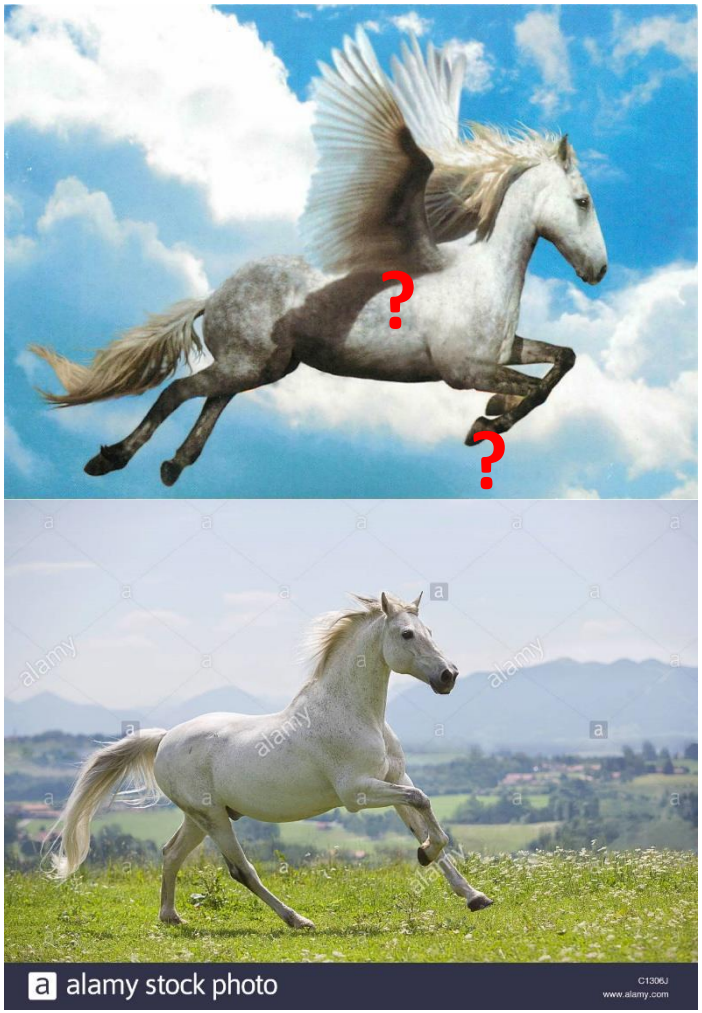


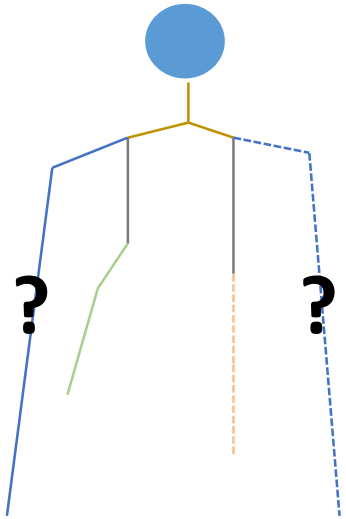


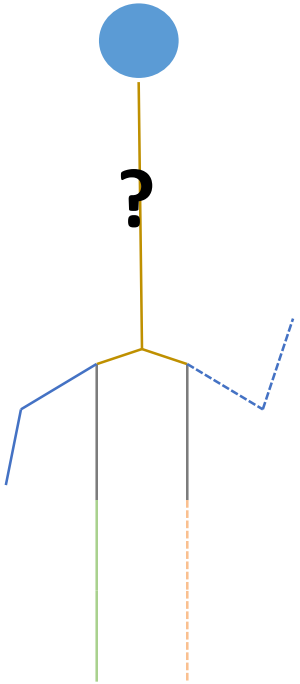
Structure **in data**

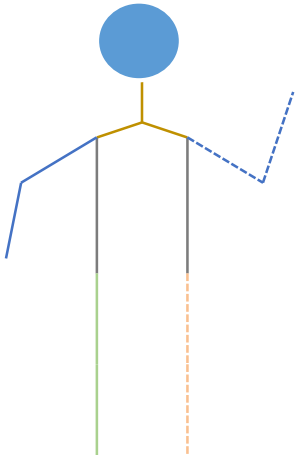


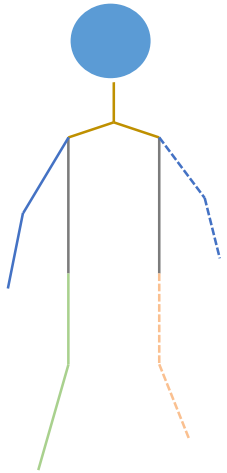
Structure **in data**



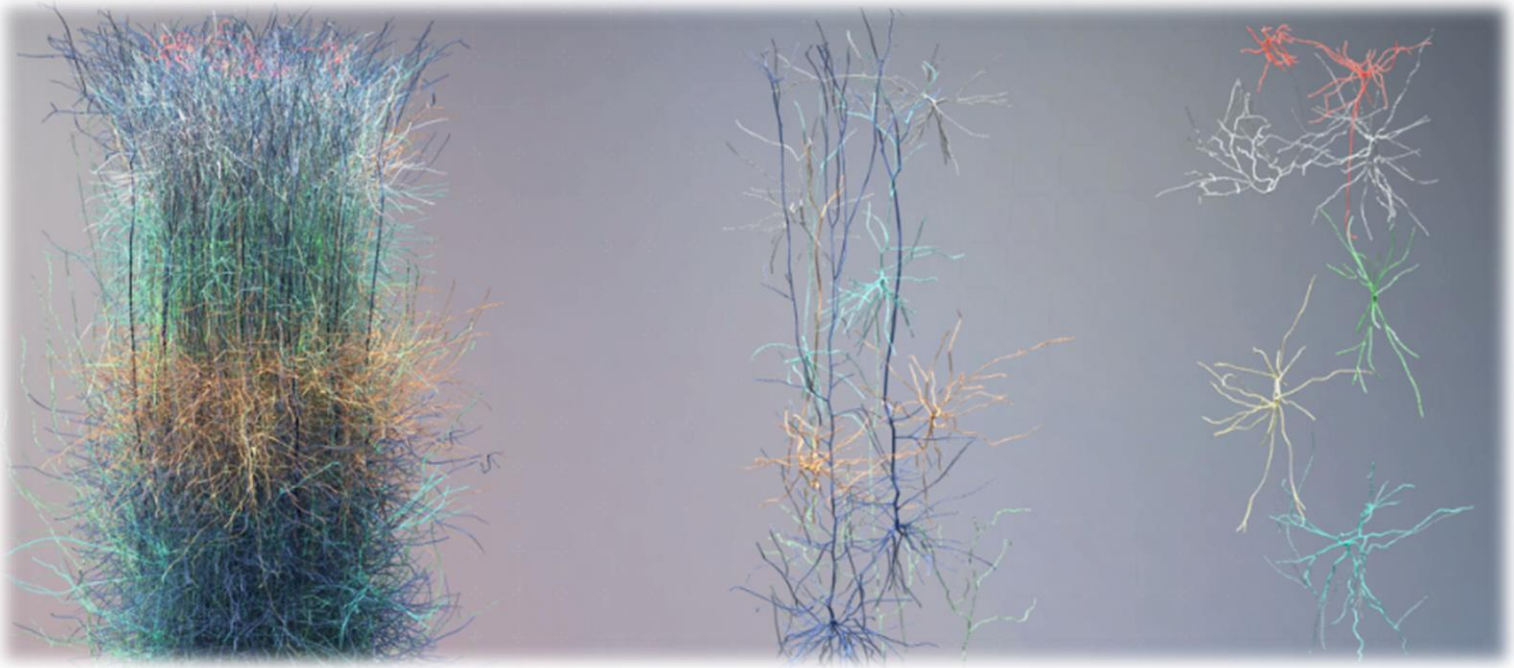








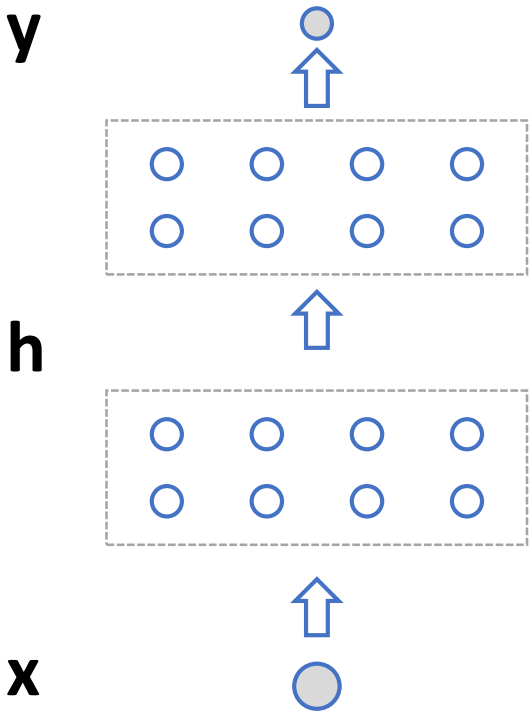
Structure in our brain



https://bgsmath.cat/meet-algebraic-topologist-helps-biologists-analyse-brain/

Structure in neurons

- Conventional neural networks
  - Only one type of correlation, neurons in adjacent layers. Neurons in the same layer have no connection

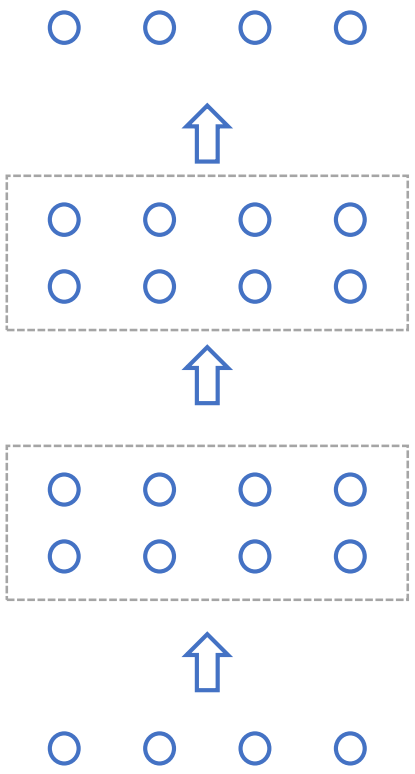


Structure exists in brain

Structure

- Structure: the arrangement （布局） of and relations （关系） between the parts or elements of something complex.

- Elements are correlated.
- For deep learning, we should learn the correlation between ?? so that they can refine each other.

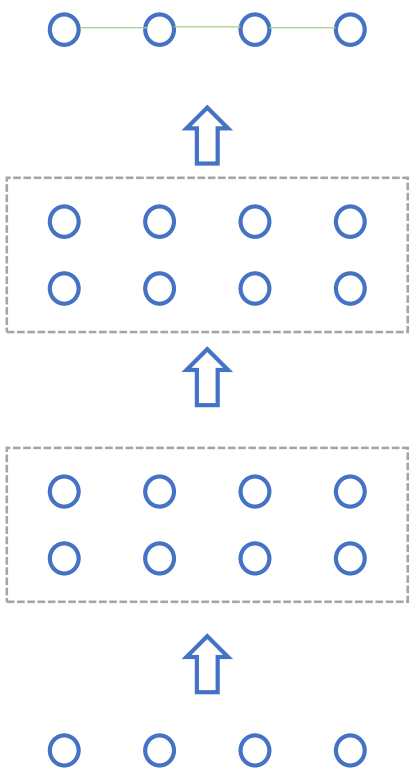


Vanilla Network

Structure

- Structure: the arrangement （布局） of and relations （关系） between the parts or elements of something complex.

- Elements are correlated.
- For deep learning, we should learn the correlation between ?? so that they can refine each other.
  - ??: Predicted output

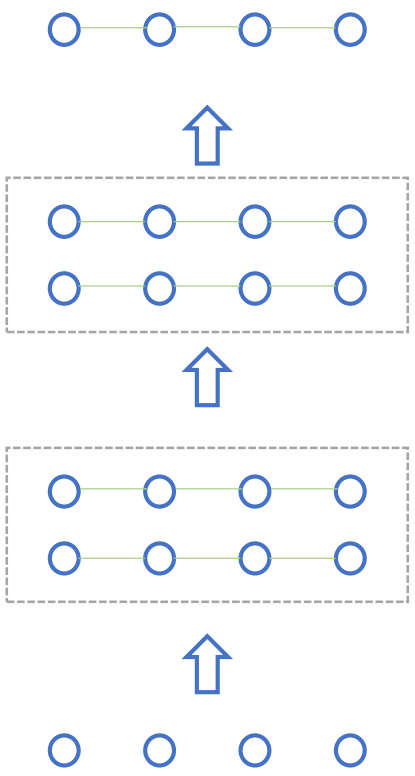


Structured output

Structure

- Structure: the arrangement （布局） of and relations （关系） between the parts or elements of something complex.

- Elements are correlated.
- For deep learning, we should learn the correlation between ?? so that they can refine each other.
  - ??: Predicted output
  - ??: Feature

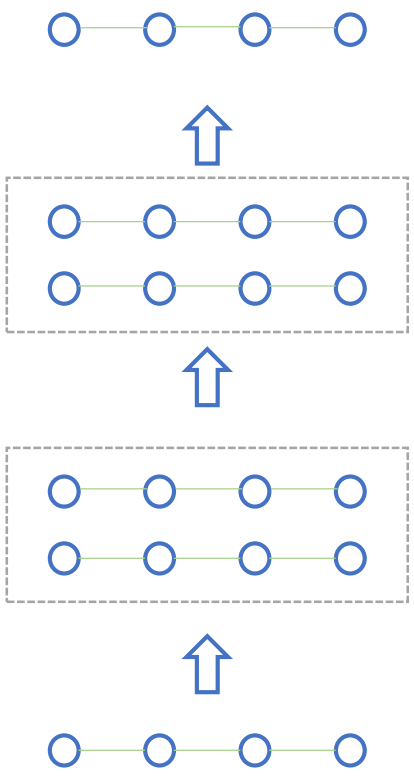


Structured feature

Structure

- Structure: the arrangement （布局） of and relations （关系） between the parts or elements of something complex.

- Elements are correlated.
- For deep learning, we should learn the correlation between ?? so that they can refine each other.
  - ??: Predicted output
  - ??: Feature
  - ??: Input



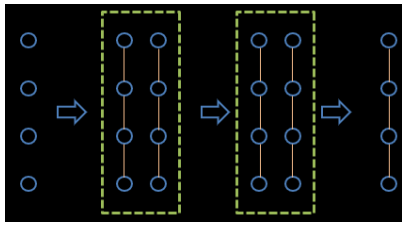
Structured input

Outline

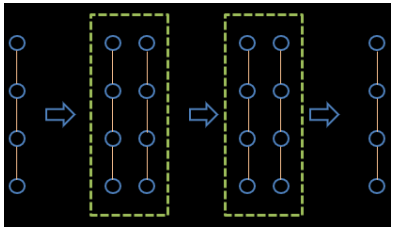
Introduction

Structured deep learning

Structured output and features



Structured input



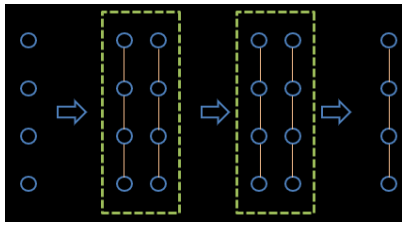
Conclusion

Outline

Introduction

Structured deep learning

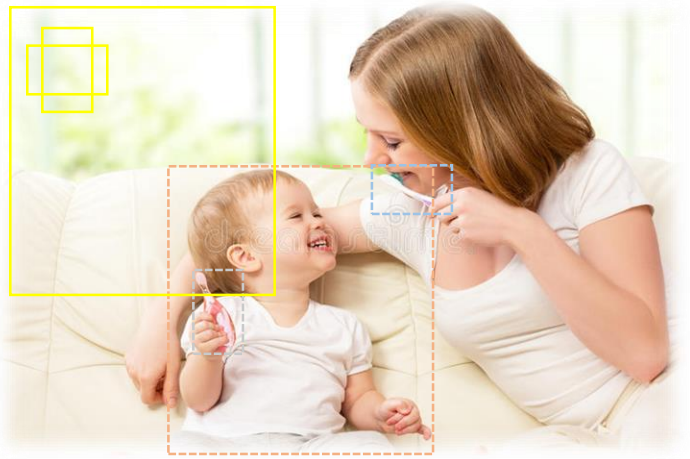
Structured output and features



Conclusion

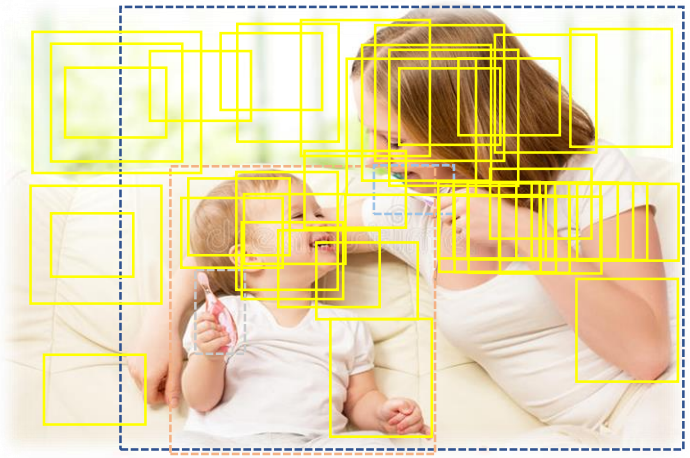
Object detection

- Sliding window
- Variable window size



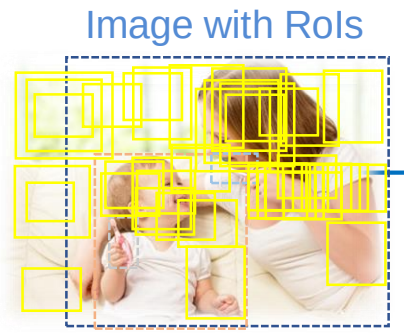
Motivation

- Much more negative samples than positive samples
- Easy to tell some regions do not contain any object



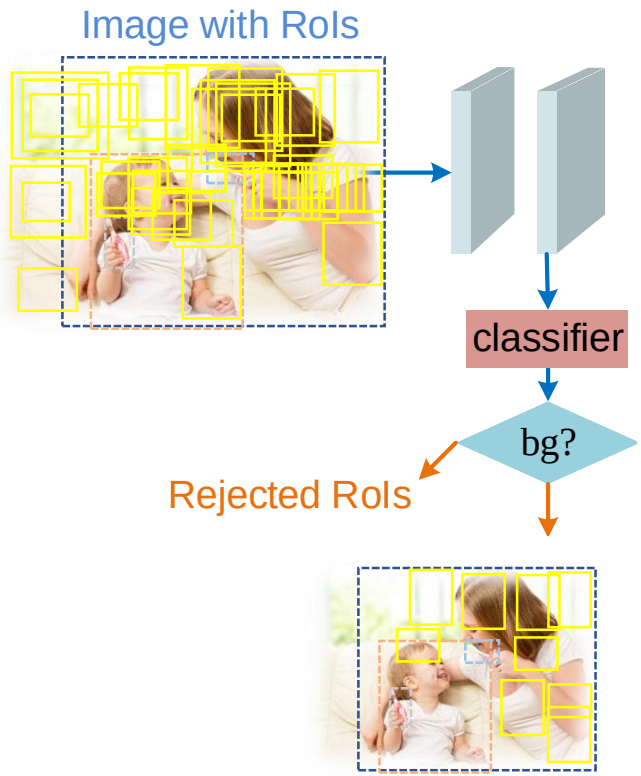
Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

Cascade Network



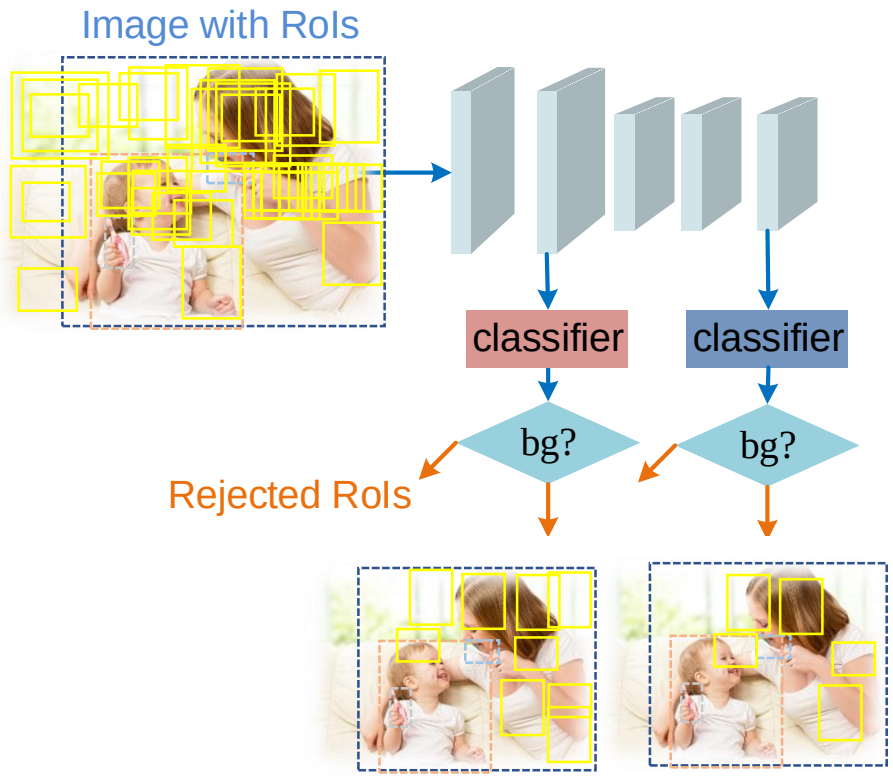
Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

Cascade Network



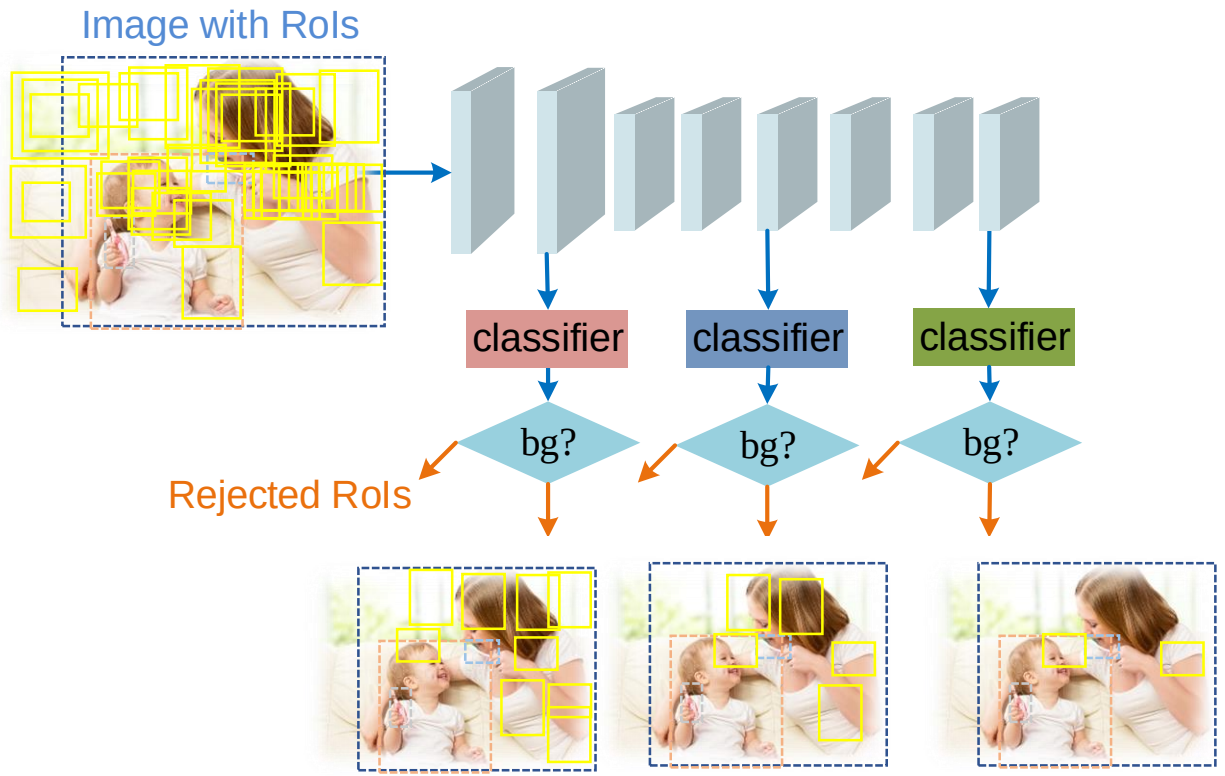
Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

Cascade Network



Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

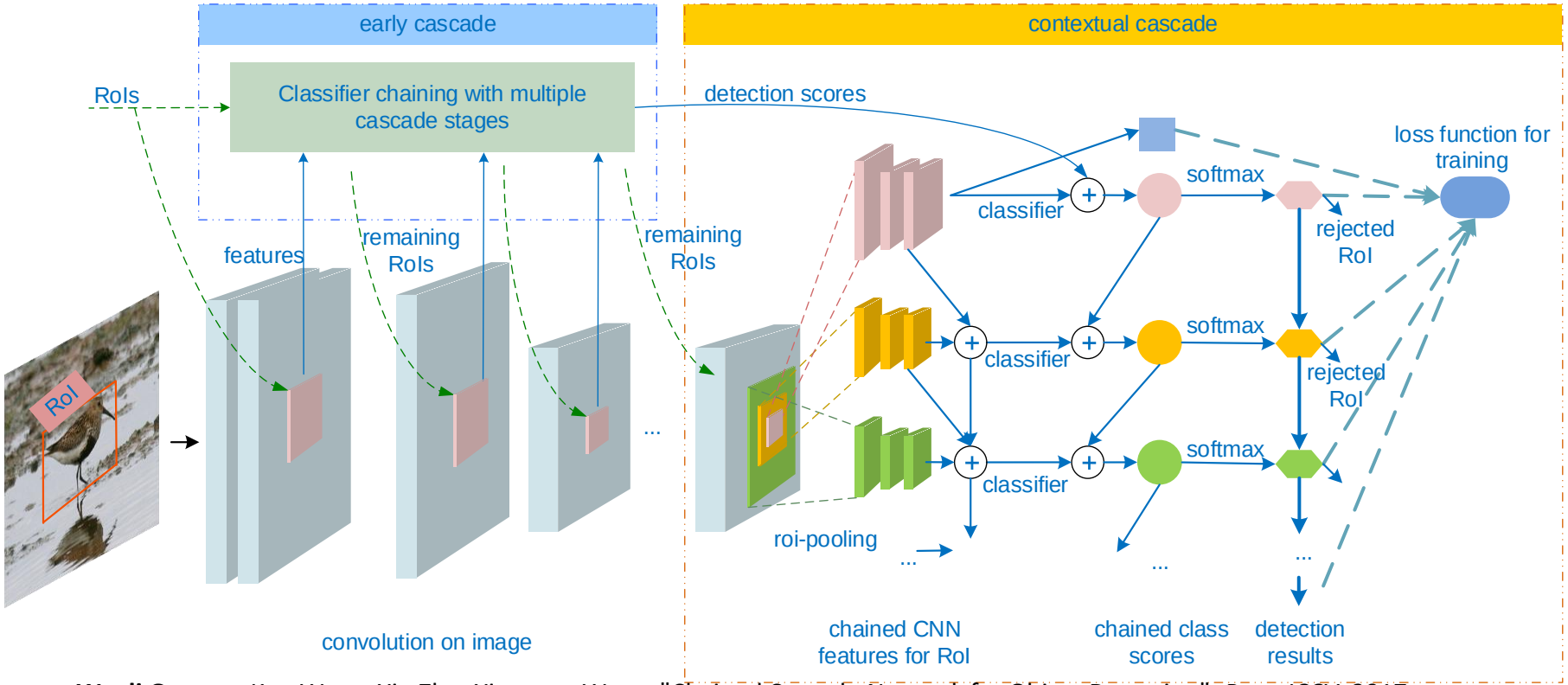
Cascade Network



Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

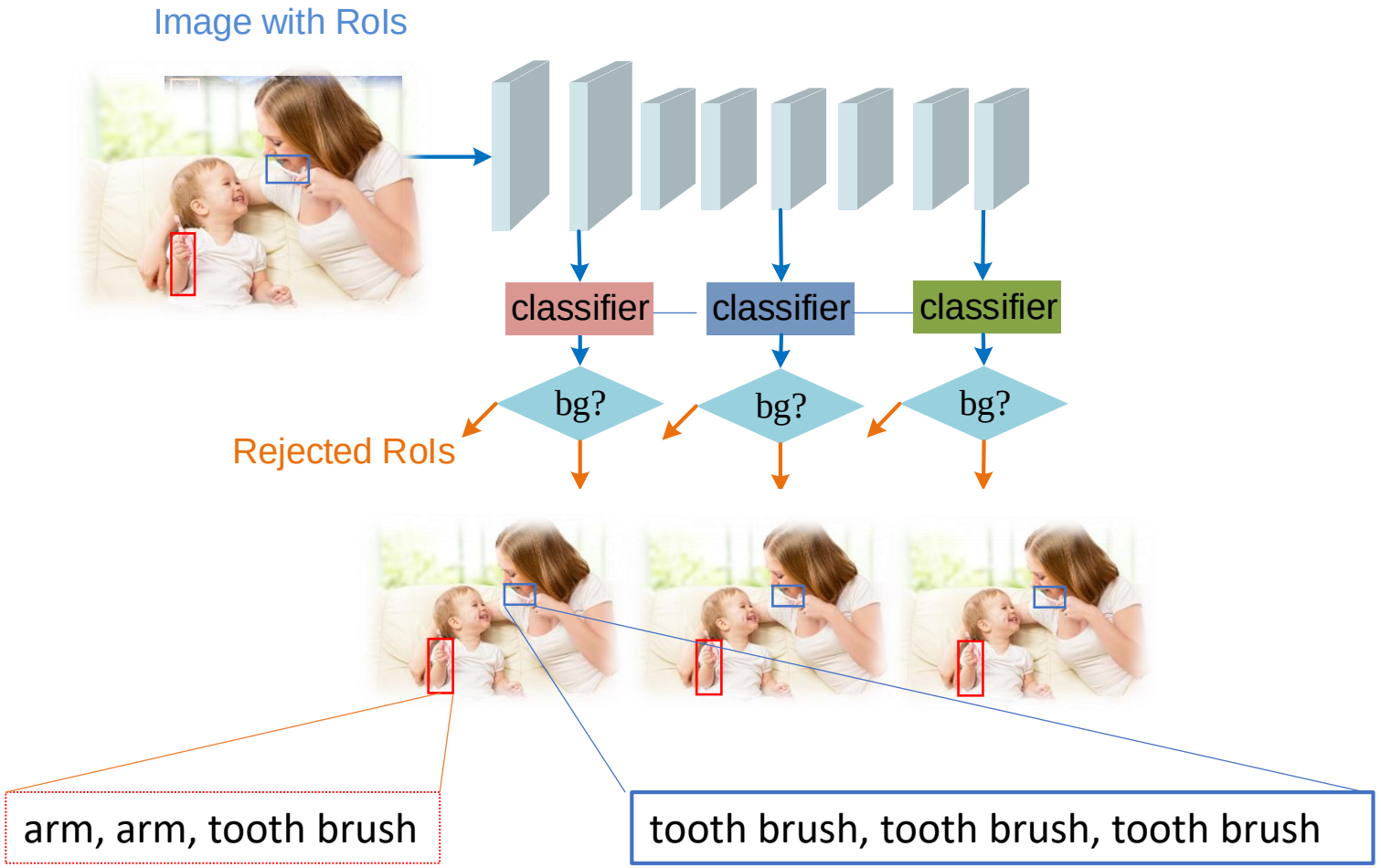
Model structures among classifiers at different stages

- Build up cascade at several stages in one network



Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

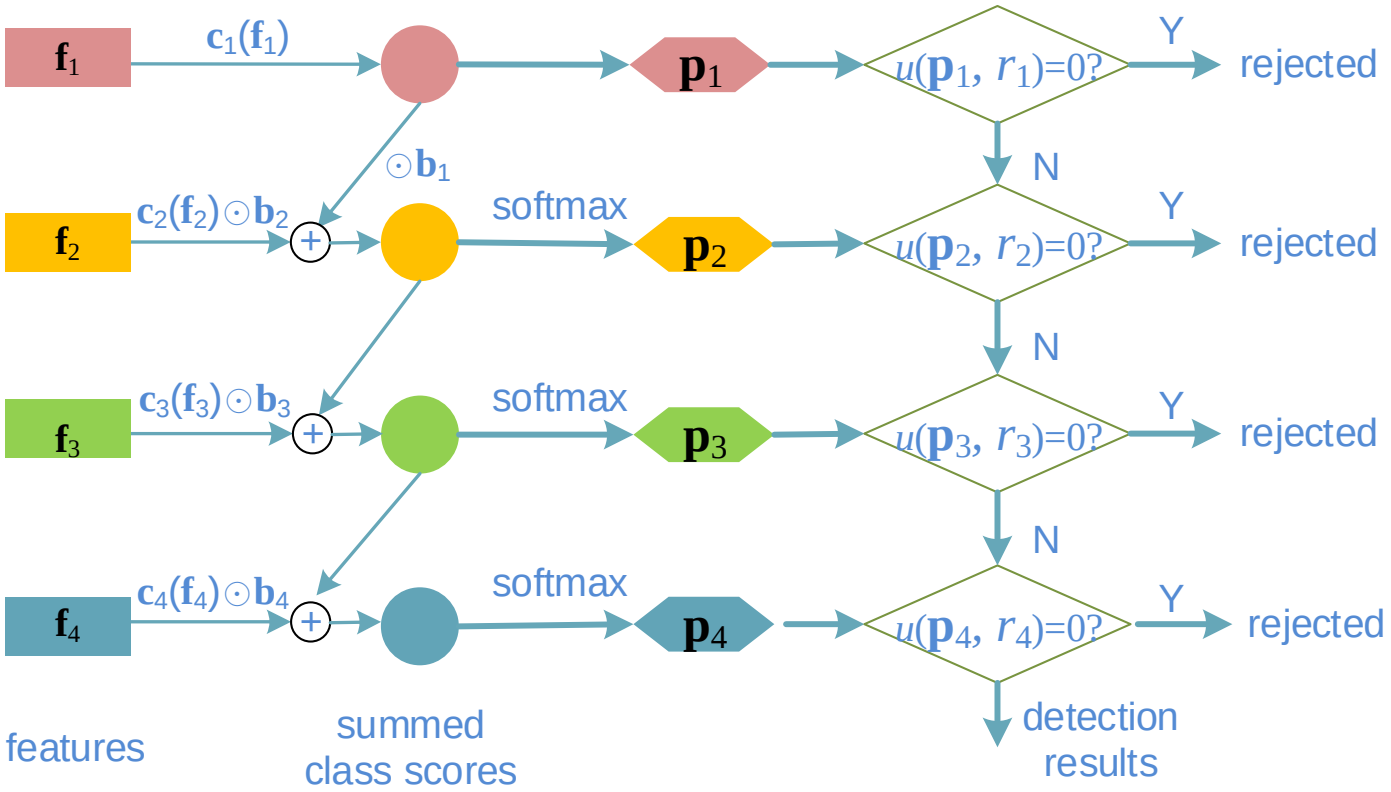
Model structures among classifiers at different stages



Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

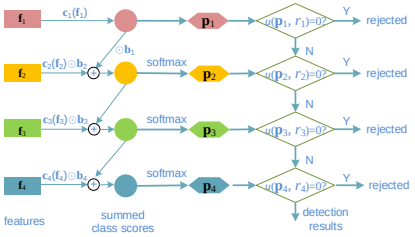
Model structures among classifiers at different stages with different context

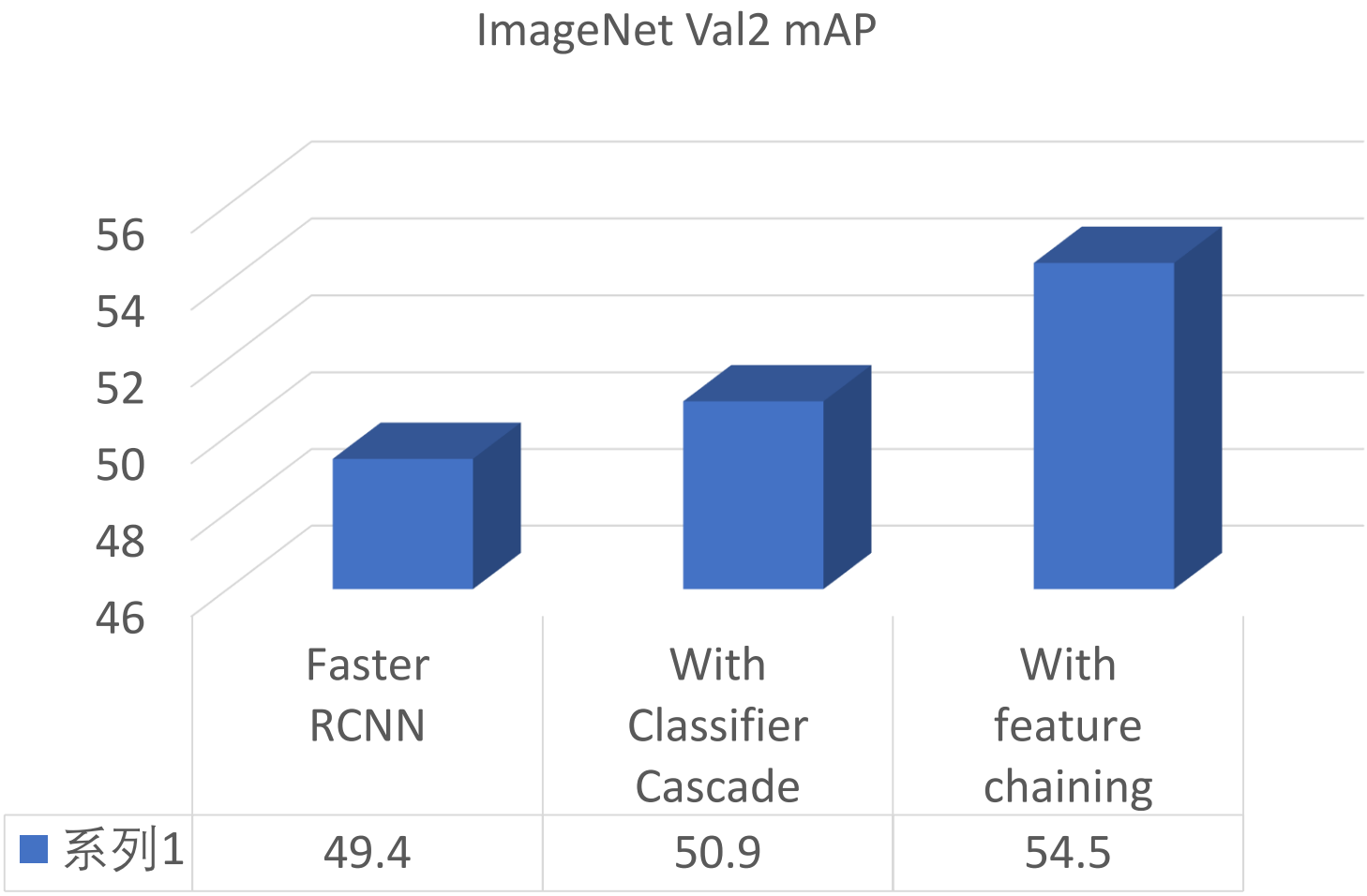
- Build up structure among classifiers $c_i(*)$ at different stages



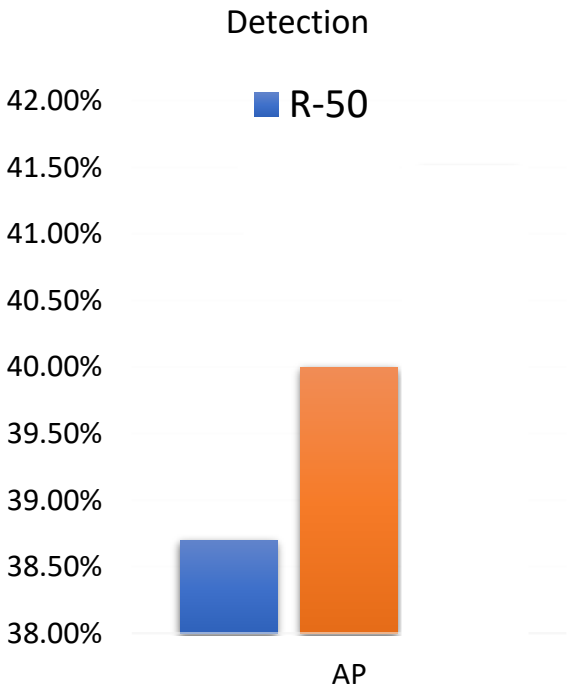
Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

Experimental results





Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

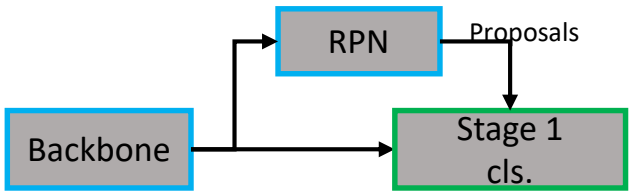


Code

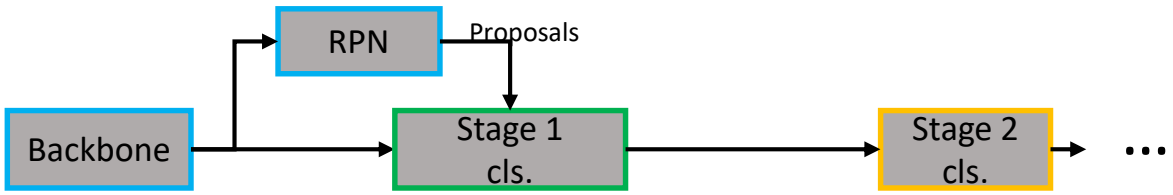
https://github.com/kevin-ssy/FishNet

Shuyang Sun, Jiangmiao Pang, Jianping Shi, Shuai Yi, **Wanli Ouyang** , "FishNet: A Versatile Backbone for Image, Region, and Pixel Level Prediction," *NurIPS. (Previously called NIPS)*, 2018.

Faster R-CNN

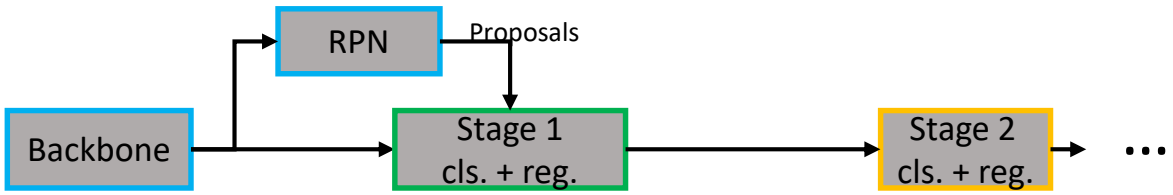


Chained Cascade



Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

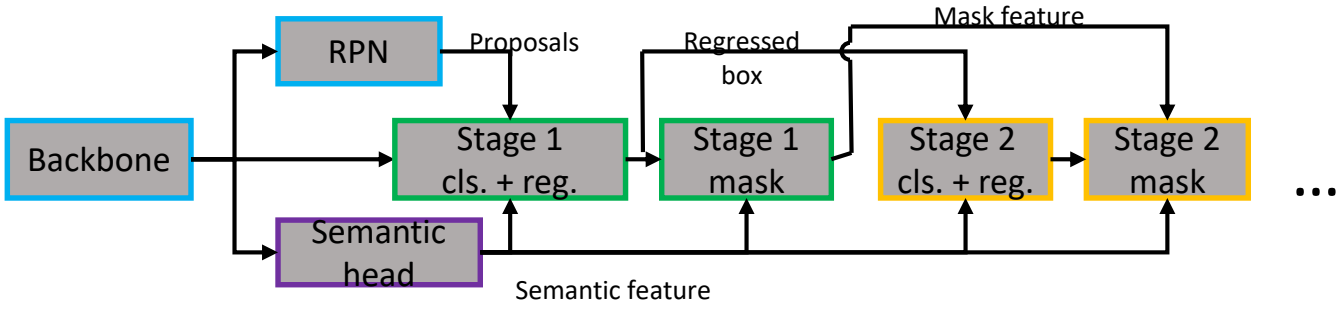
Cascade RCNN



Cai, Zhaowei, and Nuno Vasconcelos. "Cascade r-cnn: Delving into high quality object detection." *CVPR*. 2018.

Hybrid Task Cascade for Instance Segmentation

A hybrid architecture with interleaved task branching and cascade.



Chen, Kai, et al. "Hybrid task cascade for instance segmentation." *CVPR* 2019.

Codebase

- **Comprehensive**
  - ☑ RPN
  - ☑ Fast/Faster R-CNN
  - ☑ Mask R-CNN
  - ☑ FPN
  - ☑ Cascade R-CNN
  - ☑ RetinaNet
  - ☑ More … …

- **High performance**
  - ☑ Better performance
  - ☑ Optimized memory consumption
  - ☑ Faster speed

- **Handy to develop**
  - ☑ Written with PyTorch
  - ☑ Modular design

GitHub: mmdet

The entries ranking 1, 2, and 3 of iMaterialist (Fashion) 2019 at FGVC6 (CVPR 2019 Workshop) are based on HTC. Here is the post of the winner.

Codebase

Miras Amir
1st place

[Update] 1st place solution with code

posted in iMaterialist (Fashion) 2019 at FGVC6 24 days ago

95

Hi Kagglers,

My solution is based on the COCO challenge 2018 winners article: 07518.

Code:

https://github.com/amirassov/kaggle-imaterialist

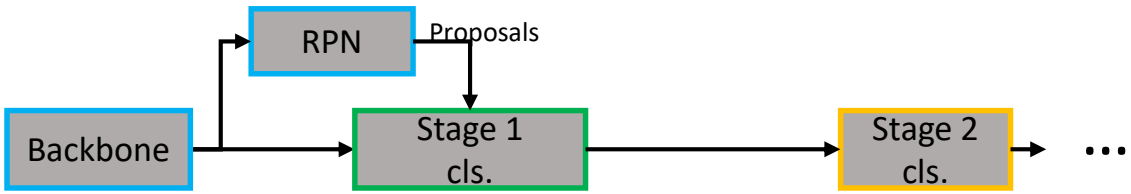
Model:

Hybrid Task Cascade with ResNeXt-101-64x4d-FPN backbone. This model has a metric Mask mAP = 43.9 on COCO dataset. This is SOTA for instance segmentation.

GitHub: mmdet

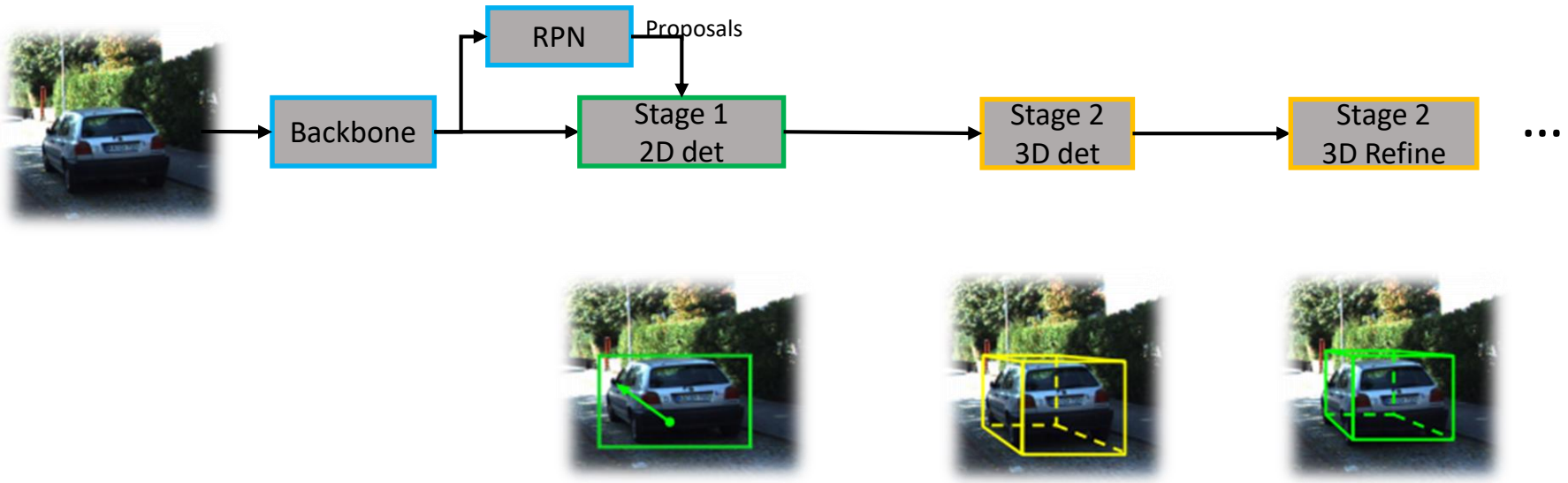
The entries ranking 1, 2, and 3 of iMaterialist (Fashion) 2019 at FGVC6 (CVPR 2019 Workshop) are based on HTC. Here is the post of the winner.

Chained Cascade



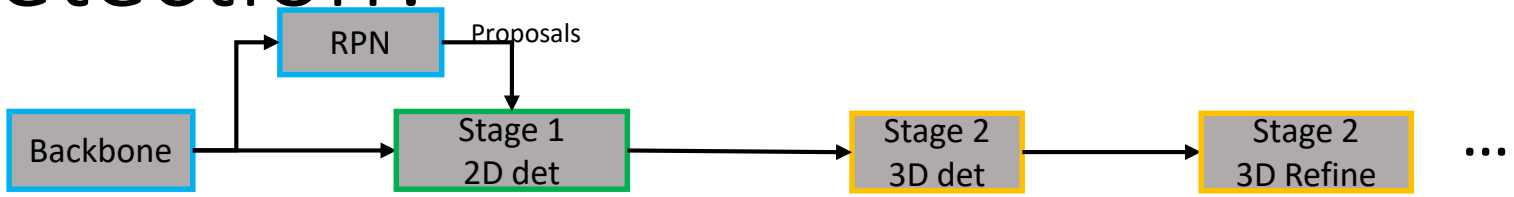
Wanli Ouyang, Kun Wang, Xin Zhu, Xiaogang Wang. "Chained Cascade Network for Object Detection", *Proc. ICCV*, 2017.

GS3D

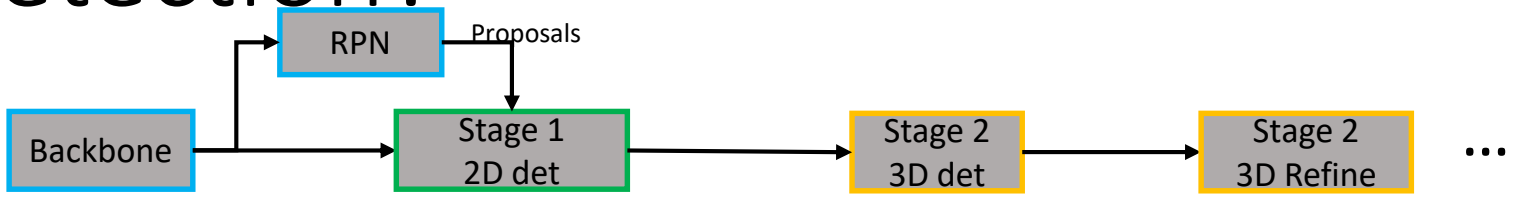


Buyu Li, **Wanli Ouyang**, Lu Sheng, et. al. "GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving", CVPR 2019

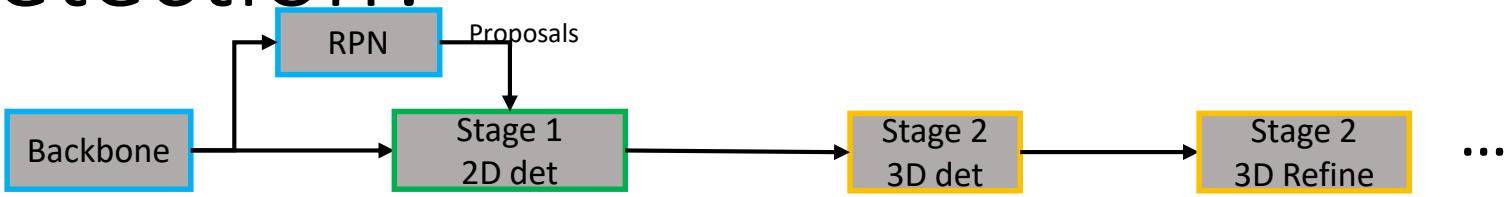
Why do we need multi-stage 3D detection?

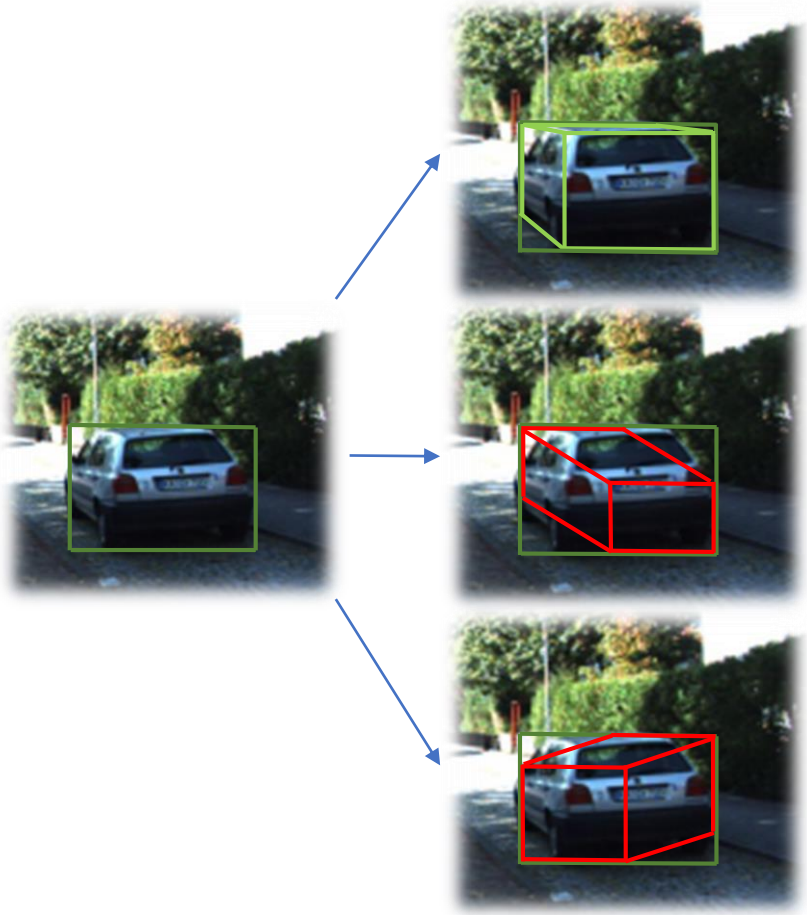


Why do we need multi-stage 3D detection?



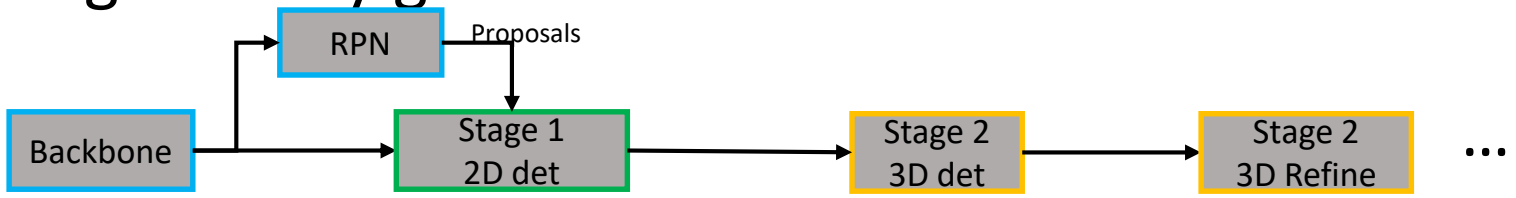
Why do we need multi-stage 3D detection?

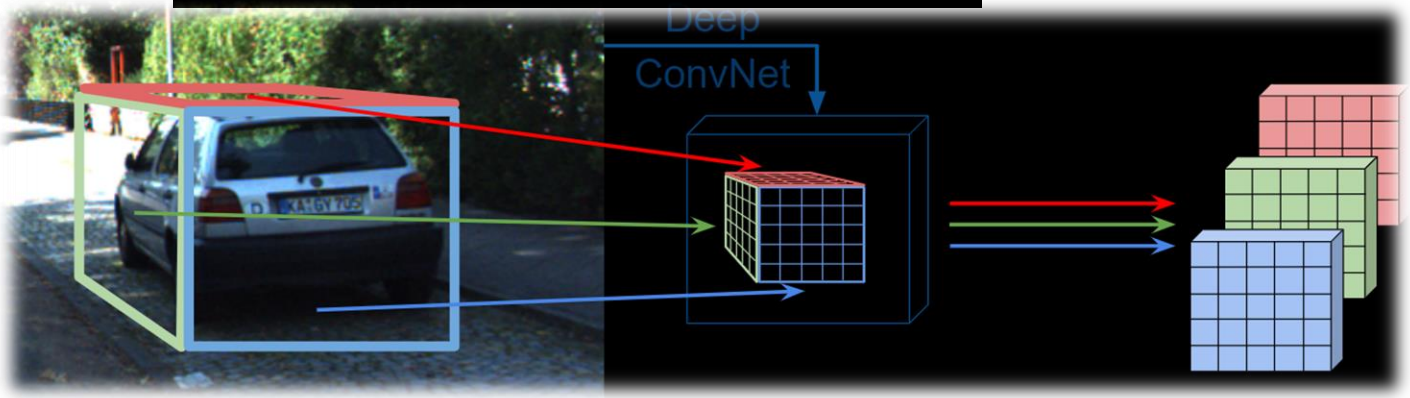




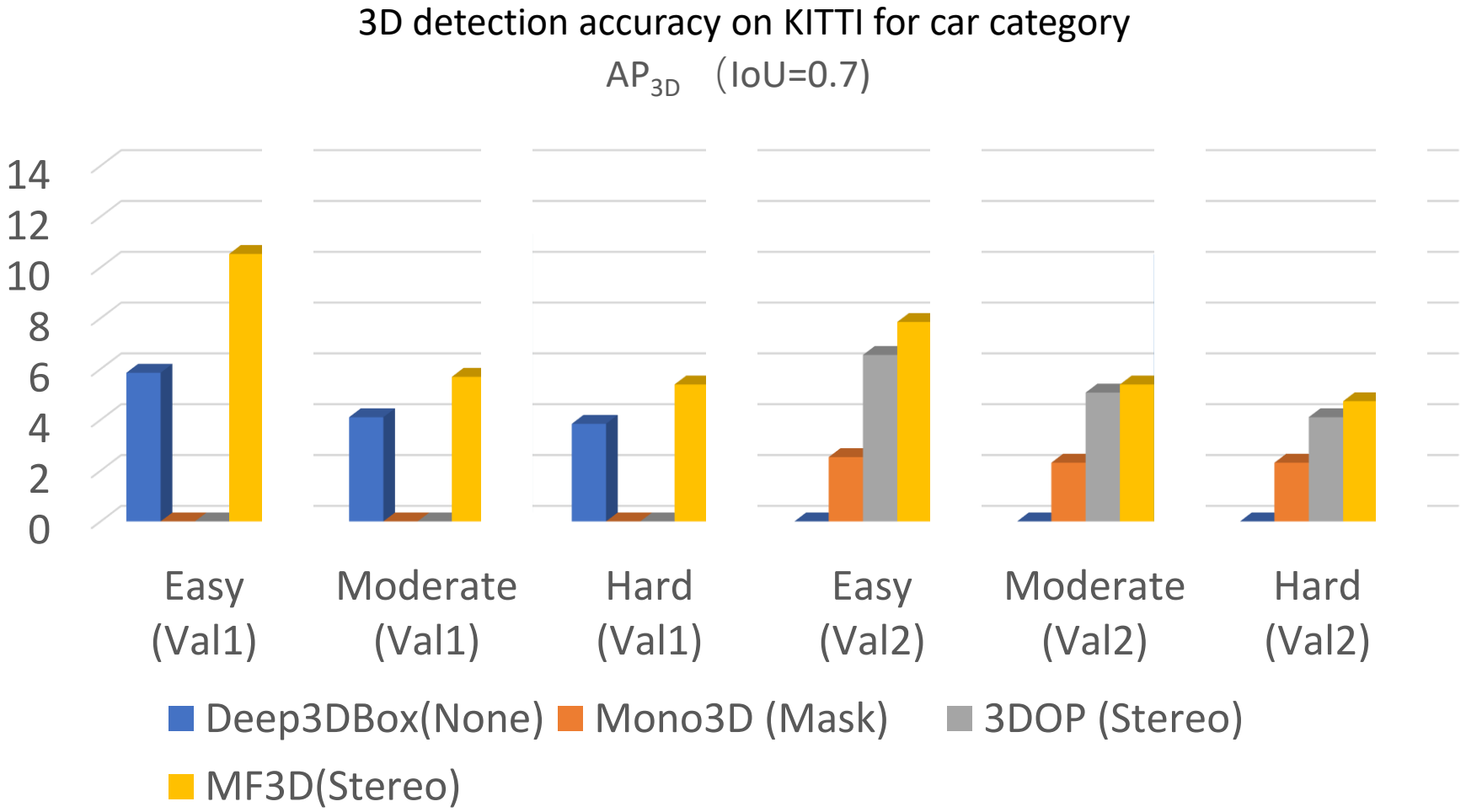
Surface feature extraction

- 3D geometry guided feature extraction





Experimental results

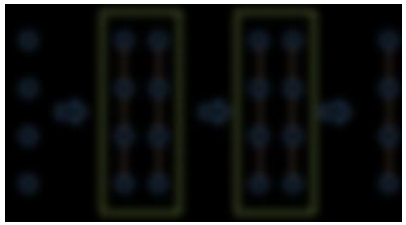


Buyu Li, **Wanli Ouyang**, Lu Sheng, et. al. "GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving", CVPR 2019

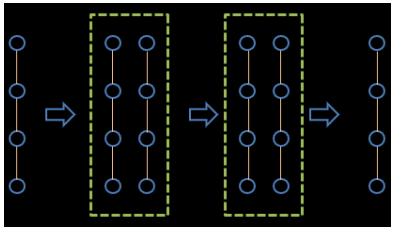
Outline

Introduction

Structured deep learning



Structured input



Conclusion

3D object detection

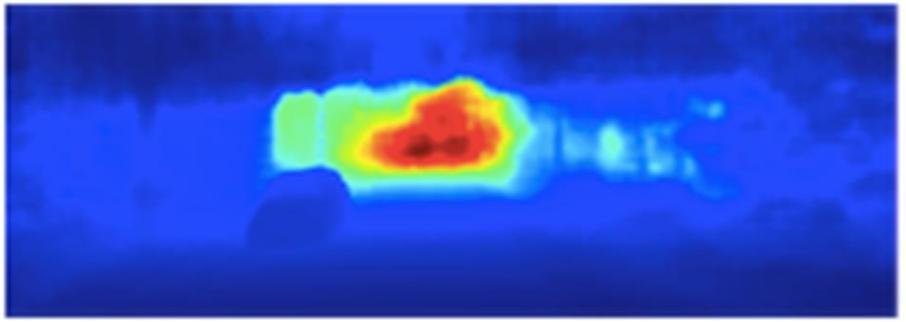
Image

Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

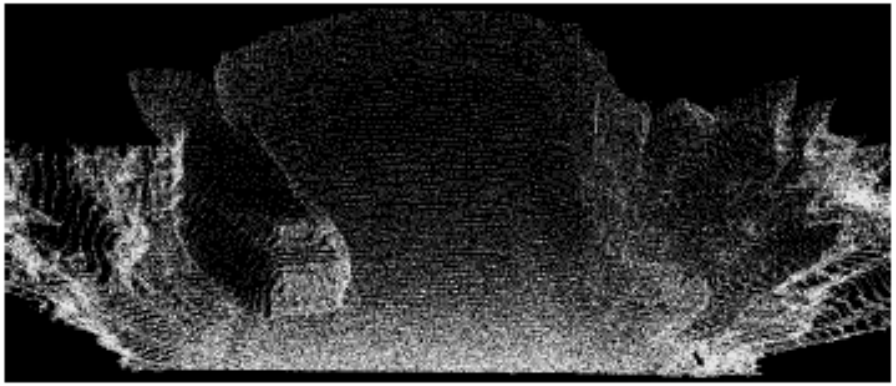
3D object detection

Image

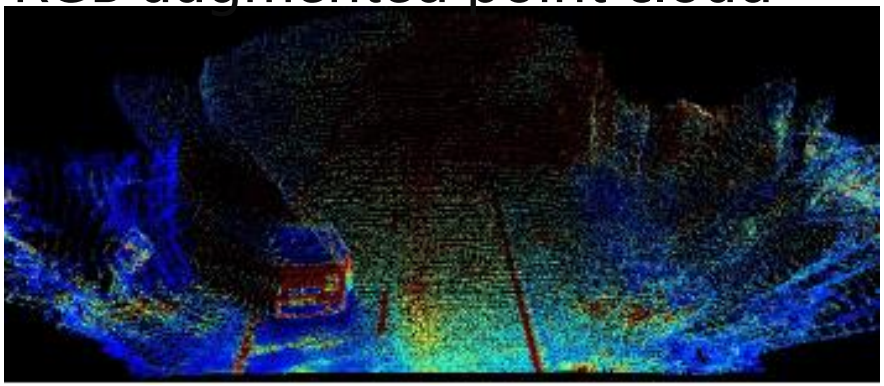
Depth



Point cloud



RGB augmented point cloud

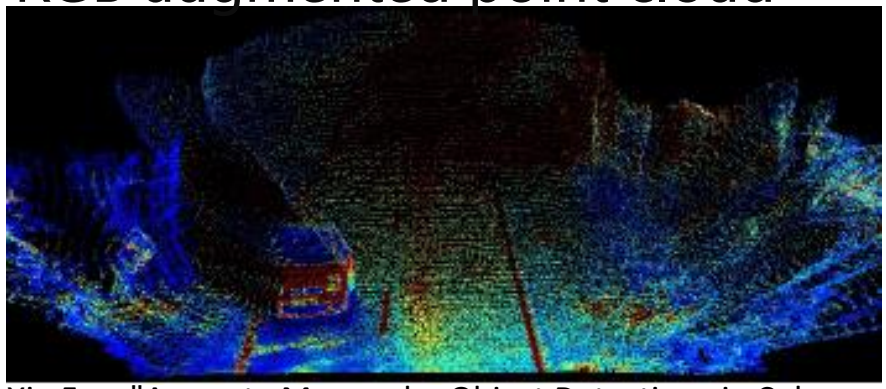


Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

3D object detection

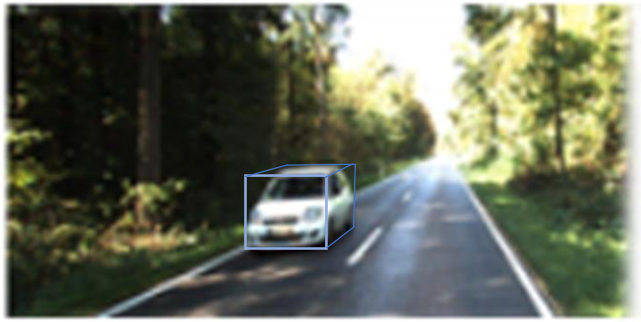
Image

RGB augmented point cloud



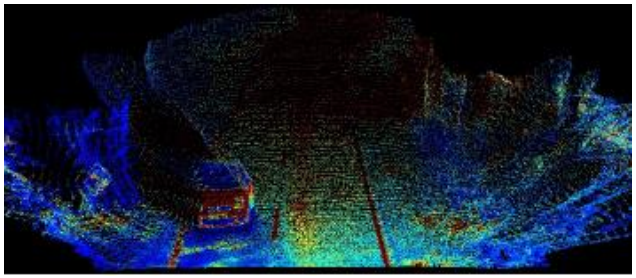
Network

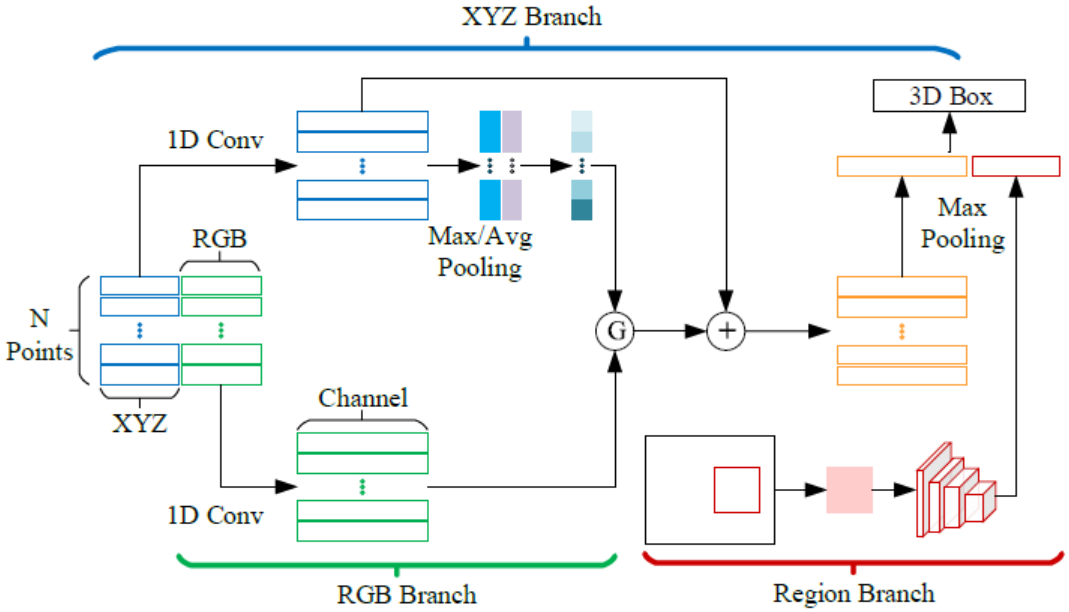
3D results



Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

- 3D box estimation (Det-Net) with RGB features fusion





Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

3D object detection

- 3D **detection** performance Average Precision (AP_{3D})

Method	Data	IoU=0.5			IoU=0.7		
		Easy	Moderate	Hard	Easy	Moderate	Hard
Mono3D [3]	Mono	25.19	18.20	15.52	2.53	2.31	2.31
Deep3DBox [21]	Mono	27.04	20.55	15.88	5.85	4.10	3.84
Multi-Fusion [30]	Mono	47.88	29.48	26.44	10.53	5.69	5.39
ROI-10D [18]	Mono	-	-	-	10.25	6.39	6.18
MonoGRNet [25]	Mono	50.51	36.97	30.82	13.88	10.19	7.62
Ours	Mono	**68.86**	**49.19**	**42.24**	**32.23**	**21.09**	**17.26**

Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

3D object detection

- 3D **localization** performance: Average Precision (AP_{loc})

Method	Data	IoU=0.5			IoU=0.7		
		Easy	Moderate	Hard	Easy	Moderate	Hard
Mono3D [3]	Mono	30.50	22.39	19.16	5.22	5.19	4.13
Deep3DBox [21]	Mono	30.02	23.77	18.83	9.99	7.71	5.30
Multi-Fusion [30]	Mono	55.02	36.73	31.27	22.03	13.63	11.60
ROI-10D [18]	Mono	-	-	-	14.76	9.55	7.57
Ours	Mono	**72.64**	**51.82**	**44.21**	**43.75**	**28.39**	**23.87**

Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, W. Ouyang, Xin Fan. "Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving", Proc. ICCV, 2019.

Is structured learning only effective for object detection?

Application of structured feature learning

- Haze removal (ICCV19)
- Depth estimation (TPAMI 18)
- Contour estimation (NIPS 17)
- Detection (TPAMI17, TPAMI18, …)
- Human pose estimation (CVPR16)
- Person re-identification (CVPR18)
- Relationship estimation (ICCV17)
- Image captioning (ICCV17)

Low-level vision

High-level vision

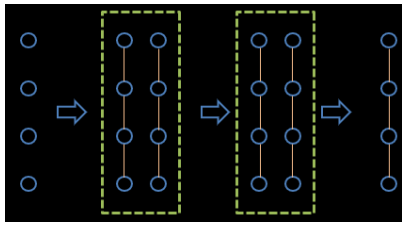
Vision + Language

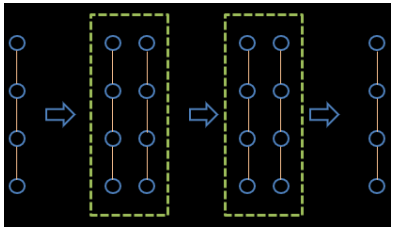
D. Xu, *et al.*, "Monocular Depth Estimation using Multi-Scale Continuous CRFs as Sequential Deep Networks," *TPAMI* 2018.
W. Ouyang, *et al.*, " Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection," *TPAMI* 2018.
W. Ouyang, *et. al*. "DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks", *TPAMI* 2017.
X. Chu, **W. Ouyang**, *et. al*. "Structured feature learning for pose estimation". *CVPR* 2016.
Y. Li, **W. Ouyang**, *et. al*. "Scene Graph Generation from Objects, Phrases and Region Captions", *ICCV*, 2017.

Outline

Introduction

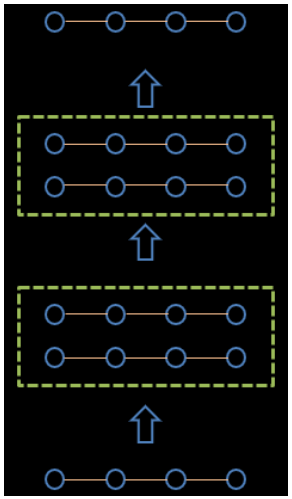
Structured deep learning





Conclusion

Take home message



- Cascade Network
  - Structured output and feature
  - Faster inference
  - Cascade enables the network to handle more and more difficult examples
  - Classifiers and features collaborate by structure modelling
  - Can be extended to instance segmentation and 3D detection
- Color-Embedded 3D Reconstruction for 3D detection
  - Structured input
  - Connect RGB, depth, and point cloud by color augmented point cloud, a better representation

Thank you!