

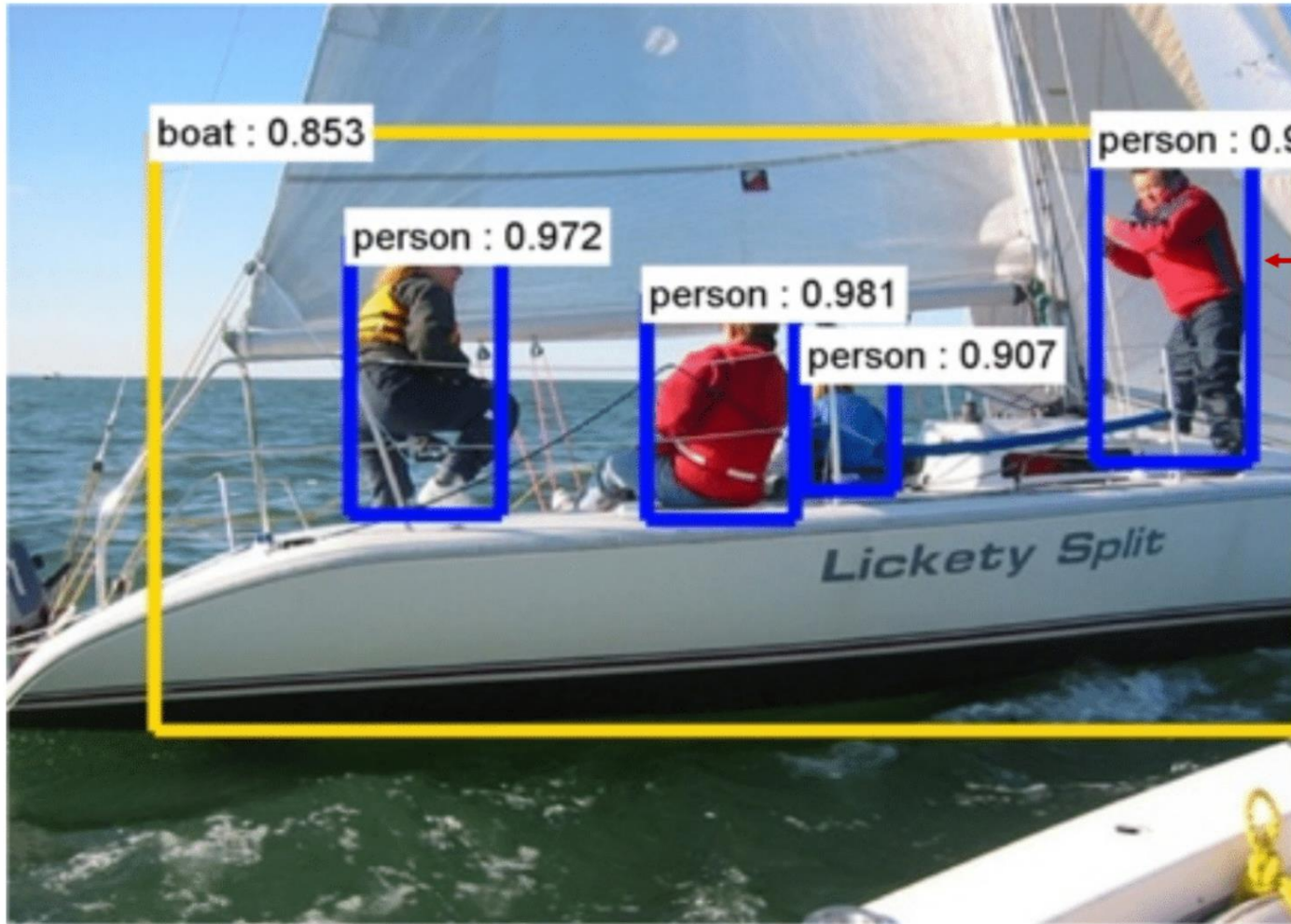
# Recent Progress in Object Detection

Jifeng Dai

Visual Computing Group

Microsoft Research Asia

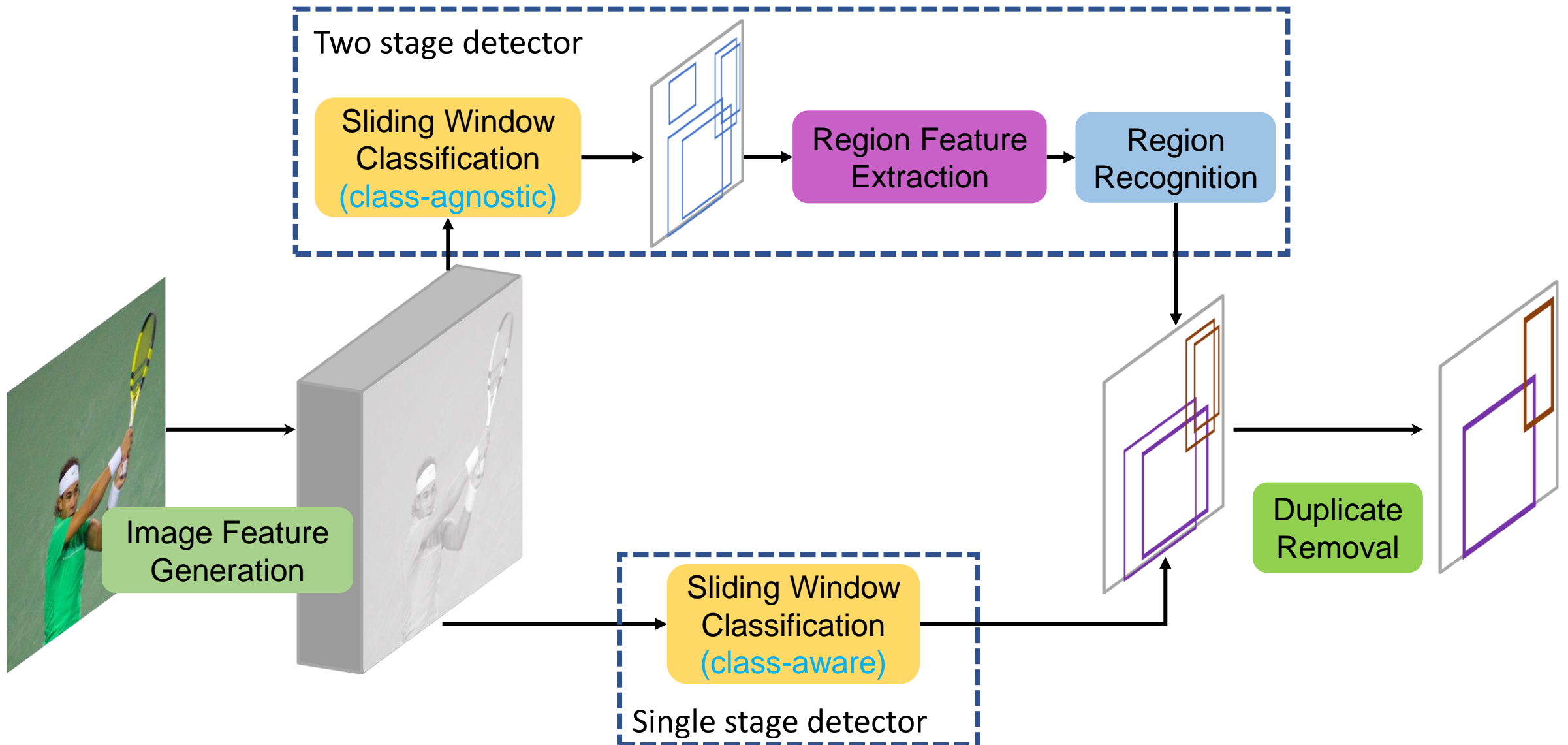
# Problem Definition



← What?

← Where?

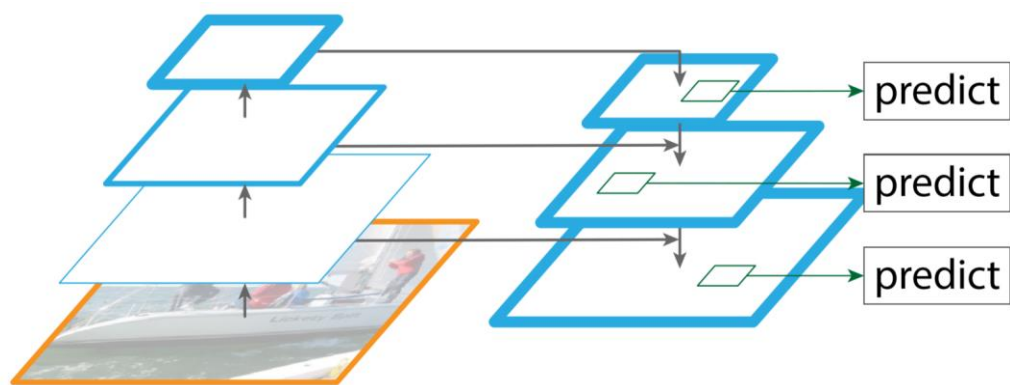
# A General Pipeline



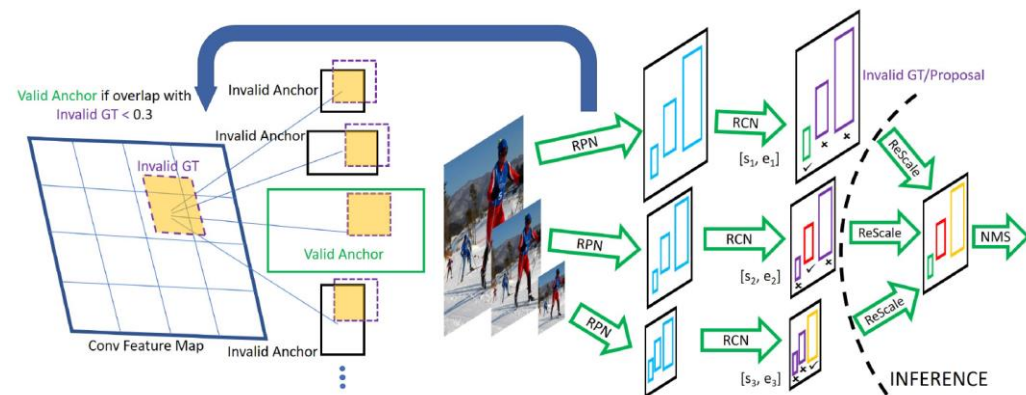
Main Progress since CVPR 2017

# Image Feature Extractor (1/2)

- Multiple scale representation
  - From single-scale image (inspired by U-Net [Ronneberger et al, 2015])
    - Feature Pyramid Networks (FPN) [Lin et al, CVPR'2017]
    - Deconvolution Single Shot Detection (DSSD) [Fu et al, 2017]
    - Top-down Module (TDM) [Shrivastava et al, CVPR'2017]
  - From multi-scale image pyramid
    - Scale Normalization for Image Pyramids (SNIP) [Singh and Davis, CVPR'2018]



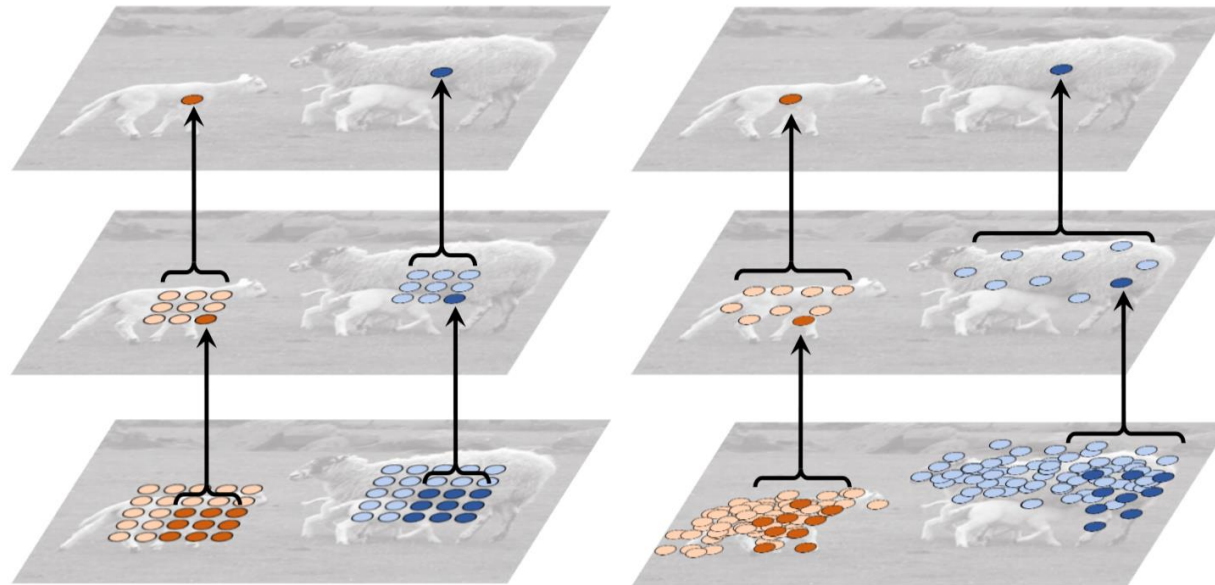
Feature Pyramid Networks [Lin et al. CVPR'2017]



Scale Normalization for Image Pyramids [Lin et al. CVPR'2017]

# Image Feature Extractor (2/2)

- Spatial deformation modeling
  - Deformable Convolutional Networks [Dai et al, ICCV'2017]



(a) standard convolution

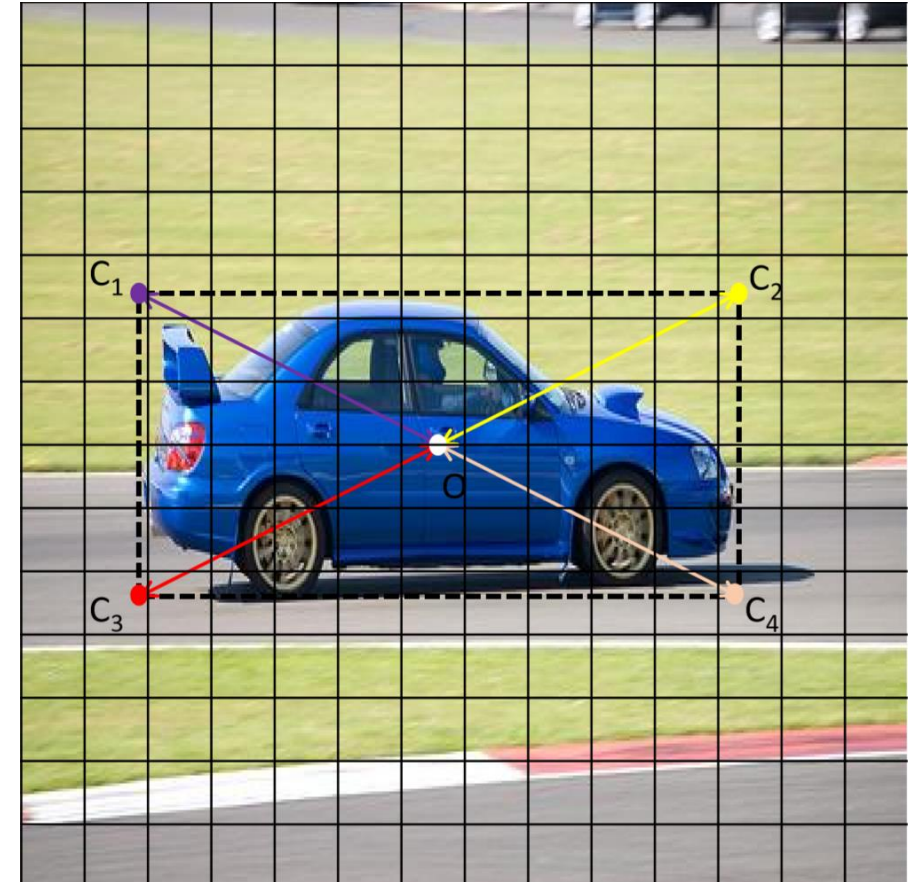
(b) deformable convolution

Stacked Deformable Convolution [Dai et al, ICCV'2017]



# Sliding Window Classification

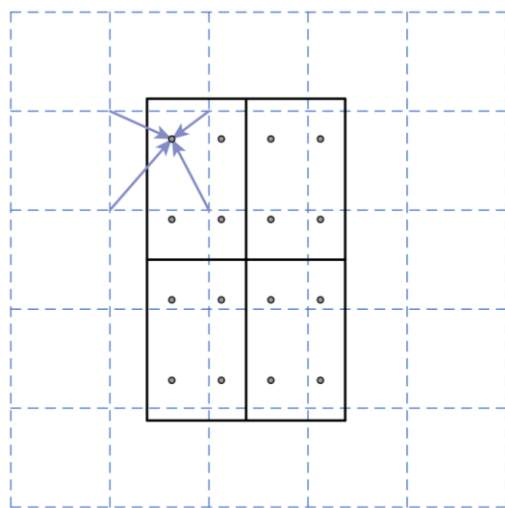
- Anchor based
  - Focal Loss [Lin et al, ICCV'2017]
    - Denser anchors
      - 12 (Faster R-CNN) -> 18 (FPN) -> 54 (Focal Loss)
    - Sparse Sampling -> Dense Sampling
    - Training: biased initialization, focal loss
- Point based
  - Point Linking [Wang et al, 2017]
  - DeNet [Tychsen-Smith et al, ICCV'2017]



Point Linking [Wang et al, 2017]

# Region Feature Extractor (1/2)

- Bilinear feature interpolation
  - Subpixel precision, gradient w.r.t. bin offset
    - Aligned RoI pooling [He et al, ICCV'2017]
    - Deformable RoI pooling [Dai et al, ICCV'2017]



Aligned RoI pooling [He et al, ICCV'2017]



Deformable RoI pooling [Dai et al, ICCV'2017]



# Region Feature Extractor (2/2)

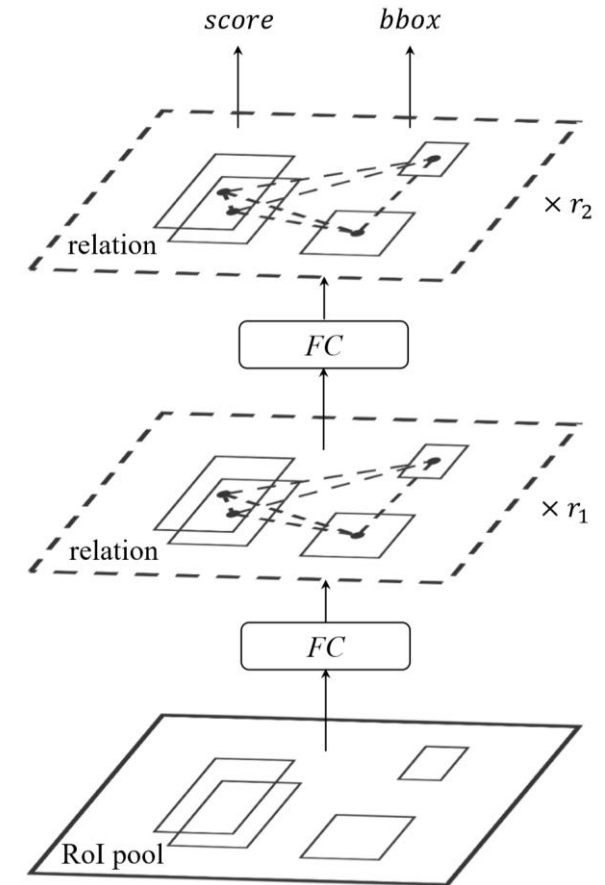
- Spatially adaptive bins
  - Deformable RoI pooling [Dai et al, ICCV'2017]
  - Fully learnable region feature extraction [Gu et al, 2018]



Fully learnable region feature extraction [Gu et al, 2018]

# Region Recognition

- Light-weight detection heads
  - Light-head R-CNN [Li et al, 2017]
- Modeling interaction among instances
  - Relation Networks [Hu et al, CVPR'2018]
  - Iterative Visual Reasoning [Chen et al, CVPR'2018]
- Multi-stage region recognition
  - Cascade R-CNN [Cai and Vasconcelos, CVPR'2018]
  - Chained Cascade Network [Ouyang et al, ICCV'2017]



Relation Networks [Hu et al, CVPR'2018]

# Duplicate Removal

- Towards learnable duplicate removal
  - Soft-NMS [Bodla et al, ICCV'2017]
  - Fitness-NMS [Smith et al, CVPR'2018]
  - Learning NMS [Hosang et al, ICCV'2017]
  - Relation Networks [Hu et al, CVPR'2018]
    - The first practical duplicate removal learner
    - Give rise to a fully end-to-end object detector

# Other Aspects

- More classes
  - YOLOv2 (9000 classes) [Redmon and Farhadi, CVPR'2017]
  - R-FCN-3000 (3000 classes) [Singh et al, CVPR'2018]
- Training
  - Enable BN in training detectors [Peng et al, CVPR'2018, Wu and He, 2018]
  - ImageNet pre-training free [Shen et al, ICCV'2017, Li et al, 2018]

# We are hiring!

- Our team targets at fundamental research in computer vision
  - R-FCN, Deformable ConvNets
  - Winning COCO challenge 2015, 2016
- Openings:
  - Short-term intern
  - Joint Ph.D. program
  - Full-time researcher

Email: Jifeng Dai <jifdai@Microsoft.com>