CornerNet: Detecting Objects as Paired Keypoints

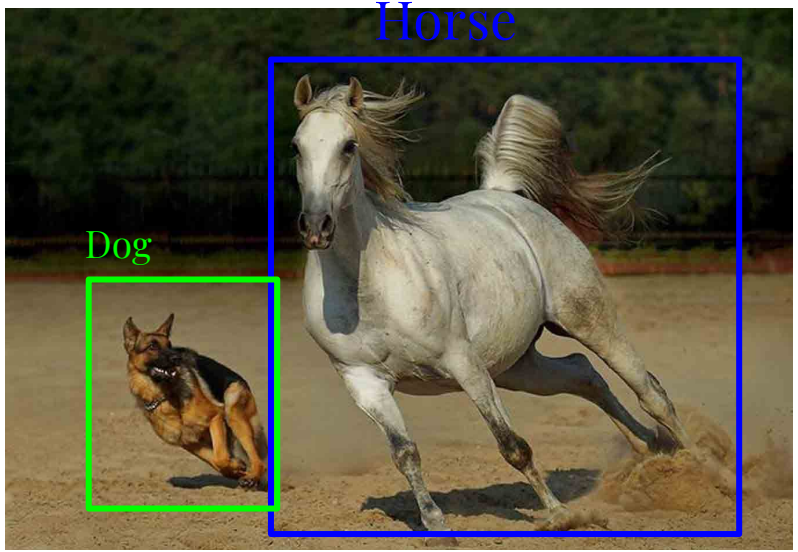
Hei Law · Jia Deng [ECCV 2018]

A new method for One Stage Object Detection

Object Detection

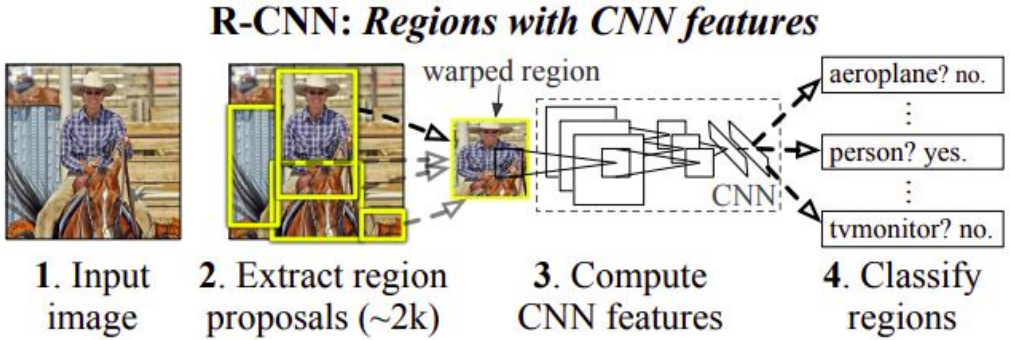
Task: Bounding Box & Classification
Challenges:

- Localization
- Unknown amount of outputs



Two Stage Object Detection [R-CNN Girshick 2014]

- First Stage:
  - Get region proposals and Bounding Boxes
- Second Stage:
  - Classify each region using some classifier
  - Choose the best scoring bounding box

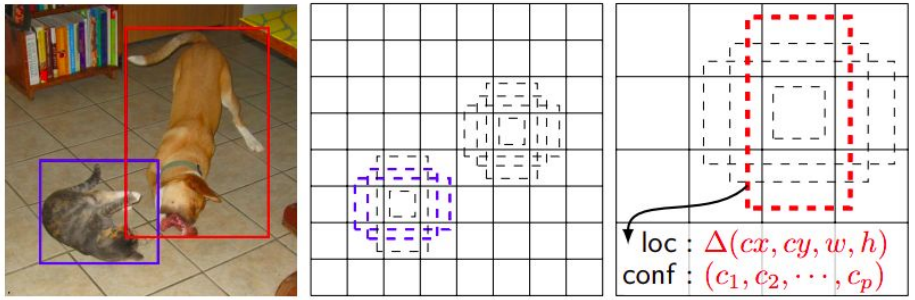


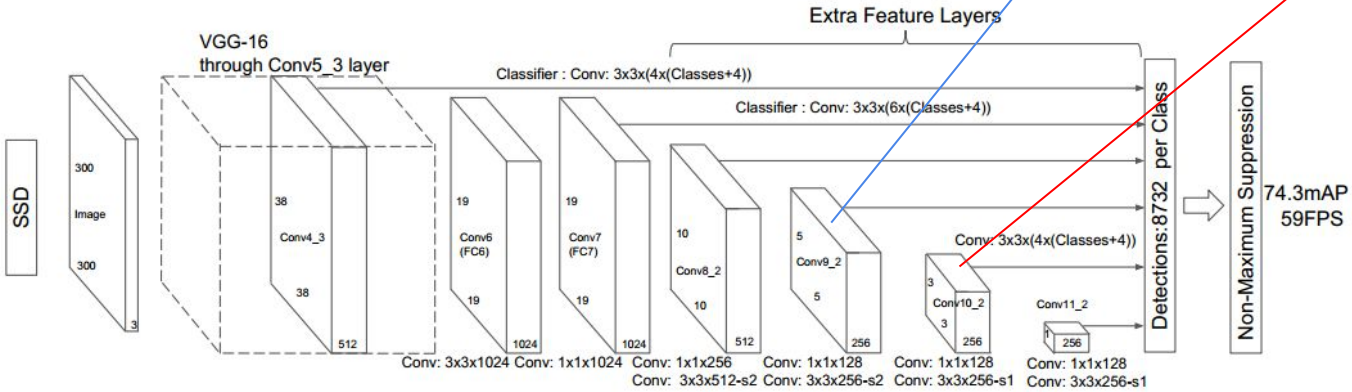
Disadvantage:

- Slow (Faster-RCNN ~7 FPS)
  - No feature sharing

Single Stage, Anchor Based, Detection (SSD - Wei at el 2015)

- High Speed (SSD ~ 59FPS on VOC2007)
- Prior (Anchor) based





RetinaNet ~ 100K Anchors

Problems With Anchor Based Methods

- Imbalance between Positive and Negative samples
  - Positive Anchor - Anchor that it's IoU with an Object > 0.5
  - "Easy" Negatives: Anchor that include only background, which the network easily identifies

- Many Hyper-Parameters:
  - How Many Anchors
  - Anchors Size
  - Anchors Aspect Ratio

Enter Corner Net [Law et al 2019]

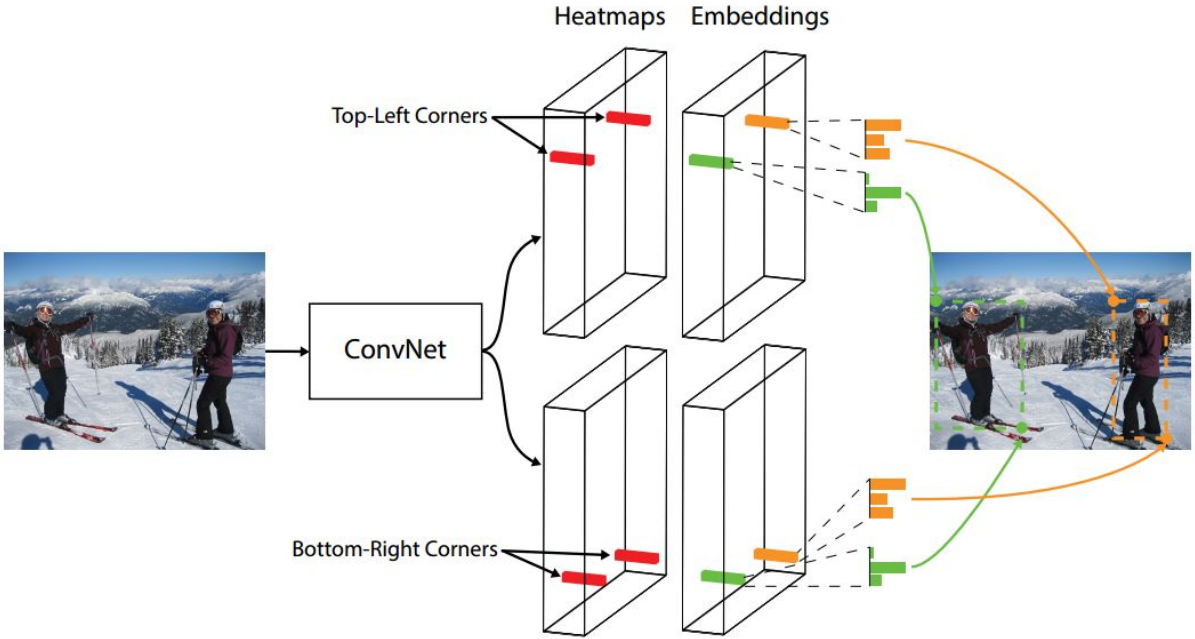
- Predicts Left and Right Corners

- Innovations:

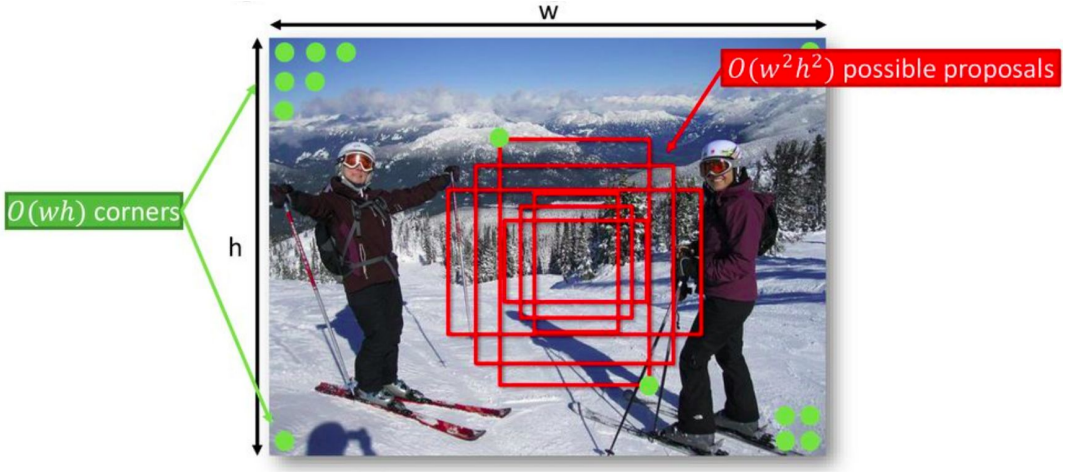
  - No Anchor Used

  - Using Associative Embeddings

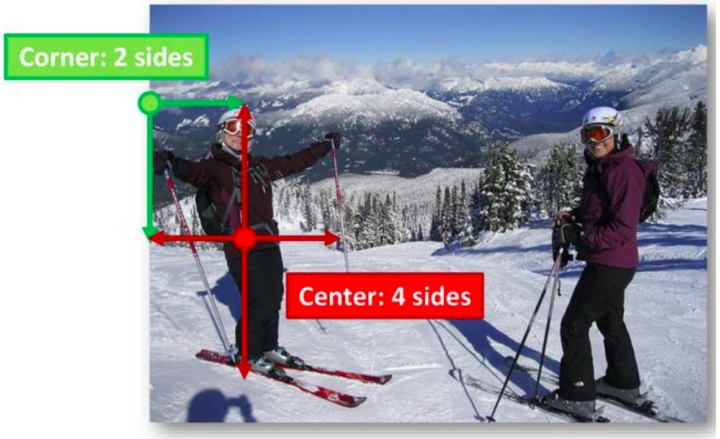
  - Corner Pooling



Advantage in Predicting Corners

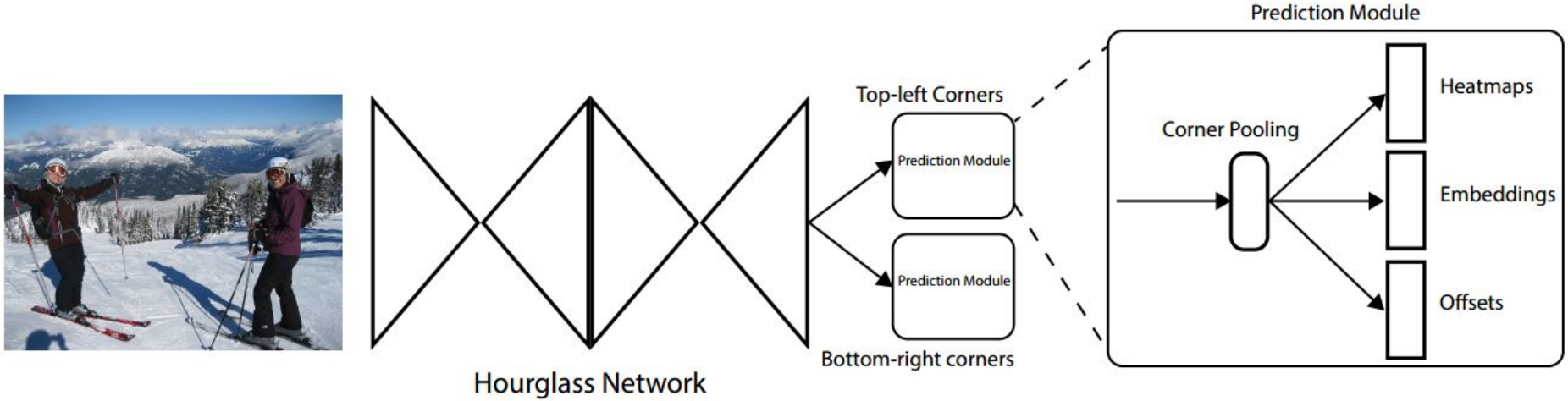


Represent $O(w^2h^2)$ possible proposals using only $O(wh)$ corners



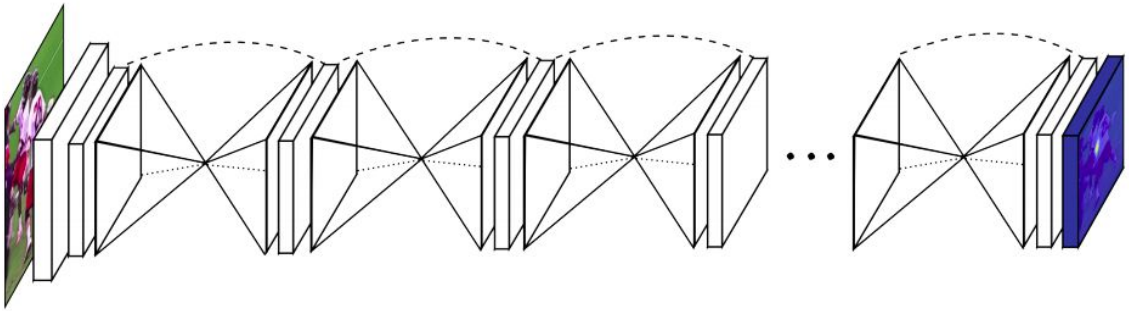
Detecting corner is easier than detecting center

Corner Net Architecture



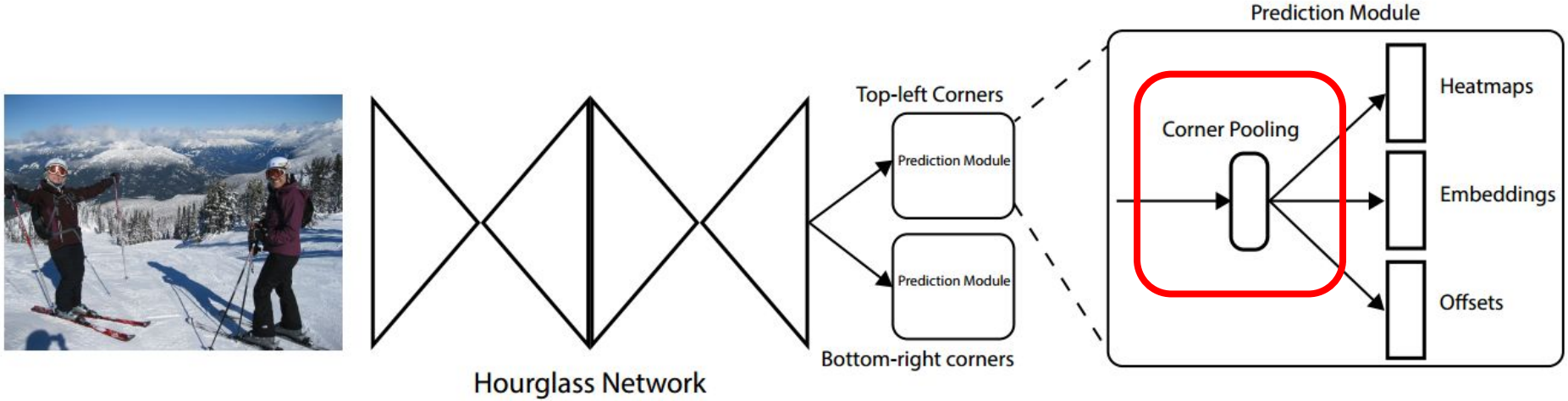
Backbone - Hourglass Network [Newell et al. 2016]

- First used for human pose estimation task
- Catch objects in different sizes
- Single output layer
- Intermediate Supervision
- Importance



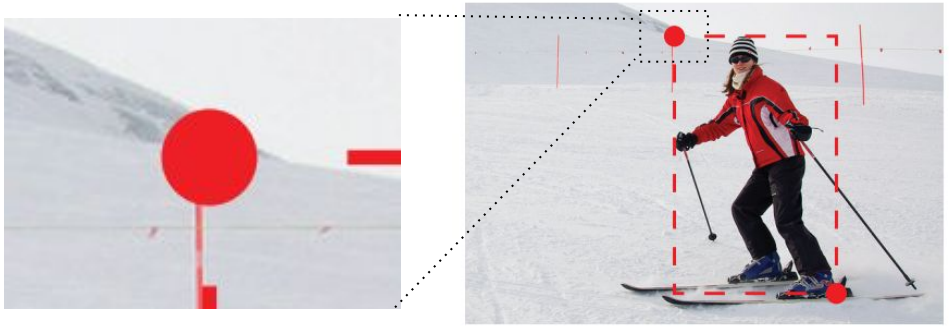
	AP	AP^{50}	AP^{75}	AP^s	AP^m	AP^l
FPN (w/ ResNet-101) + Corners	30.2	44.1	32.0	13.3	33.3	42.7
Hourglass + Anchors	32.9	53.1	35.6	16.5	38.5	45.0
Hourglass + Corners	38.4	53.8	40.9	18.6	40.5	51.8

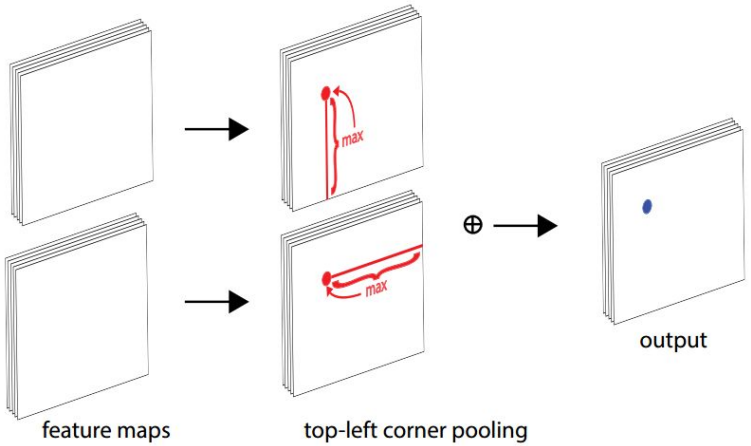
Corner Pooling



Corner Pooling

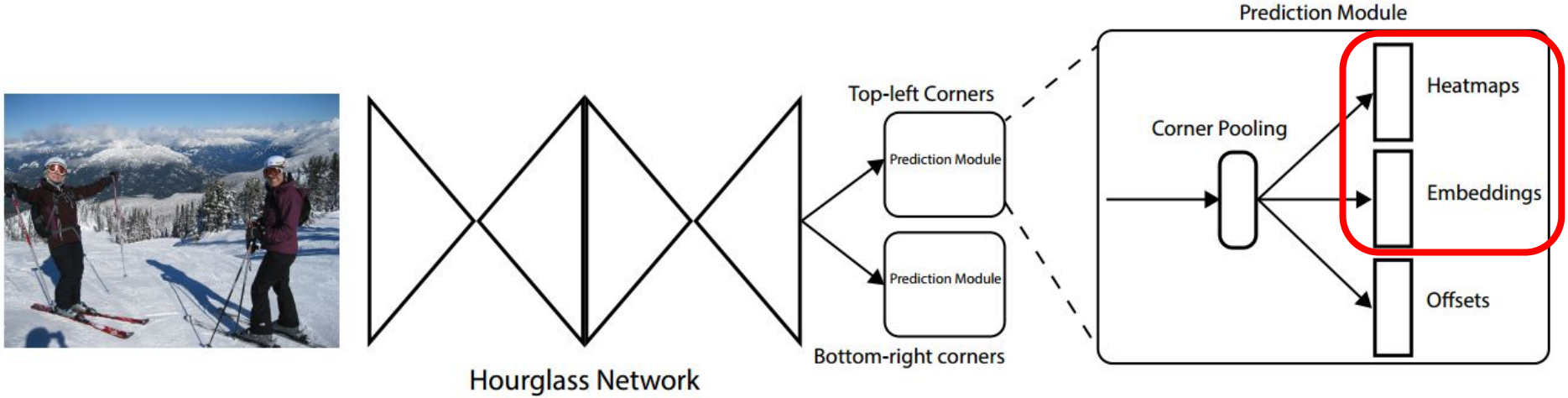
- No local evidence to where should the corner be
- TL going right, and going down should not meet object pixel
- Stability



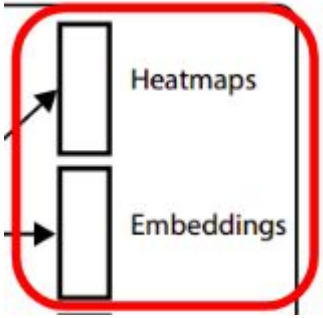


	mAP w/o pooling	mAP w/ pooling	improvement
Top-Left Corners			
Top-Left Quad.	66.1	69.2	+3.1
Bottom-Right Quad.	60.8	63.5	+2.7
Bottom-Right Corners			
Top-Left Quad.	53.4	56.2	+2.8
Bottom-Right Quad.	65.0	67.6	+2.6

Associative Embedding (and Heatmaps)



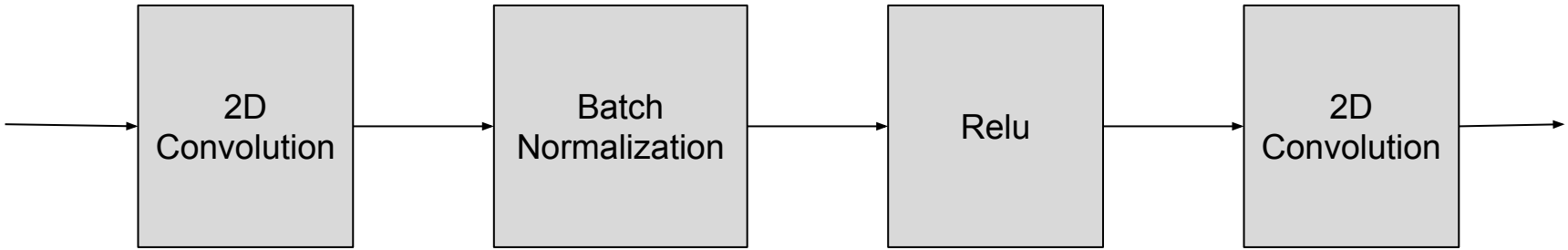
Associative Embedding [Newell et al. 2017]



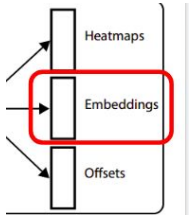
- Used for joint Detection in multi-persion environment
- Produces detection heatmaps, and embedding (“tag”) maps per class
- Embeddings are used to group joints of the same person

Detection Tagging

Embedding & Heatmap Layers



Embeddings Loss Function



- Same embedding for corners of the same object
- No explicit ground truth
- e_[OBJ][OBJ]t_k: embedding for top left corner for object k
  e_b_k: embedding for bottom right corner for object k
  e_k: average of e_t_k and e_b_k

$$L_{pull} = \frac{1}{N}\sum_{k=1}^{N}\left[(e_{t_k} - e_k)^2 + (e_{b_k} - e_k)^2\right],$$

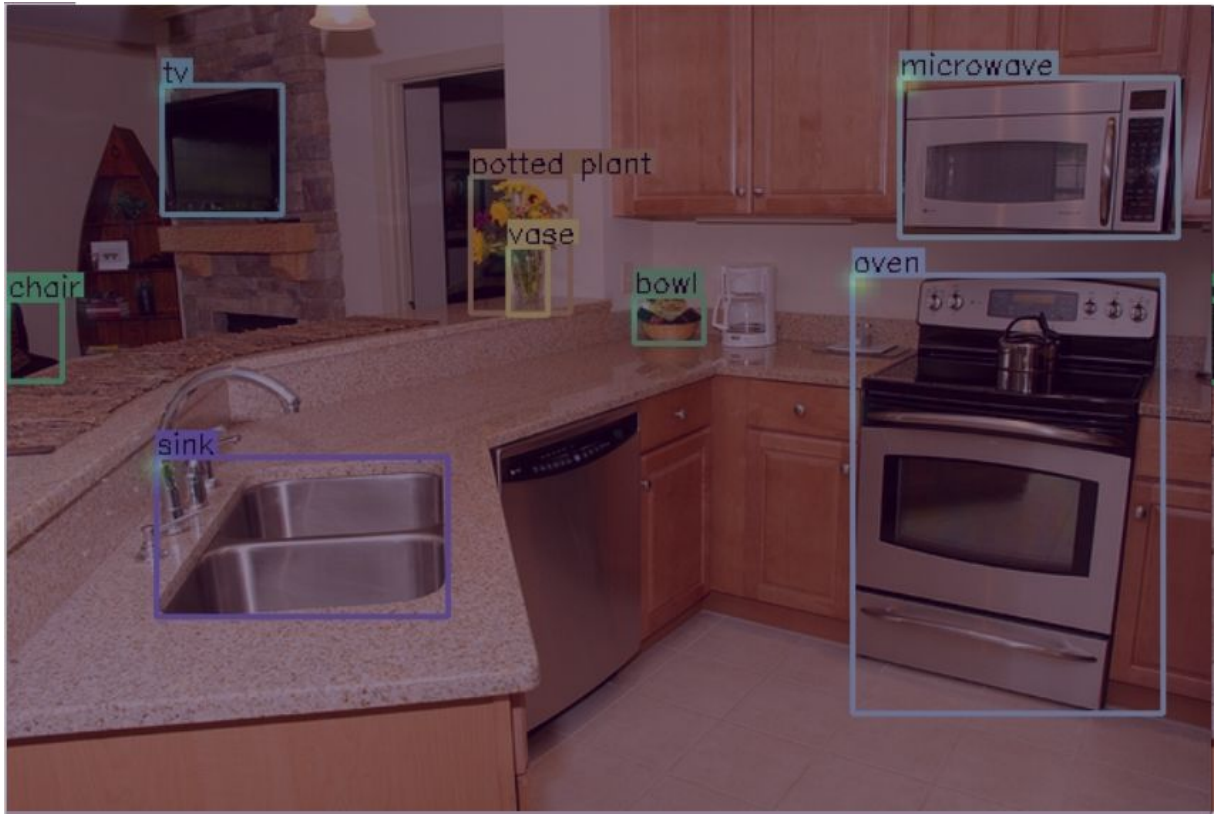
$$L_{push} = \frac{1}{N(N-1)}\sum_{k=1}^{N}\sum_{\substack{j=1 \\ j\neq k}}^{N}\max\left(0, \Delta - |e_k - e_j|\right),$$

Heatmaps

Heatmaps mark the corners of the boxes

Issue:

- Unbalanced Data

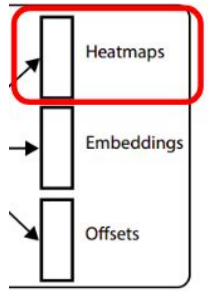


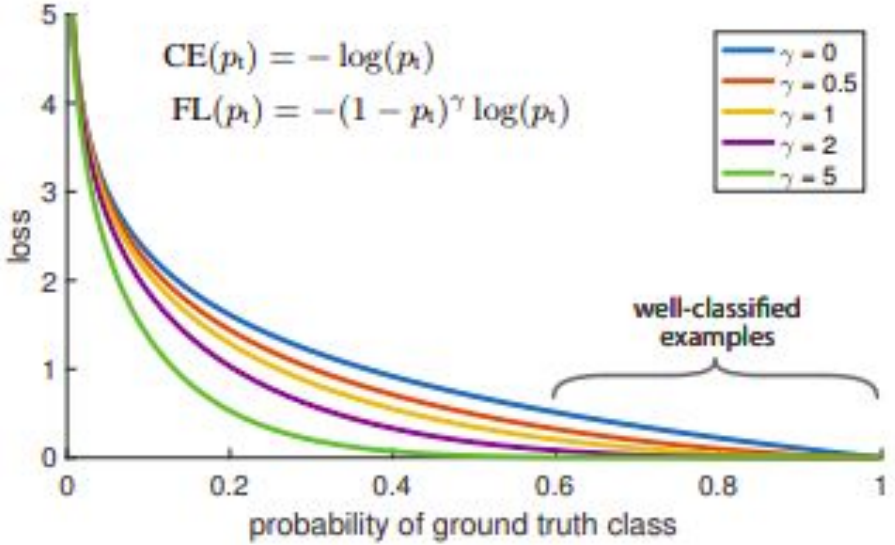
Focal Loss [Lin et al 2017]

- Used for anchor based single stage detectors

- Solves problem of imbalanced data: CE loss for p=0.9 is small (0.1), but with 7000 such examples it highly influences the total loss

- CE loss for p=0.9 ~ 0.1
  FL loss for p=0.9 ~ 0.001

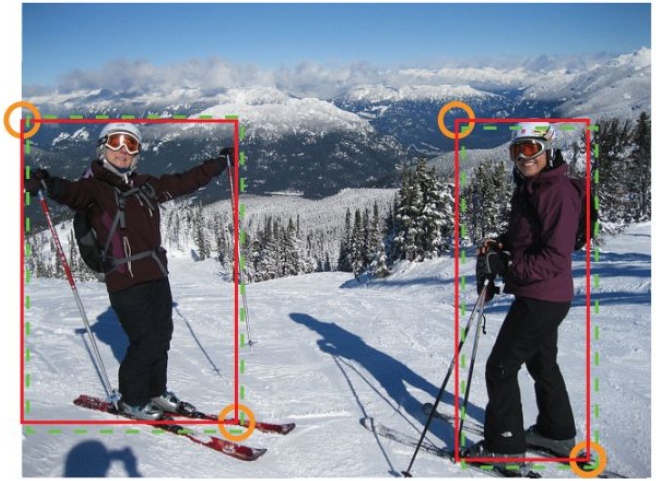




Heatmaps Loss

- Another Improvement: Reduce the penalty within a radius of the positive location. ensures 0.3 IoU with ground truth

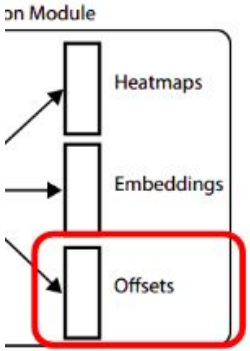
- Final Solution: Gaussian altered Focal Loss
  - A Gaussian in each positive location
  - Gives bigger weights to hard cases



$$L_{det} = \frac{-1}{N} \sum_{c=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} \begin{cases} (1 - p_{cij})^{\alpha} \log(p_{cij}) & \text{if } y_{cij} = 1 \\ (1 - y_{cij})^{\beta} (p_{cij})^{\alpha} \log(1 - p_{cij}) & \text{otherwise} \end{cases}$$

	AP	AP^{50}	AP^{75}	AP^{s}	AP^{m}	AP^{l}
w/o reducing penalty	32.9	49.1	34.8	19.0	37.0	40.7
fixed radius	35.6	52.5	37.7	18.7	38.5	46.0
object-dependent radius	38.4	53.8	40.9	18.6	40.5	51.8

Offset Fixes



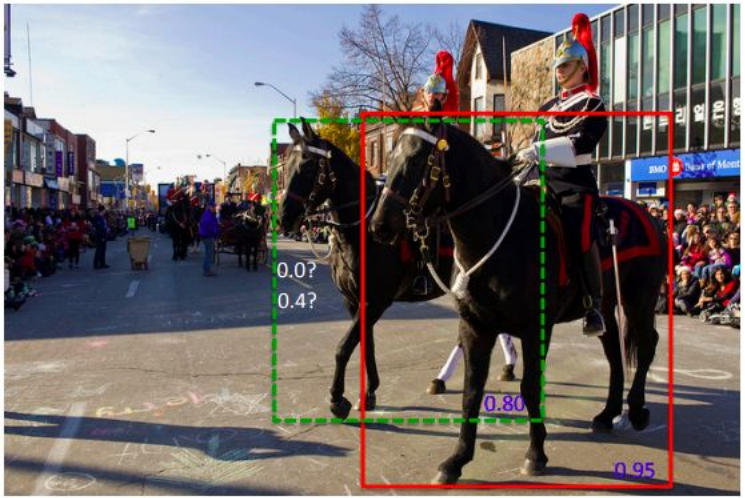
- Fix errors caused by remapping smaller scale heat maps to image size
- Uses Smooth L1 Loss [Girshick, 2015]

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

- L2 pluses without L2 exploding gradients

Post Processing

- Non Maximal Suppression by Max Pooling
- Pick top 100 TL and 100 BR
- Adjust corners according to offset layer
- Calculate L1 Distance between each two corners tags
- Remove corners with distance > 0.5
- Give each couple a score according to average heatmap score
- Apply Soft-NMS [Bodla et al 2017]



Input : $\mathcal{B} = \{b_1, .., b_N\}, \mathcal{S} = \{s_1, .., s_N\}, N_t$
$\mathcal{B}$ is the list of initial detection boxes
$\mathcal{S}$ contains corresponding detection scores
N_t is the NMS threshold

begin
  $\mathcal{D} \leftarrow \{\}$
  while $\mathcal{B} \neq$ *empty* **do**
    $m \leftarrow$ argmax $\mathcal{S}$
    $\mathcal{M} \leftarrow b_m$
    $\mathcal{D} \leftarrow \mathcal{D} \bigcup \mathcal{M}; \mathcal{B} \leftarrow \mathcal{B} - \mathcal{M}$
    for b_i *in* $\mathcal{B}$ **do**
      if $iou(\mathcal{M}, b_i) \geq N_t$ **then**
        $\mathcal{B} \leftarrow \mathcal{B} - b_i; \mathcal{S} \leftarrow \mathcal{S} - s_i$
      end — NMS
      $s_i \leftarrow s_i f(iou(\mathcal{M}, b_i))$ — Soft-NMS
    end
  end
  return $\mathcal{D}, \mathcal{S}$
end

Results

Method	Backbone	AP	AP^{50}	AP^{75}	AP^s	AP^m	AP^l	AR^1	AR^{10}	AR^{100}	AR^s	AR^m	AR^l
Two-stage detectors													
DeNet (Tychsen-Smith and Petersson, 2017a)	ResNet-101	33.8	53.4	36.1	12.3	36.1	50.8	29.6	42.6	43.5	19.2	46.9	64.3
CoupleNet (Zhu et al., 2017)	ResNet-101	34.4	54.8	37.2	13.4	38.1	50.8	30.0	45.0	46.4	20.7	53.1	68.5
Faster R-CNN by G-RMI (Huang et al., 2017)	Inception-ResNet-v2 (Szegedy et al., 2017)	34.7	55.5	36.7	13.5	38.1	52.0	-	-	-	-	-	-
Faster R-CNN+++ (He et al., 2016)	ResNet-101	34.9	55.7	37.4	15.6	38.7	50.9	-	-	-	-	-	-
Faster R-CNN w/ FPN (Lin et al., 2016)	ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2	-	-	-	-	-	-
Faster R-CNN w/ TDM (Shrivastava et al., 2016)	Inception-ResNet-v2	36.8	57.7	39.2	16.2	39.8	52.1	31.6	49.3	51.9	28.1	56.6	71.1
D-FCN (Dai et al., 2017)	Aligned-Inception-ResNet	37.5	58.0	-	19.4	40.1	52.5	-	-	-	-	-	-
Regionlets (Xu et al., 2017)	ResNet-101	39.3	59.8	-	21.7	43.7	50.9	-	-	-	-	-	-
Mask R-CNN (He et al., 2017)	ResNeXt-101	39.8	62.3	43.4	22.1	43.2	51.2	-	-	-	-	-	-
Soft-NMS (Bodla et al., 2017)	Aligned-Inception-ResNet	40.9	62.8	-	23.3	43.6	53.3	-	-	-	-	-	-
LH R-CNN (Li et al., 2017)	ResNet-101	41.5	-	-	25.2	45.3	53.1	-	-	-	-	-	-
Fitness-NMS (Tychsen-Smith and Petersson, 2017b)	ResNet-101	41.8	60.9	44.9	21.5	45.0	57.5	-	-	-	-	-	-
Cascade R-CNN (Cai and Vasconcelos, 2017)	ResNet-101	42.8	62.1	46.3	23.7	45.5	55.2	-	-	-	-	-	-
D-RFCN + SNIP (Singh and Davis, 2017)	DPN-98 (Chen et al., 2017)	45.7	67.3	51.1	29.3	48.8	57.1	-	-	-	-	-	-
One-stage detectors													
YOLOv2 (Redmon and Farhadi, 2016)	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5	20.7	31.6	33.3	9.8	36.5	54.4
DSOD300 (Shen et al., 2017a)	DS/64-192-48-1	29.3	47.3	30.6	9.4	31.5	47.0	27.3	40.7	43.0	16.7	47.1	65.0
GRP-DSOD320 (Shen et al., 2017b)	DS/64-192-48-1	30.0	47.9	31.8	10.9	33.6	46.3	28.0	42.1	44.5	18.8	49.1	65.0
SSD513 (Liu et al., 2016)	ResNet-101	31.2	50.4	33.3	10.2	34.5	49.8	28.3	42.1	44.4	17.6	49.2	65.8
DSSD513 (Fu et al., 2017)	ResNet-101	33.2	53.3	35.2	13.0	35.4	51.1	28.9	43.5	46.2	21.8	49.1	66.4
RefineDet512 (single scale) (Zhang et al., 2017)	ResNet-101	36.4	57.5	39.5	16.6	39.9	51.4	-	-	-	-	-	-
RetinaNet800 (Lin et al., 2017)	ResNet-101	39.1	59.1	42.3	21.8	42.7	50.2	-	-	-	-	-	-
RefineDet512 (multi scale) (Zhang et al., 2017)	ResNet-101	41.8	62.9	45.7	25.6	45.1	54.1	-	-	-	-	-	-
CornerNet511 (single scale)	Hourglass-104	40.6	56.4	43.2	19.1	42.8	54.3	35.3	54.7	59.4	37.4	62.4	77.2
CornerNet511 (multi scale)	Hourglass-104	42.2	57.8	45.2	20.7	44.8	56.6	36.6	55.9	60.3	39.5	63.2	77.3

Error Analysis

	AP	AP^{50}	AP^{75}	AP^{s}	AP^{m}	AP^{l}
	38.4	53.8	40.9	18.6	40.5	51.8
w/ gt heatmaps	73.1	87.7	78.4	60.9	81.2	81.8
w/ gt heatmaps + offsets	86.1	88.9	85.5	84.8	87.2	82.0

Questions?