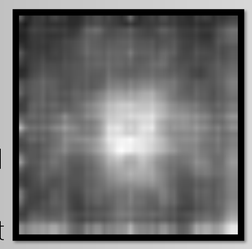
One Model to Store Them All

Spatial Memory for Context Reasoning in Object Detection

Xinlei Chen Abhinav Gupta

Spatial Memory

- ✓ Generic framework to store all kinds of **instance-**, pixel-, & image-level knowledge
- ✓ Models the **joint** layout distribution of all instances and learns **de-duplication**
- ✓ Efficient context reasoning with ConvNet for object detection



Learned Prior

Faster RCNN (ConvNet) Results [Chen & Gupta, 2017]

Object Detection

Formulation

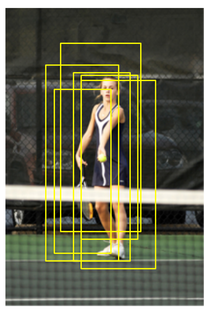
$$\arg\max_{\mathcal{M}} \mathcal{L} = \log \mathbb{P}(O_{1:N}|\mathcal{M}, I) = \sum_{n=1:N} \log \mathbb{P}(O_n|\mathcal{O}_{:n-1}, \mathcal{M}, I)$$

all objects · image · previous detections · model · current detection

Approximation (Detectors Now) $\mathcal{L} \approx \sum \log \mathbb{P}(O_n|\mathcal{M}, I)$

breakdown

Isolated Classification



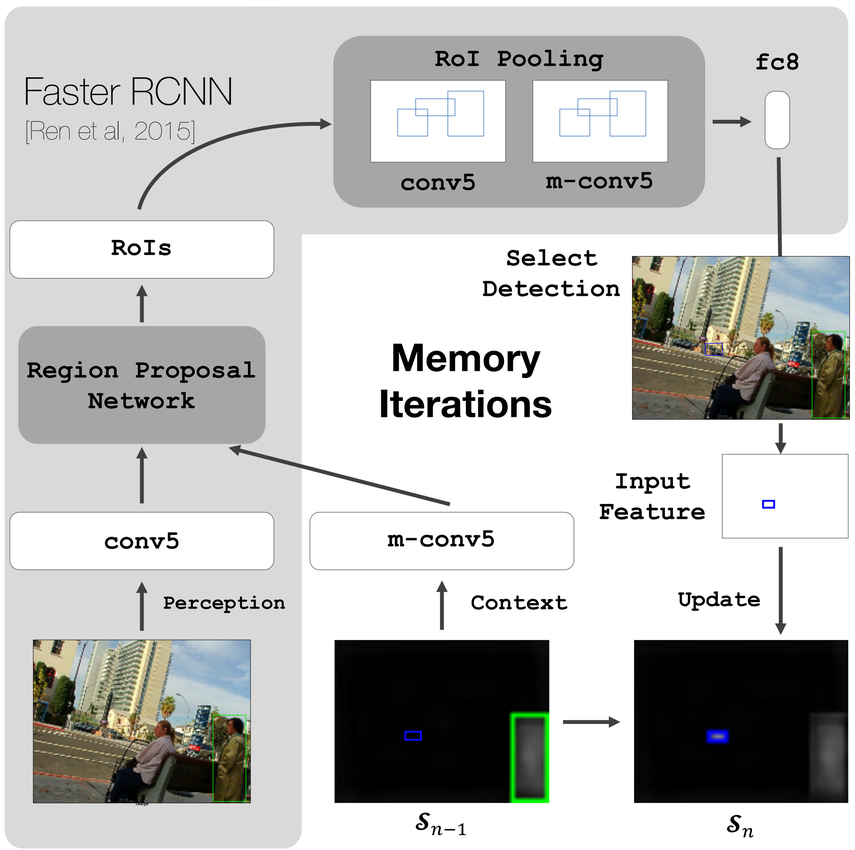
Non-Maximal Suppression

✗ Suboptimal

✗ Little semantic info

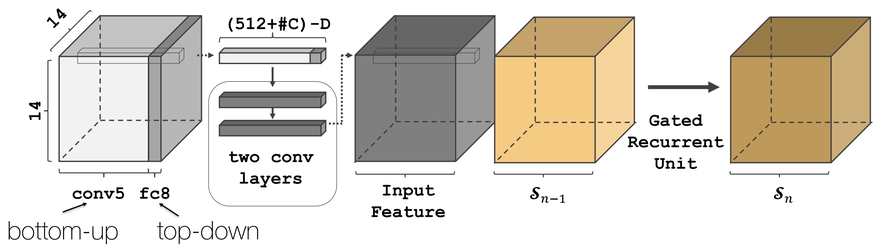
Memory-Based Detection

$$\mathcal{L} \approx \sum \log \mathbb{P}(O_n|\mathcal{S}_{n-1}, \mathcal{M}, I)$$



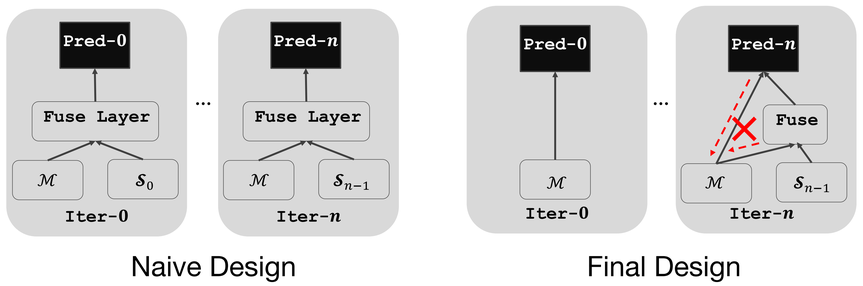
Input Zoom-In

`crop_and_resize` to perform (inverse) RoI-Pooling



Output for De-Duplication

remove ground truth during training if detected



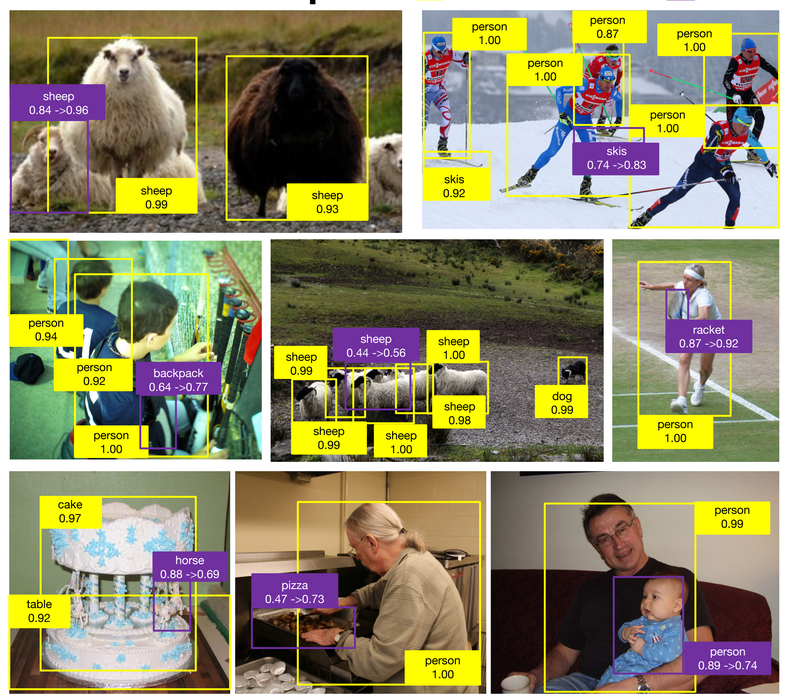
Recall Pitfall

1. Top RoIs 2. Guessing game: **ski** or **snowboard**?

$\mathcal{M}$: hmm, 40% ski, 60% snowboard ---- Should **NOT** throw away detections if not max

Qualitative Examples

: previous detections; : current



COCO Results

Baseline: [Chen & Gupta 2017] on VGG16
MLP: Extension of Baseline with more layers

Method	AP	AP-.5	AP-.75	AP-S	AP-M	AP-L	AR-S	AR-M	AR-L
Baseline	29.4	50.0	30.9	12.2	33.7	43.8	18.5	45.5	58.9
MLP	30.1	50.8	31.7	12.5	34.2	44.5	19.2	47.0	59.8
SMN	**31.6**	**52.2**	**33.2**	**14.4**	**35.7**	**45.8**	**20.5**	**48.8**	**63.2**

Future Work

- ❑ Speed-Up
- ❑ Learning the selection process with REINFORCE