Fast R-CNN

Anton Semenkin

Department of Computer Science
Higher School Of Economics

April, 2018

Outline

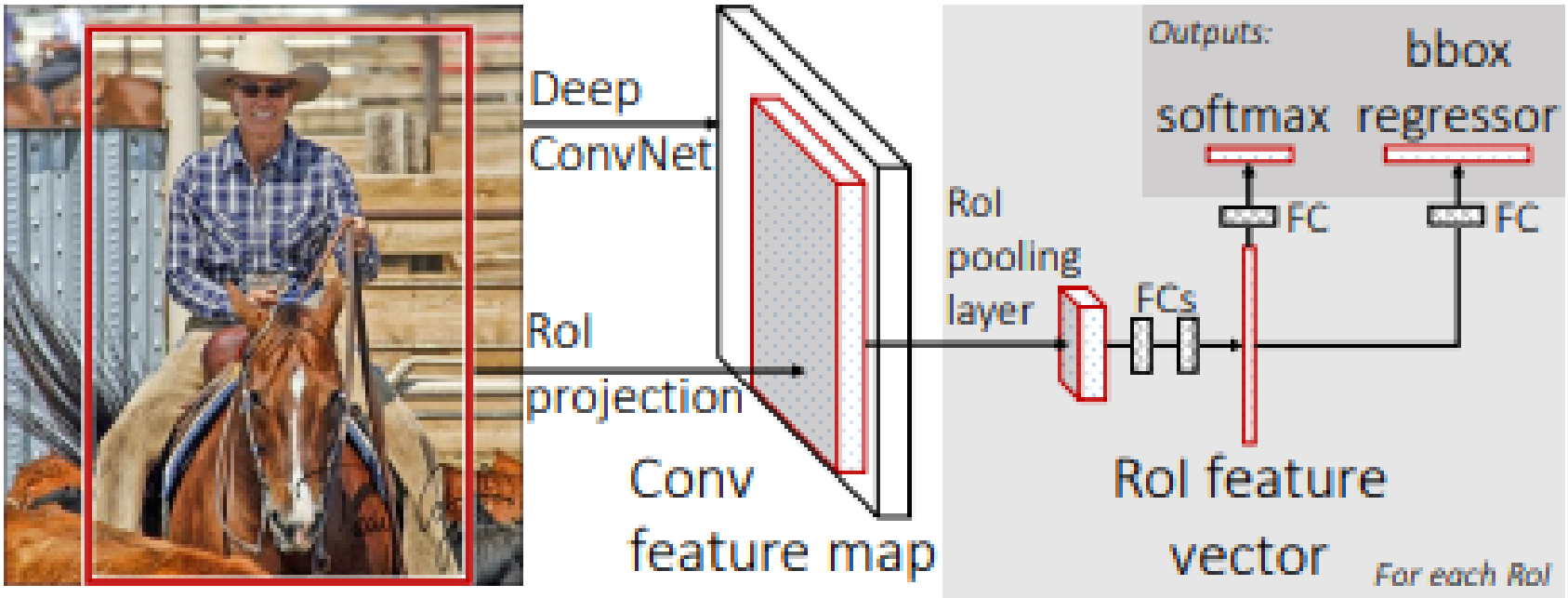
Outline

R-CNN

- R-CNN architecture
  - DeepConv
  - SVM
  - Bbox regression
- Spatial pyramid pooling networks
  - Pooling for matching
  - Increased speed
- Disadvantages
  - Memory
  - Speed

Outline

General architecture



RoI pooling layer

- RoI is described by four numbers (r, c, h, w)
- Pooling grid cell size $H \times W$
- Applied independently to each feature map channel

Pre-trained networks

- 3 networks (5 poolng layers, 5-13 convolutional layers)
- Network preparation
  - Last pooling layer is replaced by RoI pooling layer
  - Last fully connected layer and softmax are replaced with the 2 sibling layers
  - Two data inputs: images and a list of RoIs

Fine-tuning

- 128 Images with one RoI VS 2 images, 128 RoIs each
- One optimization for bounding box regression and softmax output (multitask loss)
- Two outputs: $p = \{p_0, ..., p_K\}$ and $t^k = \{t_x^k, t_y^k, t_w^k, t_h^k\}$

Multitask loss

- Loss function: $L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v)$
- Classification loss: $L_{cls}(p, u) = -\log p_u$
- Bbox regressoin loss: $L_{loc}(t^u, v) = \sum_{i \in \{x,y,w,h\}} smooth_{L_1}(t_i^u - v_i)$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1; \\ |x| - 0.5 & otherwise; \end{cases}$$

Outline

Experiments 1

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	t
SPPnet BB [11]†	07 \ diff	73.9	72.3	62.5	51.5	44.4	74.4	73.0	74.4	42.3	73.6	
R-CNN BB [10]	07	73.4	77.0	63.4	45.4	**44.6**	75.1	78.1	79.8	40.5	73.7	
FRCN [ours]	07	74.5	78.3	69.2	53.2	36.6	77.3	78.2	82.0	40.7	72.7	
FRCN [ours]	07 \ diff	74.6	**79.0**	68.6	57.0	39.3	79.5	**78.6**	81.9	**48.0**	74.0	
FRCN [ours]	07+12	**77.0**	78.1	**69.3**	**59.4**	38.3	**81.6**	**78.6**	**86.7**	42.8	**78.8**	

table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
57.7	70.3	74.6	74.3	54.2	34.0	56.4	56.4	67.9	73.5	63.1
62.2	79.4	78.1	73.1	64.2	**35.6**	66.8	67.2	70.4	**71.1**	66.0
67.9	79.6	79.2	73.0	69.0	30.1	65.4	70.2	75.8	65.8	66.9
67.4	80.5	80.7	74.1	69.6	31.8	67.1	68.4	75.3	65.5	68.1
68.9	**84.7**	**82.0**	**76.6**	**69.9**	31.8	**70.1**	**74.8**	**80.4**	70.4	**70.0**

Experiments 2

method	train set	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	t
BabyLearning	Prop.	78.0	74.2	61.3	45.7	42.7	68.2	66.8	80.2	40.6	70.0	
NUS_NIN_c2000	Unk.	80.2	73.8	61.9	43.7	**43.0**	70.3	67.6	80.7	41.9	69.7	
R-CNN BB [10]	12	79.6	72.7	61.9	41.2	41.9	65.9	66.4	84.6	38.5	67.2	
FRCN [ours]	12	80.3	74.7	66.9	46.9	37.7	73.9	68.6	87.7	41.7	71.1	
FRCN [ours]	07++12	**82.3**	**78.4**	**70.8**	**52.3**	38.7	**77.8**	**71.6**	**89.3**	**44.2**	**73.0**	

table	dog	horse	mbike	persn	plant	sheep	sofa	train	tv	mAP
49.8	79.0	74.5	77.9	64.0	35.3	67.9	55.7	68.7	62.6	63.2
51.7	78.2	75.2	76.9	65.1	**38.6**	**68.3**	58.0	68.7	63.3	63.8
46.7	82.0	74.8	76.0	65.2	35.6	65.4	54.2	67.4	60.3	62.4
51.1	86.0	77.8	79.8	69.8	32.1	65.5	63.8	76.4	61.7	65.7
55.0	**87.5**	**80.5**	**80.8**	**72.0**	35.1	**68.3**	**65.7**	**80.4**	**64.2**	**68.4**

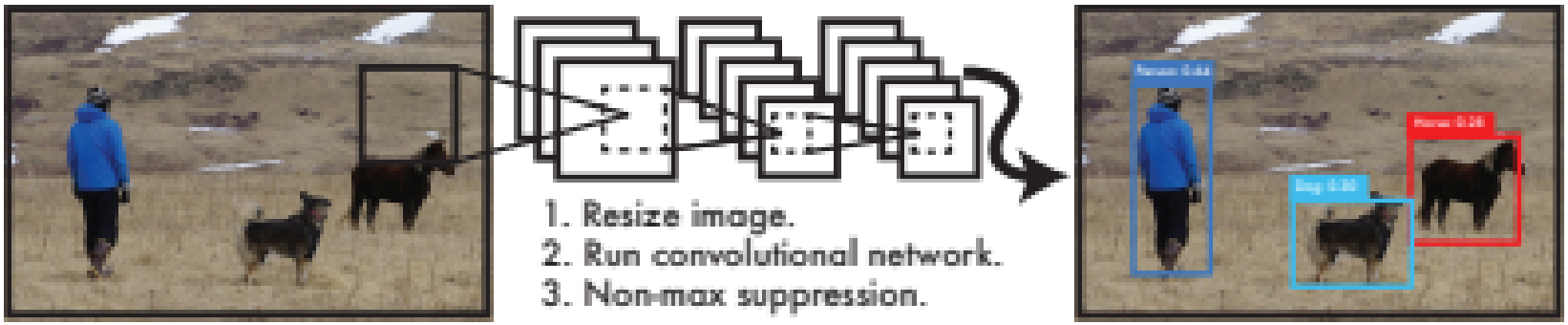
Summary

- The first main message of your talk in one or two lines.
- The second main message of your talk in one or two lines.
- Perhaps a third message, but not more than that.

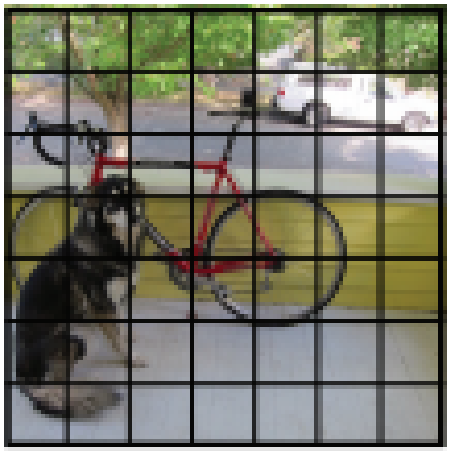
- Outlook
  - Something you haven't solved.
  - Something else you haven't solved.

Outline

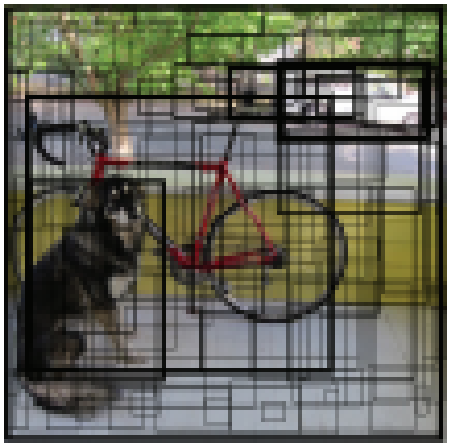
YOLO


1. Resize image.
2. Run convolutional network.
3. Non-max suppression.

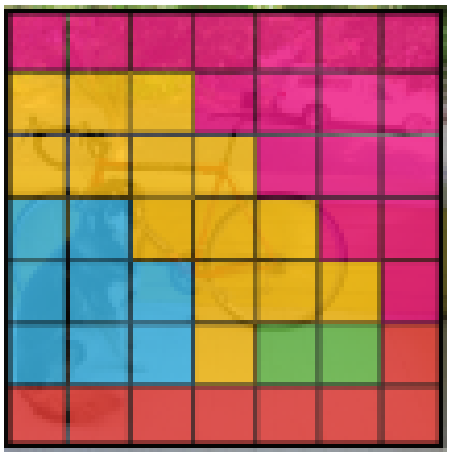
YOLO



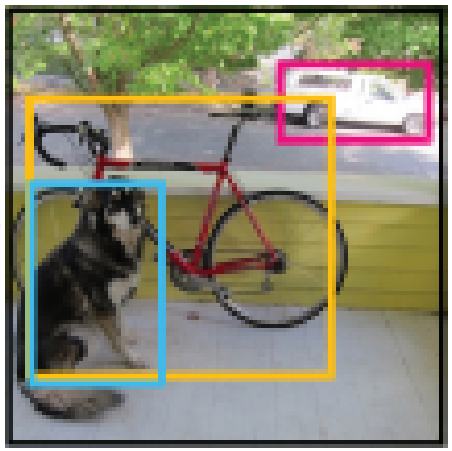
S × S grid on input



Bounding boxes + confidence



Class probability map



Final detections

YOLO