北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

REAL-TIME OBJECT DETECTION BY A MULTI-FEATURE FULLY CONVOLUTIONAL NETWORK

Yajing Guo, Xiaoqiang Guo, Zhuqing Jiang, Aidong Men, Yun Zhou

Email：gyj@bupt.edu.cn






School of Information and Communication Engineering
Beijing University of Posts and Telecommunications

1 Object Detection Overview

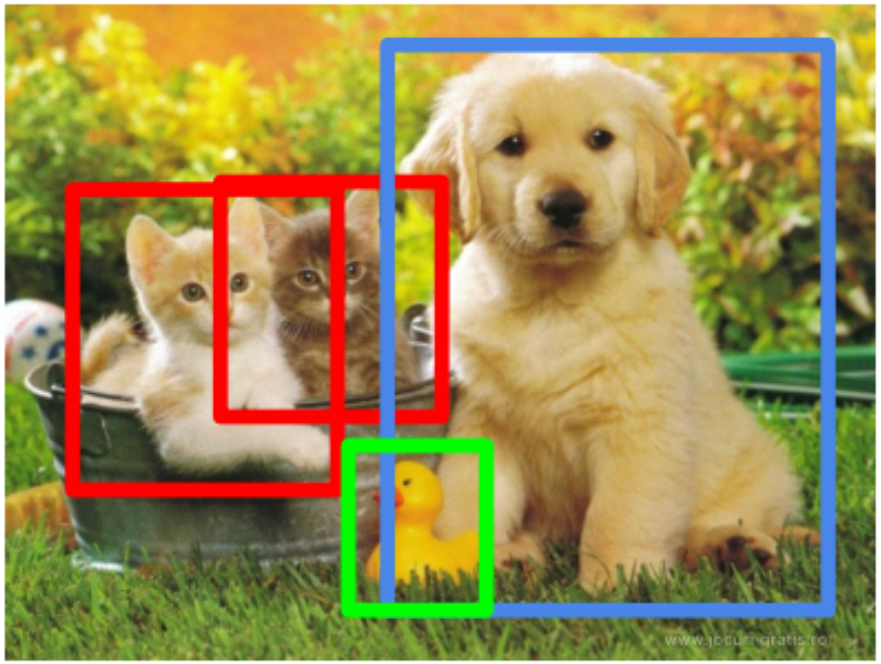
2 Related Work

3 Proposed Method

4 Experiments & Results

1 Object Detection Overview

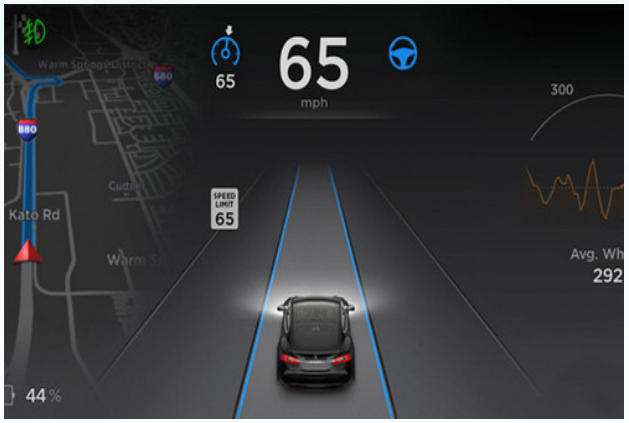
- Object detection aims to predict both the category and location of objects in terms of a bounding box.
- Object detection is indispensable for many applications such as video surveillance, autonomous vehicles, augmented reality, and human-computer interaction



CAT, DOG, DUCK

1 Object Tracking Overview

- Object detection aims to predict both the category and location of objects in terms of a bounding box.
- Object detection is indispensable for many applications such as video surveillance, autonomous vehicles, augmented reality, and human-computer interaction



1 Object Tracking Overview

This paper focus on designing a multi-feature fully convolutional network, aiming to achieve real-time object detection.

Contributions:

Step1
- A single fully convolutional network treating detection task as a regression problem.

step2
- Multi-feature concatenation fusing shallow and deep information and increasing the detection confidence.

step3
- Anchor boxes mechanism discretizing the space of output box shapes.

1 Object Detection Overview

2 **Related Work**

3 Proposed Method

4 Experiments & Results

2 Related Work

- Detectors based on region proposals:
  - R-CNN *R.Girshick, 2014*
  - Fast R-CNN *R.Girshick, 2015*
  - Faster R-CNN *S. Ren, 2015*
  - MSCNN *S. Gidaris,2015*

- Detectors free from region proposals:
  - YOLO *J. Redmon, 2015*
  - SSD *W.Liu, 2016*
  - G-CNN *M. Najibi, 2016*

1 Object Detection Overview

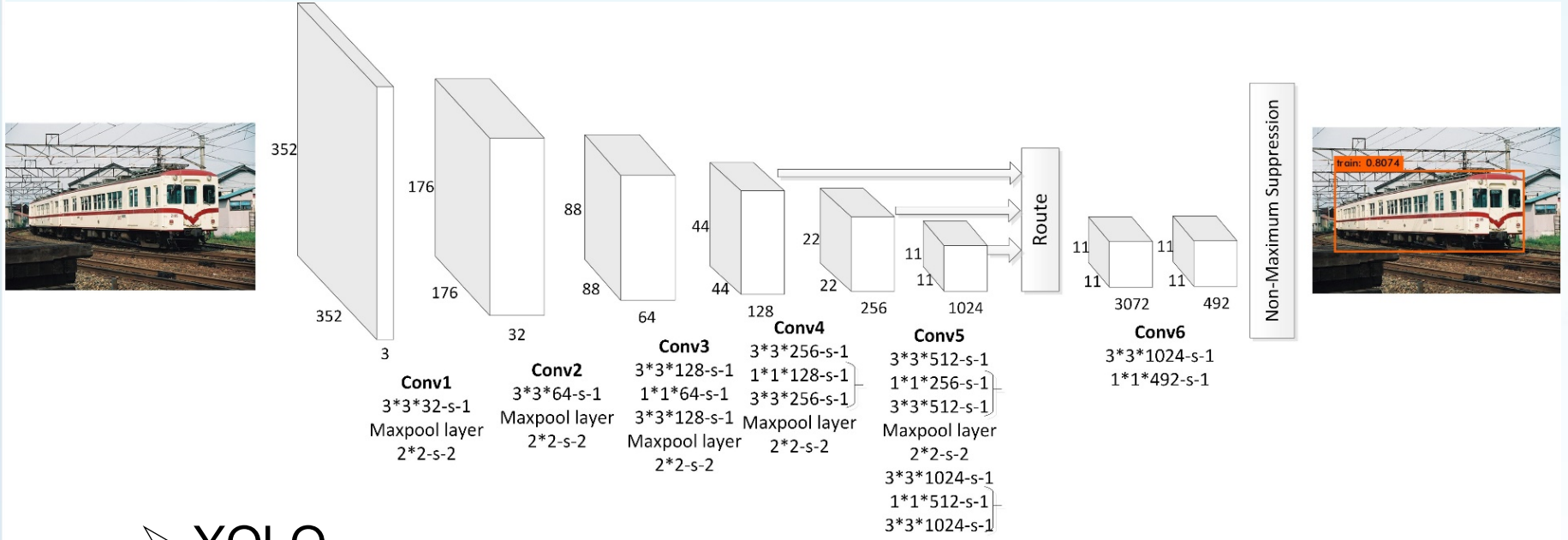
2 Related Work

3 **Proposed Method**

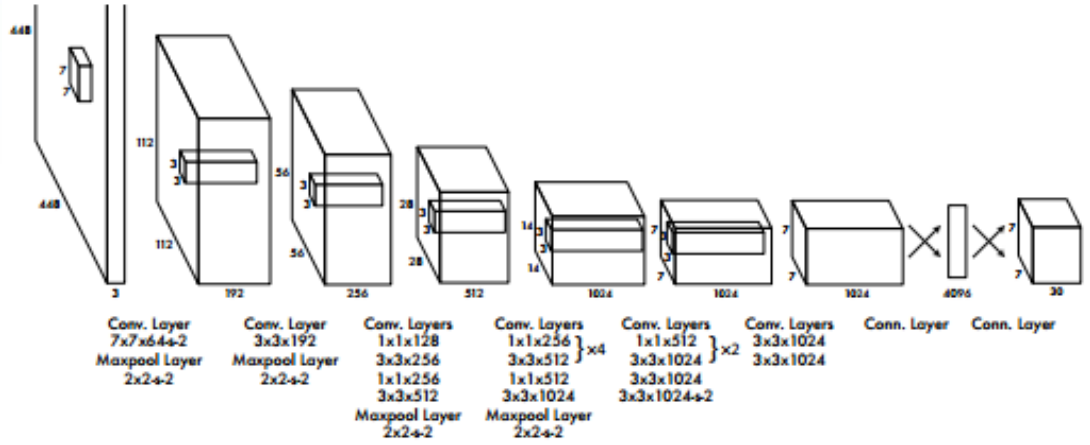
4 Experiments & Results

3 Proposed Method

- **System Overview**
  - The framework of proposed system

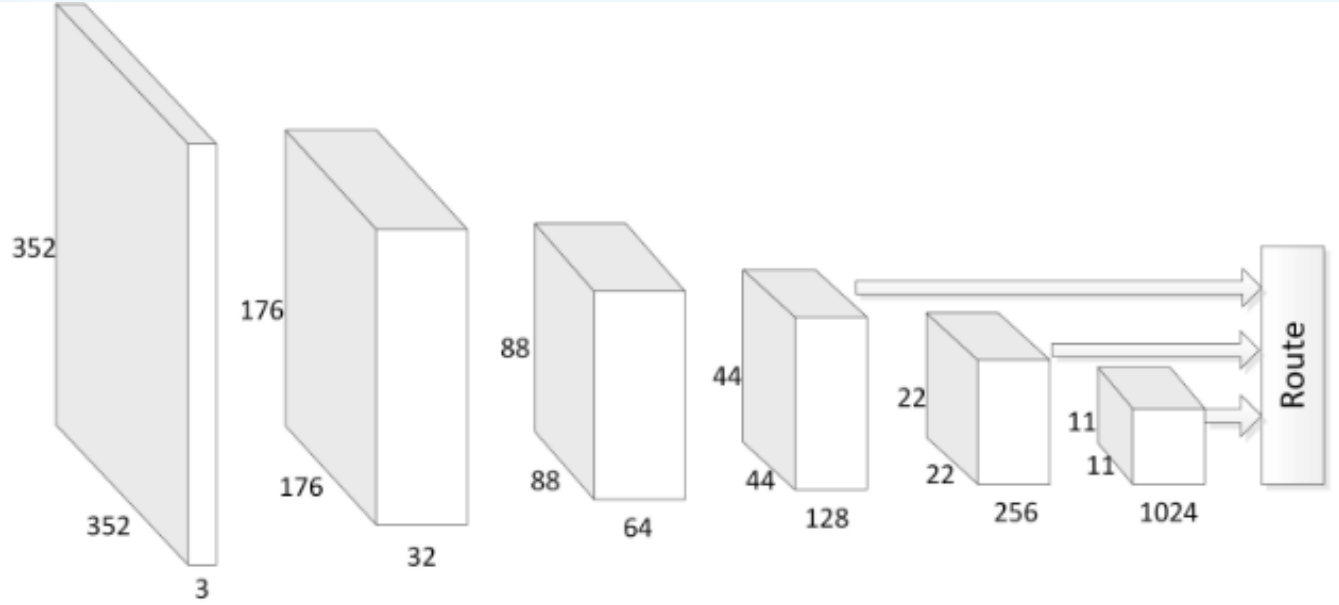


- YOLO



3 Proposed Method

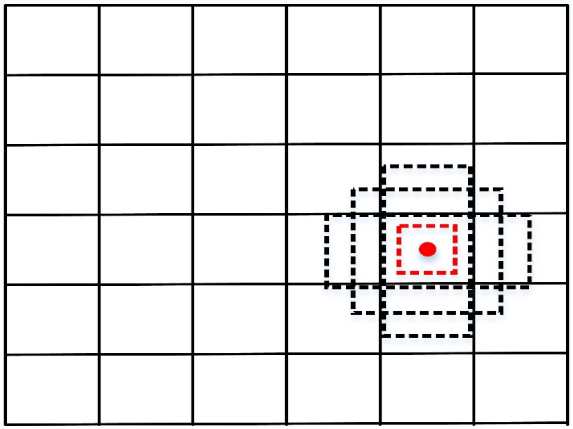
- **Multi-feature concatenation**



44*44*128 → 11*11*2048
22*22*256 → 11*11*1024

3 Proposed Method

- **Anchor boxes mechanism**



3 Proposed Method

- **Anchor boxes mechanism**

Each bounding box contains 4+C values

$$b_x = \sigma(t_x) + c_x$$
$$b_y = \sigma(t_y) + c_y$$
$$b_w = p_w e^{t_w}$$
$$b_h = p_w e^{t_h}$$
$$P(object) * IoU_b^{truth} = \sigma(t_c)$$

1 Object Detection Overview

2 Related Work

3 Proposed Method

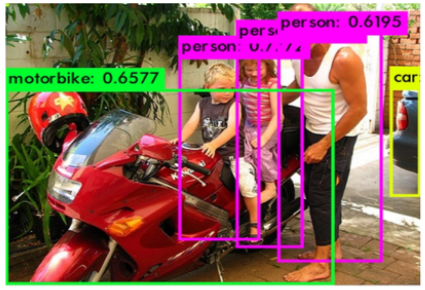
4 **Experiments & Results**

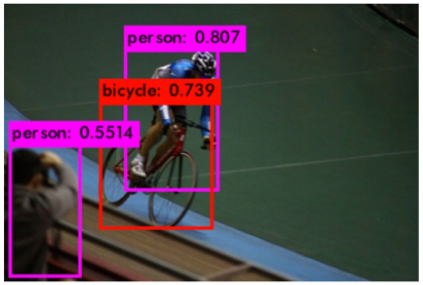
4 Experiments & Results

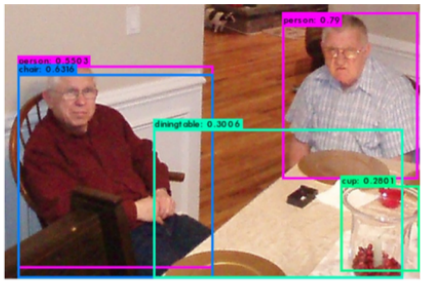
- Evaluating the proposed method on PASCAL VOC dataset with comparisons to other methods.

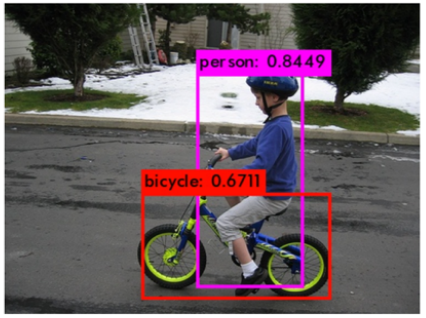
4 Experiments & Results

- **Qualitative Evaluation**









4 Experiments & Results

- **Quantitative Evaluation**
  - mean average precision (mAP) and frame per second (FPS) are used to quantitatively evaluate the method

Table 1. PASCAL VOC2012 test detection results. Each model was trained on PASCAL VOC2012 trainval and VOC2007 trainval and test set. Fast and Faster R-CNN use images with minimum dimension 600, while the image size for YOLO and SSD300 is 448 × 448 and 300 × 300 respectively.

Method	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast R-CNN [7]	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	87.5	80.5	80.8	72.0	35.1	68.3	65.7	80.4	64.2
Faster R-CNN [8]	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
YOLO [12]	57.9	77.0	67.2	57.7	38.3	22.7	68.3	55.9	81.4	36.2	60.8	48.5	77.2	72.3	71.3	63.5	28.9	52.2	54.8	73.9	50.8
SSD300 [13]	72.4	85.6	80.1	70.5	57.6	46.2	79.4	76.1	89.2	**53.0**	77.0	**60.8**	87.0	**83.1**	82.3	79.4	45.9	75.9	**69.5**	81.9	67.5
our MFCN	**73.2**	**86.1**	**82.0**	**74.4**	**59.2**	**50.8**	**79.6**	**76.2**	**90.2**	52.1	**78.2**	58.1	**89.0**	82.5	**83.4**	**81.1**	**48.5**	**77.1**	62.4	**83.6**	**68.2**

Table 2. Detection performance for speed on PASCAL VOC2012 test set. Our MFCN is faster and more accurate than prior detection methods.

Method	mAP (%)	Time (ms)	FPS
Fast R-CNN [7]	68.4	1830	0.5
Faster R-CNN [8]	70.4	142	7
YOLO [12]	57.9	22	45
SSD300 [13]	72.4	21	46
SSD500 [13]	**74.9**	52	19
our MFCN	73.2	**13**	**75**

4 Experiments & Results

- **Error Analysis**

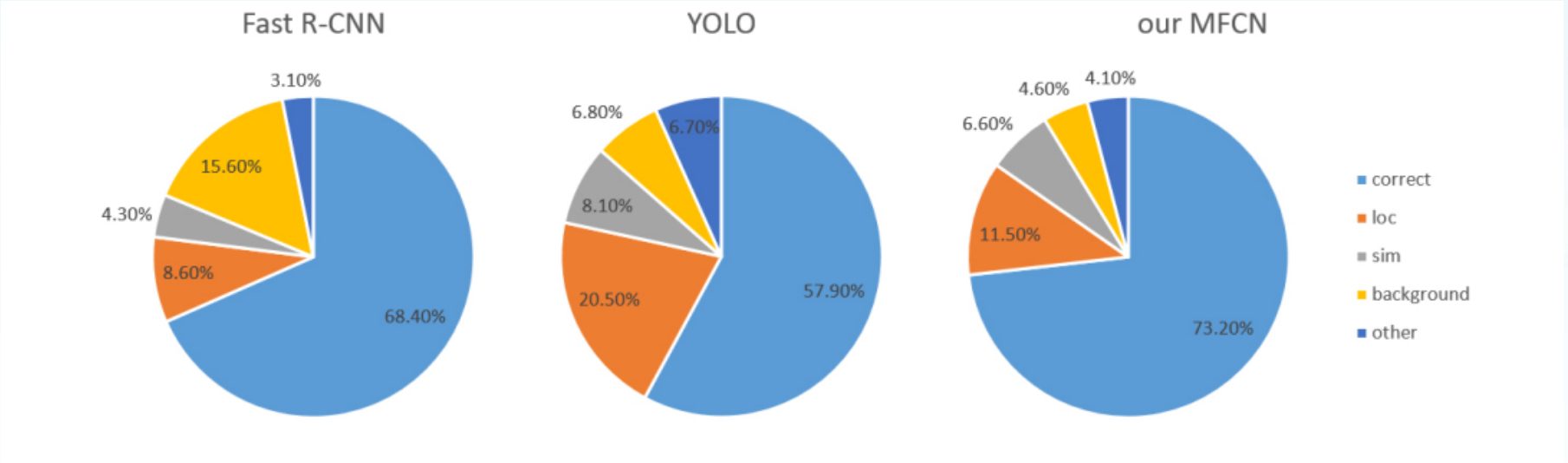


Fig. 3. **Error analysis from PASCAL VOC2012 test.** These charts show the percentage of detections that are correct or false positive due to poor localization (loc), confusion with similar categories (sim), with others, or with background. **Best viewed in color**.

4 Experiments & Results

- **Conclusion**
  - framing object detection as a regression problem can simplify detection pipeline and improve the detection speed
  - multi-feature concatenation can efficiently fuse shallow and deep information and in- crease the detection confidence
  - anchor boxes mechanism helps to discretize the space of output box shapes

- **Future Work**
  - more attention will be paid for tradeoff between accuracy and speed.

北京郵電大學

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

Thank You !