Nanjing University
of Aeronautics & Astronautics

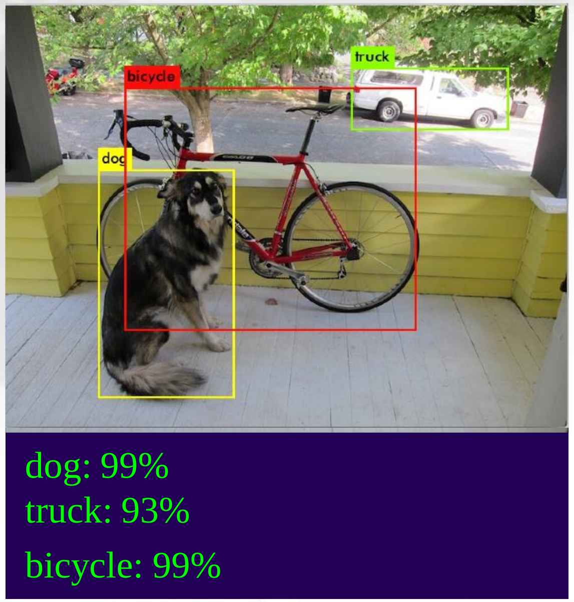
# Small Object Detection Based on Modified FSSD and Model Compression

Qingcai Wang, Hao Zhang, Xianggong Hong, and Qinqin Zhou

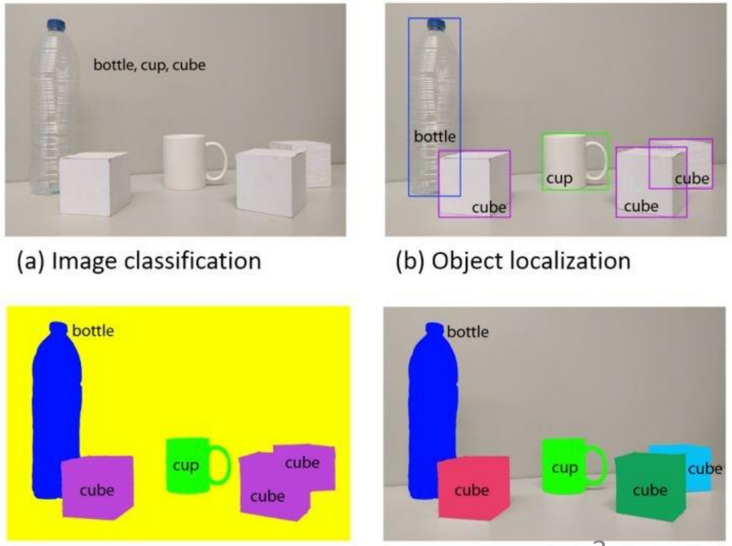
# Content

# Object Detection

Object detection to identify and locate specific objects in images or videos. It is widely used in industrial fields such as autonomous driving, video surveillance, robot vision, and new retail [1].








(a) Image classification

(b) Object localization

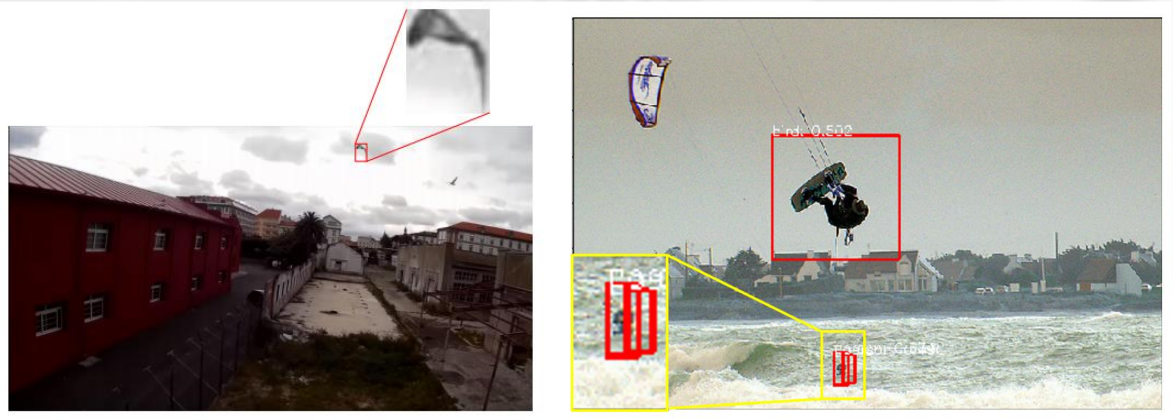
(c) Semantic segmentation

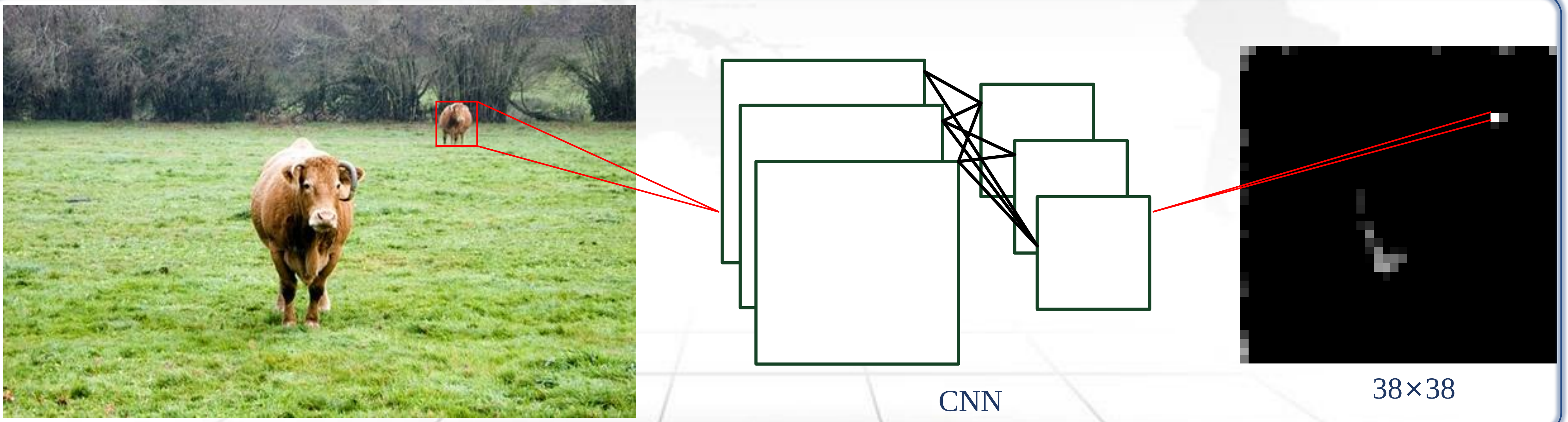
(d) Instance segmentation

# Small object detection

Small Object: <32×32 (Lin .*et al*, 2014)

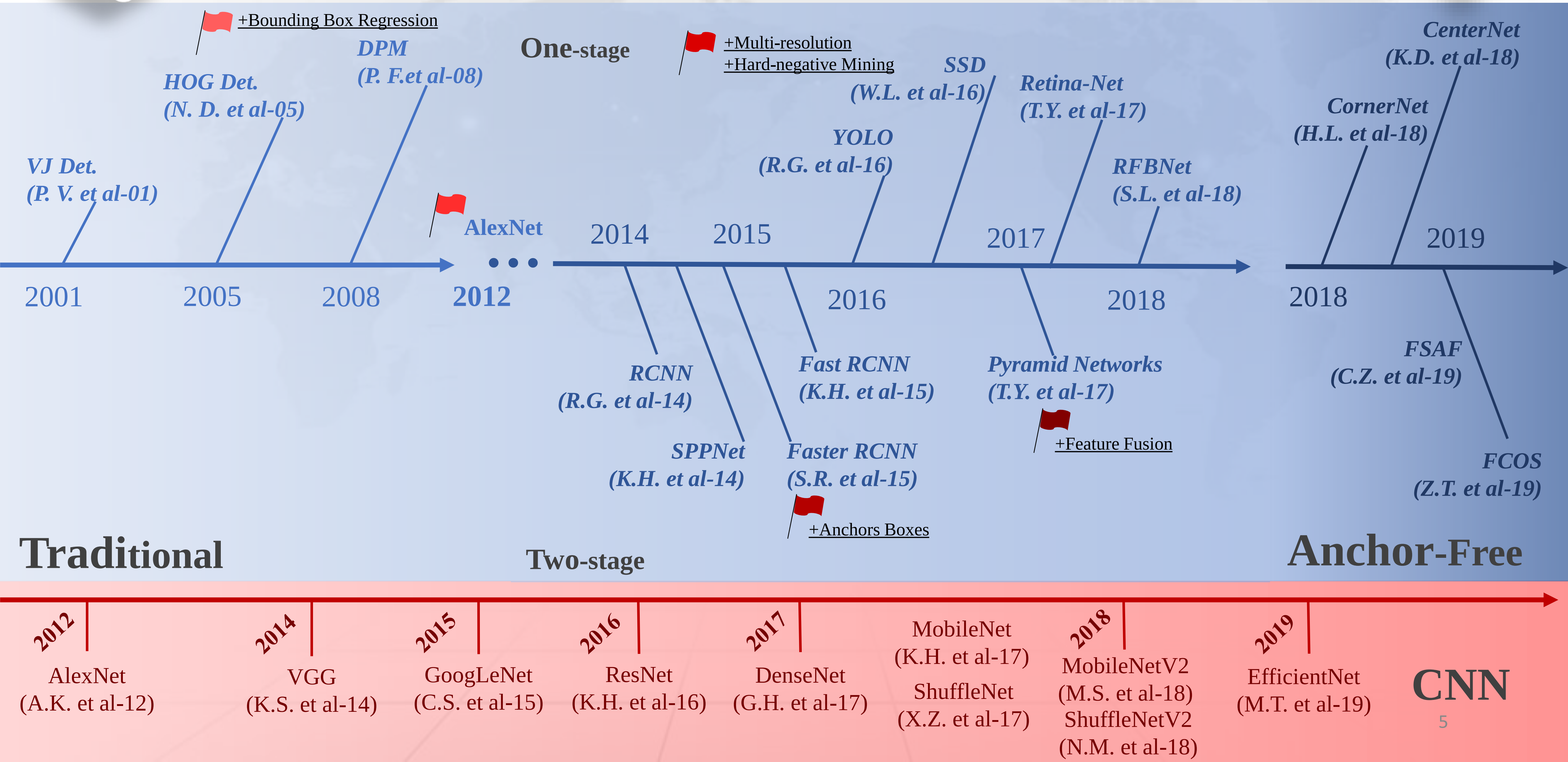
- **Low resolution**
- **Unobvious feature**



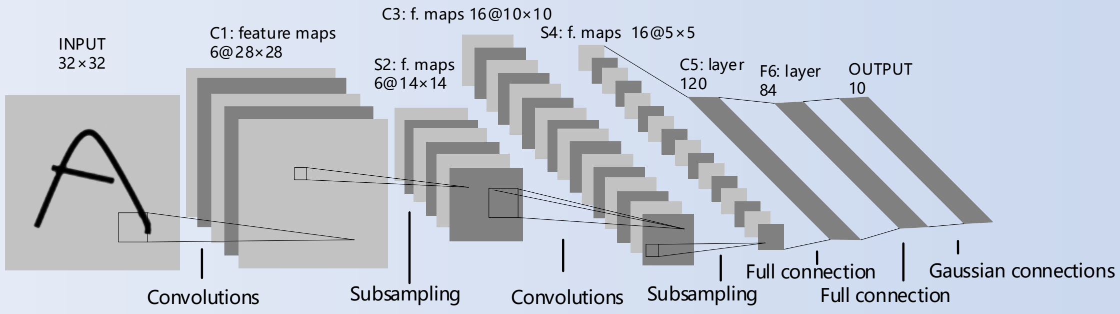



Feature extraction

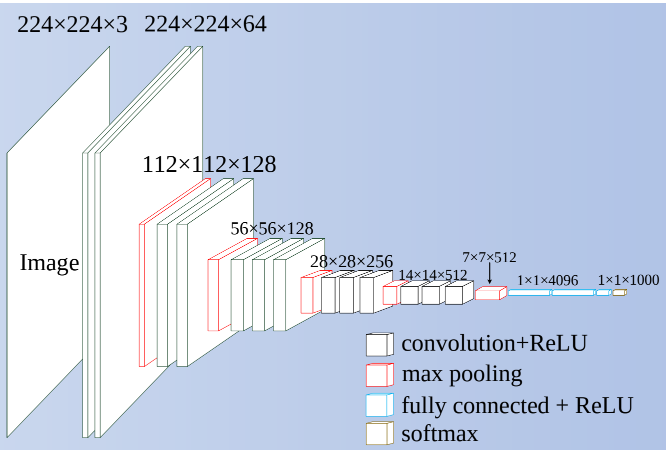
# The brief history



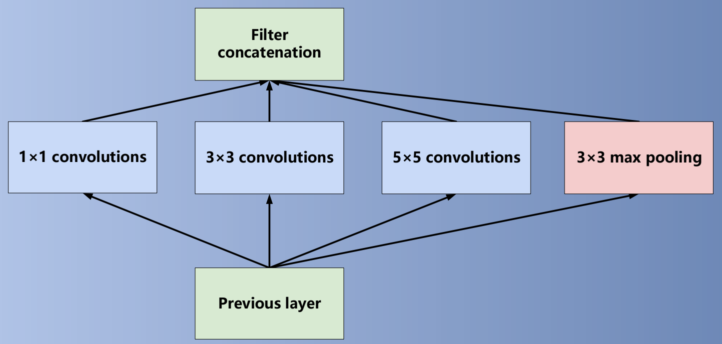
# CNN



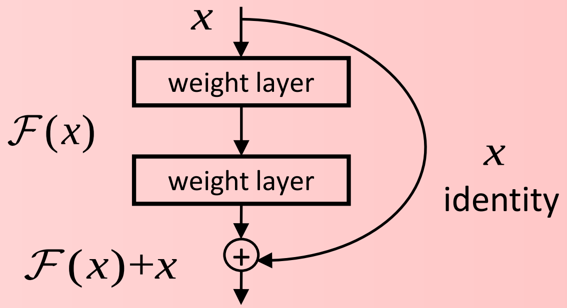

**LeNet**
**(Y.L. et al-98)**



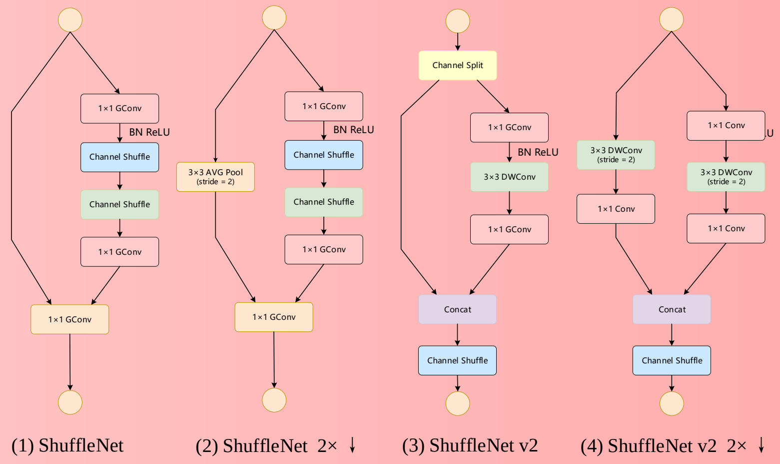

**VGG**
**(K.S. et al-14)**




**GoogLeNet**
**(C.S. et al-15)**

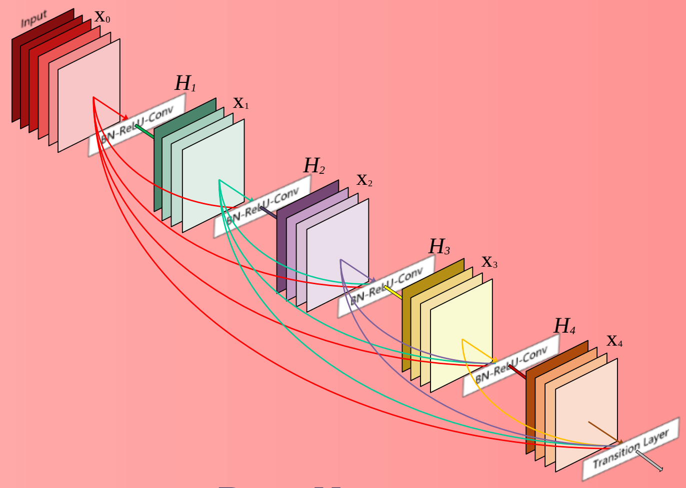



**ResNet**
**(K.H. et al-16)**



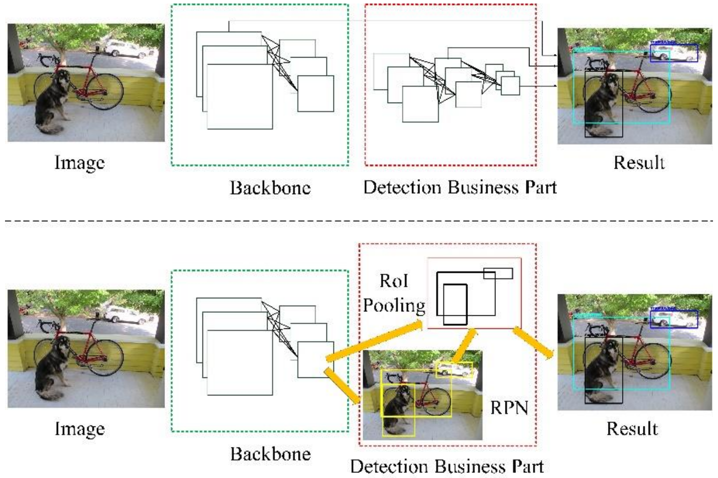

**ShuffleNet**
**(X.Z. et al-17)**

**ShuffleNetV2**
**(N.M. et al-18)**

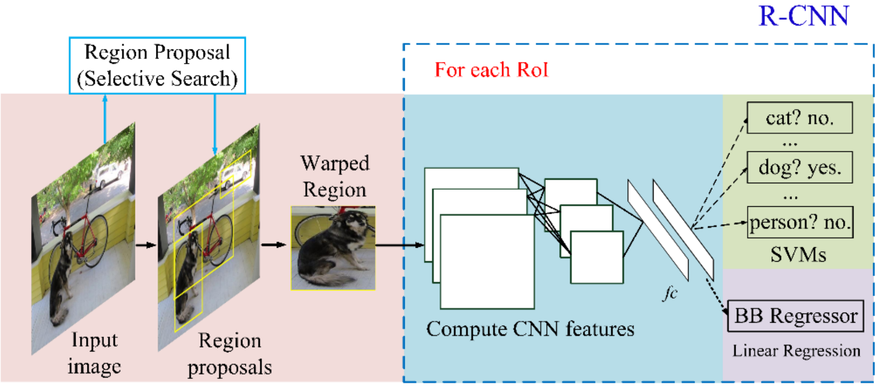


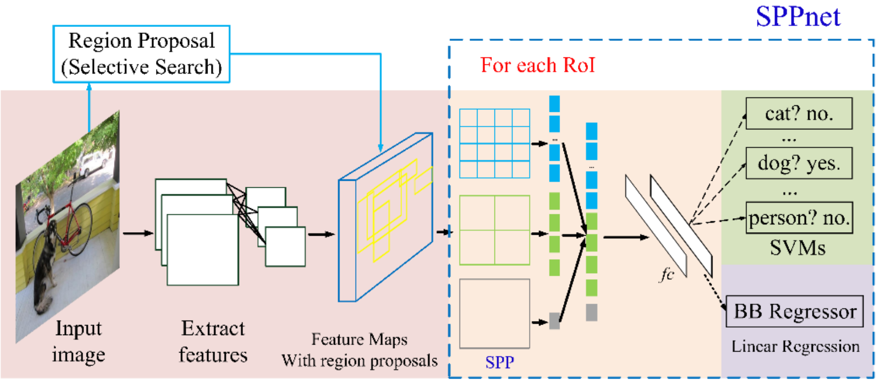

**DenseNet**
**(G.H. et al-17)**

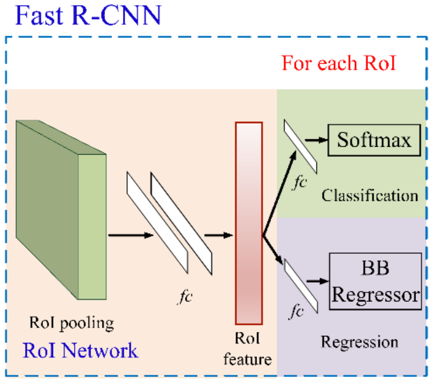
# Object detection

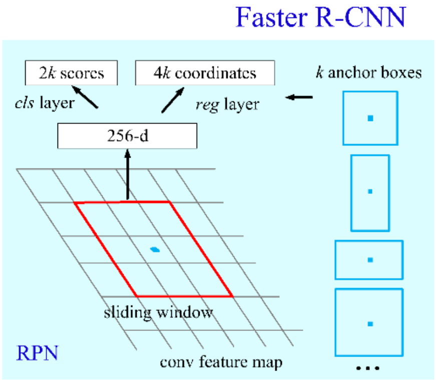


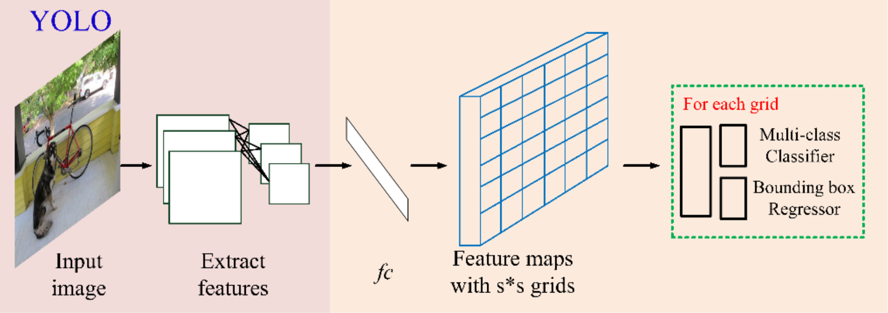

(H.Zhang. et al-19)

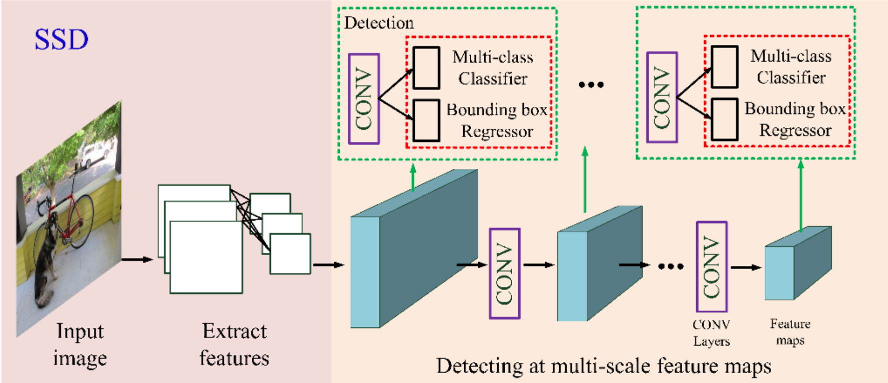


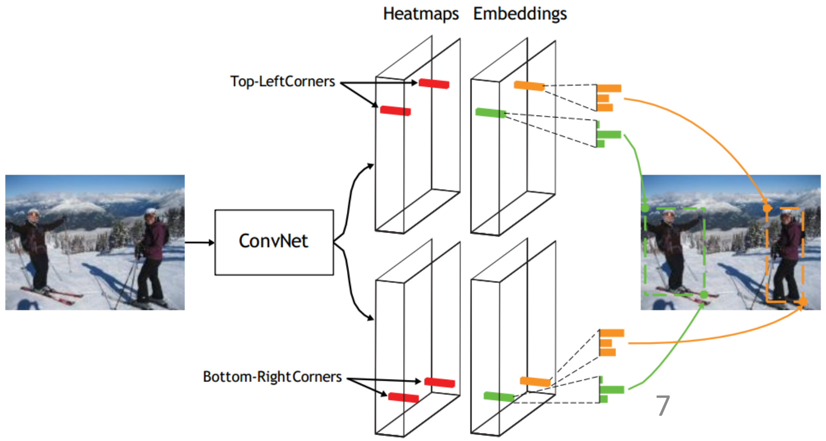



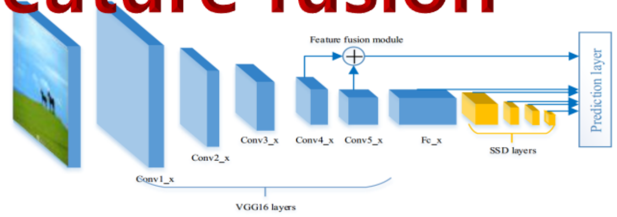



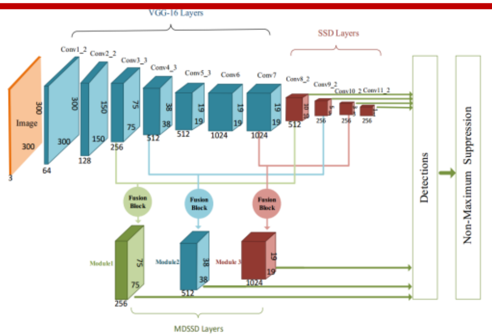



# Small object detection models

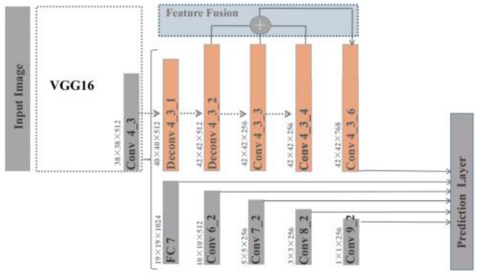
## Feature fusion




Feature-fused SSD (G.C. et al-18)



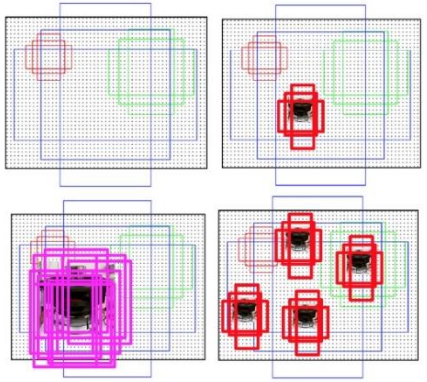

MDSSD (M.X. et al-18)




SFE-SSD(H.T. et al-19)

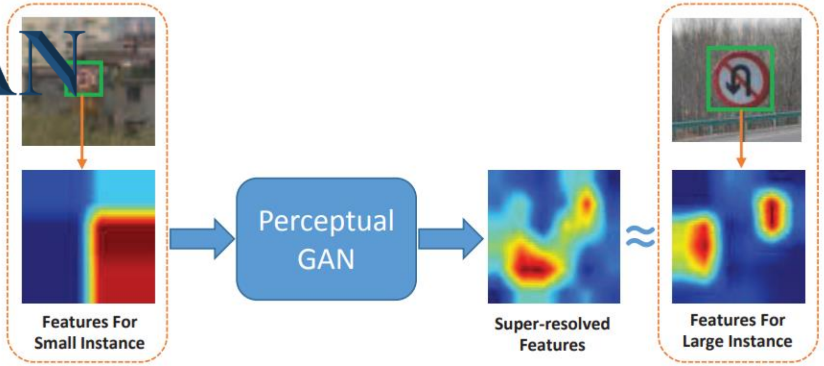

FPN (T.L. et al-17)

## Data

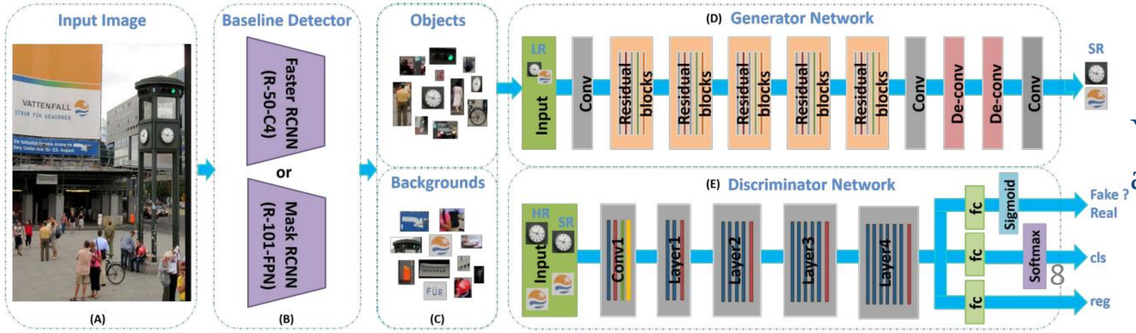



M.K. et al-19

## GAN

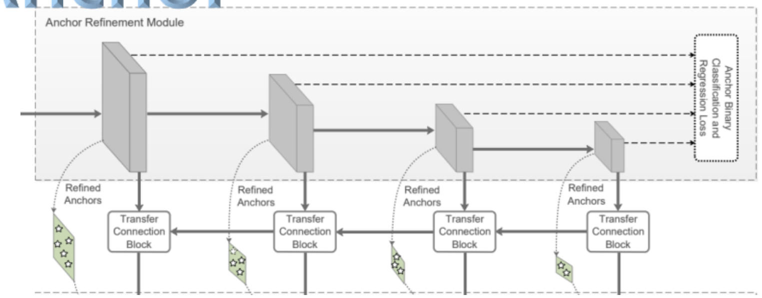



J.L. et al-17

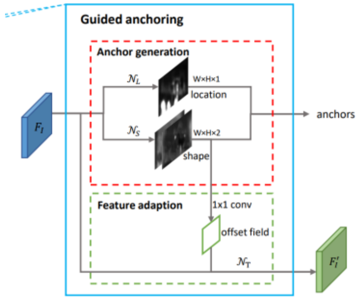



Y.B. et al-18

## Anchor




RefineDet(S.Z. et al-18)

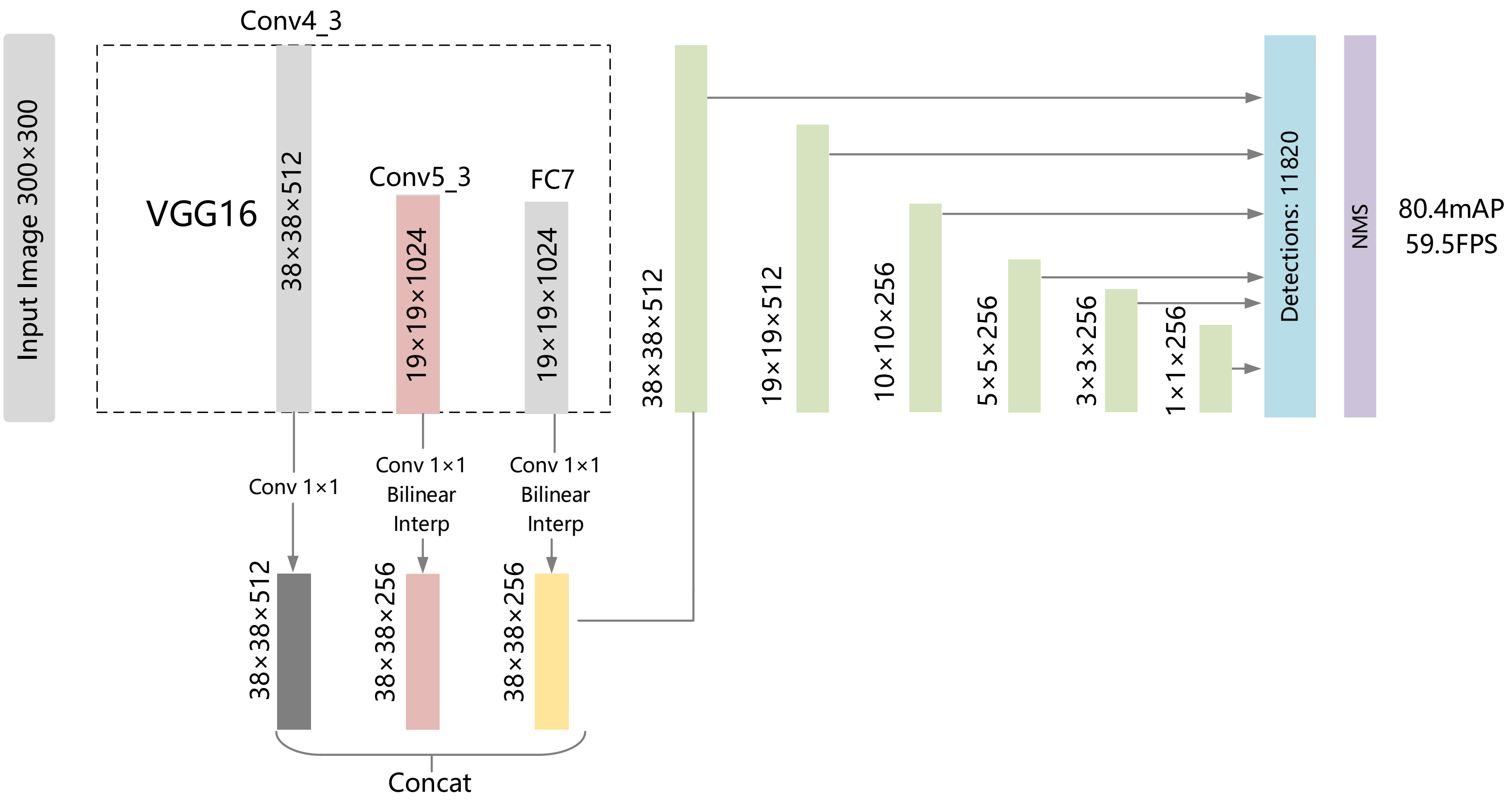



(J.W. et al-19)

# MFSSD

## MFSSD

1. combine the features from the original Conv4_3, Conv5_3 and FC7 layers from VGG-16 [21] to get better and robust features
2. set the channel number of Conv4_3 layer to 512
3. Using the BN layer during training
4. Model compression after the training



# Model compression

After the BN layer, each channel of the feature map will correspond to a scaling factor, and the size of the scaling factor represents the importance of the channel. During training, the network weights and scaling factors are jointly trained, and finally the channels with smaller scaling factors are removed to achieve network pruning. W is the network weight, γ is the set scaling factor, and λ is a hyper-parameter used to balance the loss caused by the network weight and the loss introduced by the scaling factor:

$$L=\sum_{(x,y)} l(f(x,W),y)+\lambda\sum_{\gamma\in f} g(\gamma)$$

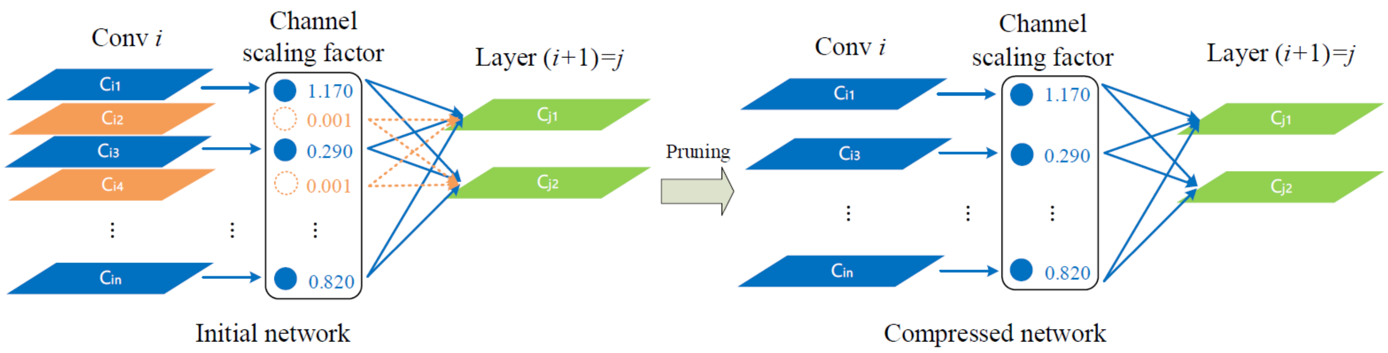



Fig.2 Net Slimming

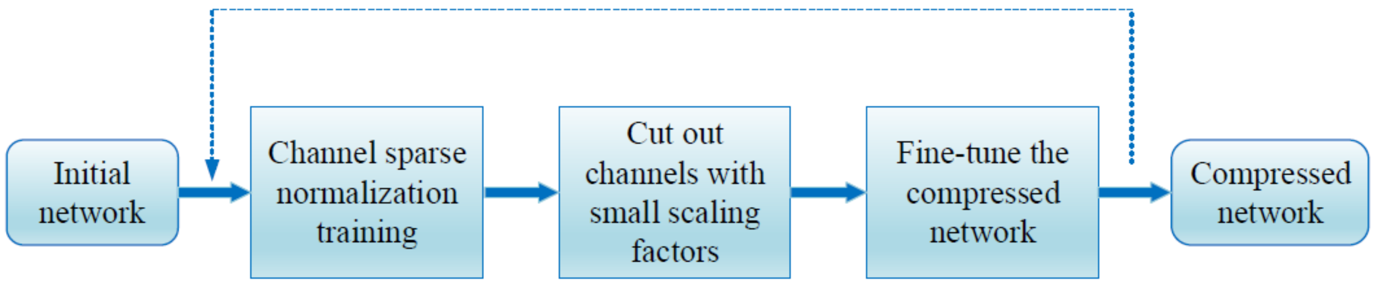



Fig. 3. The procedure of model compression through net slimming.

# Results

TABLE I. PARAMETERS COMPARISION BEFORE/AFTER PRUNING

| Model | Parameters | Size (MB) | After pruning |
|---|---|---|---|
| FSSD[15] | 34,132,960 | 130.21 | - |
| MFSSD | 33,709,664 | 128.59 | - |
| MFSSD-BN | 33,713,888 | 128.61 | - |
| MFSSD-Pruned | **22,785,625** | **86.92** | **66.7%** |

TABLE III. COMPARISION OF SPPED N PASCAL VOC2007

| Model | bone | mAP | FPS | #Proposals |
|---|---|---|---|---|
| SSD | VGG | 77.4 | 72 | 8732 |
| MDSSD | VGG | 78.6 | 38.5 | 8732 |
| FSSD | VGG | 78.8 | 65.8 | 8732 |
| SSD | VGG | 77.7 | 70.8 | 11620 |
| FFSSD | VGG | 78.6 | 60.8 | 11620 |
| RFB-Net | VGG | **80.5** | 49 | 11620 |
| MFFSD | VGG | **80.4** | 59.5 | 11620 |
| MFSSD-P | VGG | 79.9 | **79.5** | 11620 |

MFSSD can achieve **80.4%** on PASCAL VOC2007, and after pruning, MFSSD_P can get **33% acceleration** in speed with lite lost in accuracy, reaching a balance between inference speed and accuracy.

# Results

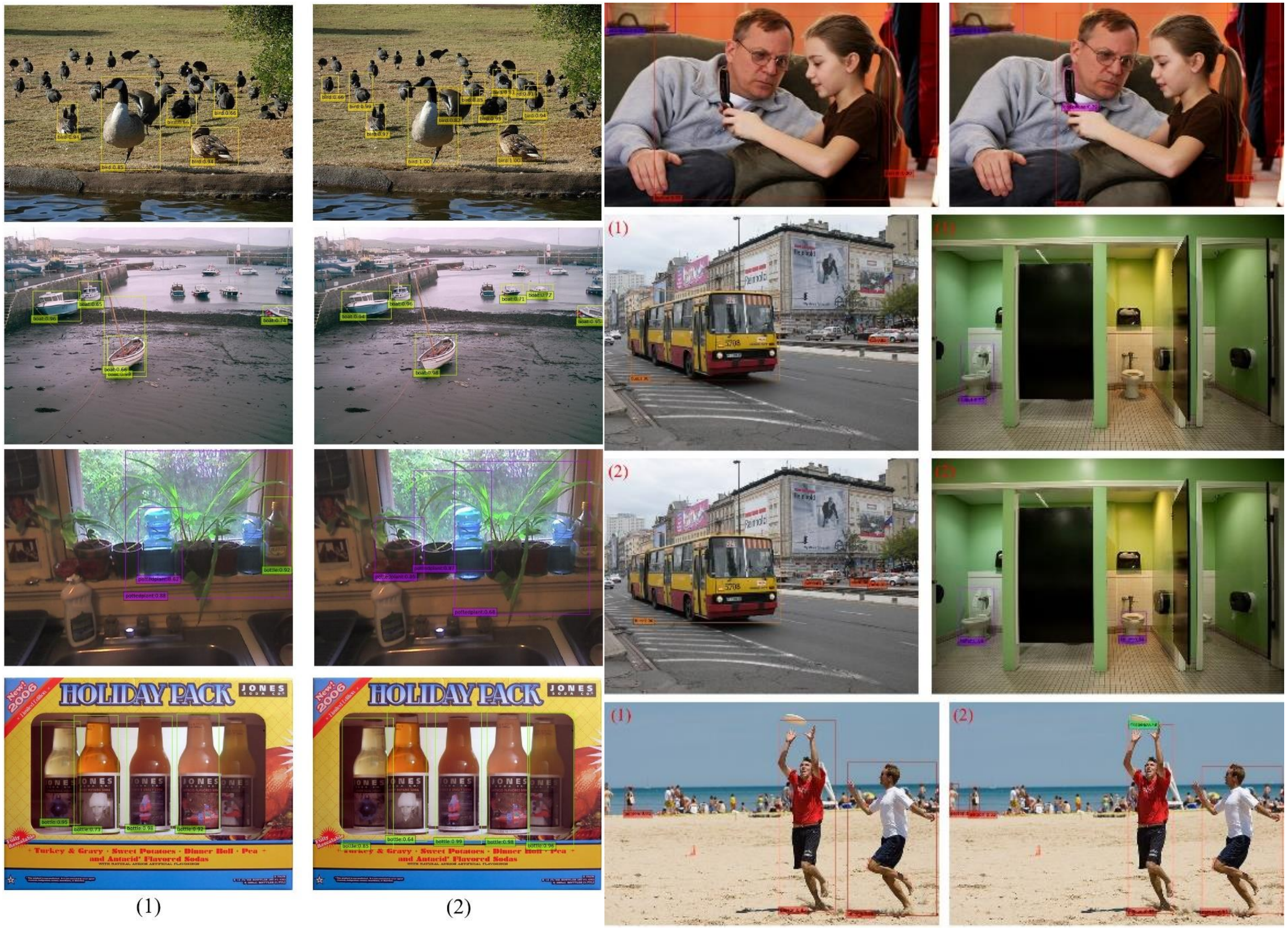
TABLE III. ACCURACY COMPARISON ON MS COCO

| model | **bone** | **Avg. Precision, IoU** | | | **Avg. Precision, Area:** | | | **Avg. Recall, #Dets:** | | | **Avg. Recall, Area:** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.5:0.95 | 0.5 | 0.75 | S | M | L | 1 | 10 | 100 | S | M | L |
| FSSD | VGG | 27.1 | 47.7 | 27.8 | 8.7 | 29.2 | 42.2 | 24.6 | 37.4 | 40 | 15.9 | 44.2 | 58.6 |
| MDSSD | VGG | 26.8 | 45.9 | 27.7 | 10.8 | 27.5 | 40.8 | 24.3 | 36.6 | 38.8 | 15.8 | 42.3 | 56.3 |
| DSSD | ResNet | 28 | 46.1 | 29.2 | 7.4 | 28.1 | **47.6** | 25.5 | 37.1 | 39.4 | 12.7 | 42 | 62.6 |
| RefineDet | VGG | 29.4 | 49.2 | 31.3 | 10 | 32 | 44 | - | - | - | - | - | - |
| RFB-Net | VGG | **30.3** | 49.3 | **31.8** | 11.8 | 31.9 | 45.9 | - | - | - | - | - | - |
| MFSD | VGG | 29.4 | **49.8** | 30.6 | **12.1** | **33.5** | 44.1 | **26** | 39.3 | 41.9 | **18.7** | 48.1 | 59.7 |

The average accuracy of MFSSD in this paper is **29.4%**

the performance on small objects reaches the best accuracy of **12.1%**

# Results



# Conclusion

This paper proposes a small object detection algorithm MFSSD based on FSSD, meanwhile, in order to reduce the computational cost and storage space, pruning is carried out to achieve model compression.

- Firstly, the semantic information contained in the features of different layers can be used to detect different scale objects, and the feature fusion method is improved to obtain more information beneficial to small objects;
- secondly, batch normalization layer is introduced to accelerate the training of neural network and make the model sparse;
- finally, the model is pruned by scaling factor to get the corresponding compression model.

The experimental results show that the algorithm can not only improve the detection accuracy of small objects, but also greatly improves the detection speed, which reaches a balance between speed and accuracy.

# Thank you !