



# 军事场景下的目标定位

赵祈杰

2018.11.24

Final-year Master candidate

实验室：Visual Data Interpreting and Generation Lab(VDIG)

单位：北京大学计算机科学与技术研究所

导师：王勇涛副研究员

个人主页：[qijiezhao.github.io](https://github.com/qijiezhao)

邮箱：[zhaoqijie@pku.edu.cn](mailto:zhaoqijie@pku.edu.cn)



北京大学  
PEKING UNIVERSITY



北京大学计算机科学与技术研究所  
INSTITUTE OF COMPUTER SCIENCE & TECHNOLOGY OF PEKING UNIVERSITY

# 团队介绍

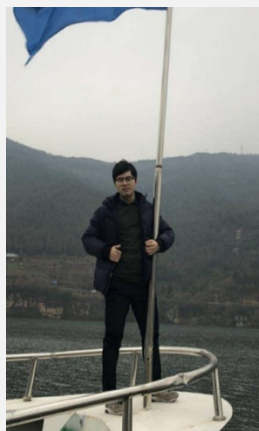
- 队名：VDIG
- 姓名：倪烽，赵祈杰，王勇涛
- 单位：北京大学，计算机科学技术研究所，VDIG lab
- 成绩：检测（第一），分割（第一）



倪烽

北京大学软件与微电子学院  
硕士一年级学生

研究方向：深度学习与计算机视觉，  
物体识别与目标检测



赵祈杰

北京大学计算机科学技术研究所  
硕士三年级学生

研究方向：深度学习与计算机视觉，  
目标检测，语义分割，行为识别等



导师：王勇涛

北京大学计算机科学技术研究所  
副研究员

研究方向：复杂文档图像理解，深度学习与计算机视觉

联系方式：[{nifeng, zhaoqijie, wyt}@pku.edu.cn](mailto:{nifeng, zhaoqijie, wyt}@pku.edu.cn)

# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望

# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望

# 赛题介绍



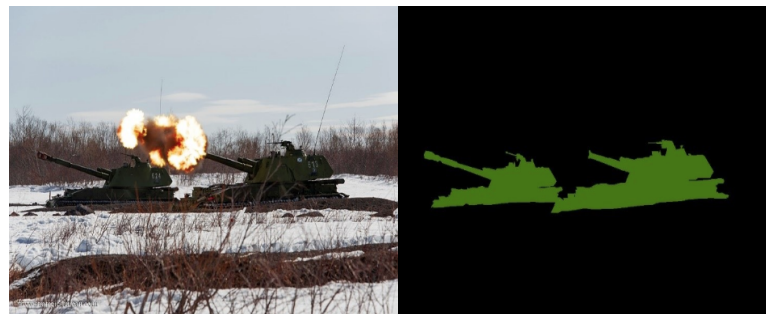
- 背景：在军事领域，由于战场环境复杂、实际战争存在“战争迷雾”等原因导致的**真实数据获取困难和数据噪声极大**等问题，给军事智能的发展带来了极大的挑战。

- 任务：

- **目标检测**：定位出**坦克和装甲车**的目标框位置
- **语义分割**：分割出图像中属于**坦克和装甲车**的区域



(1)图像的目标检测结果



(2)图像的语义分割结果

# 赛题介绍

- 数据集：

- 20000张人工标注的图像，包含坦克和军用装甲车目标的位置矩形框
- 训练集12000张，验证集3000张，测试集5000张

- 问题难点：

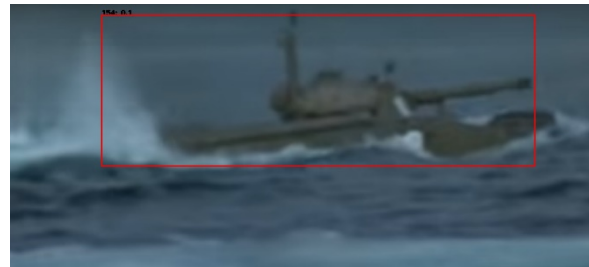
(1) 图像类型不统一，包含3通道、单通道、4通道的图像



(2) 目标尺度变化极大



(3) 难以精准定位边界框



# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望

# 相关工作

## ➤ 目标检测

- 按**回归步骤**可分为：1,单步法; 2,双步法; 3,其他
- 按**功能目标**可分为：1,快速目标检测器; 2,精准优先检测器; 3,其他特定任务的检测器
- 单步法:  
YOLO[1],SSD[2],DSSD[3],RefineDet[4],RetinaNet[5],CFENet[25],M2Det[26]...
- 双步法:  
Faster R-CNN[6],R-FCN[7],Mask R-CNN[8],PANet[9],SNIP[10]...
- 快速目标检测器：  
YOLO(v1,2,3),PELEE[11],RFBNet[12],Light-head R-CNN[13],CFENet...
- 精准优先检测器：  
ResNeXt152 + Mask R-CNN + Cascade R-CNN[14](Cascade rcnn - Detectron)



## 小物体的目标检测



Feature Pyramid Networks

Beyond Skip Connections  
Top Down Modulation

小物体检测

Deformable Convolutional  
Networks

不规则形状物体的检测

Focal Loss

Chained Cascade Network

RON

改进正负样本不均衡的问题

DetNet

refinedet

Peele

RFBNet

修改网络结构以提升效果



soft-NMS

RRC

被遮挡物体检测

MegDet

改进检测mini-batch过小的问题

Relation Networks  
for Object Detection

关注物体之间关联性信息

## Object Detection

## 单步法vs双步法

Stage

one-stage

YOLO

YOLO9000

YOLOv3

Object detection  
at 200 Frames Per Second

SSD

DSSD

DSOD

two-stage

Fast-RCNN/Faster-RCNN

Mask-RCNN

R-FCN

R-FCN-3000

logo检测

Scalable Object Detection  
for Stylized Objects

实例分割

PAN

目标检测用于视频分段

Rethinking the Faster R-CNN Architecture  
for Temporal Action Localization

调研速度和准确率平衡的综述

Speed/accuracy trade-offs for  
modern convolutional object detectors



表示可有效提升检测器至State-of-the-art精度的方法

# 相关工作

## ➤ 语义分割

- 分割精度优先

FCN[15], DeepLab(v1,v2,v3,v3+)[16], RefineNet[17], PSPNet[18], Large Kernel Matters[19], ExFuse[20], UperNet[21]

- 分割效率优先：

ICNet[22], ESPNet[23], BiSENet[24]

## ➤ 实例分割、姿态估计(关键点预测)等:

- 基于RoI的方法都适用，比如Mask R-CNN[8], PANet[9]

# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望

# 方法介绍：CFENet

- 介绍：

CFENet是一个基于SSD设计的单步检测器，通过增加专门设计的**CFE模块**在网络中对检测物体最敏感的位置进行特征增强，同时对检测器的速度影响很小。即：一个**快速而精准**的单步检测器。

- Paper:

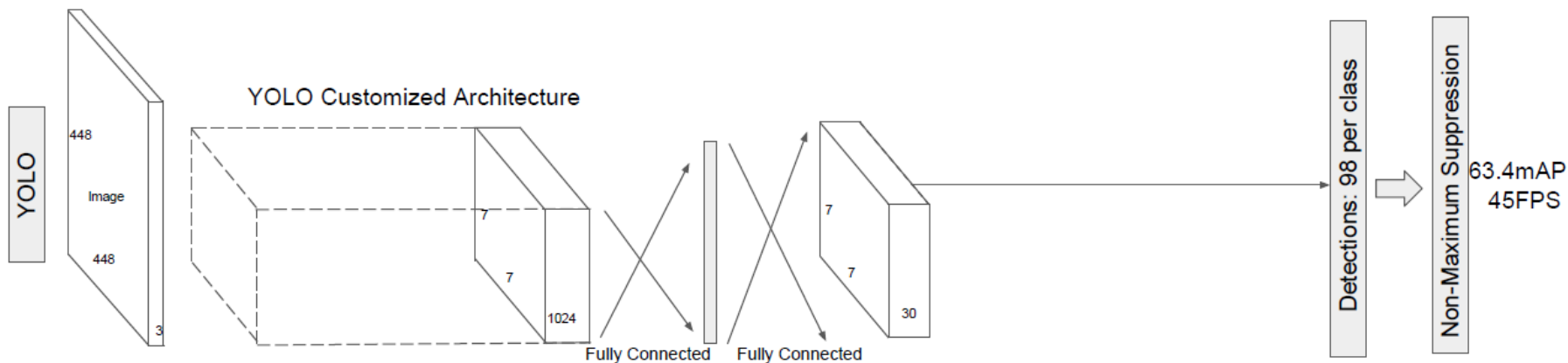
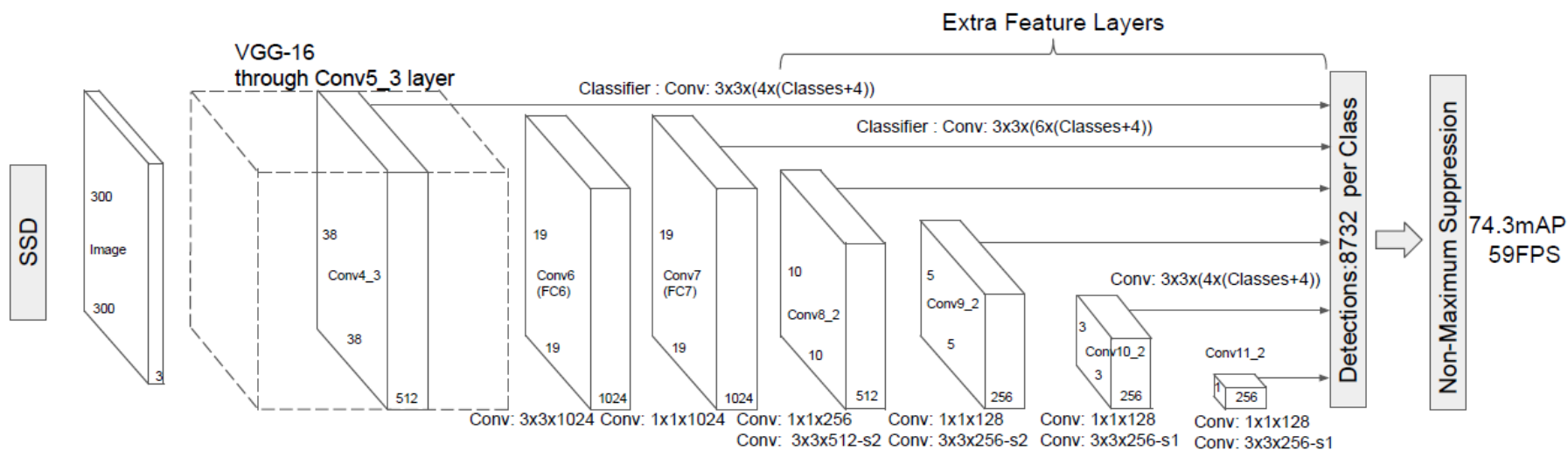
**Qijie Zhao, Yongtao Wang, Tao Sheng et. al. Comprehensive Feature Enhancement Module for Single-Shot Object Detector.** *Asian Conference on Computer Vision (ACCV-18)*. [\[pdf\]](#)

**Qijie Zhao, Tao Sheng, Feng Ni, et. al. CFENet: An Accurate and Efficient Single-Shot Object Detector for Autonomous Driving.** *Computer Vision and Pattern Recognition (CVPR2018 Workshop), Workshop of Autonomous Driving technique report.*

- Code:

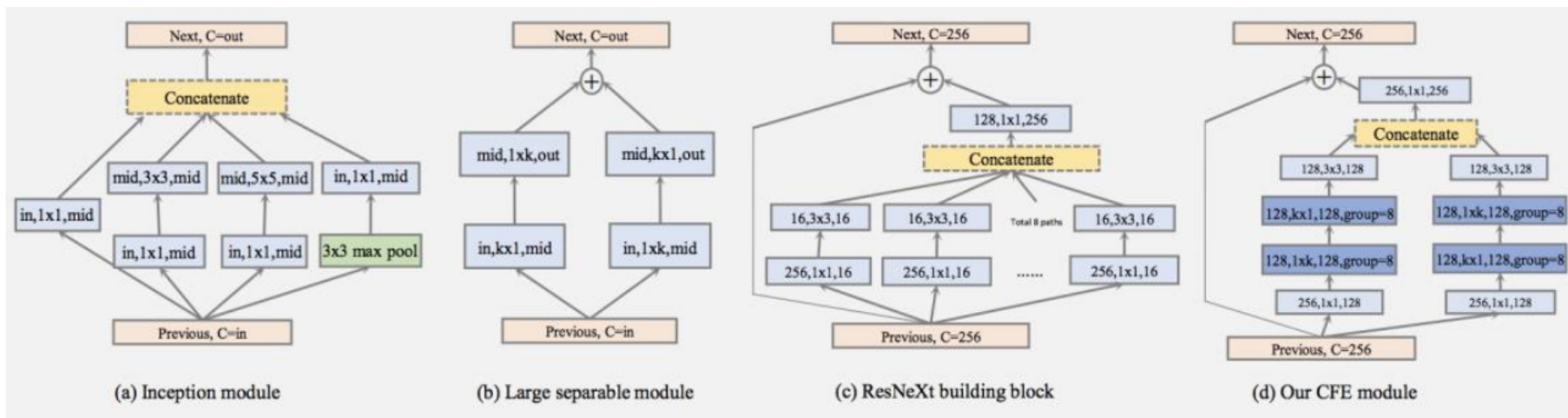
<https://github.com/qijiezhao/CFENet>

# Review the architecture of SSD and YOLO



# 方法介绍：CFENet

## • CFE module



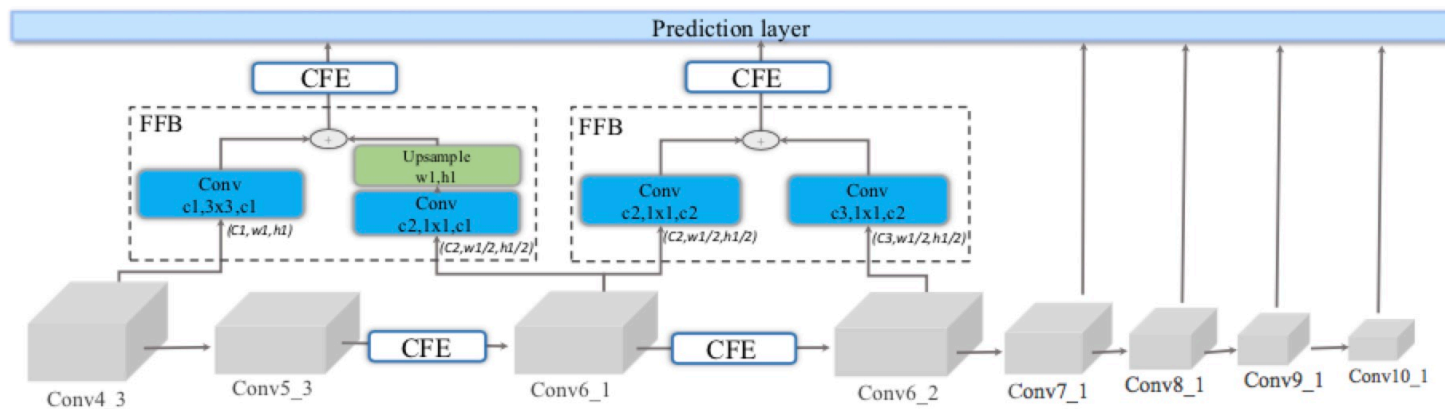
➤ Inception:  
Multi-branch  
Short cut learning  
Feature ensemble

➤ Large kernel:  
Symmetric learning  
Receptive field

➤ Group Conv:  
Less redundancy  
Cardinality  
Efficiency

CFE module:  
**All**

# 方法介绍：CFENet



**How** to design the CFENet ? (怎么设计CFENet)

通过在我们提出的拓扑结构里插入CFE模块，使得网络可以得到表达能力更强的特征。

**Where** to assemble CFE modules? (应该把CFE模块插在什么位置)

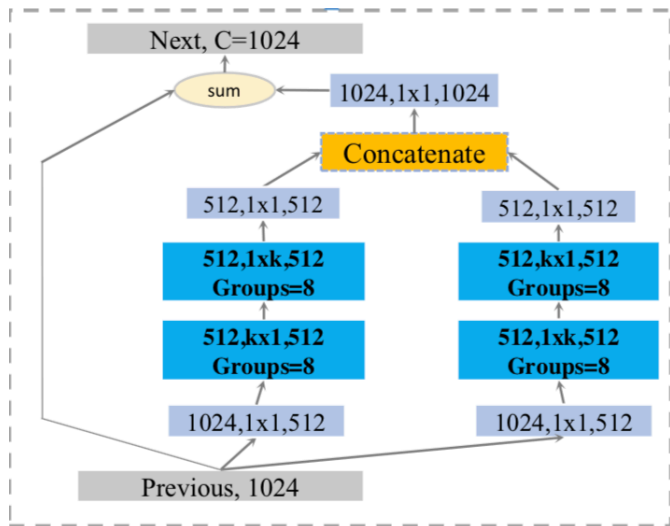
检测小物体的大尺度特征图处于网络中较浅位置，缺乏语义信息，感受野有限。  
利用特征融合。  
利用浅层的获益补充深层特征。

**Why** no more CFE modules? (为什么不增加更多的CFE模块)

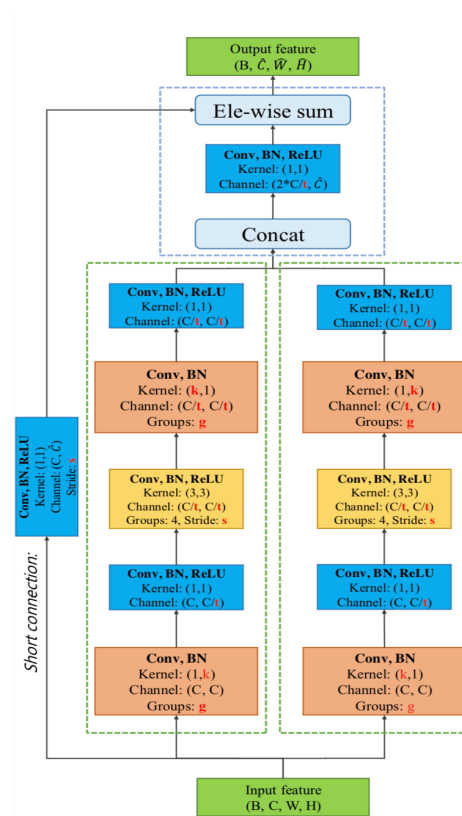
CFE模块不能加太多，因为这会让模型速度变慢、会让网络的训练变得困难。  
当前的方案只丢失一定的速度，但提升了很大的精度。

# 方法介绍：CFENet

我们的后续工作在继续升级CFE模块：  
CFEv1 v.s. CFEv2



(a).CFEv1模块

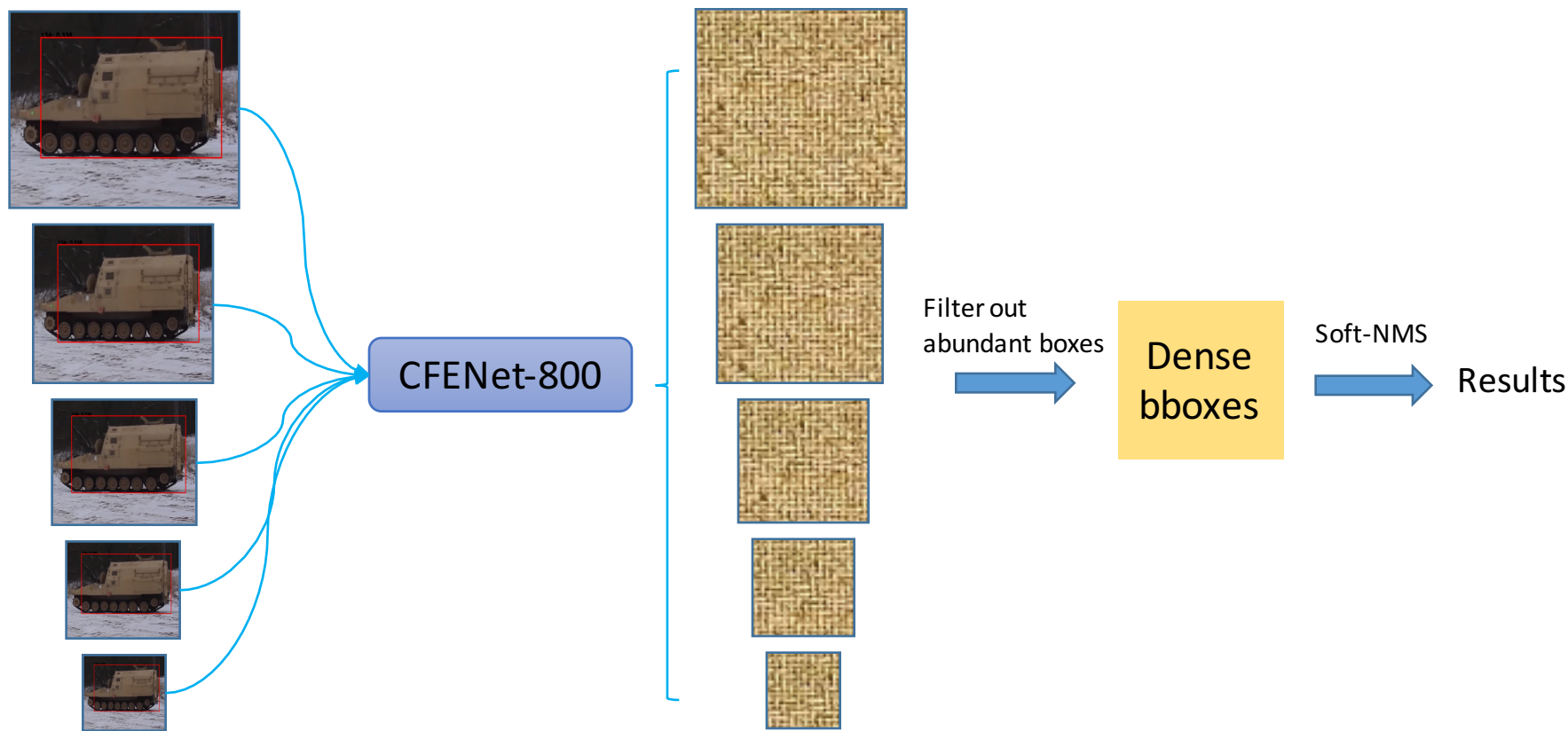


(a).CFEv2模块

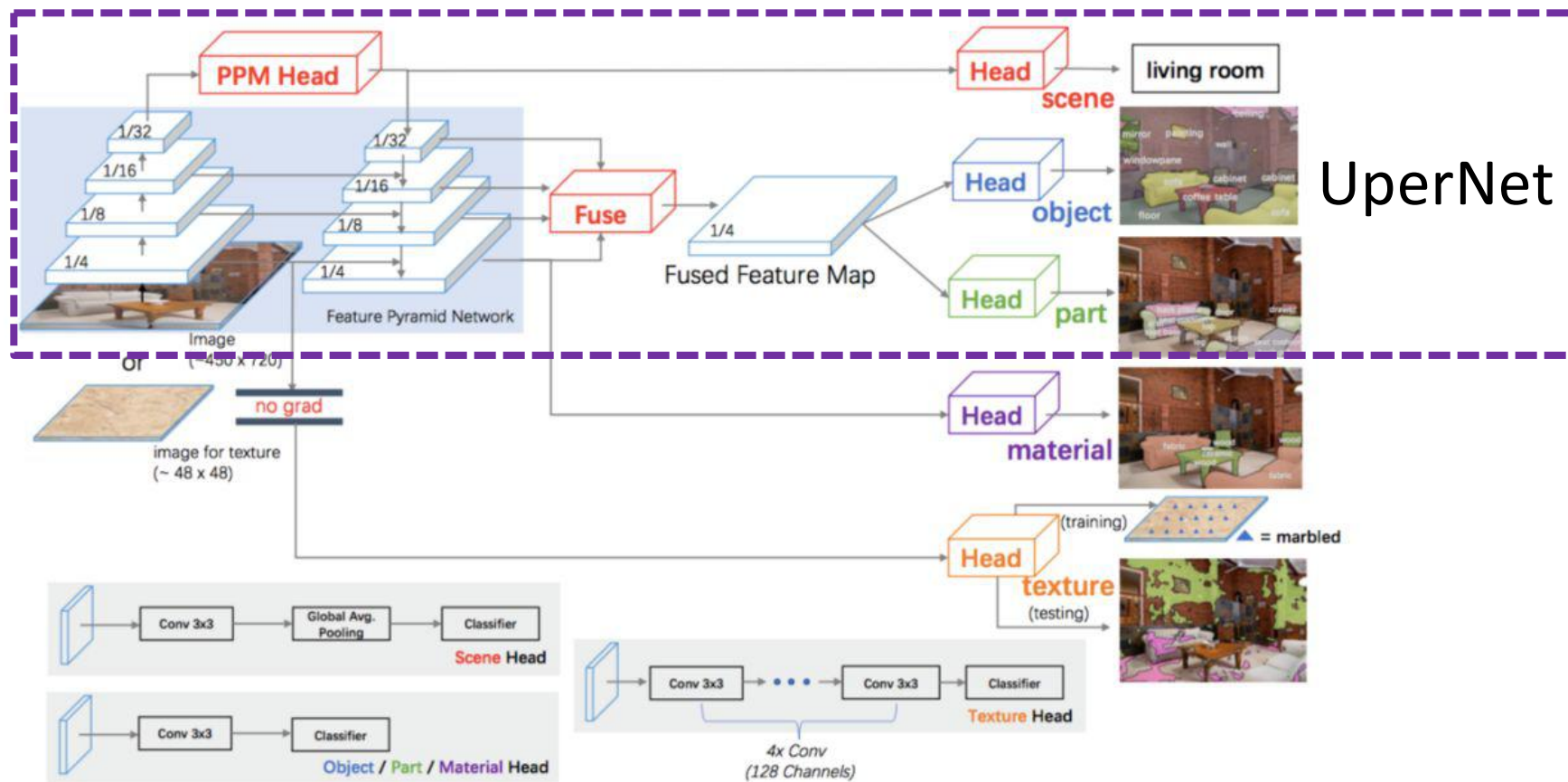


# 方法介绍：多尺度测试

- 单尺度的CFENet很快，精度相比SSD等单步检测器已经足够高。但是要和state-of-the-art比精度，需要用到多尺度测试。



## 方法介绍：UperNet

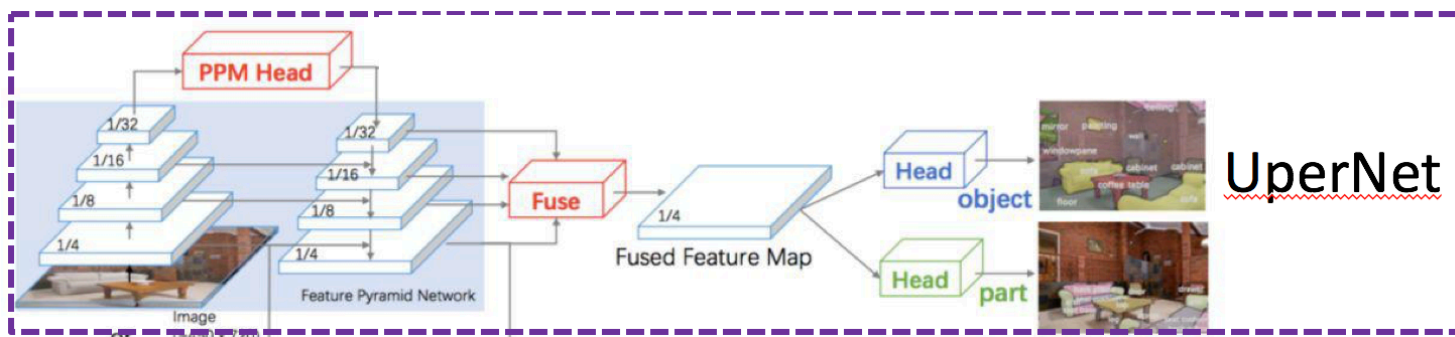


Xiao T, Liu Y, Zhou B, et al. Unified Perceptual Parsing for Scene Understanding[J]. 2018.

# 方法介绍：UperNet

## ➤ UperNet为何擅长语义分割？

- 整体来看，UperNet结合了PSPNet和FPN，来得到表达能力（浅+深，感受野，非线性，全局+局部）更强的特征
- 多尺度特征融合，计算机视觉问题中广泛有效的方法
- 深度监督，由PSPNet提出，较浅层（Res5）也加入监督训练



代码已开源：[www.github.com/CSAILVision/semantic-segmentation-pytorch](https://www.github.com/CSAILVision/semantic-segmentation-pytorch)

# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望

# 实验

- 实验平台和实验环境：
  - 4\*Titan XP GPUs, 48 Gigabytes memory.
  - Ubuntu 16.04, cuda 9.0, cuDNN7.1.4
  - PyTorch 0.4.0
- 训练，测试时长：
  - 训练阶段：~3天（75小时）
  - 测试阶段：单尺度：20图/秒；多尺度：1图/秒

# 实验

- 最后实验结果，检测：

名次	队名	所在单位/高校	成绩(%)
1	VDIG	北京大学	84.1
2	山支队	北京邮电大学	82.2
3	SYSU-XZC	中山大学	81.6

# 实验

- 最后实验结果，分割：

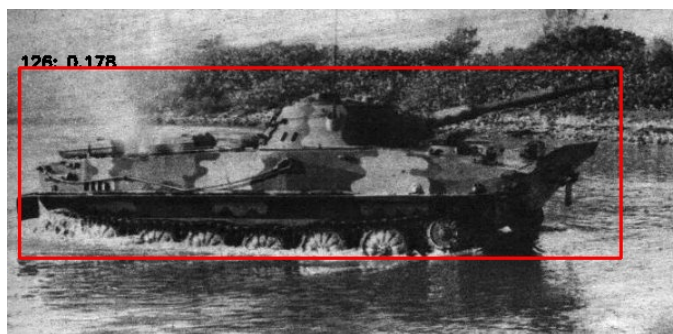
名次	队名	所在单位/高校	成绩(%)
1	VDIG	北京大学	73.55
2	仰望楼上dalao	中山大学	58.76
3	陆战天眼	中国人民解放军陆军工程大学	57.63

# 内容

- 赛题介绍
- 相关工作总结
- 方法介绍
- 实验
- 展望



# 展望—现阶段技术的可视化

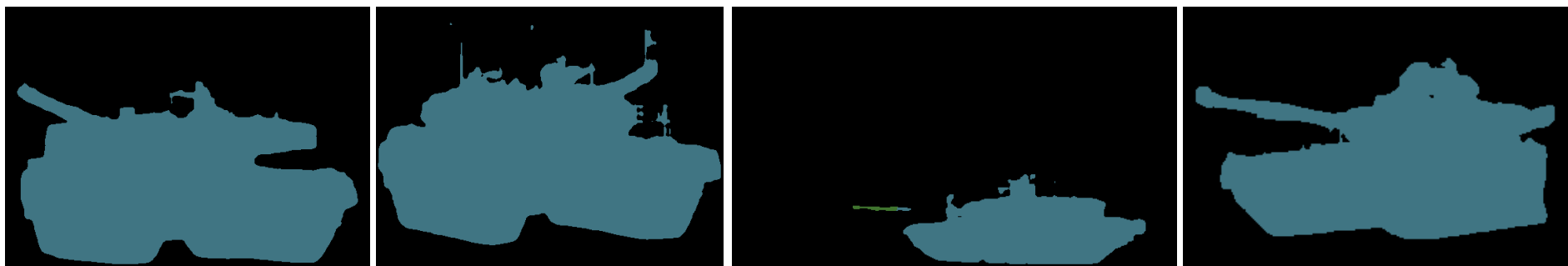


# 展望—现阶段技术的可视化

标注图像



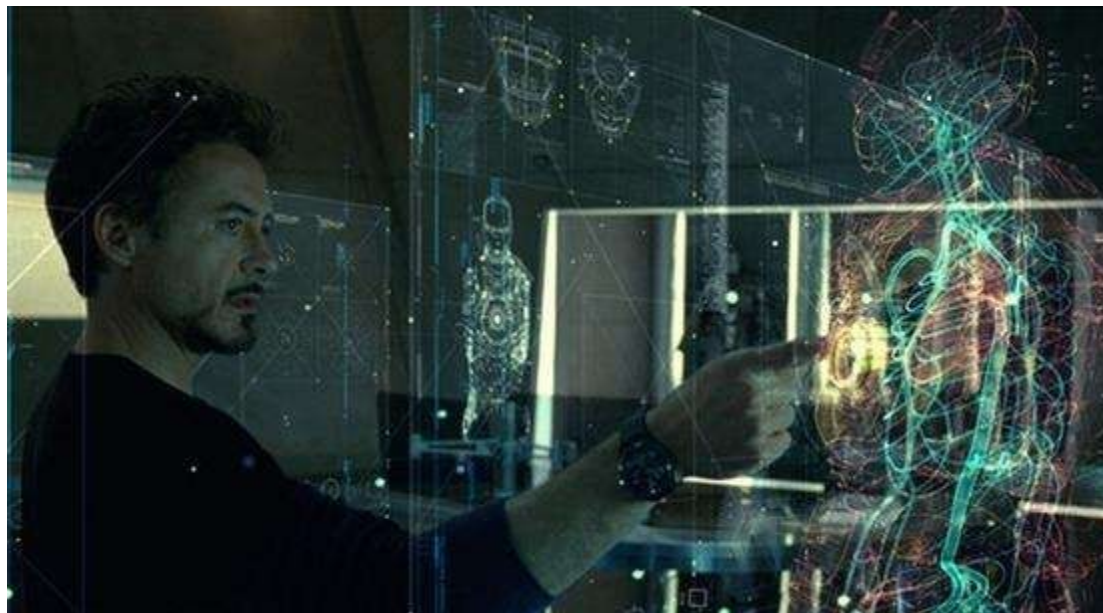
算法的分割结果



# 展望—人工智能(计算机视觉)与军事

➤现阶段的计算机视觉技术对于大部分领域的意义都在于其赋予了一双更明亮的眼睛，可以帮助系统更好的做决策，但仍然还有很长的路要走：

- 单任务单模型问题严重，模型迁移困难
- 过度依赖训练数据
- 现实中的特殊情形(Corner case)比实验环境多得多
- (\*)真正会思考的人工智能尚未起步



The End  
Q&A

\*此Slides 已可在 [qijiezhao.github.io/imgs/prcv18\\_VDIG\\_slides.pdf](http://qijiezhao.github.io/imgs/prcv18_VDIG_slides.pdf) 下载

## References:

- [1]Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2016:779-788.
- [2]Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S. E.; Fu, C.; and Berg, A. C. 2016. SSD: single shot multibox detector. In ECCV 2016, 21–37.
- [3]Fu, C.; Liu, W.; Ranga, A.; Tyagi, A.; and Berg, A. C. 2017. DSSD : Deconvolutional single shot detector. CoRR abs/1701.06659.
- [4]Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; and Li, S. Z. 2018. Single-shot refinement neural network for object detection. In IEEE CVPR.
- [5]Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [6]Ren, S.; He, K.; Girshick, R. B.; and Sun, J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. In NIPS 2015, 91–99.
- [7]Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[J]. 2016.
- [8]He, K.; Gkioxari, G.; Dollar, P.; and Girshick, R. 2017. Mask R-CNN. In ICCV 2017, 2980–2988. IEEE.
- [9]Liu S , Qi L , Qin H , et al. Path Aggregation Network for Instance Segmentation[J]. 2018.
- [10]Singh, B., and Davis, L. S. 2018. An analysis of scale invariance in object detection–snip. In CVPR 2018, 3578–3587.
- [11]Wang R J , Li X , Ao S , et al. Pelee: A Real-Time Object Detection System on Mobile Devices[J]. 2018.
- [12]Liu S, Huang D, Wang Y. Receptive Field Block Net for Accurate and Fast Object Detection[J]. 2017.
- [13]Li, Zeming, Peng, Chao, Yu, Gang, et al. Light-Head R-CNN: In Defense of Two-Stage Object Detector[J]. 2017.
- [14]Cai, Z., and Vasconcelos, N. 2018. Cascade r-cnn: Delving into high quality object detection. In CVPR 2018.

- [15]Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, PP(99):1-1.
- [16]Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. TPAMI 40(4), 834{848 (2018)
- [17]Lin, G., Milan, A., Shen, C., Reid, I.: Refinenet: Multi-path refinement networks for high-resolution semantic segmentation (2016)
- [18]Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR. pp. 6230{6239 (2017)
- [19]Peng C, Zhang X, Yu G, et al. Large Kernel Matters — Improve Semantic Segmentation by Global Convolutional Network[J]. 2017.
- [20]Zhang Z, Zhang X, Peng C, et al. ExFuse: Enhancing Feature Fusion for Semantic Segmentation[J]. 2018.
- [21]Xiao T, Liu Y, Zhou B, et al. Unified Perceptual Parsing for Scene Understanding[J]. 2018.
- [22]Zhao H, Qi X, Shen X, et al. ICNet for Real-Time Semantic Segmentation on High-Resolution Images[J]. 2017.
- [23]Duggan J, Browne J. ESPNET: expert-system-based simulator of Petri nets[J]. IEE Proceedings D - Control Theory and Applications, 1988, 135(4):239-247.
- [24]Yu C, Wang J, Peng C, et al. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation[J]. 2018.
- [25]Qijie Zhao et al. Comprehensive Feature Enhancement Module for Single-Shot Object Detector. Asian Conference on Computer Vision (ACCV-18)
- [26]Qijie Zhao et al. M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network. AAAI Conference on Artificial Intelligence (AAAI-19)