

# 基于深度卷积网络的目标检测综述

吴 帅<sup>1</sup> 徐 勇<sup>1</sup> 赵东宁<sup>1,2</sup>

**摘 要** 在基于区域的卷积神经网络提出后,深度卷积网络开始在目标检测领域普及,更快的基于区域的卷积神经网络将整个目标检测过程合成在一个统一的深度网络框架上。随后 YOLO 和 SSD 等目标检测框架的提出进一步提升目标检测的效率。文中系统总结基于深度网络的目标检测方法,归为 2 类:基于候选窗口的目标检测框架和基于回归的目标检测框架。基于候选窗口的目标检测框架首先需要在输入的图像上产生很多的候选窗口,然后对这些候选窗口进行判别。这里的判别包括:对窗口包含物体的类别(包括背景)进行判断、对窗口的位置进行回归。基于回归的目标检测方法将图像目标检测看作是一个回归的过程。在此基础上,在 PASCAL\_VOC 和 COCO 等主流数据库上对比目前两类目标检测框架中的主流方法,分析两类方法各自的优势。最后根据当前深度网络目标检测方法的发展趋势,对目标检测方法未来的研究热点做出合理预测。

**关键词** 深度卷积网络,目标检测,候选窗口,感兴趣区域(ROI)池化

**引用格式** 吴 帅,徐 勇,赵东宁.基于深度卷积网络的目标检测综述.模式识别与人工智能,2018,31(4):335-346.

**DOI** 10.16451/j.cnki.issn1003-6059.201804005

中图法分类号 TP 391.4

## Survey of Object Detection Based on Deep Convolutional Network

WU Shuai<sup>1</sup>, XU Yong<sup>1</sup>, ZHAO Dongning<sup>1,2</sup>

**ABSTRACT** Deep convolutional network is prevalent in object detection task. Region-based convolutional neural network (RCNN) bridges the gap between the classification of deep convolutional network and the object detection task well. Then the whole object detection process is aggregated into a unified deep framework by Faster-RCNN. You only look once (YOLO) and single shot multibox detector (SSD) effectively improve the efficiency of object detection. Different deep object detection frameworks are comprehensively analyzed and divided into two categories: the proposal based framework and the regression based framework. The proposal based framework is utilized to generate thousands of candidate proposals and then classification and bounding box regression are conducted on these proposals. The regression based framework outputs the bounding box position through some special iterations directly. Furthermore, the advantage for different kinds of frameworks is demonstrated through adequate experiments on the mainstream database like PASCAL\_VOC and COCO. Finally, the development direction of object detection is discussed.

**Key Words** Deep Convolutional Network, Object Detection, Candidate Proposals, Region of Interest (ROI) Pooling

**Citation** WU S, XU Y, ZHAO D N. Survey of Object Detection Based on Deep Convolutional Network. Pattern Recognition and Artificial Intelligence, 2018, 31(4): 335-346.

收稿日期: 2018-01-15; 录用日期: 2018-02-06

Manuscript received January 15, 2018;

accepted February 6, 2018

本文责任编辑 黄华

Recommended by Associate Editor HUANG Hua

1. 哈尔滨工业大学 深圳研究生院 智能感知与生物信息学创新团队 深圳 518000

2. 深圳大学 信息工程学院 深圳 518000

1. IntelliSense and Bioinformatics Innovation Team, HIT Institute of Technology Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055

2. College of Information Engineering, Shenzhen University, Shenzhen 518000

目标检测是进行场景内容理解等高级视觉任务的前提,已应用于智能视频监控、基于内容的图像检索、机器人导航和增强实现等任务中<sup>[1-8]</sup>。相比视频中的目标检测,静态图像目标检测具有更大的难度、更高的挑战性<sup>[9-16]</sup>。传统的目标检测方法主要使用方向梯度直方图(Histogram of Oriented Gradient, HOG)<sup>[17-19]</sup>、尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)<sup>[20-22]</sup>特征对滑动窗口进行判别,主要代表方法为部位形变模型(Deformable Part Model, DPM)及其扩展<sup>[23-26]</sup>。由于滑动窗口需要大量的计算开销,基于候选窗口的目标检测方法后来居上,目前较通用的候选窗口产生方法包括选择提取(Selective Search)<sup>[27]</sup>、边缘窗口(Edge-Box)<sup>[28]</sup>等。

近年来,深度卷积网络在计算机视觉领域取得突破性发展。深度卷积网络主要通过权值共享策略将网络的层次不断加深,使网络具有更强的解析能力。ImageNet<sup>[29]</sup>、COCO<sup>[30]</sup>等大规模图像数据库的建立极大促进深度网络的发展。具有 7 层的 Alex-Net<sup>[31]</sup>卷积网络在 ImageNet 图像分类大赛中以绝对的优势获得冠军,其有效性也得到越来越多的验证<sup>[32-36]</sup>。随后,VGG<sup>[37]</sup>网络、GoogleNet<sup>[38]</sup>及残差网络<sup>[39]</sup>等将卷积网络推向更深层次,大幅提高网络的性能,使大规模图像分类的准确率提升到很高的水平。

同时,研究人员也开始寻求深度卷积网络在其它领域的扩展。在目标检测方面,基于区域的卷积神经网络(Region Based Convolutional Neural Network, RCNN)<sup>[40]</sup>成功连接目标检测与深度卷积网络,将目标检测的准确率提升到一个新的层次。RCNN 由 3 个独立的步骤组成:产生候选窗口、特征提取、SVM 分类及窗口回归。RCNN 主要采用 Selective Search 的方法产生许多候选窗口,随后将所有产生的候选窗口一次送入深度网络提取特征,最后训练 SVM 分类器对所有候选窗口进行分类及窗口回归。由于 RCNN 分为 3 个独立的过程,所以检测效率很低。基于此种情况,学者们改进 RCNN,提出尺度金字塔池化网络(Spatial Pyramid Pooling Net, SPPnet)<sup>[41]</sup>和快速基于区域的神经网络(Fast Region Based Convolutional Neural Network, Fast-RCNN)<sup>[42]</sup>。不需要将所有的候选窗口送入网络,只需将图像送入深度网络一次,再将所有的候选窗口在网络中某层上进行映射,大幅提升模型的检测速度。

更快的基于区域的卷积神经网络(Faster Region-Based Convolutional Neural Network, Faster-RCNN)<sup>[43]</sup>利用候选窗口网络(Region Proposal Network, RPN)

产生候选窗口,利用与 Fast-RCNN 相同的结构进行分类及窗口回归,RPN 和 Fast-RCNN 共享主要的深度网络。Faster-RCNN 将目标检测合成一个统一的深度网络框架。基于区域的全卷积网络(Region Based Fully Convolutional Network, RFCN)<sup>[44]</sup>在此基础上进行进一步改进,分析发现感兴趣区域(Region of Interest, ROI)池化后的网络层不再具有平移不变性,并且 ROI 池化后层数的多少会直接影响检测效率。因此,RFCN 设计位置敏感的 ROI 池化层,直接对此池化之后的结果进行判别,大幅提升检测效率。YOLO(You Only Look One)<sup>[45]</sup>与 SSD(Single Shot Multibox Detector)<sup>[46]</sup>的提出旨在提升目标检测的检测效率,试图使目标检测达到实时检测的程度。特别地,SSD 能在提升目标检测效率的同时保持检测精度,是在检测精度与检测效率上双赢的算法。

相比传统目标检测方法,基于深度网络的目标检测方法在准确率方面具有明显优势。首先,神经网络是模拟人脑的一种具有自学习功能的网络结构,深度网络的前向计算可看作将物体不断抽象的一个过程,深度网络的高层(靠近输出层)记录事物更多的感念信息<sup>[34-35]</sup>。其次,深度网络结构能较好地契合大规模的训练样本。目标检测的较大难点在于目标物体的多变性,目标物体在不同场景下具有不同颜色、大小及形态。深度网络具有海量的参数数量,使其具有强大的学习能力,大量的训练样本有利于激活深度网络神经元,使其记忆式地存储解析目标物体不同颜色、形态、环境下的状态<sup>[36]</sup>。基于上述 2 点,能看到深度网络在目标检测领域取得优异成果的必然性,目标检测的准确率远高于普通方法。

本文的主要目的在于系统性地总结近年来基于深度卷积网络的目标检测框架,并根据不同的侧重点将这些方法进行归类介绍。同时,对比不同方法在 PASCAL\_VOC 及 COCO 等主流数据库上的实验结果,分析不同类别方法的优劣性。最后分析当前主流深度网络目标检测框架的发展脉络,预测未来深度网络目标检测框架的发展方向。

## 1 主流深度网络的发展

Krizhevsky 等<sup>[31]</sup>提出的 AlexNet 以绝对的优势拿下了 2012 年 ImageNet 图片分类竞赛的冠军。ZF-Net<sup>[47]</sup>通过对网络进行可视化分析,进一步改进 AlexNet 的网络结构。内嵌网络(Network-In-Network, NIN)<sup>[48]</sup>主要通过在网络中建立子网络结构,提升网络性能。VGG<sup>[37]</sup>网络通过“栈”的方式将网络推向更深层次。上述网络为初期的网络,其后 Inception 和

恒等连接结构成为深度网络的主流机构。

### 1.1 稀疏连接 Inception 结构

GoogleNet<sup>[38]</sup> 分析提升网络性能最直接的办法就是增加网络的深度和广度。但直接增加网络的层数与单层元素个数会极大增加网络参数的个数, 很容易造成过拟合现象, 特别是在训练样本个数被限定及类别之间具有很小差异时。GoogleNet 解决上述两个缺点的根本方法是将全连接层转化为稀疏连接, 其依据如下: 一方面实现生物神经系统的连接也是稀疏的, 另一方面有文献表明, 对于大规模稀疏的神经网络, 可以通过分析输出值的统计特性和对高度相关的输出进行聚类, 逐层构建一个最优网络。因此 GoogleNet 设计的 Inception 结构既能保持网络结构的稀疏性, 又能利用密集矩阵的高计算性能。

Inception 不同版本结构如图 1 所示。(a) 为最初的结构, 通过  $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$  大小的卷积核分别提取不同尺度大小的局部结构。在  $3 \times 3$  和  $5 \times 5$  的卷积核前设置  $1 \times 1$  的卷积核, 主要是为了通过通道数减少参数数量。在 Inception 结构基础上, 进一步拆解较大的卷积核, 得到 Inception-v2 及 Inception-v3 版本<sup>[49]</sup>, 见 (b)、(c)、(d)。Inception 结构还能进一步升级且能与残差网络 (Residual Network, ResNet) 相结合形成 ResNet-Inception<sup>[50]</sup> 结构, 进一步提升网络的解析能力。

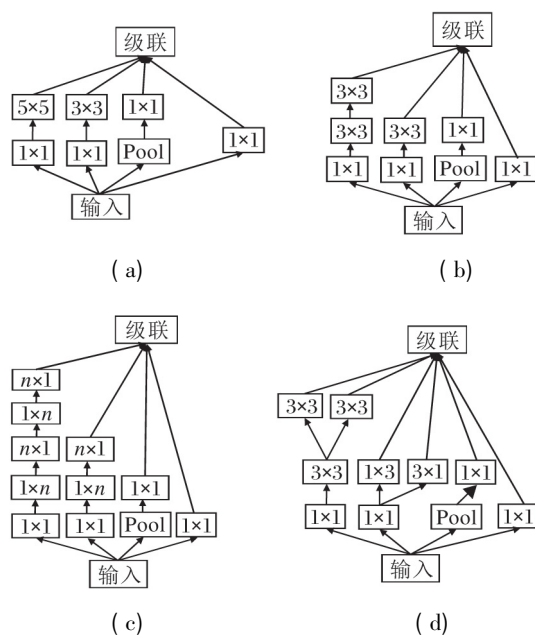


图 1 Inception 不同版本的结构

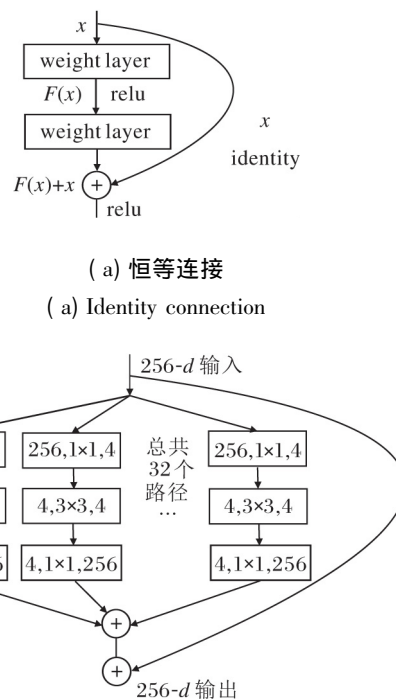
Fig.1 Different versions of Inception

### 1.2 恒等连接结构

残差网络 (ResNet)<sup>[39]</sup> 指出, 网络深度对于网络的性能至关重要, 但深度网络需避免梯度爆炸问题。

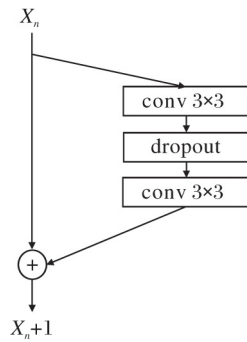
目前, 使用 Normalization、Dropout、ReLU 等手段解决梯度爆炸问题。另外, 网络的性能 (准确率) 会随着网络深度的增加迅速达到饱和, 称其为退化问题。如图 2 所示, 残差网络主要通过恒等连接解决退化问题, 恒等连接主要通过额外的连接将  $x$  直接传递到若干层后, 中间不经过卷积。见 (a) 在进行反向回传时, 前面的导数会直接传递给  $x$ , 这样可以较好地解决退化问题。 $F(x) + x$  为最终目标,  $F(x)$  可看作是目标与输入  $x$  之间的残差, 称为残差网络。(b) 为残差网络的一个变种, 称为聚合残差网络 (ResNext)<sup>[51]</sup>, 它将卷积的通道进行拆分, 网络引入 “cardinality” 参数, 主要用于控制分组的个数。(c) 中总共有 32 个路径 (32 paths), 就有 32 个 group, 因此 cardinality 为 32。广度残差网络<sup>[52]</sup> 通过拓宽残差网络的宽度提升网络性能, 即拓宽网络的通道数提升网络性能。(c) 为广度残差网络的基本结构, 由于卷积层的通道很多, 需要使用 Dropout 防止过拟合。

密集连接网络 (Densely Connected Network, DenseNet)<sup>[53]</sup>、双路径网络<sup>[54]</sup> 及提取激发网络 (Squeeze and Excitation Network, SENet)<sup>[55]</sup> 是当前提出的网络结构, 分别从不同视角提升网络性能。特别地, SENet 获得 17 年 ImageNet 大赛分类的冠军, 其主旨思想是对同一层的不同通道进行加权, 筛选较重要的通道。



(b) 聚合残差网络<sup>[51]</sup>

(b) ResNext



(c) 广度残差网络

(c) Wide residual network

图 2 ResNet 及其变种

Fig.2 ResNet and its major releases

## 2 基于深度卷积网络的目标检测框架

### 2.1 基于候选窗口的目标检测框架

基于候选窗口的目标检测框架需要在检测图像上产生很多候选窗口, 判别这些候选窗口, 判别其类别并进行窗口回归, 最终得到的包含物体的候选窗口就是目标窗口。

#### 2.1.1 RCNN 系列基础目标检测框架

RCNN 将深度网络与目标检测相连. 将目标检测分解成若干独立的步骤, 包括: 1) 候选窗口的提取. 2) 窗口的特征提取. 3) 特征分类. 4) 窗口回归 (Bounding-Box-Regression). Faster-RCNN 将这些独立的步骤合成为一个整体的深度网络框架. Faster-RCNN<sup>[43]</sup> 使用 RPN 网络结构产生的目标窗口如图 3 所示. 针对每个神经元预测  $N$  个目标窗口. 在产生目标检测窗口后, 使用与 Fast-RCNN 相同的结构进行分类与 Bounding\_Box 回归. Faster-RCNN 的候选窗口的产生及目标检测是 2 个既独立又统一的过程. 说其统一是因为从输入图像开始到最后一个卷积层结束, Fast-RCNN 与 RPN 的计算都是共享的. 说其独立是因为从最后一个卷积层开始, 分为 2 个独立的步骤. 以该层特征层 (Feature Map) 为基础 RPN 网络, 利用这些 Feature Map 产生不同位置、大小尺度的候选窗口. 这些产生的窗口会在同一层上进行映射, 再进行 ROI 池化, 池化结果会送入后续的子网络进行分类及窗口回归。

RFCN<sup>[44]</sup> 分析 Faster-RCNN 中所有候选窗口经过 ROI 池化后都会送入后续子网络进行分类及窗

口回归, 子网络的大小根据主体网络设定. 这是主要影响 Faster-RCNN 检测效率的地方. RFCN 设计位置敏感的 ROI 池化层 (Position Sensitive ROI Pooling), 使池化后的输出结果不需再经过子网络直接分类及窗口回归. 位置敏感的 ROI 池化层与 ROI 池化的根本区别在于 ROI 池化分块候选区域, 并在一个 Feature Map 上输出每块的最大值, 而位置敏感 ROI 池化不同块的截取是在不同的 Feature Map 上完成, 即不同的 Feature Map 只对应目标的一个区域, 所以叫位置敏感池化。

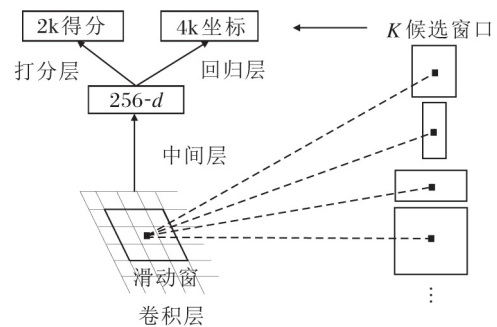


图 3 RPN 网络结构

Fig.3 RPN structure

#### 2.1.2 提升目标检测准确率的辅助方法

很多提升性能的辅助手段可提升基于 RCNN 框架的方法, 现对此进行介绍. 多层特征融合的方法在深度卷积网络里一直是重要的提升网络性能的手段. 超网络 (HyperNet)<sup>[56]</sup> 通过融合不同网络层的特征, 提高检测准确率. HyperNet 的主要思想是对 Hyper Feature 的精细设计, Hyper Feature 结合高层语义信息、中层的辅助信息及浅层的具有高分辨率的信息. 图 4 给出 HyperNet 的结构, 介绍 Hyper Feature 的具体产生细节. Hyper Feature 融合 Conv1、Conv3 及 Conv5 层特征, 在此基础上使用 RPN 网络产生候选窗口, 再使用 Fast-RCNN 进行分类及窗口回归。

特征金字塔网络 (Feature Pyramid Network, FPN) 继承 HyperNet 的思想, 认为处在网络 Top 的网络层具有较强的语义信息, 而处在网络 Bottom 的网络层通常具有较高的分辨率, 即能够提供较准确的位置信息. FPN 设计一个 Top-to-Down 的额外路径, 通过 Upsample 的方式扩张网络顶端的 Feature Map, 如图 5 所示. 这些经过顶层 Feature Map 扩张得到的新的网络层能够具有较强的语义信息, 这些新的网络层与原网络对应尺度的网络层相融合能使高分辨率的网络层也具有较强的语义信息。

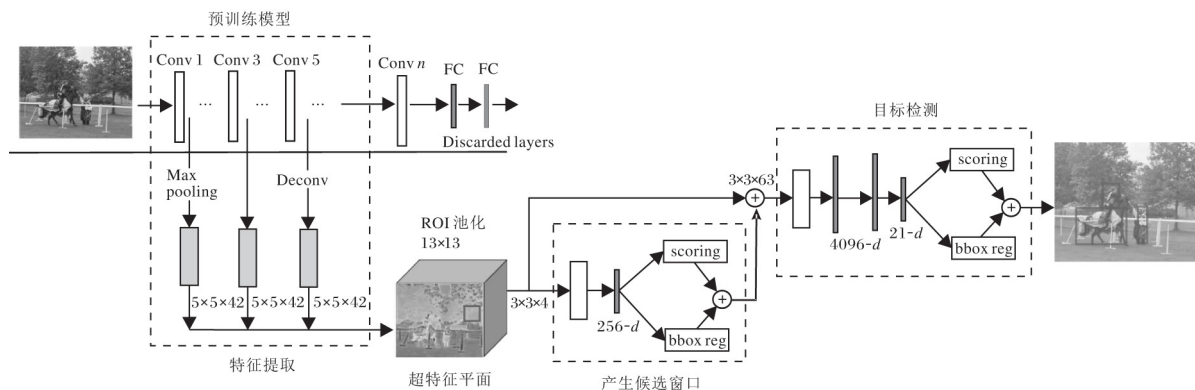


图4 HyperNet 网络结构

Fig.4 HyperNet framework

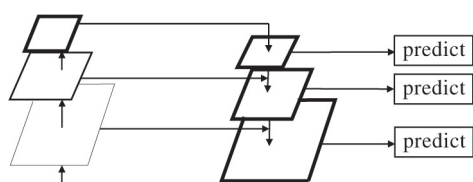


图5 FPN 网络结构

Fig.5 FPN structure

内外网络(Inside-Outside Net, ION)<sup>[57]</sup>不只关注目标窗口内的信息,还关注目标窗口外的上下文信息.ION采用Fast-RCNN的网络构架,首先使用Selective Search方法在输入的图像上产生候选窗口.图6给出ION网络整体框架,使用循环神经网络(Recurrent Neural Network, RNN)的网络层获取窗口上下左右4个方向上的信息,然后共同决定.ION网络在VOC2007数据集上的检测准确率(包括分类精度和定位精度)高于Fast-RCNN.

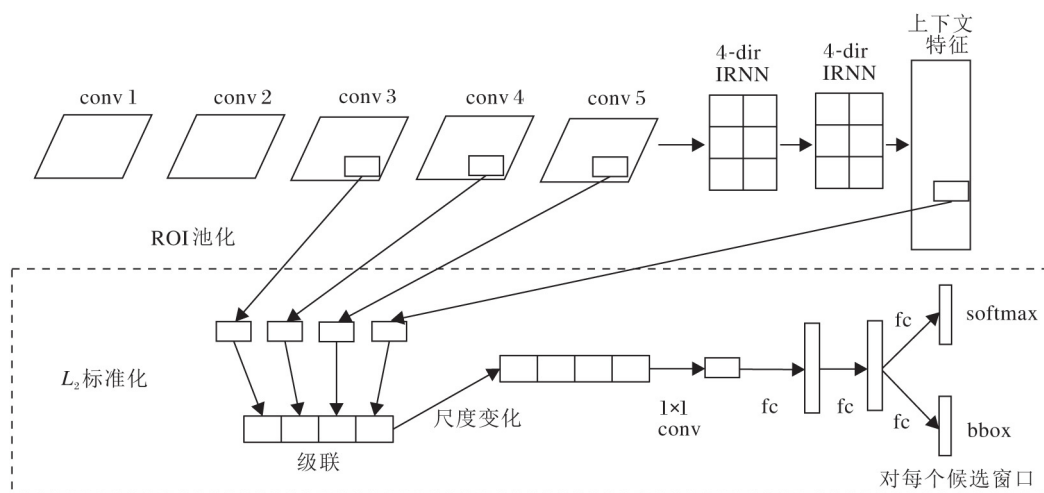


图6 ION 网络整体框架

Fig.6 ION framework

对抗网络目前在目标检测中得到广泛应用,主要思想如下:若要提升目标检测的性能,不仅可以分析对目标检测有用的信息,还可以分析对目标检测不利的信息. A-Fast-RCNN (Adversary Fast-RCNN)<sup>[58]</sup>就是利用对抗网络提升目标检测性能的

方法. A-Fast-RCNN的主旨是利用网络本身产生遮挡、形变的样本,进而提高网络处理这两类问题的能力.其产生这两类样本所用的主要方法是使用两种类型的对抗网络 ASDN (Adversarial Spatial Dropout Network) 及 ASTN (Adversarial Spatial Transformer



Network). ASTN 借鉴 STN (Spatial Transformer Network) [59-61] 网络的思想. ASDN 网络利用一个 Mask 对候选窗口的不同区域进行判别, 判断其是否具有较好的目标特征. ASDN 会对目标特征较强的区域进行 Dropout, 如此产生遮挡的样本.

### 2.1.3 处理物体形变问题

物体形变一直是目标检测中的难题, 处理这个问题的有效方法就是搜集足够的训练样本. 但是, 由于工作量巨大, 研究人员开始在网络中建立某种机制处理物体形变的问题.

STN [59-61] 网络能在网络中对物体进行空间变换. 空间变化的参数能直接通过网络生成, 通过训练样本进行端到端的训练. 图 7 给出 STN 网络的示意图, 首先通过卷积网络生成一组空间变化参数, 这些空间变换的参数可以是不同的类型, 如放射变换、薄板样条等. 得到转换参数后需对输入进行空间变换, 如图 8 所示. STN 网络在手写体识别上能取得较好效果.

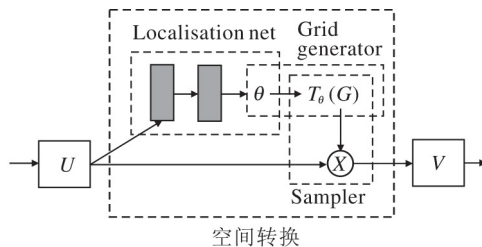


图 7 STN 网络构架

Fig.7 STN structure

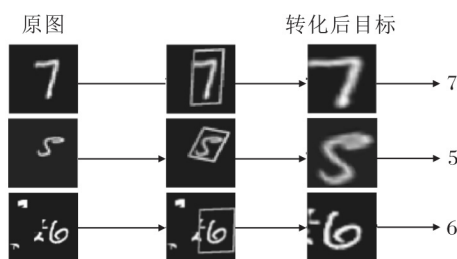


图 8 STN 网络效果

Fig.8 STN results

DPM 在处理物体形变方面取得较好效果, 将 DPM 与深度网络相结合的模型越来越多. DeepID [62] 采用 RCNN 基础框架, 将每个候选窗口送入网络, 在网络的最后产生分支, 一个分支得到 200 个类别的分类得分. 另外一个分支得到 1 000 个类别的分类得分, 此 1 000 个分类得分会对之前产生的 200 分类得分进行微调, 首先使用 Selective Search 产生 2

000 幅候选窗口. DeepID 使用不同大小的卷积核继续进行卷积, 构成第 6 个卷积层. 在此卷积层后引入 Deformation Pooling 层, 主要负责计算位移损失.

形变卷积网络 (Deformation Convolutional Network, DCN) [63] 以一个新的视角处理物体的形变问题, 主要设计 2 种结构: 形变卷积层、形变池化层. 这两种结构的核心思想是通过网络得到一个偏置, 使网络能进行不规则形状的卷积及池化. 标准的卷积通常会在规则的区域内做卷积, 而形变卷积会对这些区域的每个位置生成一个偏置, 使卷积的位置发生偏移, 进而得到不规则形状的卷积. 同理, 形变池化会对 ROI 池化划分的每个块生成一个偏移量, 使块的位置发生偏移. 对于 DCN 网络具体的实现细节这里不再详述. 图 9 分别给出形变卷积与形变池化的可视化效果.



(a) 目标分割可视化结果

(a) Visualization for segmentation



(b) 目标检测可视化结果

(b) Visualization for detection

图 9 形变卷积网络

Fig.9 Deformation convolutional network

## 2.2 基于回归的目标检测框架

基于回归的目标检测框架不需要产生候选窗口, 直接将待检测的图像送入目标窗口, 得到目标物体所在的位置.

### 2.2.1 基于 Mask 的目标检测回归模型

图 10 给出基于 Mask 的目标检测模型<sup>[16]</sup>的具体过程. 将一张图像输入深度网络, 经过深度网络后生成一个  $d \times d$  大小的 Mask, Mask 中的值为 0 或 1, 表示原图对应的区域内是否为目标物体的一部分. 利用此 Mask 最终确定目标物体的具体位置. 然而单纯这样的过程存在一个问题, 即当不同的实例距离很近或相互遮挡时, Mask 模型不能正确区分不同的实例物体. 因此, 为了解决此问题, 作者设计 {full, bottom, top, left, right} 5 种不同的 Mask 模型. 同时, 作者还采用多尺度的检测, 首先在小图像上进行检测, 然后将得到的目标窗口不断送入网络进行精简, 最终确定目标所在区域.

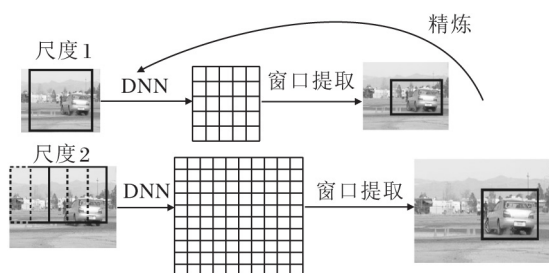


图 10 基于 Mask 的目标检测模型

Fig.10 Mask based object detection model

### 2.2.2 窗口位置的回归

AttentionNet<sup>[64]</sup>旨在对目标窗口左上角和右下角的两个点不断地进行回归, 最终确定目标物体的位置. 对于一幅输入图像, 设置图像左上角及右下角为初始位置. 然后将此图像送入一个 8 层的深度卷积网络, 卷积网络最后一层有 2 个分支, 分别使用  $1 \times$

1 的卷积核得到 5 个输出. 这 5 个输出分别表示左上角的点和右下角的点的移动状态. 左上角的点的 5 个状态分别是右移、右下移、下移、不动、F. 右下角的点对应 5 个状态分别为上移、左移、左上移、不动、F. 若 2 点的状态都为 F, 表示该图像不包含任何物体, 否则 2 点会根据对应的状态进行移动. 最终当 2 点的状态都为不动时, 说明找到目标物体.

图 11 显示 AttentionNet 进行目标检测的可视化示意图, 从图中可以看出, 以原图左上角和右下角为起点, AttentionNet 通过移动最终确定目标的位置. 当然, AttentionNet 只能检测单个目标, 无法检测到多个目标. 文中对多目标进行检测的方法是要产生一些可能包含物体的候选窗口, 然后对这些窗口应用 AttentionNet, 最后整合结果, 最终确定窗口位置.

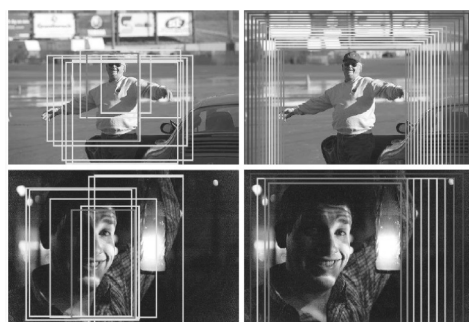


图 11 AttentionNet 可视化结果

Fig.11 Visual results of AttentionNet

G-CNN<sup>[65]</sup>是对窗口位置不断进行回归的方法, 首先定义少量的 Grid, 然后不断将初始的 Grid 向目标窗口靠拢. 图 12 给出 G-CNN 的迭代过程, 通过不断迭代将初始的 Grid 回归成包含目标的窗口.

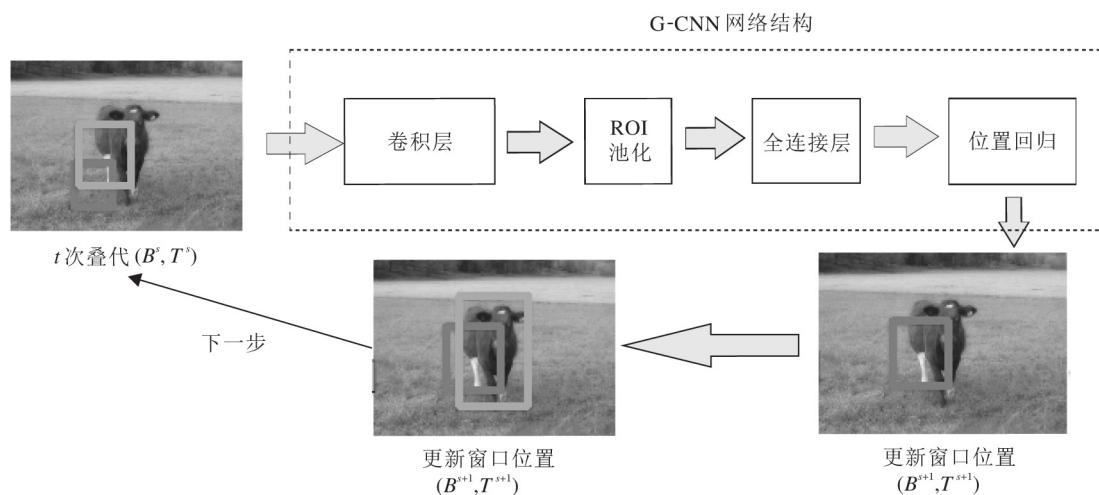


图 12 G-CNN 网络结构

Fig.12 G-CNN framework

### 2.2.3 神经元预测窗口框架

为了提高目标检测的效率, YOLO<sup>[45]</sup> 快速模型能达到 155 帧/秒的速度, 完全满足实时性的要求。如图 13 所示, YOLO 首先将图像分成  $S \times S$  个格子, 如果一个物体中心落在这个格子中, 这个格子就负责该物体的检测。每个格子会对  $B$  个窗口进行预测, 给出位置信息, 同时还会给出  $B$  个窗口的置信度得分, 每个置信度得分表示该窗口包含物体的可信度。每个格子除要对  $B$  个窗口进行预测之外, 也要预测  $C$  个类别概率, 这些类别概率预测该格子包含不同物体的概率。需要指出的是, 这里的类别概率是相对格子而言, 而不是相对格子对应的窗口而言。无论格子对应多少个窗口, 只需要为格子预测一组类别概率。

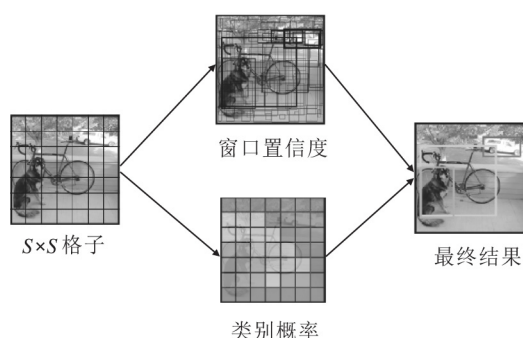


图 13 YOLO 的主体框架

Fig.13 YOLO structure

SSD<sup>[46]</sup> 在 YOLO 基础上进行改进, 使用不同尺度的 Feature Map 对应不同的 Anchors. SSD 没有窗口的产生过程, 不同尺度的 Feature Map 上每个神经元都预测  $n$  个候选窗口 (Anchor Boxes). 尺度较大的 Feature Map 上神经元的感受野较小, 因此尺度大的 Feature Map 上的神经元对应的 Anchors 也较小, 反之较大。

SSD 与 YOLO 的共同特点都是利用 Feature Map 上的每个神经元预测  $n$  个窗口, 这在基于回归方法及基于候选窗口的方法之间是一个较好的承接过渡。DSSD<sup>[66]</sup> 是最近提出的基于 SSD 的改进方法, 与 FPN 类似, 利用反卷积 (Deconv) 建立 Top-to-Down 路径, 产生新的不同尺度的 Feature Map, 这些 Feature Map 比原来的语义信息更强, DSSD 在检测精度上有大幅提升。

基于候选窗口的目标检测算法通常检测准确率较高, 而基于回归窗口的目标检测方法在检测效率上结果较好, 目前也有方法试图在这两类之间建立联

系, RON<sup>[67]</sup> 网络就试图利用 RPN 建立目标先验以提升回归模型的检测效率。

## 3 实验及结果分析

目前基于深度网络的目标检测框架使用的主流数据库包括 PASCAL\_VOC、IMAGENET 和 COCO 数据库。不同方法在 PASCAL\_VOC 数据集上的实验结果对比如下。

表 1 给出不同的目标检测框架在 PASCAL\_VOC 2007 数据集上的实验结果, 其中上半部分给出大多数基于回归的目标检测方法的实验结果, 下半部分给出大部分基于候选窗口的目标检测框架的实验结果。总体上, 基于候选窗口的目标检测框架在准确率上高于基于回归的目标检测框架。DetectionNet 是早期利用深度网络进行目标检测的方法, 是基于 Mask 的方法, 这个方法在 PASCAL\_VOC 2007 数据集上的准确率并不高, 而且针对基于 Mask 的方法还存在很多争论。AttentionNet 和 G-CNN 都是对窗口位置不断进行回归的模型, 在 PASCAL\_VOC 数据集上的准确率也基本类似。相比其它的回归方法, YOLO 和 SSD 的准确率更高, 特别是 SSD, 在 PASCAL\_VOC 2007 数据集上的准确率达到 75.1%。

RCNN 系列方法是目前基于候选窗口的主流框架, 可以看到从 RCNN 到 Faster-RCNN, 目标检测的准确率不断提升, 在 PASCAL\_VOC 2007 数据集上的准确率已达 73.2%。同时, 不同的主体网络对目标检测性能的影响很大, 应用 resnet101 的 Faster-RCNN 的准确率高于应用 VGG16 的 Faster-RCNN 框架。同时, OHEM (Online Hard Examples Mining) 对目标检测性能提升很大, 在 resnet101 主体网络下, 应用 OHEM 的 Faster-RCNN 在准确率上提升 2.8%。相比 Faster-RCNN, RFCN 与其在准确率上不相上下, 但在检测准确率上可以提升很多。HyperNet 与 ION 这些辅助的方法对准确率的提升也有很大的帮助, 特别是 ION, 在引入目标窗口上下文信息后, 能将 Fast-RCNN 准确率提升 7%。

表 2 给出不同方法在 PASCAL\_VOC 2012 测试数据集上的检测效率。对比表 1, 基于回归方法 (SSD 和 YOLO) 检测效率较高, SSD 不但达到较高的检测效率, 在准确率上也不会低于其它方法太多, 可以说是具有最好平衡性的方法。



表 1 不同方法在 PASCAL\_VOC 数据集上的实验结果对比

Table 1 Experimental results comparison of different methods on PASCAL\_VOC dataset

方法	测试数据集	网络	OHEM	平均 准确率/%
DPM <sup>[24]</sup>	VOC 2007	-	-	29.2
DetectionNet <sup>[16]</sup>	VOC 2007	DNN	-	30.5
AttentionNet <sup>[64]</sup>	VOC 2007	ANet	no	65.0
G-CNN <sup>[65]</sup>	VOC 2007	VGG16	no	66.8
YOLO <sup>[45]</sup>	VOC 2007	VGG16	no	66.4
YOLO <sup>[45]</sup>	VOC 2012	VGG16	no	57.9
SSD <sup>[46]</sup>	VOC 2007	VGG16	no	75.1
SSD <sup>[46]</sup>	VOC 2012	VGG16	no	73.1
RCNN <sup>[40]</sup>	VOC 2007	VGG16	no	66.0
Fast-RCNN <sup>[42]</sup>	VOC 2007	VGG16	no	70.0
Faster-RCNN <sup>[43]</sup>	VOC 2007	VGG16	no	73.2
Faster-RCNN <sup>[43]</sup>	VOC 2007	resnet101	no	76.4
Faster-RCNN <sup>[43]</sup>	VOC 2012	resnet101	no	73.4
Faster-RCNN <sup>[43]</sup>	VOC 2007	resnet101	yes	79.2
RFCN <sup>[44]</sup>	VOC 2007	resnet101	yes	79.5
RFCN <sup>[44]</sup>	VOC 2012	resnet101	yes	77.6
HyperNet <sup>[56]</sup>	VOC 2007	VGG16	no	76.3
ION <sup>[57]</sup>	VOC 2007	VGG16	no	77.6
Deep-ID <sup>[62]</sup>	VOC 2007	VGG16	no	64.1

表 2 不同方法在 PASCAL\_VOC 2012 测试集上的实验结果

Table 2 Experimental results of different methods on PASCAL\_VOC 2012 testing dataset

方法	平均 准确率/%	FPS
Faster-RCNN( resnet101) <sup>[43]</sup>	76.4	3
Faster-RCNN( VGG16) <sup>[43]</sup>	73.2	7
Faster-RCNN( ZF) <sup>[43]</sup>	62.1	17
RFCN( resnet101) <sup>[44]</sup>	79.5	8
YOLO <sup>[45]</sup>	63.4	45
Fast YOLO <sup>[45]</sup>	52.7	155
SSD300 <sup>[46]</sup>	72.1	58
SSD500 <sup>[46]</sup>	75.1	23

4 结 束 语

近几年基于深度卷积网络的目标检测框架发展迅猛,具有如下研究热点.

1) 多层网络特征融合结构.目标检测是进行场景内容理解等高级视觉任务的前提.静态图像的目标检测存在诸多难点,如物体形变、物体遮挡及小目标检测等.如何更有效合理解决这些问题是未来重要的研究方向.基于功能性子网络的目标检测研究

主要致力于将传统的在处理某些特殊问题(物体形变、遮挡等)有效的目标检测方法引入深度网络,增加深度网络处理这些特殊问题的能力,提升目标检测性能.处理过程是首先深入分析某种技术或方案,拆解主要步骤.以此解析主要步骤,并使用尽量简单的数学语言进行描述.然后针对不同的步骤设计对应的网络层.确定具体针对的神经元(如有的操作需要针对单一的神经元进行,而有的操作需要在 Feature Map 的不同区域上进行),最终合理组织这些神经元,得到想要的结果.在反向计算过程中,需要进行求导.对于求解导数的方程,可以使用极限手段得到近似导数.最后结合不同的网络层操作形成局部的网络结构,完成最终的子网络嵌入.

2) 基于有效感受野的目标检测.讨论深度卷积网络,就涉及到感受野.感受野是指网络中的某个神经元在输入上对应的区域范围.该类方法首先量化感受野中所有元素对于神经元的贡献程度,然后将其中贡献程度较大的区域称为有效感受野,并证明该量化值呈现高斯分布,即有效感受野在感受野中只是很小的一部分.有效感受野在目标检测中是一个重要的概念.如果元素的有效感受野在目标中占有面积很大,说明这个神经元会在目标检测中起重要作用.因此,如何合理利用有效感受野的概念是后续重要的研究方向.

3) 压缩目标检测模型.提高检测效率.由于深度网络模型会占有大量的存储空间,在移动端的应用很受限制,因此,压缩目标检测模型也会是未来重要的研究思路.基于模拟网络的研究旨在设计特殊的有监督与无监督任务,使目标检测的训练更高效.其主要思想就是利用已训练好的深度网络目标检测框架为目标,训练较小的目标检测模型.通过额外任务的训练使小框架的训练变得更高效,因此可在保证目标检测性能的前提下压缩目标检测的整体框架.具体地,可以以深层训练好的目标检测框架为基准对浅层的目标检测框架进行训练.此训练模式主要采用 Logits 训练,即以深度网络最后一层的 Logits 单元为目标训练浅层网络向这些 Logits 单元进行收敛.另外,也可以增加无监督的训练任务,加强网络训练.通过设置编码与解码器对网络较浅层进行额外任务的训练,进一步提升网络训练效率.

参 考 文 献

[1] DOLLÁR P, APPEL R, BELONGIE S, *et al.* Fast Feature Pyramids for Object Detection. IEEE Transactions on Pattern Analysis

- and Machine Intelligence , 2014 , 36( 8) : 1532–1545.
- [2] GALL J , LEMPITSKY V. Class-Specific Hough Forests for Object Detection // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2013: 1022–1029.
  - [3] KALA R. Advanced Driver Assistance Systems // SURHONE L M , TENNOE M T , HENSSONOW S F , eds. On-Road Intelligent Vehicles. Cambridge , USA: Elsevier , 2016: 59–82.
  - [4] JAZAYERI A , CAI H Y , ZHENG J Y , *et al.* Vehicle Detection and Tracking in Car Video Based on Motion Model. IEEE Transactions on Intelligent Transportation Systems , 2011 , 12( 2) : 583–595.
  - [5] CARAFFI C , VOJÍŘ T , TREFNÝ J , *et al.* A System for Real-Time Detection and Tracking of Vehicles from a Single Car-Mounted Camera // Proc of the 15th IEEE International Conference on Intelligent Transportation Systems. Washington , USA: IEEE , 2012: 975–982.
  - [6] WANG M , DAAMEN W , HOOGENDOORN S P , *et al.* Driver Assistance Systems Modeling by Model Predictive Control // Proc of the 15th IEEE International Conference on Intelligent Transportation Systems. Washington , USA: IEEE , 2012: 1543–1548.
  - [7] CHO H , SEO Y W , KUMAR B V K V , *et al.* A Multi-sensor Fusion System for Moving Object Detection and Tracking in Urban Driving Environments // Proc of the IEEE International Conference on Robotics and Automation. Washington , USA: IEEE , 2014: 1836–1843.
  - [8] LEVINSON J , ASKELAND J , BECKER J , *et al.* Towards Fully Autonomous Driving: Systems and Algorithms // Proc of the IEEE Intelligent Vehicles Symposium. Washington , USA: IEEE , 2011: 163–168.
  - [9] ZHOU X W , YANG C , YU W C. Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2013 , 35( 3) : 597–610.
  - [10] OREIFEJ O , LI X , SHAH M. Simultaneous Video Stabilization and Moving Object Detection in Turbulence. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2013 , 35( 2) : 450–462.
  - [11] KULCHANDANI J S , DANGARWALA K J. Moving Object Detection: Review of Recent Research Trends // Proc of the International Conference on Pervasive Computing. Washington , USA: IEEE , 2015. DOI: 10.1109/PERVASIVE.2015.7087138.
  - [12] CHAVEZ-GARCIA R O , AVCARD O. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking. IEEE Transactions on Intelligent Transportation Systems , 2016 , 17( 2) : 525–534.
  - [13] HU W C , CHEN C H , CHEN T Y , *et al.* Moving Object Detection and Tracking from Video Captured by Moving Camera. Journal of Visual Communication & Image Representation , 2015 , 30: 164–180.
  - [14] WANG J D , JIANG H Z , YUAN Z J , *et al.* Salient Object Detection: A Discriminative Regional Feature Integration Approach // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2013: 2083–2090.
  - [15] SZEGEDY C , REED S , ERHAN D , *et al.* Scalable , High-Quality Object Detection [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1412.1441v2.pdf>.
  - [16] SZEGEDY C , TOSHEV A , ERHAN D. Deep Neural Networks for Object Detection // BURGESS C J C , BOTTOU L , WELLING M , *et al.* , eds. Advances in Neural Information Processing Systems 26. Cambridge , USA: The MIT Press , 2013: 2553–2561.
  - [17] WANG X Y , HAN T X , YAN S C. An HOG-LBP Human Detector with Partial Occlusion Handling // Proc of the 12th IEEE International Conference on Computer Vision. Washington , USA: IEEE , 2009: 32–39.
  - [18] DALAL N , TRIGGS B. Histograms of Oriented Gradients for Human Detection // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2005: 886–893.
  - [19] ZHU Q , YEH M C , CHENG K T , *et al.* Fast Human Detection Using a Cascade of Histograms of Oriented Gradients // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2006: 1491–1498.
  - [20] LOWE D G. Distinctive Image Features from Scale-Invariant Key-points. International Journal of Computer Vision , 2004 , 60( 2) : 91–110.
  - [21] LUO J , GWUN O. A Comparison of SIFT , PCA-SIFT and SURF. International Journal of Image Processing , 2013 , 3( 4) : 143–152.
  - [22] LIU C , YUEN J , TORRALBA A. SIFT Flow: Dense Correspondence across Scenes and Its Applications. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2011 , 33( 5) : 978–994.
  - [23] DIVVALA S K , EFROS A A , HEBERT M. How Important Are “Deformable Parts” in the Deformable Parts Model? // Proc of the 12th European Conference on Computer Vision. London , UK: Springer-Verlag , 2012 , III: 31–40.
  - [24] FELZENSZWALB P F , GIRSHICK R B , MCALLESTER D , *et al.* Object Detection with Discriminatively Trained Part Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence , 2010 , 32( 9) : 1627–1645.
  - [25] GIRSHICK R , IANDOLA F , BARRELL T , *et al.* Deformable Part Models Are Convolutional Neural Networks [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1703.06211.pdf>.
  - [26] OUYANG W L , WANG X G. Joint Deep Learning for Pedestrian Detection // Proc of the IEEE International Conference on Computer Vision. Washington , USA: IEEE , 2013: 2056–2063.
  - [27] UIJLINGS J R R , VAN DE SANDE K E A , GEVERS T , *et al.* Selective Search for Object Recognition. International Journal of Computer Vision , 2013 , 104( 2) : 154–171.
  - [28] ZHU G , PORIKLI F , LI H D. Tracking Randomly Moving Objects on Edge Box Proposals [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1507.08085.pdf>.
  - [29] DENG J , DONG W , SOCHER R , *et al.* ImageNet: A Large-Scale Hierarchical Image Database // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2009: 248–255.
  - [30] LIN T Y , MAIRE M , BELONGIE S , *et al.* Microsoft Coco: Common Objects in Context // Proc of the 13th European Conference

- on Computer Vision. New York, USA: Springer, 2014: 740–755.
- [31] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 2017, 60(6): 84–90.
- [32] SZEGEDY C, ERHAN D, TOSHEV A T. Object Detection Using Deep Neural Networks [P/OL]. [2017–12–10]. <http://www.freepatentsonline.com/9275308.pdf>.
- [33] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651.
- [34] HANSEN L K, SALAMON P. Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993–1001.
- [35] KINGSBURY B, SAINATH T N, SOLTAU H. Scalable Minimum Bayes Risk Training of Deep Neural Network Acoustic Models Using Distributed Hessian-Free Optimization // *Proc of the 13th Annual Conference of International Speech Communication Association*. New York, USA: ACM, 2012: 10–13.
- [36] XU L, REN J S J, LIU C, *et al.* Deep Convolutional Neural Network for Image Deconvolution // GHAHRAMANI Z, WELLING M, CORTES C, *et al.*, eds. *Advances in Neural Information Processing Systems 27*. Cambridge, USA: The MIT Press, 2014: 1790–1798.
- [37] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J/OL]. [2017–12–10]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [38] SZEGEDY C, LIU W, JIA Y Q, *et al.* Going Deeper with Convolutions [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1409.4842.pdf>.
- [39] HE K M, ZHANG X Y, REN S Q, *et al.* Deep Residual Learning for Image Recognition // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 770–778.
- [40] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2013: 580–587.
- [41] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916.
- [42] GIRSHICK R. Fast R-CNN // *Proc of the IEEE International Conference on Computer Vision*. Washington, USA: IEEE, 2015: 1440–1448.
- [43] REN S Q, HE K M, GIRSHICK R B, *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149.
- [44] DAI J F, LI Y, HE K M, *et al.* R-FCN: Object Detection via Region-Based Fully Convolutional Networks // LEE D D, SUGIYAMA M, LUXBURG U V, *et al.*, eds. *Advances in Neural Information Processing Systems 29*. Cambridge, USA: The MIT Press, 2016: 379–387.
- [45] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You Only Look Once: Unified, Real-Time Object Detection // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 779–788.
- [46] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single Shot Multibox Detector // *Proc of the 14th European Conference on Computer Vision*. New York, USA: Springer, 2016, 1: 21–37.
- [47] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Neural Networks // *Proc of the 13th European Conference on Computer Vision*. New York: USA: Springer, 2014, 1: 818–833.
- [48] LIN M, CHEN Q, YAN S C. Network in Network [J/OL]. [2017–12–10]. <https://arxiv.org/pdf/1312.4400.pdf>.
- [49] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al.* Rethinking the Inception Architecture for Computer Vision // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 2818–2826.
- [50] SZEGEDY C, IOFFE S, VANHOUCKE V, *et al.* Inceptionv4, Inception-Resnet and the Impact of Residual Connections on Learning [J/OL]. [2017–12–10]. <https://arxiv.org/pdf/1602.07261.pdf>.
- [51] XIE S N, GIRSHICK R B, DOLLÁR P, *et al.* Aggregated Residual Transformations for Deep Neural Networks // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2017: 5987–5995.
- [52] ZAGORUYKO S, KOMODAKIS N. Wide Residual Networks [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1605.07146.pdf>.
- [53] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely Connected Convolutional Networks // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2017: 2261–2269.
- [54] CHEN Y P, LI J N, GE H X, *et al.* Dual Path Networks // GUYON I, LUXBURG U V, BENGION S, *et al.*, eds. *Advances in Neural Information Processing Systems 30*. Cambridge, USA: The MIT Press, 2017: 4470–4478.
- [55] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1709.01507.pdf>.
- [56] KONG T, YAO A B, CHEN Y R, *et al.* HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 845–853.
- [57] BELL S, ZITNICK L C, BALA K, *et al.* Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 2874–2883.
- [58] WANG X L, SHRIVASTAVA A, GUPTA A. A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2017: 3039–3048.
- [59] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al.* Spatial Transformer Networks // CORTES C, LAWRENCE N D, LEE D O, *et al.*, eds. *Advances in Neural Information Processing Systems*

28. Cambridge , USA: The MIT Press , 2015: 2017–2025.
- [60] LIN C H , LUCEY S. Inverse Compositional Spatial Transformer Networks // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2016: 2252 – 2260.
- [61] SØNDERBY S K , SØNDERBY C K , MAALØE L , *et al.* Recurrent Spatial Transformer Networks [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1509.05329.pdf>.
- [62] OUYANG W L , WANG X G , ZENG X Y , *et al.* Deepid-Net: Deformable Deep Convolutional Neural Networks for Object Detection // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2015: 2403–2412.
- [63] DAI J F , QI H Z , XIONG Y W , *et al.* Deformable Convolutional Networks // Proc of the IEEE International Conference on Computer Vision. Washington , USA: IEEE , 2017: 764–773.
- [64] YOO D , PARK S , LEE J Y , *et al.* Attentionnet: Aggregating Weak Directions for Accurate Object Detection // Proc of the IEEE International Conference on Computer Vision. Washington , USA: IEEE , 2015: 2659–2667.
- [65] NAJIBI M , RASTEGARI M , DAVIS L S. G-CNN: An Iterative Grid Based Object Detector // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2016: 2369–2377.
- [66] FU C Y , LIU W , RANGA A , *et al.* DSSD: Deconvolutional Single Shot Detector [C/OL]. [2017–12–10]. <https://arxiv.org/pdf/1701.06659.pdf>.
- [67] KONG T , SUN F C , YAO A B , *et al.* RON: Reverse Connection

with Objectness Prior Networks for Object Detection // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington , USA: IEEE , 2017: 5244–5252.

## 作者简介



吴 帅 , 博士研究生 , 主要研究方向为模式识别、深度学习. E-mail: 949766996 @ qq.com.

( **WU Shuai** , Ph.D. candidate. His research interests include pattern analysis and deep learning.)



徐 勇 , 博士 , 教授 , 主要研究方向为模式识别、人工智能、图像处理等. E-mail: yongxu@ymail.com.

( **XU Yong** , Ph.D. , professor. His research interests include pattern analysis , artificial intelligence and image processing.)



赵东宁 , 博士 , 主要研究方向为多媒体信息处理、大数据技术、人工智能. E-mail: 582101@ qq.com.

( **ZHAO Dongning** , Ph.D.. Her research interests include multimedia information processing , big data and artificial intelligence.)

\*\*\*\*\*  
( 上接 334 页)

## 二、论文出版

大会将出版 CAC2018 论文集( U 盘版) 。 2013 年以来的历届会议英文论文全被 IEEE Xplore 收录 , 并被 EI 检索。经过专家评审 , 本届大会部分优秀论文将被推荐到国内外 SCI、EI 收录权威期刊发表( IEEE/CAA Journal of Automatica Sinica ( JAS) 、《自动化学报》、《控制理论与应用》、《机器人》、《信息与控制》等期刊) 。

## 三、重要时间

投稿截止日期: 2018 年 6 月 1 日

录用通知日期: 2018 年 8 月 1 日

终稿提交截止日期: 2018 年 9 月 1 日

## 四、联系方式

咨询信箱: caa@ ia.ac.cn ( 中国自动化学会邮箱) cac2018@ mail.xjtu.edu.cn

咨询电话: 010-82544542 010-62522472 010-6252248( FAX) ( 中国自动化学会)  
029-82664437 029-82663421

会议官网: www.cac2018.cn