## CVPR 2019 | PoolNet：基于池化技术的显著性目标检测

原创： 文永亮 [PaperWeekly](javascript:void(0);)

**PaperWeekly**

微信号paperweekly

功能介绍PaperWeekly是一个推荐、解读、讨论和报道人工智能前沿论文成果的学术平台，致力于让国内外优秀科研工作得到更为广泛的传播和认可。社区：http://paperweek.ly | 微博：@PaperWeekly

5月27日

作者丨文永亮

学校丨哈尔滨工业大学（深圳）

研究方向丨目标检测、GAN





# 研究动机

这是一篇发表于 CVPR 2019 的关于显著性目标检测的 paper，

在 U 型结构的特征网络中，高层富含语义特征捕获的位置信息在自底向上的传播过程中可能会逐渐被稀释，另外卷积神经网络的感受野大小与深度是不成正比的。目前很多流行方法都是引入 Attention（注意力机制），但是**本文是**

**基于 U 型结构的特征网络研究池化对显著性检测的改进**，具体步骤是引入了两个模块

GGM (Global Guidance Module，全局引导模块) 和

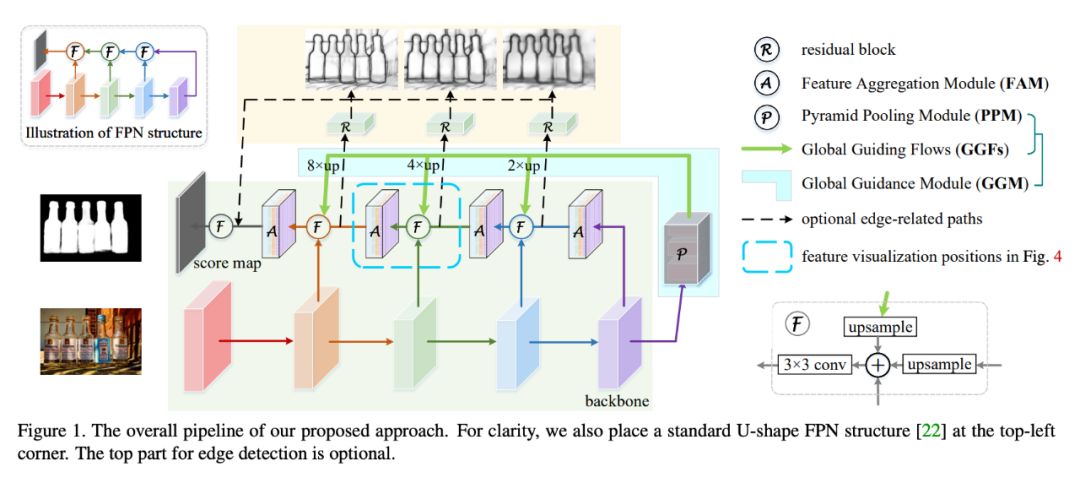
FAM (Feature Aggregation Module，特征整合模块)，进而

锐化显著物体细节

并且检测速度能够达到 30FPS。因为这两个模块都是基于池化做的改进所以作者称其为 ***PoolNet***，并且放出了源码：

https://github.com/backseason/PoolNet

# 模型架构



**两个模块**

**GGM（全局引导模块）**

我们知道高层语义特征对挖掘显著对象的详细位置是很有帮助的，但是中低层的语义特征也可以提供必要的细节。因为在 top-down 的过程中，**高层语义信息被稀释，而且实际上的感受野也是小于理论感受野**，所以对于全局信息的捕捉十分的缺乏，导致显著物体被背景吞噬。

因此作者提出了 GGM 模块，

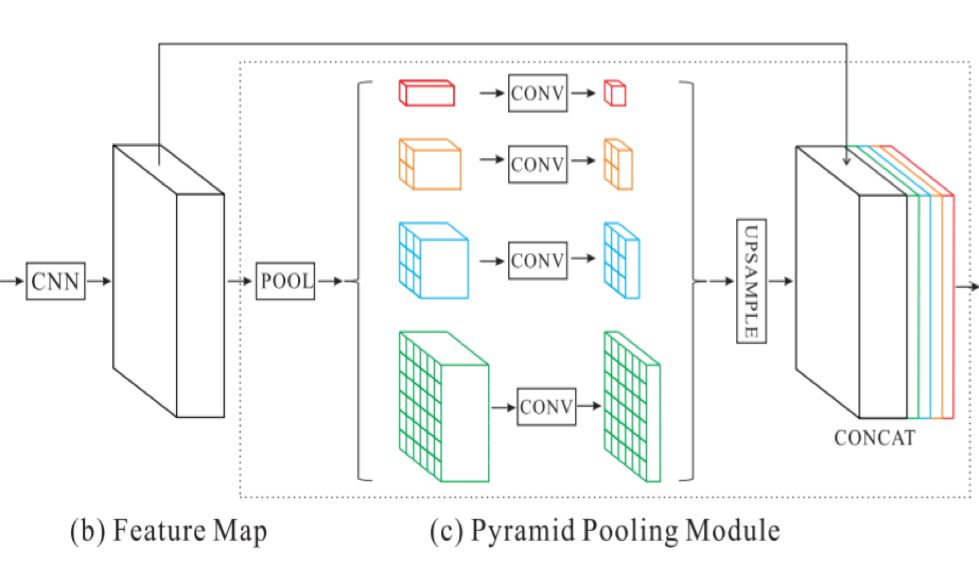
GGM 其实是 **PPM（Pyramid Pooling module，金字塔池化模块）**的改进并且

加上了一系列的 **GGFs（Global Guiding Flows，全局引导流）**，

这样做的好处是，在特征图上的每层都能关注到显著物体，另外不同的是，GGM 是

一个独立的模块，而 PPM 是在 U 型架构中，在基础网络（backbone）中参与引导全局信息的过程。

其实这部分论文说得并不是很清晰，没有说 **GGM** 的详细结构，我们可以知道 **PPM** [7] 的结构如下：



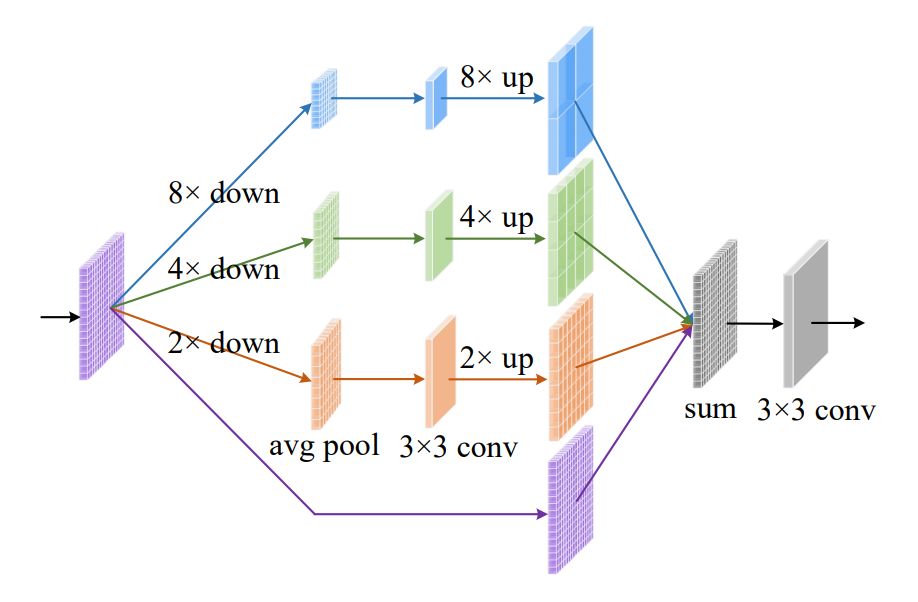
该 PPM 模块融合了 4 种不同金字塔尺度的特征，第一行红色是最粗糙的特征–全局池化生成单个 bin 输出，后面三行是不同尺度的池化特征。为了保证全局特征的权重，如果金字塔共有 N 个级别，则在每个级别后使用 1×1 的卷积将对于级别通道降为原本的 1/N。再通过双线性插值获得未池化前的大小，最终 concat 到一起。

如果明白了这个的话，其实 **GGM** 就是在 **PPM** 的结构上的改进，**PPM**是对每个特征图都进行了金字塔池化，所以作者说是嵌入在 U 型结构中的，但是他加入了 **global guiding flows（GGFs）**，即 Fig1 中绿色箭头，引入了对每级特征的不同程度的**上采样映射**（文中称之为 identity mapping），所以可以是个独立的模块。

简单地说，作者想要 FPN 在 top-down 的路径上不被稀释语义特征，**所以在每次横向连接的时候都加入高层的语义信息**，这样做也是一个十分直接主观的想法。

**FAM（特征整合模块）**

**特征整合模块也是使用了池化技巧的模块**，如下图，先把 **GGM** 得到的高层语义与该级特征分别上采样之后横向连接一番得到 **FAM** 的输入 b，之后采取的操作是先把 b 用 {2,4,8} 的三种下采样得到蓝绿红特征图然后 avg pool（平均池化）再上采样回原来尺寸，最后蓝绿红紫（紫色是 FAM 的输入 b）四个分支像素相加得到整合后的特征图。



**FAM 有以下两个优点：**

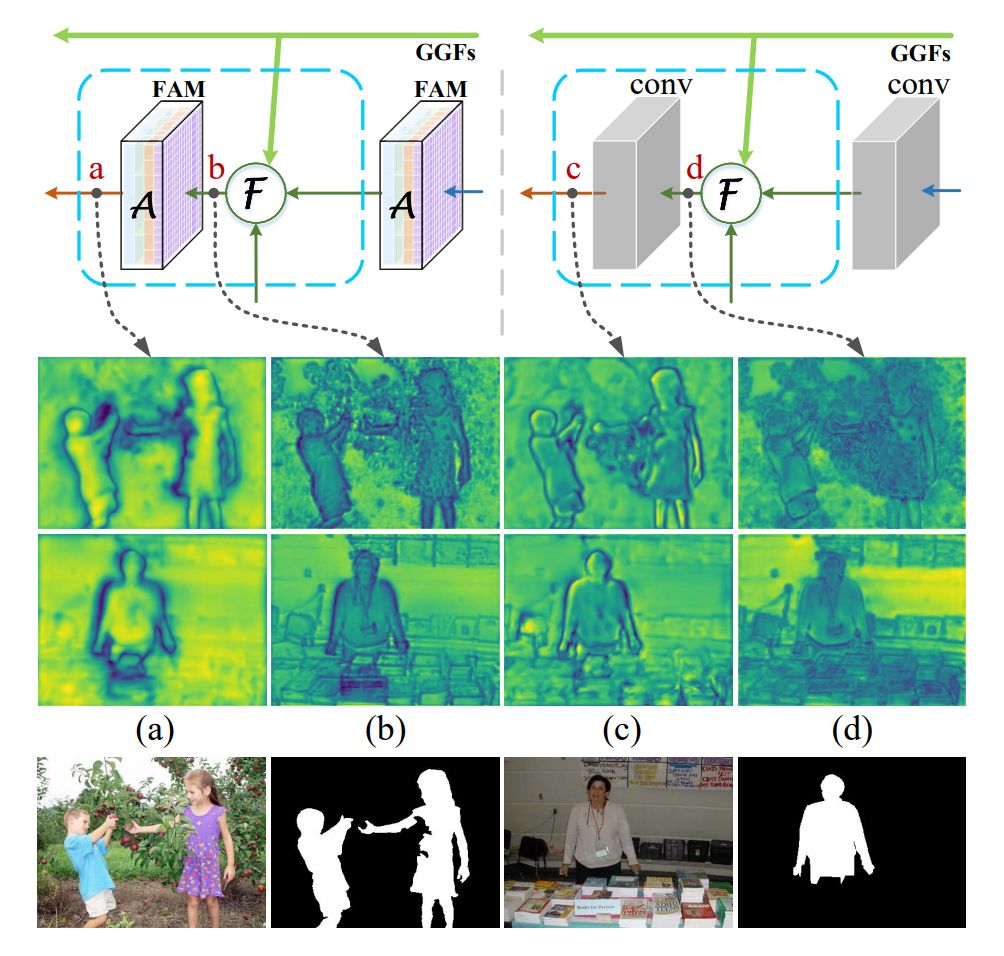
1. 帮助模型降低上采样（upsample）导致的混叠效应（aliasing）；

对于第一点降低混叠效应的理解，用明珊师姐说的话，混叠效应就相当于引入杂质，GGFs 从基础网络最后得到的特征图经过金字塔池化之后需要最高是 8 倍上采样才能与前面的特征图融合，这样高倍数的采样确实容易引入杂质。

2. 从不同的多角度的尺度上纵观显著物体的空间位置，放大整个网络的感受野。

第二点很容易理解，从不同角度看，不同的放缩尺度看待特征，能够放大网络的感受野。

作者就是因为这样才会提出 FAM，进行特征整合，**先把特征用不同倍数的下采样，池化之后，再用不同倍数的上采样，最后叠加在一起。**因为单个高倍数上采样容易导致失真，所以补救措施就是高倍数上采样之后，再下采样，再池化上采样**平均下来可以弥补错误**。

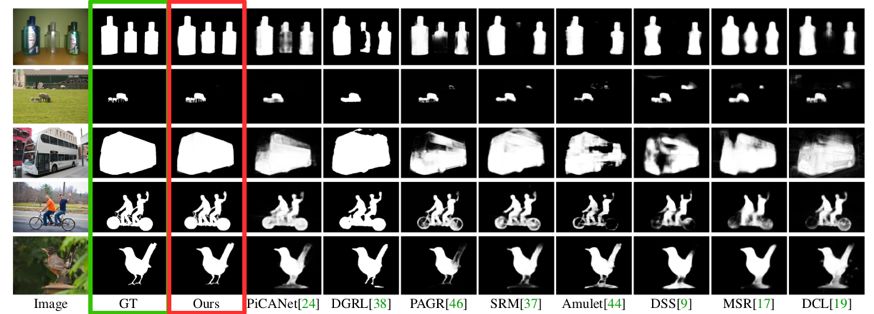


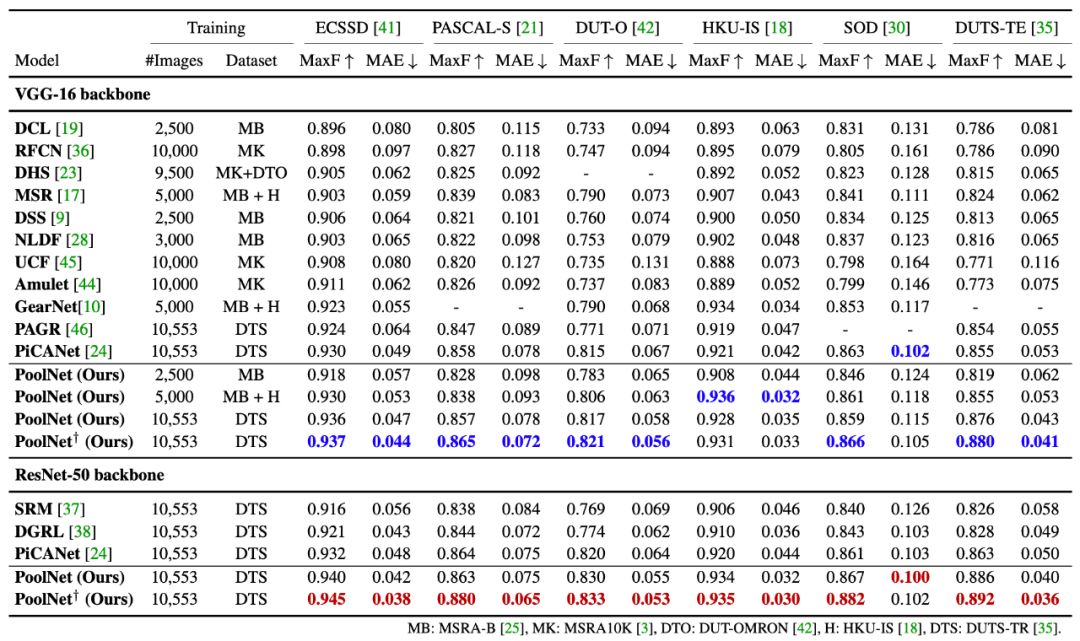
上图就是为了说明 FAM 的优点的，经过高倍上采样之后的图像（b）和（d）容易引入许多杂质，致使边缘不清晰，但是经过 FAM 模块之后的特征图就能**降低混叠效应**。

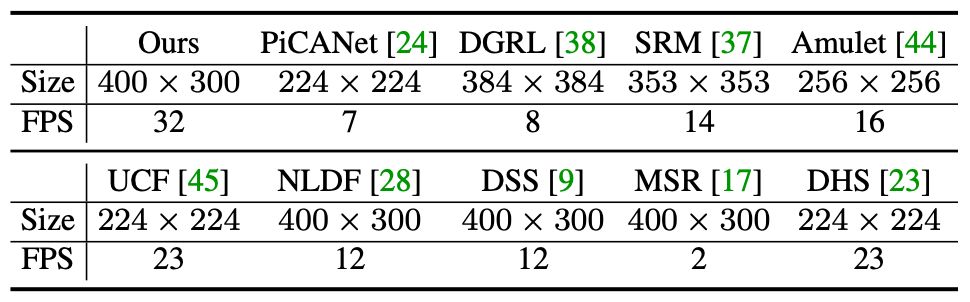
# 实验结果

论文在常用的 6 种数据集上做了实验，有 ECSSD [8], PASCALS [9], DUT-OMRON [10], HKU-IS [11], SOD [12] 和 DUTS [13], 使用二值交叉熵做显著性检测，平衡二值交叉熵（balanced binary cross entropy）[14] 作为边缘检测（edge detection）。

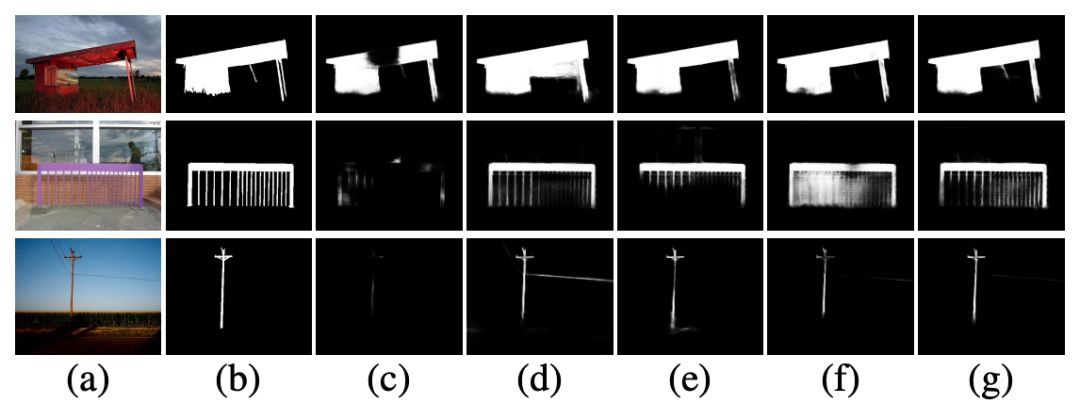
以下是文章方法跟目前 state-of-the-arts 的方法的对比效果，绿框是 GT，红框是本文效果。可以看到无论在速度还是精度上都有很大的优势。

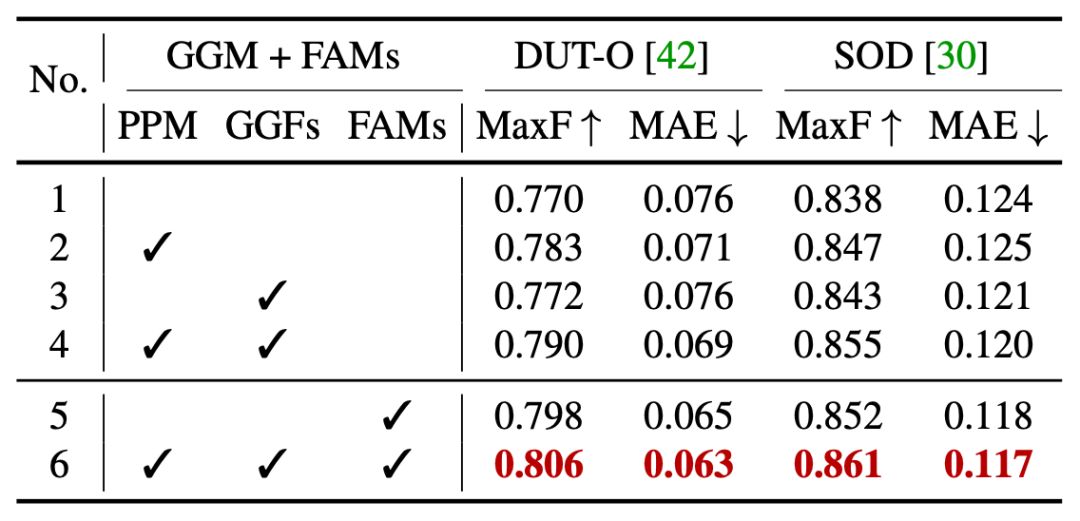






论文还针对三个改进的技术 PPM、GGFs 和 FAMs 的不同组合做了实验，(a) 是原图，(b) 是 Ground truth，(c) 是 FPN 的结果，(d) 是 FPN+FAMs，(e) 是 FPN+PPM，(f) 是 FPN+GGM，(g) 是 FPN+GGM+FAMs。





# 总结

该 paper 提出了两种基于池化技术的模块 GGM（全局引导模块）和 FAM（特征整合模块），改进 FPN 在显著性检测的应用，而且这两个模块也能应用在其他金字塔模型中，具有普遍性，但是 FAM 的整合过程我认为有点像是**用平均中和了上采样带来的混叠效应**，但是不够优雅，**先下采样池化再上采样带来的损失可能代价太大**。

# 参考文献

[1]. Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017. 1, 3.

[2]. Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. A stagewise refinement model for detecting salient objects in images. In ICCV, pages 4019–4028, 2017. 1, 3, 6, 7, 8.

[3].Nian Liu and Junwei Han. Dhsnet: Deep hierarchical saliency network for salient object detection. In CVPR, 2016.1, 2, 3, 7, 8.

[4]. Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. IEEE TPAMI, 41(4):815–828, 2019. 1, 2, 3, 5, 6, 7, 8.

[5]. Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. A stagewise refinement model for detecting salient objects in images. In ICCV, pages 4019–4028, 2017. 1, 3, 6, 7, 8.

[6]. Tiantian Wang, Lihe Zhang, Shuo Wang, Huchuan Lu, Gang Yang, Xiang Ruan, and Ali Borji. Detect globally, refine locally: A novel approach to saliency detection. In CVPR, pages 3127–3135, 2018. 1, 3, 6, 7, 8.

[7]. Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, 2017. 1, 3.

[8]. Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In CVPR, pages 1155–1162, 2013. 1, 5, 8.

[9]. Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille. The secrets of salient object segmentation. In CVPR, pages 280–287, 2014. 5, 7, 8.

[10]. Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In CVPR, pages 3166–3173, 2013. 5, 6, 7, 8.

[11]. Guanbin Li and Yizhou Yu. Visual saliency based on multiscale deep features. In CVPR, pages 5455–5463, 2015. 2, 5, 6, 7, 8.

[12]. Vida Movahedi and James H Elder. Design and perceptual validation of performance measures for salient object segmentation. In CVPR, pages 49–56, 2010. 5, 6, 7, 8.

[13]. Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan. Learning to detect salient objects with image-level supervision. In CVPR, pages 136–145, 2017. 5, 7, 8.

[14]. Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In ICCV, pages 1395–1403, 2015. 6.

https://mmbiz.qpic.cn/mmbiz_png/VBcD02jFhgmPEF4lW0pL5weJia5y4xhJbog2pIZZ3ZCgVUDynvus6rCzNKGAAAI6R8jaXTpYPISCMicpFegVdG0g/640?wxfrom=5&wx_lazy=1&wx_co=1

**点击以下标题查看更多往期内容：**

* [PFLD：简单高效的实用人脸关键点检测算法](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496432&idx=1&sn=e61134b51b2422ed2d79fc3ba17c6bf5&chksm=96ea2d70a19da466e09732832c09983c2ede9dec55d6f5b59d3e856c30f62895e423197a32f4&scene=21#wechat_redirect)
* [CVPR 2019 | 实体零售场景下密集商品的精确探测](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247497260&idx=1&sn=15b47d4fff58cb9f58d0ec924d0f65c8&chksm=96ea29aca19da0ba2df95bfba810b3386d049d77055fdc9d6ad1e2f6ac934fd46effbbdba001&scene=21#wechat_redirect)
* [CVPR 2019 | STGAN: 人脸高精度属性编辑模型](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247497236&idx=1&sn=9bdf1220087291cbb97017ec42225608&chksm=96ea2994a19da082f7b39c9a280e936a46ab31f8a6a7f3bb65af17de3ff65e116f20638d124c&scene=21#wechat_redirect)
* [从动力学角度看优化算法：GAN的第三个阶段](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496992&idx=1&sn=6589bed77fc3d6c9a2b0bfa1f562082a&chksm=96ea2aa0a19da3b6a3d8d944b32259df238c32b797a48a67cd0e60d4fdc1364f8f8277f9f0e8&scene=21#wechat_redirect)
* [CVPR 2019 | 基于高清表示网络的人体姿态估计](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496906&idx=1&sn=c2d064d54e6dd0c4ca3fa4d0c3576a3e&chksm=96ea2b4aa19da25cd75bb33169fb707fa14b27088b84ddb6bee0fdd8deb623b275d98b83ed3e&scene=21#wechat_redirect)
* [免费中文深度学习全书：理论详解加代码分析](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496408&idx=1&sn=da933f6f69dd21aecaaa21d700d8e7df&chksm=96ea2d58a19da44e9b3d5e2ecc038f8a26ae2d48cbc27d7207e14ead75721dbc0b562b26145d&scene=21#wechat_redirect)
* [目标检测小tricks之样本不均衡处理](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247496155&idx=1&sn=b720e982a8e99a5db93a48c59ecad8d5&chksm=96ea2e5ba19da74dff62f9e57043423e9dd4ac4109933aad530b96fed279d80e5f7ad516eb50&scene=21#wechat_redirect)
* [小米拍照黑科技：基于NAS的图像超分辨率算法](http://mp.weixin.qq.com/s?__biz=MzIwMTc4ODE0Mw==&mid=2247495166&idx=1&sn=a158e603651bc4f26836151a9113e856&chksm=96ea327ea19dbb68b8987aca041bb21579a35b1c679e91fd2368c7f2fb7acd58508cd531bdfe&scene=21#wechat_redirect)