

Beyond Skip Connections: Top-Down Modulation for Object Detection

笔记本: Technique

创建时间: 2019/10/31 21:48

更新时间: 2019/10/31 10:18

作者: ming71

这篇论文是FPN的同期研究，也是通过多尺度特征融合改善检测精度以及小目标的性能。

作者提出了如下的TDM网络：

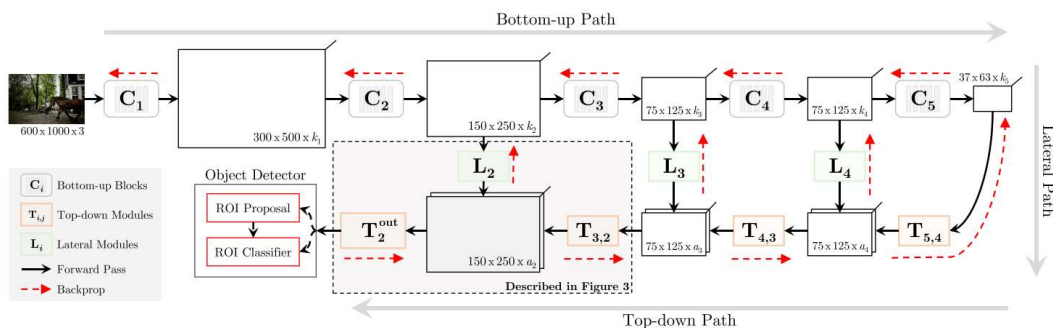


Figure 2. The illustration shows an example of **Top-Down Modulation (TDM)** Network, which is integrated with the bottom-up network with lateral connections. C_i are bottom-up, feedforward feature blocks, L_i are the lateral modules which transform low level features for the top-down contextual pathway. Finally, $T_{j,i}$, which represent flow of top-down information from index j to i . Individual components are explained in Figure 3 and 4.

也是类似于将前向不同层级的特征构建top-down连接融合不同尺度特征图的信息。为了分析上图具体结构，看其中的一个模块如下。如下图左边，从模块 C_i 出来的特征图为 $H \times W \times K_i$ ，经过 L_i 模块卷积得到 $H \times W \times l_i$ 维度，然后与经过两倍上采样的 $H \times W \times T_j$ 进行**通道叠加卷积融合**（右图）。最终采用的融合了多阶特征的 T_{out} 进行预测或proposal。

L模块负责学习怎样转换底层特征；T模块负责学习上一层的哪些特征是需要保留的。

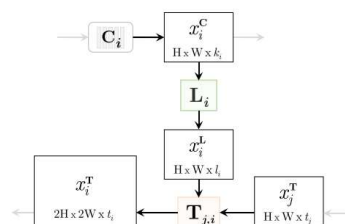


Figure 3. The basic building blocks of Top-Down Modulation Network (detailed Section 3.1).

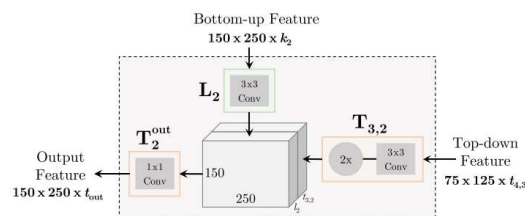


Figure 4. An example with details of top-down modules and lateral connections. Please see Section 3.1 for details of the architecture.

比较尴尬的是，作者发现一个一个地加入特征捷联模块训练的效果要比直接搭建统一训练的效果好，所以采取逐个训练的策略（麻烦）。

Ablation study

- 是不是搭建的TDM越深越好（在硬件允许的条件下用到越底层的特征越好）？

通过实验证明一般是这样，但是VGG-16却在倒数第二个卷积层达到了性能饱和，再往前融合特征不会取得更佳的效果了，原因不明（这个问题FPN在

Ablation study时没有探讨过)

该网络特点:

- 捷径连接 (L T模块) , 让网络自己决定保留不同特征图哪部分的特征
- 信息融合方式是通道叠加后卷积融合 (yolov3后来借鉴的这个)
- **缺点:** 只在融合了多尺度信息的单特征图上预测 (相比多尺度预测)