通道注意力机制发展

Subsequently, development of attention modules can be roughly divided into two directions: (1) enhancement of feature aggregation; (2)combination of channel and spatial attentions.

SE-Net presents for the first time an effective mechanism to learn channel attention and achieves promising performance.

GSoP introduces a second-order pooling for more effective feature aggregation.

GE explores spatial extension using a depth-wise convolution [5] to aggregate features.

CBAM and **scSE** compute spatial attention using a 2D convolution of kernel size k x k, then combine it with channel attention
...

通道注意力模块能够显著提升网络的学习表达能力。基于 SENet 有很多改进优化，其中改进的一个点就是 SENet 中两层全连接不仅导致通道信息丢失，同时还导致参数量与计算量变多。而在 VoVNet 中，将两层全连接替换为一层，验证了效果不会变差，但有利于轻量化。

最新论文 ECANet（CVPR2020）基于 SENet 提出了一种速度与精度 tradeoff 的结构，改进了注意力模块中全连接这一个点，利用了 1D 卷积，卷积核大小 k 自定义。实验结果表明，在各种常用 BackBone 中引入 ECANet 能够实现速度基本不变的情况下，较大幅度提升精度。

最后，鉴于 VoVNet 以及 ECANet 简单易用且效果极佳，希望大家在以后的工作中，能够多多利用哦

VoVNet2
https://github.com/youngwanLEE/vovnet-detectron2

ECANet
https://github.com/BangguWu/ECANet

SENet

详解：https://blog.csdn.net/u014380165/article/details/78006626

we first revisit the channel attention module in SENet. Specifically, given the input features,SE block first employs a **global average pooling** for each channel independently, then **two fully-connected (FC) layers** with non-linearity followed by a **Sigmoid function** are used to generate channel weights. The two FC layers are designed to capture non-linear cross-channel interaction, which involve dimensionalty reduction for controlling model complexity.

Although this strategy is widely used in subsequent channel attention modules [33, 13, 9], our **empirical studies show dimensionality reduction brings side effect on channel attention prediction**, and it is inefficient and unnecessary to capture dependencies across all channels.

添加了 SE block 后，模型的参数到底增加了多少。其实从前面的介绍可以看出增加的参数主要来自两个全连接层，两个全连接层的维度都是 C/r * C，那么这两个全连接层的参数量就是 2*C^2/r

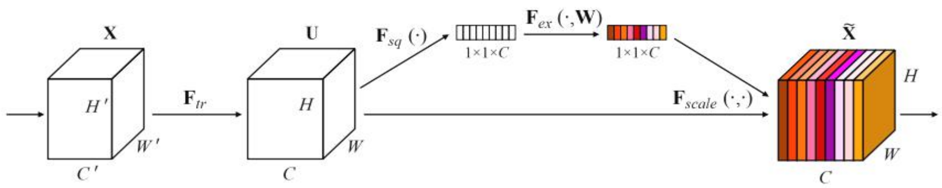


Fig. 1. A Squeeze-and-Excitation block.

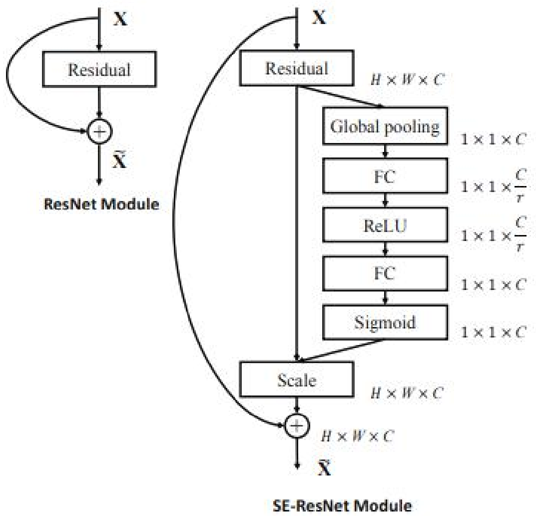


Fig. 3. The schema of the original Residual module (left) and the SE-ResNet module (right).

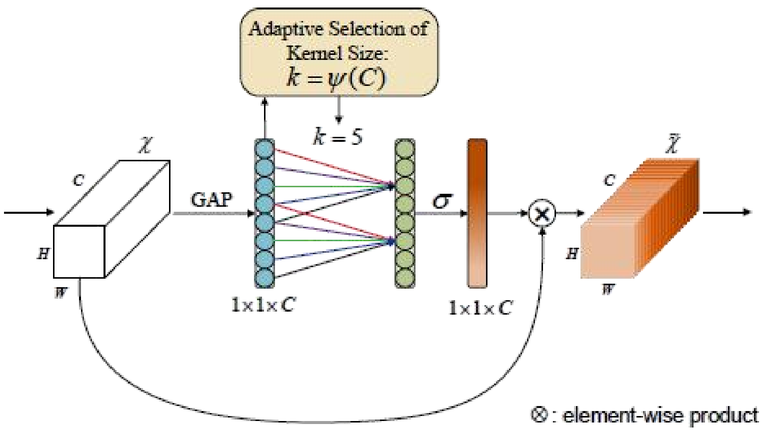
ECANet(CVPR2020)

出发点：1、SENet 中的通道压缩会导致信息丢失；2、and it is inefficient and unnecessary to capture dependencies across all channels.

创新点：1、after channelwise global average pooling without dimensionality reduction,our ECA captures local cross-channel interaction by considering every channel and its k neighbors. Such method is proven to guarantee both efficiency and effectiveness.

Note that our ECA can be efficiently implemented by fast 1D convolution of size k, where kernel size k represents the coverage of local cross-channel interaction

结果： We extensively evaluate our ECA module on image classification, object detection and instance segmentation with backbones of ResNets and MobileNetV2.
The experimental results show our module is more efficient while performing favorably against its counterparts.



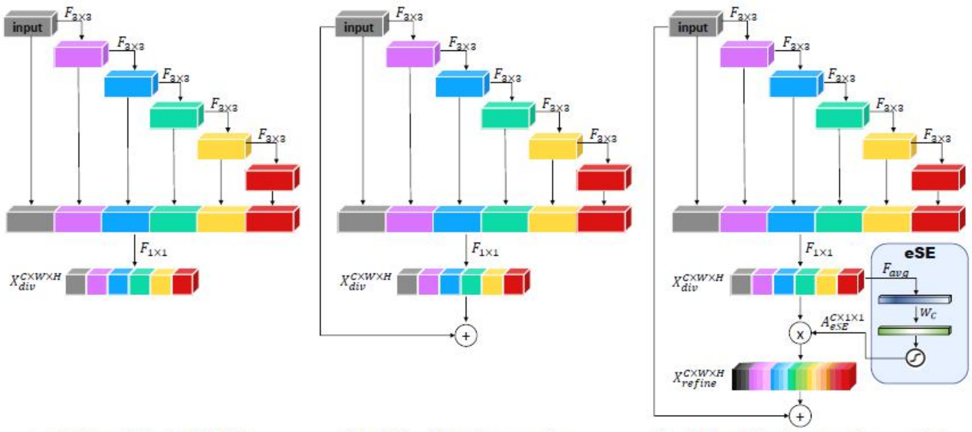

```
class eca_layer(nn.Module):
    """Constructs a ECA module.
    Args:
        channel: Number of channels of the input feature map
        k_size: Adaptive selection of kernel size
    """
    def __init__(self, channel, k_size=3):
        super(eca_layer, self).__init__()
        self.avg_pool = nn.AdaptiveAvgPool2d(1)
        self.conv = nn.Conv1d(1, 1, kernel_size=k_size, padding=(k_size - 1) // 2, bias=False)
        self.sigmoid = nn.Sigmoid()

    def forward(self, x):
        # x: input features with shape [b, c, h, w]
        b, c, h, w = x.size()

        # feature descriptor on the global spatial information
        y = self.avg_pool(x)

        # Two different branches of ECA module
        y = self.conv(y.squeeze(-1).transpose(-1, -2)).transpose(-1, -2).unsqueeze(-1)

        # Multi-scale information fusion
        y = self.sigmoid(y)

        return x * y.expand_as(x)
```

VoVNet2

出发点：SENet 中使用了两次全连接（C—>C/r—>C），通道先压缩后扩张，导致特征信息有丢失。因此，VoVNet2 中将两层全连接改为一个全连接。

Feature map (h*w*c)—>Favg (1*1*c)—>FC(1*c)—>sigmoid(c)—>相乘

优缺点：可以作为目标检测和图像分割的 Backbone 网络，结合了 DenseNet、ResNet 以及 ESNet，效果很好。



(a) OSA module in VoVNet (b) OSA + identity mapping (c) OSA + identity mapping + eSE

Figure 3: Comparison of OSA modules. $F_{1\times1}, F_{3\times3}$ denote $1\times1, 3\times3$ conv layer respectively, F_{avg} is global average pooling, W_C is fully-connected layer, A_{eSE} is channel attention map, $\otimes$ indicates element-wise multiplication and $\oplus$ denotes element-wise addition.