

一文读懂计算机视觉中的注意力机制原理及其模型发展



The banner for ProCon 2019 AI Developer Conference features a blue background with abstract light patterns. On the left, the 'AI ProCon' logo is displayed with the text '开发者大会' (Developer Conference) and the dates '2019.9.6-7' and location '北京·富力万丽酒店' (Beijing · Fuli Wanyu Hotel). To the right, it lists '9大热门技术专题' (9 major hot technology topics), '60+前沿技术专家' (60+ cutting-edge technology experts), and '百余家知名AI企业' (over 100 famous AI companies). The CSDN logo is in the top right corner. At the bottom, a ticket price of '2050元' (original price 5888) is shown with the text '限时抢购中!' (limited time purchase!). A QR code is located on the right side of the banner.

AI ProCon
开发者大会
2019.9.6-7 | 北京·富力万丽酒店

9大热门技术专题
60+前沿技术专家
百余家知名AI企业

CSDN

早鸟票已售罄，优惠票 **2050元** (原价5888) **限时抢购中!**

学生票享惊喜优惠哦!

作者&编辑 | 言有三

来源 | AI有道 (ID:yanyousan_ai)

导读：Attention机制在近几年来在图像，自然语言处理等领域中都取得了重要的突破，被证明有益于提高模型的性能。

Attention机制本身也是符合人脑和人眼的感知机制，这次我们主要以计算机视觉领域为例，讲述Attention机制的原理，应用以及模型的发展。

1 Attention机制与显著图

1.1 何为Attention机制

所谓Attention机制，便是聚焦于局部信息的机制，比如图像中的某一个图像区域。随着任务的变化，注意力区域往往会发生变化。

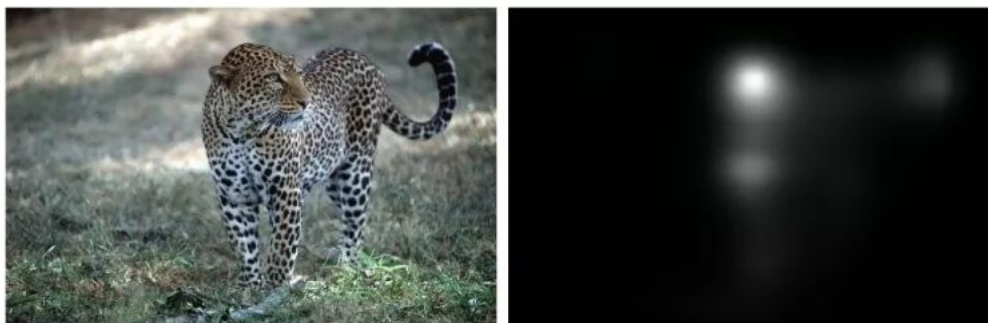


面对上面这样的一张图，如果你只是从整体来看，只看到了很多人头，但是你拉近一个一个仔细看就了不得了，都是天才科学家。

图中除了人脸之外的信息其实都是无用的，也做不了什么任务，Attention机制便是要找到这些最有用的信息，可以想见最简单的场景就是从照片中检测人脸了。

1.2 基于Attention的显著目标检测

和注意力机制相伴而生的一个任务便是显著目标检测，即salient object detection。它的输入是一张图，输出是一张概率图，概率越大的地方，代表是图像中重要目标的概率越大，即人眼关注的重点，一个典型的显著图如下：



右图就是左图的显著图，在头部位置概率最大，另外腿部，尾巴也有较大概率，这就是图中真正有用的信息。

显著目标检测需要一个数据集，而这样的数据集的收集便是通过追踪多个实验者的眼球在一定时间内的注意力方向进行平均得到，典型的步骤如下：

- (1) 让被测试者观察图。
- (2) 用eye tracker记录眼睛的注意力位置。
- (3) 对所有测试者的注意力位置使用高斯滤波进行综合。
- (4) 结果以0~1的概率进行记录。

于是就能得到下面这样的图，第二行是眼球追踪结果，第三行就是显著目标概率图。



上面讲述的都是空间上的注意力机制，即关注的是不同空间位置，而在CNN结构中，还有不同的特征通道，因此不同特征通道也有类似的原理，下面一起讲述。

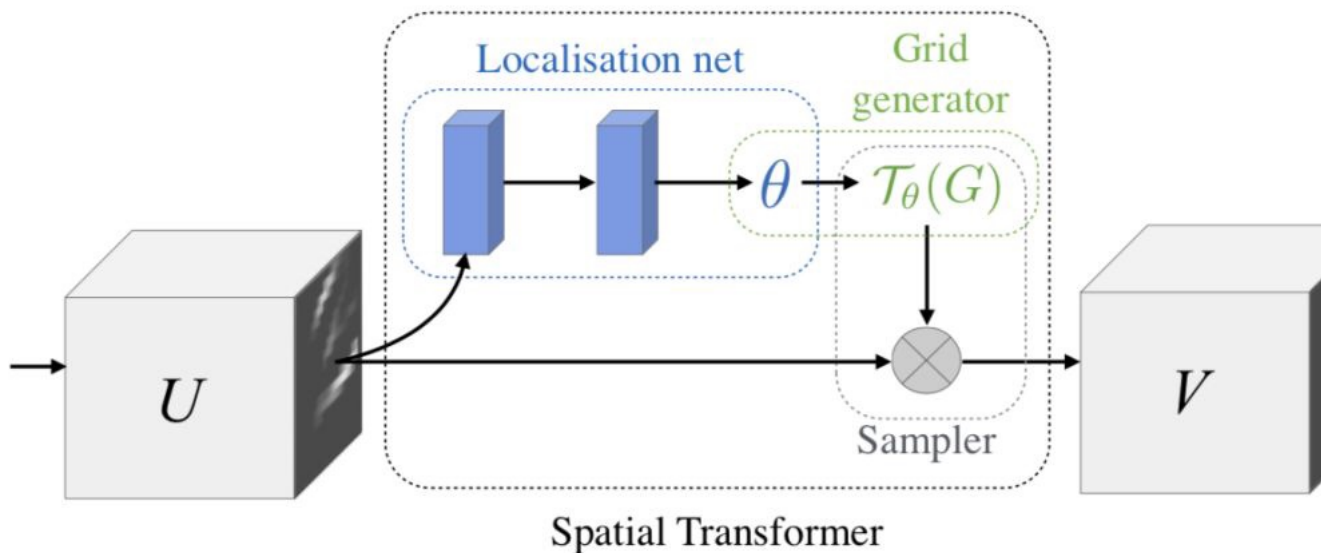
2 Attention模型架构

注意力机制的本质就是定位到感兴趣的信息，抑制无用信息，结果通常都是以概率图或者概率特征向量的形式展示，从原理上来说，主要分为空间注意力模型，通道注意力模型，空间和通道混合注意力模型三种，这里不区分soft和hard attention。

2.1 空间注意力模型(spatial attention)

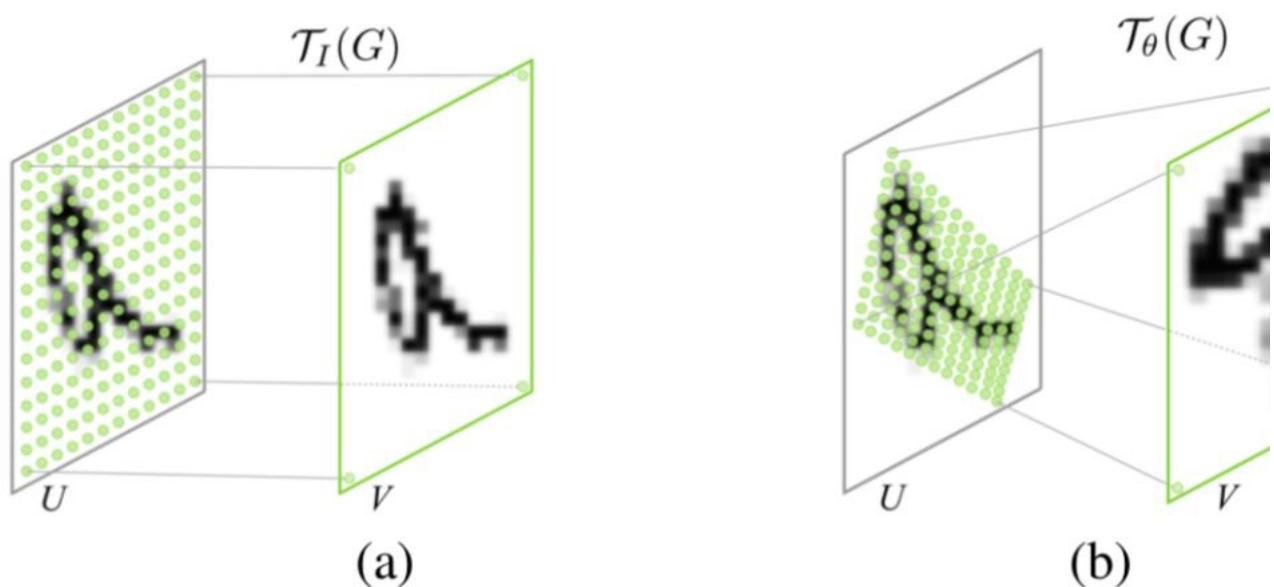
不是图像中所有的区域对任务的贡献都是同样重要的，只有任务相关的区域才是需要关心的，比如分类任务的主体，空间注意力模型就是寻找网络中最重要的部位进行处理。

我们在这里给大家介绍两个具有代表性的模型，第一个就是Google DeepMind提出的STN网络(Spatial Transformer Network[1])。它通过学习输入的形变，从而完成适合任务的预处理操作，是一种基于空间的Attention模型，网络结构如下：



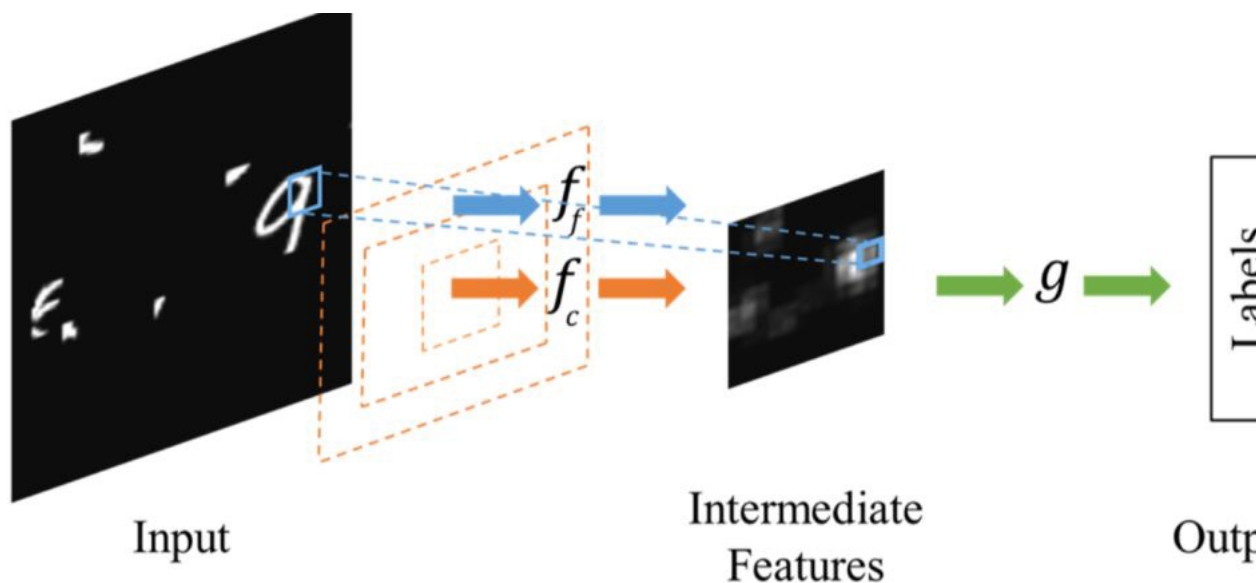
这里的Localization Net用于生成仿射变换系数，输入是 $C \times H \times W$ 维的图像，输出是一个空间变换系数，它的大小根据要学习的变换类型而定，如果是仿射变换，则是一个6维向量。

这样的一个网络要完成的效果如下图：



即定位到目标的位置，然后进行旋转等操作，使得输入样本更加容易学习。这是一种一步调整的解决方案，当然还有很多迭代调整的方案，感兴趣可以去[三知识星球](#)中阅读。

相比于Spatial Transformer Networks 一步完成目标的定位和仿射变换调整，Dynamic Capacity Networks[2]则采用了两个子网络，分别是低性能的子网络(coarse model)和高性能的子网络(fine model)。低性能的子网络(coarse model)用于对全图进行处理，定位感兴趣区域，如下图中的操作fc。高性能的子网络(fine model)则对感兴趣区域进行精细化处理，如下图的操作ff。两者共同使用，可以获得更低的计算代价和更高的精度。

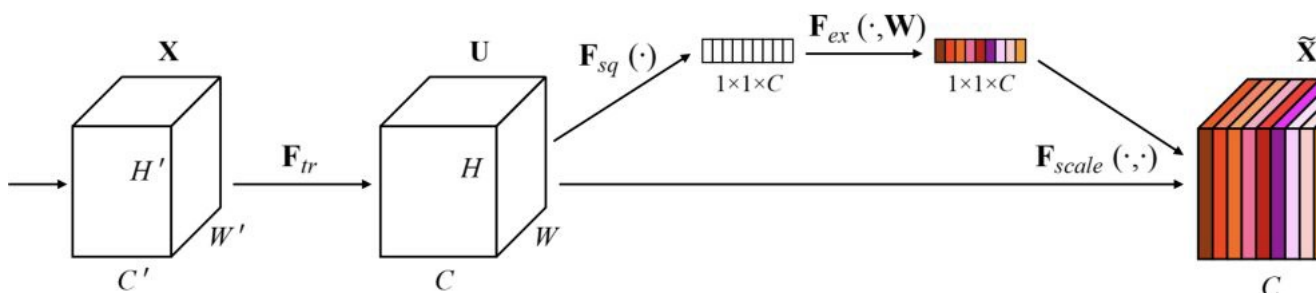


由于在大部分情况下我们感兴趣的区域只是图像中的一小部分，因此空间注意力的本质就是定位目标并进行一些变换或者获取权重。

2.2 通道注意力机制

对于输入2维图像的CNN来说，一个维度是图像的尺度空间，即长宽，另一个维度就是通道，因此基于通道的Attention也是很常用的机制。

SENet(Squeeze and Excitation Net)[3]是2017届ImageNet分类比赛的冠军网络，本质上是一个基于通道的Attention模型，它通过建模各个特征通道的重要程度，然后针对不同的任务增强或者抑制不同的通道，原理图如下。



在正常的卷积操作后分出了一个旁路分支，首先进行Squeeze操作(即图中 $F_{sq}(\cdot)$)，它将空间维度进行特征压缩，即每个二维的特征图变成一个实数，相当于具有全局感受野的池化操作，特征通道数不变。

然后是Excitation操作(即图中的 $F_{ex}(\cdot)$)，它通过参数 w 为每个特征通道生成权重， w 被学习用来显式地建模特征通道间的相关性。在文章中，使用了一个2层bottleneck结构(先降维再升维)的全连接层+Sigmoid函数来实现。

得到了每一个特征通道的权重之后，就将该权重应用于原来的每个特征通道，基于特定的任务，就可以学习到不同通道的重要性。

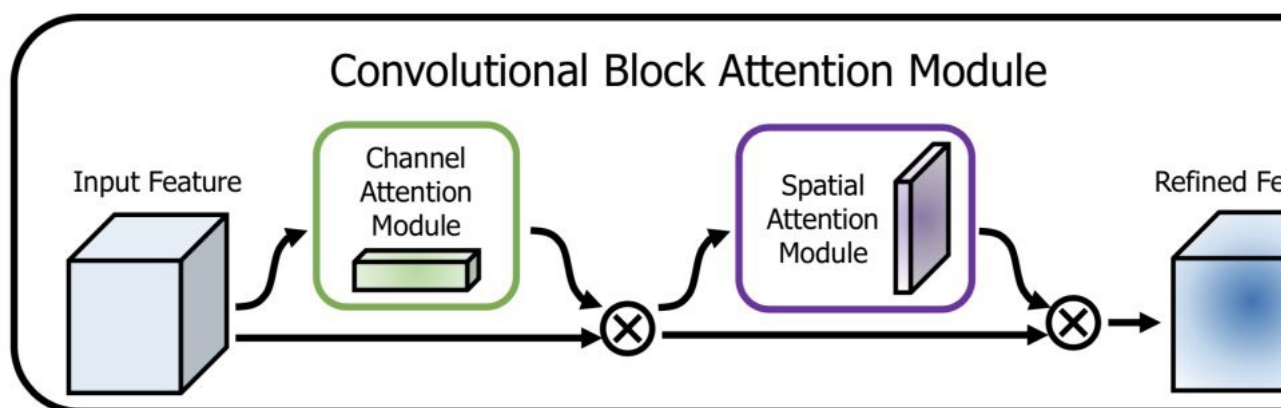
将其机制应用于若干基准模型，在增加少量计算量的情况下，获得了更明显的性能提升。作为一种通用的设计思想，它可以被用于任何现有网络，具有较强的实践意义。而后SKNet[4]等方法将这样的通道加权的思想和Inception中的多分支网络结构进行结合，也实现了性能的提升。

通道注意力机制的本质，在于建模了各个特征之间的重要性，对于不同的任务可以根据输入进行特征分配，简单而有效。

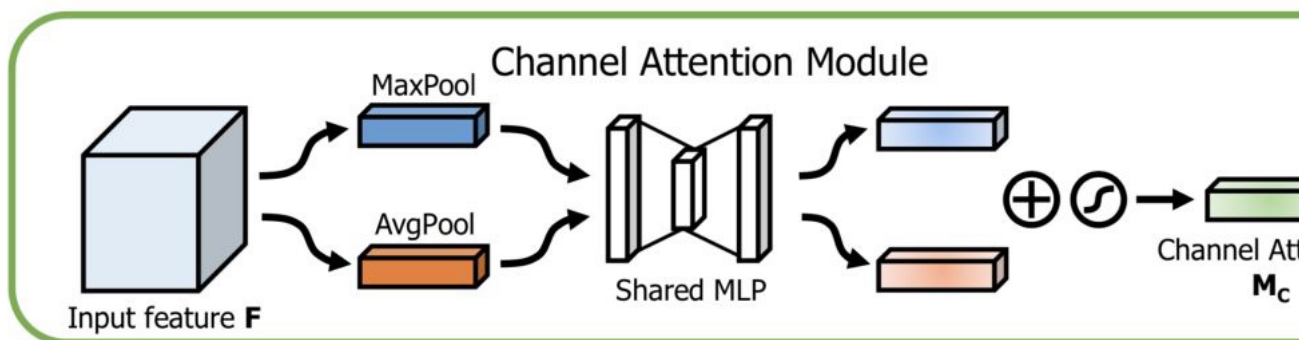
2.3 空间和通道注意力机制的融合

前述的Dynamic Capacity Network是从空间维度进行Attention，SENet是从通道维度进行Attention，自然也可以同时使用空间Attention和通道Attention机制。

CBAM(Convolutional Block Attention Module) [5]是其中的代表性网络，结构如下：

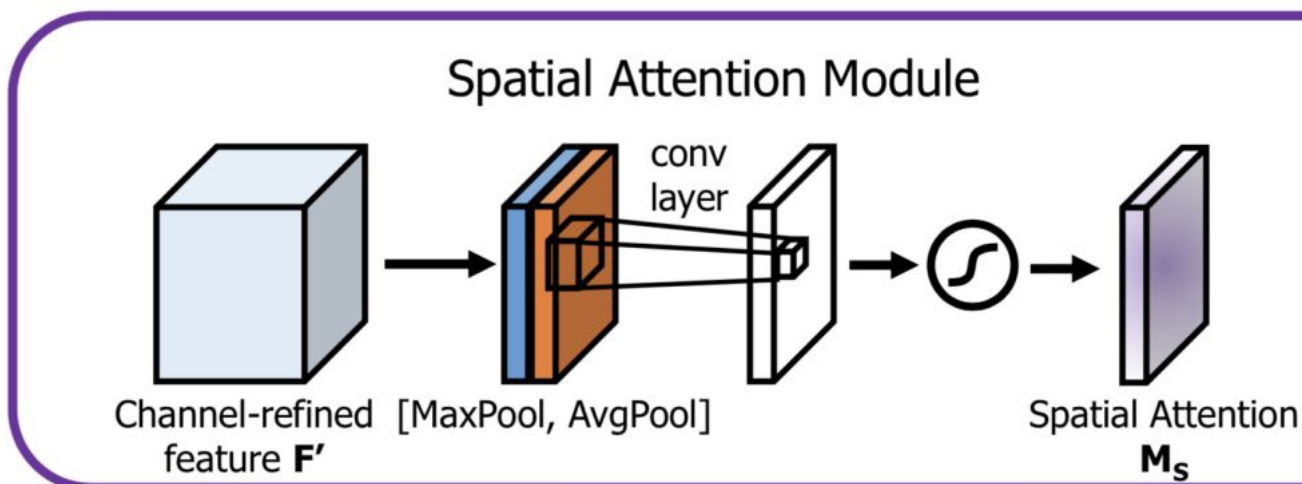


通道方向的Attention建模的是特征的重要性，结构如下：



同时使用最大pooling和均值pooling算法，然后经过几个MLP层获得变换结果，最后分别应用于两个通道，使用sigmoid函数得到通道的attention结果。

空间方向的Attention建模的是空间位置的重要性，结构如下：



首先将通道本身进行降维，分别获取最大池化和均值池化结果，然后拼接成一个特征图，再使用一个卷积层进行学习。

这两种机制，分别学习了通道的重要性和空间的重要性，还可以很容易地嵌入到任何已知的框架中。

除此之外，还有很多的注意力机制相关的研究，比如残差注意力机制，多尺度注意力机制，递归注意力机制等。

(*本文为 AI科技大本营转载文章，转载请联系原作者)

社群福利

扫码添加小助手，回复：大会，加入2019 AI开发者大会福利群，每周**一、三、五**更新技术福利，还有不定期的抽奖活动~



精彩推荐

60+技术大咖与你相约 2019 AI ProCon! 大会早鸟票已售罄，优惠票速抢进行中.....2019 AI开发者大会将于9月6日-7日在北京举行，这一届AI开发者大会有哪些亮点？一线公司的大牛们都在关注什么？AI行业的风向是什么？2019 AI开发者大会，倾听大牛分享，聚焦技术实践，和万千开发者共成长。

推荐阅读

- [华人“霸榜”ACL最佳长短论文、杰出论文一作，华为、南理工等获奖](#)
- [5G+AI重新定义生老病死?](#)
- [干货 | 20个Python教程，掌握时间序列的特征分析（附代码）](#)
- [2019 年度程序员吸金榜：你排第几？](#)
- [字节跳动入局全网搜索；思科回应中国区裁员；IntelliJ IDEA 新版发布！ | 极客头条](#)
- [AI+DevOps正当时](#)

- [知名饮料制造商股价暴涨500%惊动FBI，只因在名字中加入了“区块链”？](#)



你点的每个“在看”，我都认真当成了喜欢