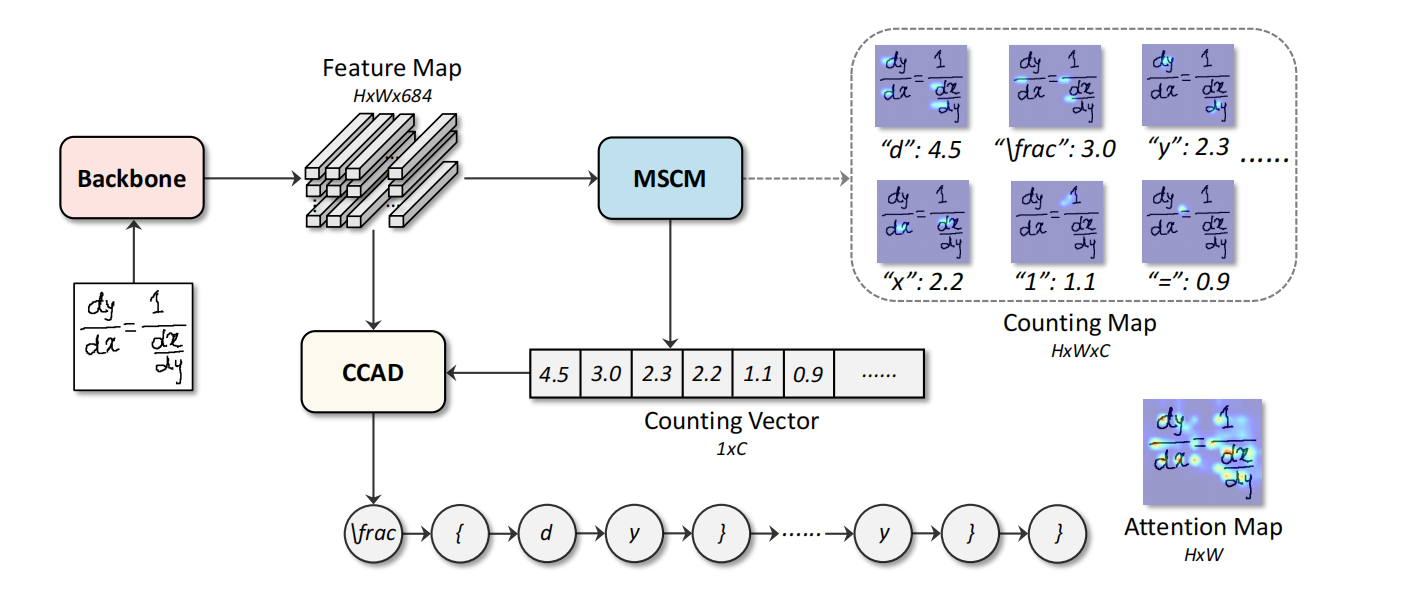
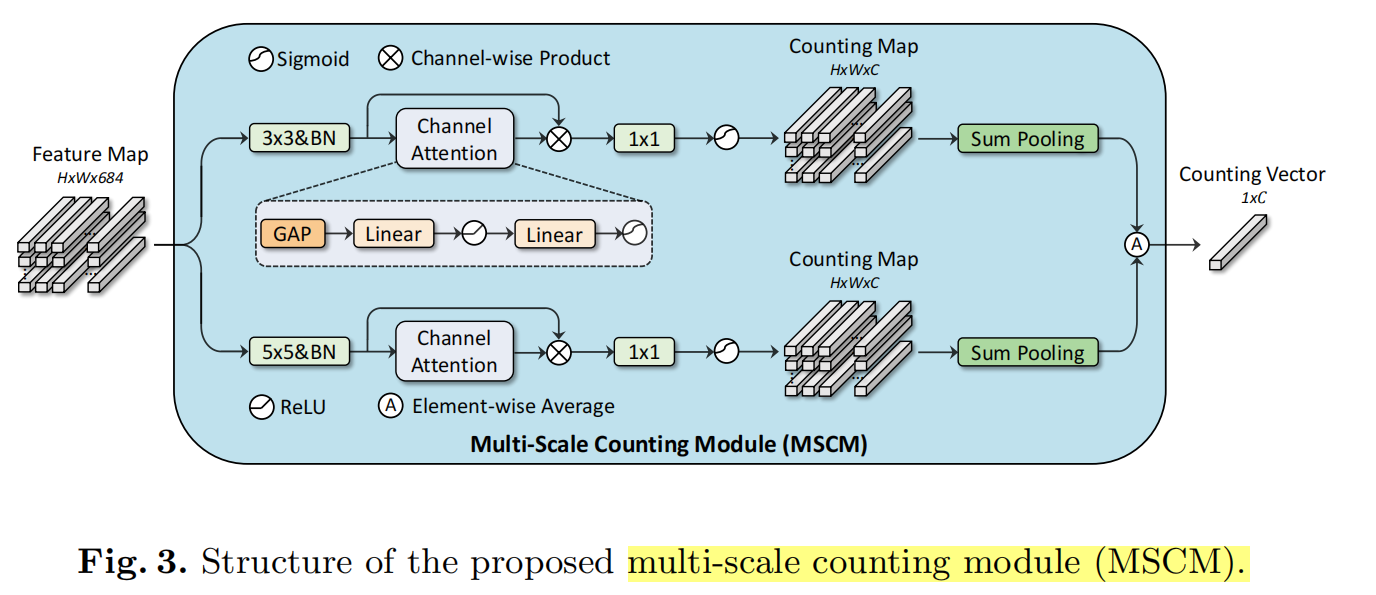
**An overview figure of CAN**



Backbone: DenseNet

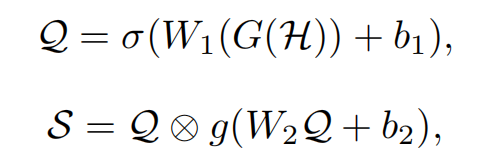
**Multi-Scale Counting Module(MSCN)**

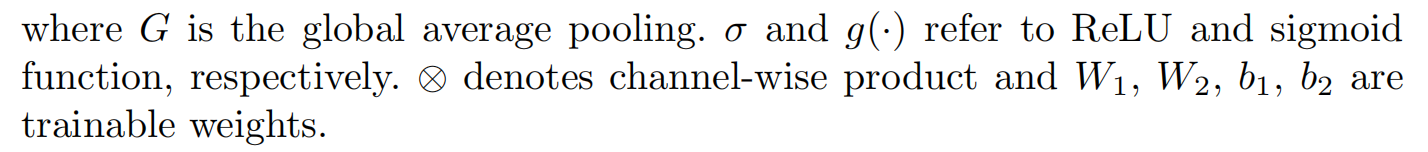


①单个内核大小不能有效地处理规模变化,利用两个并行卷积分支，通过使用不同的核大小（设置为3×3和5×5）来提取多尺度特征。

②采用Channel Attention来进一步增强特征信息

Let us denote H ∈ R(H×W×C) as the extracted feature map from the convolution (3 × 3 or 5 × 5) layer.The enhanced feature S can be written as:





**什么是Channel Attention（通道注意力模块）**[【channel attention&spatial attention】](https://blog.csdn.net/qq_42194397/article/details/122556066?spm=1001.2101.3001.6650.18&utm_medium=distribute.pc_relevant.none-task-blog-2~default~BlogCommendFromBaidu~Rate-18-122556066-blog-120884524.235^v38^pc_relevant_anti_vip_base&depth_1-utm_source=dis)



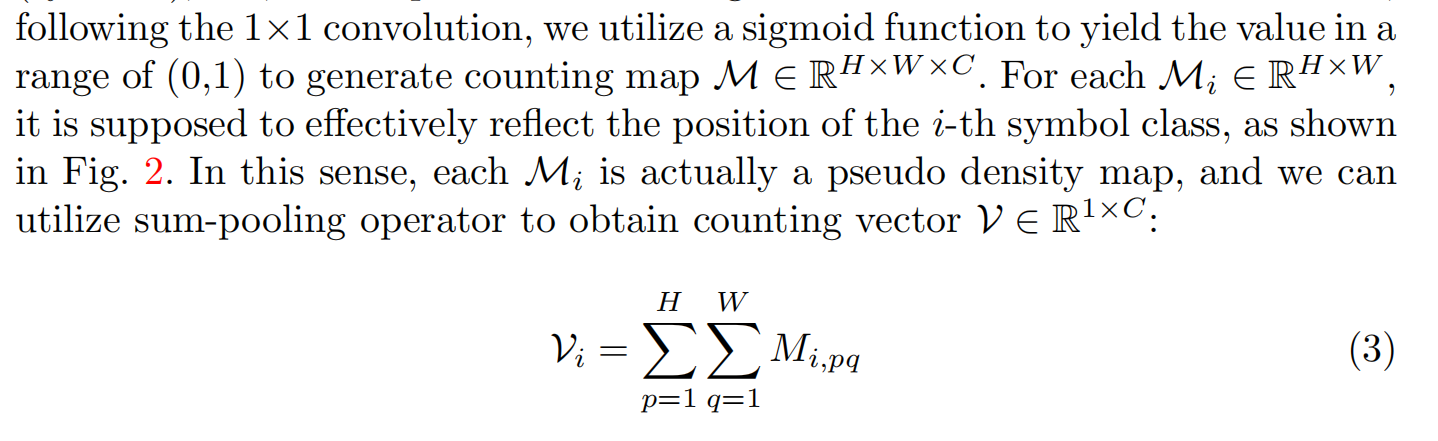
GAP:全局平均池化---HxWxC的输入图像被压缩成为1x1x``C的通道描述符

使用通道注意力模块的目的：为了让输入的图像更有意义，大概理解就是，通过网络计算出输入图像各个通道的重要性（权重），也就是哪些通道包含关键信息就多加关注，少关注没什么重要信息的通道，从而达到提高特征表示能力的目的。

③1×1 convolution layer:

reduce the channel number from C’ to C, where C is the number of symbol classes.

④：sum-pooling the counting map:



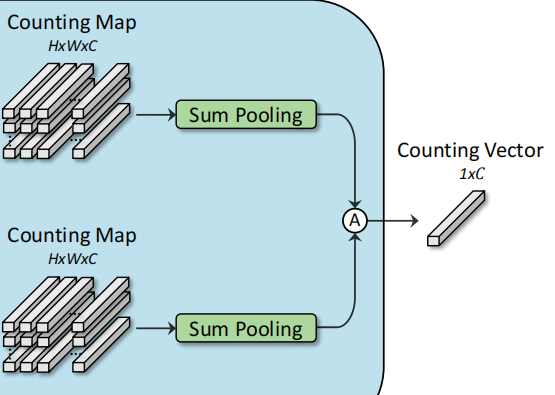
M的第i个channel的特征图反应第i个字符的分布情况

⑤：Element-wise Average

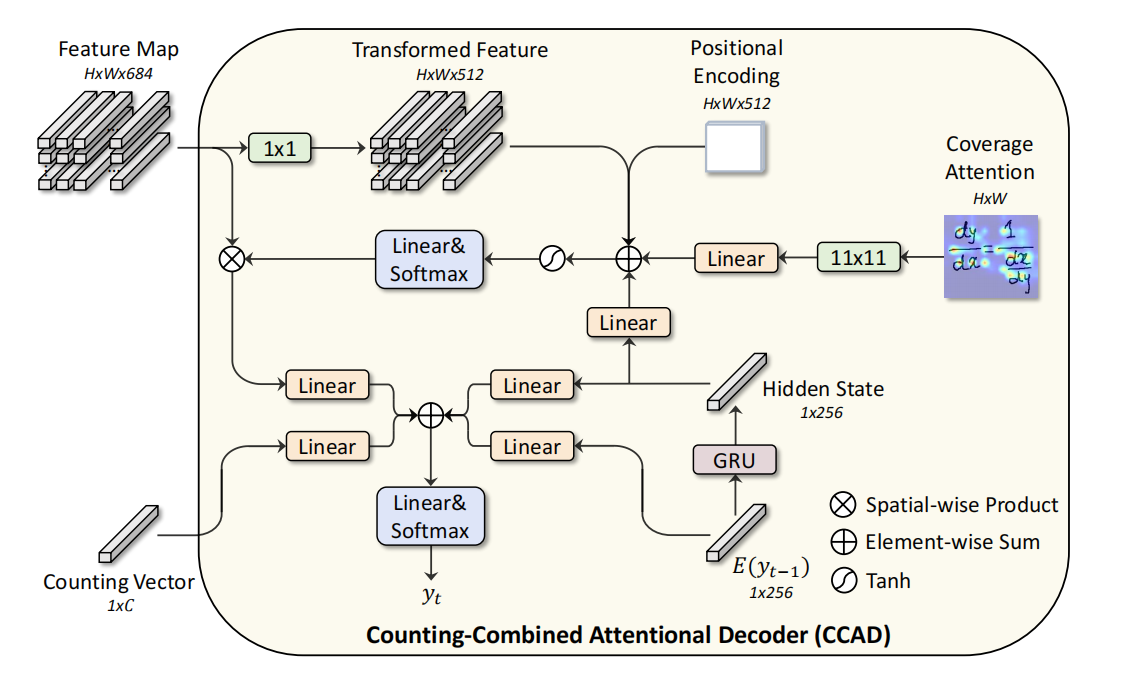
the feature maps of different branches contain different scale information

and are highly complementary. Thus, we combine the complementary counting

vectors and use the average operator to generate the fifinal result Vf∈ R(1×C),



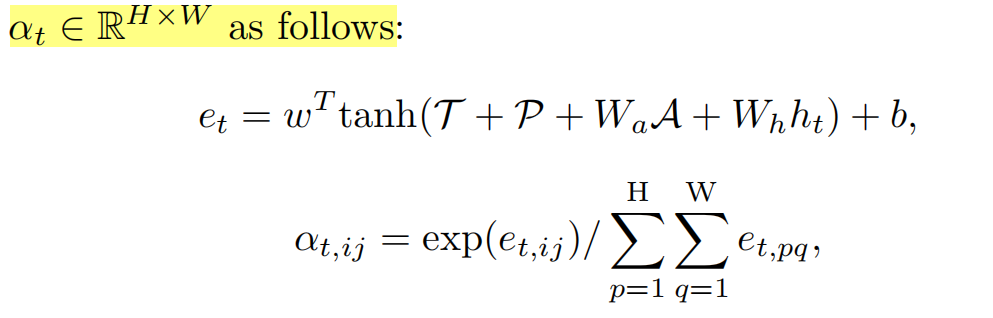
**Counting-Combined Attentional Decoder(CCAD)**



①1×1 convolution 得到 transformed feature T∈R(H×W×512) ,压缩channel，之后对T就可以使用absolute encoding[【position encoding】](https://zhuanlan.zhihu.com/p/121126531)得到 P∈R (H×W×512)来表示T中不同的空间位置

②将上一步的特征yt-1嵌入到GRU[【GRU】](https://zhuanlan.zhihu.com/p/32481747)【补充机器学习常见[product](https://blog.csdn.net/oldlybaby/article/details/108307712)】单元中，得到t步骤的隐藏特征ht∈R(1×256)

③由T、P、A(coverage attention[【coverage attention】](https://blog.csdn.net/u013250861/article/details/114241279))、ht计算得注意力权重：



④将spatial-wise product应用于注意权重αt和特征图F得到文本向量C∈R(1×256)，

⑤作者认为在之前的大多数HMER方法中，他们只使用向量C、隐藏状态ht和E（yt−1）来预测yt。实际上，C只是对应于特征图F的一个局部区域，ht和E（yt−1）也缺乏全局信息。但作者从MSCN模块中得到的计数向量V是从全局计数的角度计算出来的，具有全局信息，能使使预测更加准确。于是将它们组合在一起，预测yt如下：

