

20 个注意力机制涨点模块

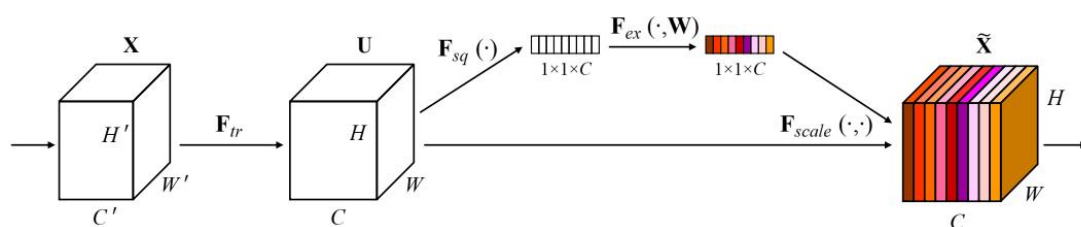
无痛涨点！轻松论文！

1、SENet：通道注意力模块	1
2、ECA：高效通道注意力模块	2
3、Selective Query Recollection（SQR）	3
4、HAT：混合注意力机制	4
5、自适应空间特征融合（ASFF）	5
6、CA 注意力	6
7、完全注意力 FANs	7
8、无参数注意力机制（simAM）	8
9、Focused Linear Attention：聚焦式线性注意力模块	9
10、卷积三重注意力模块	11
11、SAN-SAW：域泛化语义分割模型	12
12、EMA：高效多尺度注意力模块	13
13、GAM 注意力模块	14
14、BAM 双边注意力模块	15
15、CoTAttention	16
16、MobileViTAttention	17
17、ShuffleAttention	18
18、S2Attention	19
19、ParNetAttention	20
20、CoordAttention	21

1、SENet：通道注意力模块

论文：Squeeze-and-Excitation Networks

简述：本文提出了一种新的架构单元"挤压激励"（SE）块，通过明确模拟通道间的依赖关系，适应性地校准通道特征反应。在此基础上，研究人员构建了 SENet 架构，其具有出色的通用性，SE 块对于现有的先进深度架构来说，可以在极小的额外计算成本下提供重大的性能改进。



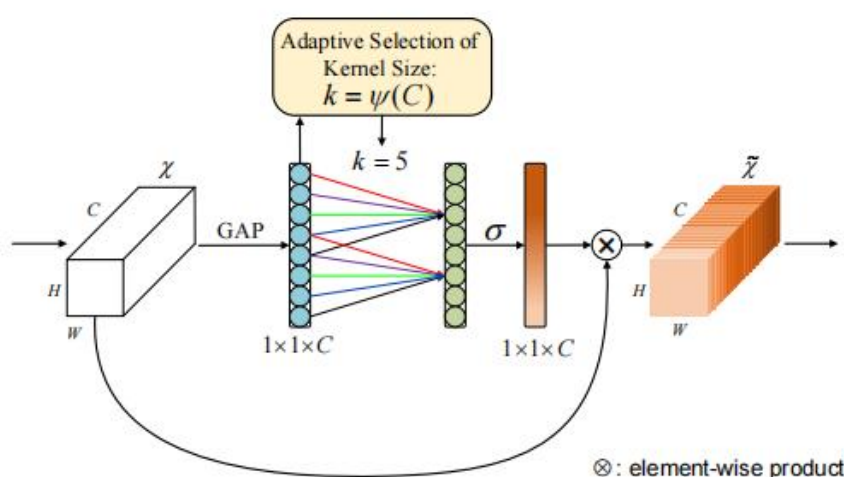
优点：能够显著提高网络对特征通道的敏感性，通过学习不同通道的重要性并据此调整特征响应，从而增强模型对关键特征的表征能力。

缺点：增加了计算复杂度和模型参数量，在训练数据不足的情况下可能导致过拟合，且在某些任务中效果提升不明显。

2、ECA：高效通道注意力模块

论文：ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks

简述：本文提出了一个名为 **ECA** 的高效通道注意力模块，它不降维，仅用少量参数显著提升性能。通过一维卷积实现局部通道交互策略，并自适应确定最佳卷积核大小。实验证明 **ECA** 模块在 ResNet50 等网络中提高准确率，同时降低参数数量和计算成本。此模块在图像分类、对象检测和实例分割等多项任务中表现出更高的效率和性能。



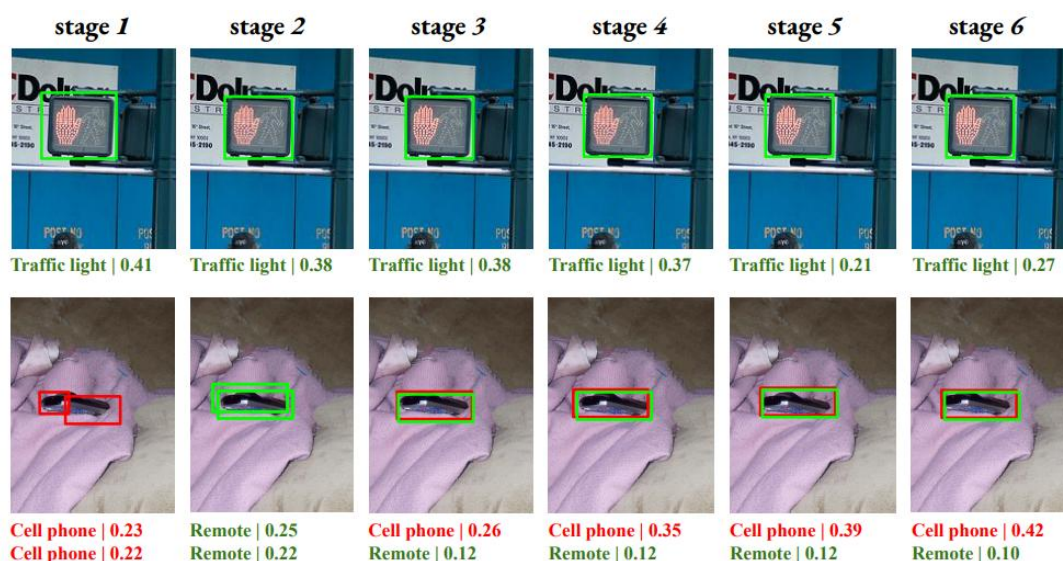
优点：避免了 SEnet 中复杂的全局池化和全连接层，从而显著减少了计算量和参数数量，使其更加高效。

缺点：使用局部跨通道交互，这可能不如 SEnet 的全局池化能够捕捉到全局的通道依赖关系。此外，ECA 的性能可能对卷积核大小和卷积层的设计参数比较敏感，需要仔细调整以获得最佳效果。

3、Selective Query Recollection (SQR)

论文：Enhanced Training of Query-Based Object Detection via Selective Query Recollection

简介：这篇论文研究了基于查询的目标检测器在最后解码阶段预测错误而在中间阶段预测正确的现象。通过回顾训练过程，作者归因该现象于两个限制：后期阶段缺乏训练强调以及解码顺序导致的级联错误。为此，作者设计了选择性查询回忆(SQR)策略来增强基于查询的目标检测器的训练。该策略累积收集中间查询，并选择性地将其直接输入后期阶段，从而强调后期阶段的训练，并让后期阶段可以直接使用中间查询。



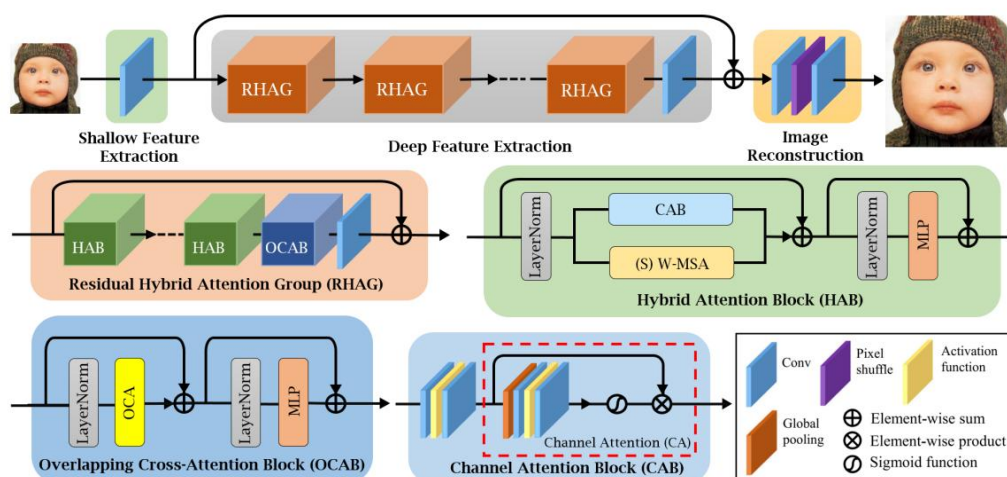
优点：通过选择性地关注和回忆过去的隐藏状态，能够有效地捕捉长期依赖关系，提高模型在处理长序列时的性能。

缺点：SQR 的实现可能相对复杂，需要精心设计选择和回忆机制。此外，选择性地回忆过去的信息可能会导致一些重要细节的丢失，特别是在处理非常长的序列时。

4、HAT：混合注意力机制

论文：Activating More Pixels in Image Super-Resolution Transformer

简述：本文提出了一种称为混合注意力 Transformer（HAT）的新型网络架构，它融合了通道注意力和窗口自注意力机制的长处，提高了模型处理全局和局部信息的能力。此外，研究人员还引入了一个跨窗口的注意力模块，用以强化邻近窗口特征间的互动。通过在训练阶段实行同任务预训练，进一步提升了模型性能。经过一系列实验，这个方法在性能上显著优于现有最先进技术，达到了 1dB 以上的提高。



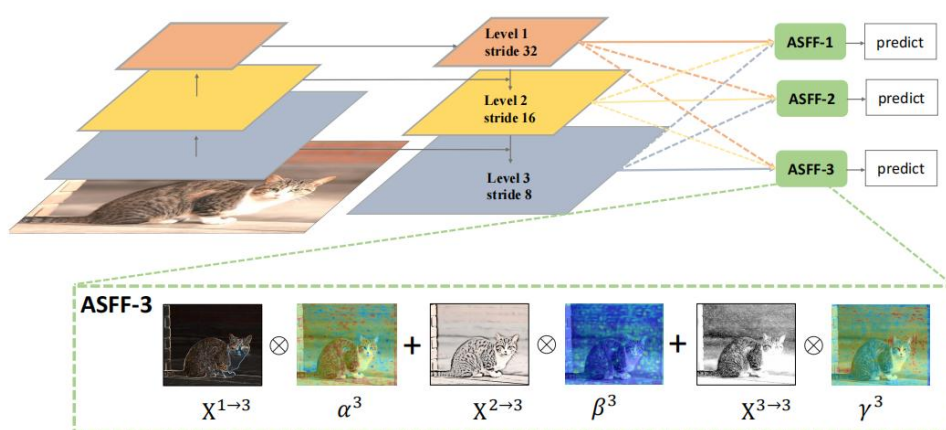
优点：通过同时捕捉空间和通道维度的依赖关系能够更全面地捕捉特征图中的重要信息，从而提高模型在图像分类、目标检测等视觉任务中的性能。

缺点：相比于单一注意力机制可能会增加计算成本和模型的复杂性，需要更多的调试和优化。

5、自适应空间特征融合（ASFF）

论文：Learning Spatial Fusion for Single-Shot Object Detection

简介：为了处理物体检测中尺度变化的挑战，金字塔特征表示是常见的做法。但是，基于特征金字塔的单阶段检测器，不同尺度特征之间的不一致性是其主要局限。本文提出了一种新颖的数据驱动的金字塔特征融合策略，即自适应空间特征融合。它可以学习空间过滤冲突信息的方式来抑制不一致性，从而提高特征的尺度不变性，并增加很小的推理开销。



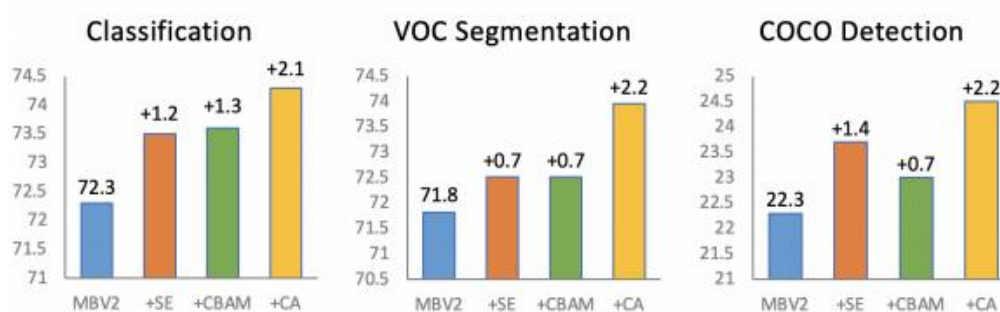
优点：能够有效地融合不同尺度的特征，提高模型对多尺度图像内容的理解。它通过学习自适应权重，增强了模型在目标检测和图像分割等任务中的性能。

缺点：引入了额外的计算成本，尤其是在处理高分辨率图像时。此外，学习融合权重增加了模型的复杂性。

6、CA 注意力

论文：Coordinate Attention for Efficient Mobile Network Design

简介：移动网络中的通道注意力机制通常会忽略空间位置信息，这对生成位置敏感的注意力图很重要。本文提出了一种坐标注意力机制，将位置编码嵌入到通道注意力中，以获得对位置敏感的注意力。它将通道注意力分解成两个方向的 1D 特征编码，每个方向聚合一维的特征，这样就可以在一个方向上捕获长程依赖，同时在另一个方向保留精确的位置信息。



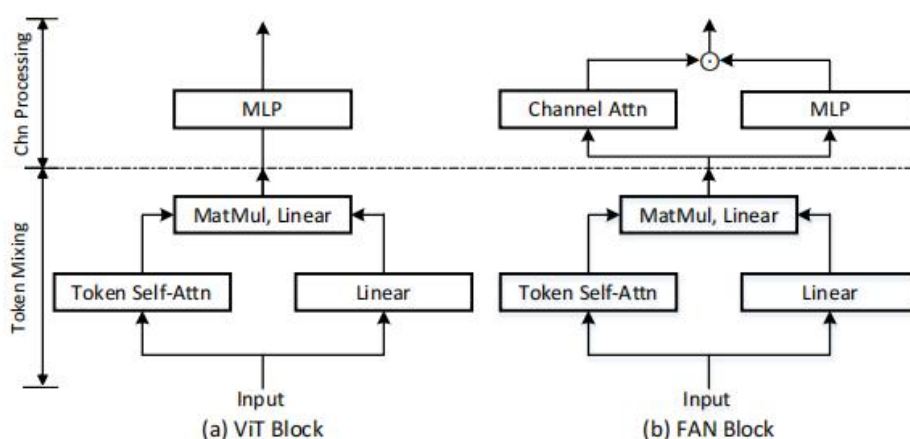
优点：能够有效地增强模型对特征通道的敏感性，提高特征表示的质量。它有助于网络更有效地利用特征信息，从而提高图像分类和其他视觉任务的性能。

缺点：引入了额外的计算成本，尤其是在处理通道数较多的特征图时。

7、完全注意力 FANs

论文：Understanding The Robustness in Vision Transformers

简介：最新的研究显示，视觉 Transformer(ViTs)在处理各种图像损坏时表现出很强的鲁棒性。尽管这种鲁棒性部分归因于自注意力机制，但我们对其中的工作原理还不是很清楚。论文通过引入全注意力网络(FANs)中的注意力通道模块，加强了自注意力在学习鲁棒特征表示方面的作用。



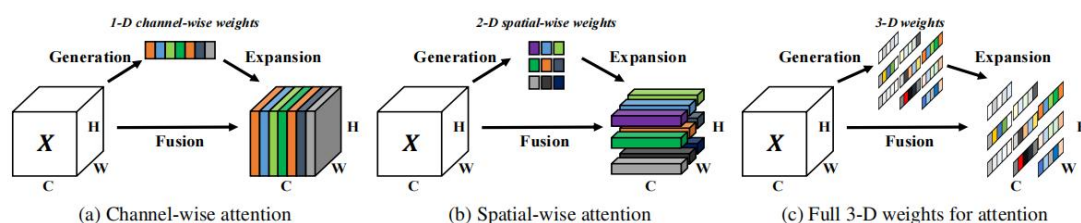
优点：通过全局注意力机制，能够更好地捕捉输入数据中的长距离依赖关系，提高模型在序列建模和图像识别等任务中的性能。

缺点：FANs 引入了额外的计算成本，尤其是在处理高维输入数据时。

8、无参数注意力机制（simAM）

论文：SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks

简介：论文提出一个简单高效的卷积神经网络注意力模块 SimAM。不同于现有逐通道或空间注意力，SimAM 不增加网络参数就可以为层内特征图推断三维注意力。具体来说，作者定义一个能量函数并导出闭式解来发现每个神经元的重要性，用少于 10 行代码实现。SimAM 运算符选择基于能量函数解得出，避免结构调优。



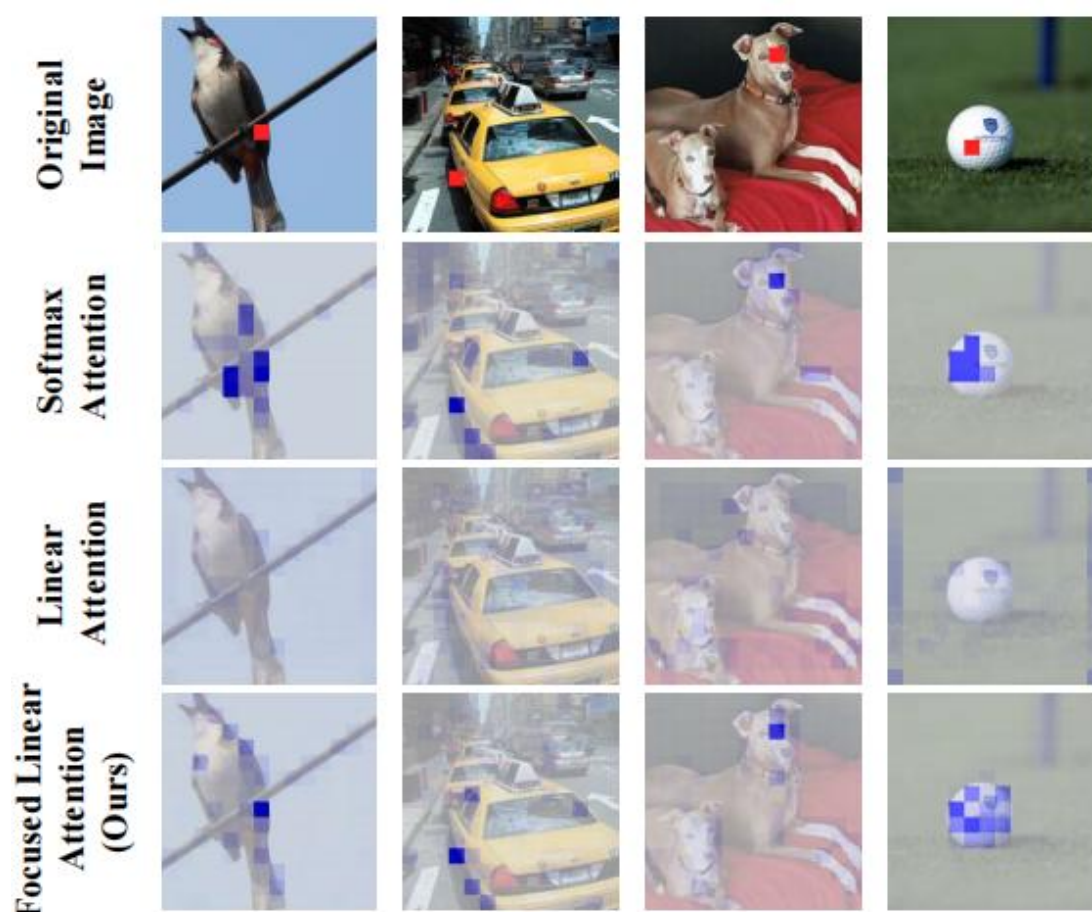
优点：设计简单，易于实现，并且可以应用于各种网络架构。能够自适应地调整特征的重要性，增强模型对关键特征的关注，提高特征表示的质量。

缺点：SimAM 的性能可能受到输入特征质量和网络深度的限制，效果可能因任务而异，对于某些任务，其性能提升可能不如其他类型的注意力机制明显。

9、Focused Linear Attention: 聚焦式线性注意力模块

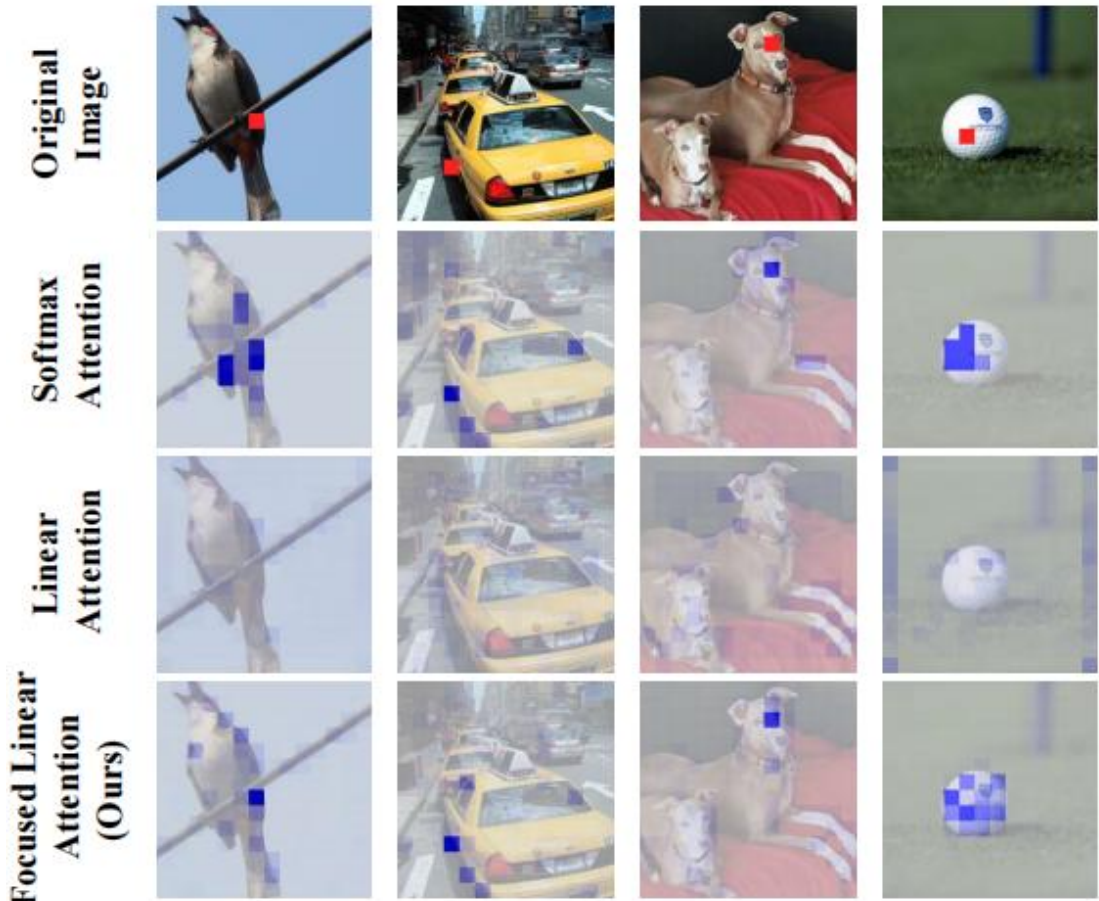
论文: FLatten Transformer: Vision Transformer using Focused Linear Attention

简述: 本文提出了一种新颖的 **Focused Linear Attention** 模块, 以实现高效率和表达性。通过分析影响性能的关键因素, 引入了一个简单而有效的映射函数和一个高效的等级恢复模块, 以增强 self-attention 的表达性, 同时保持较低的计算复杂度。大量的实验表明, 这个线性注意力模块适用于各种先进的 vision Transformers, 并在多个基准测试中取得了持续的改进性能。



优点：通过减少注意力计算中的冗余来提高效率，特别是在处理大规模数据时，同时更好地捕捉到关键特征，提高模型在序列建模和图像识别等任务中的性能，还可以显著减少内存使用。

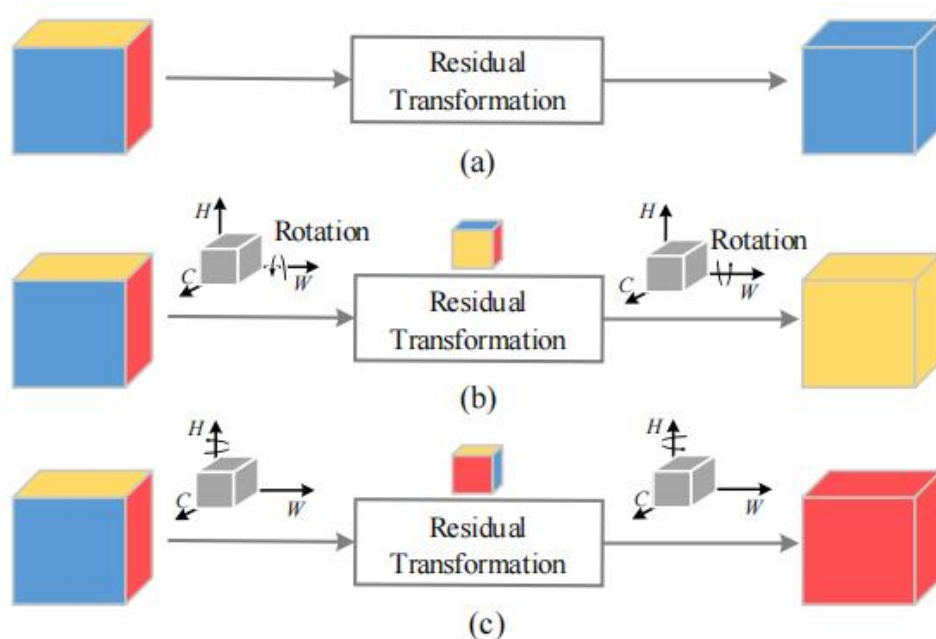
缺点：**Focused Linear Attention** 的内部机制可能相对复杂，理解和实现起来可能需要更多的专业知识，性能可能对某些超参数的选择比较敏感，需要仔细调整以达到最佳效果，且对于某些任务，其性能提升可能不如其他类型的注意力模块明显。



10、卷积三重注意力模块

论文：Rotate to Attend: Convolutional Triplet Attention Module

简介：本文研究了一种轻量级但有效的注意力机制——三元组注意力，它通过三支结构和旋转操作来捕获输入张量不同维度之间的交互，从而计算注意力权重。该方法可以轻松集成到典型的 CNN 模型中，对计算和参数量影响很小。



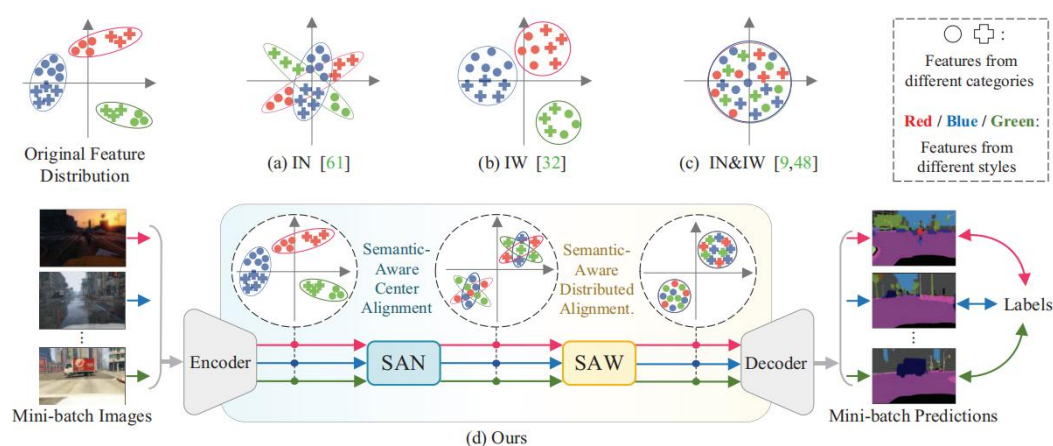
优点：通过同时捕捉空间、通道和尺度上的依赖关系，能够更全面地捕捉到特征图中的重要信息，有助于提高模型在图像分类、目标检测等视觉任务中的性能。

缺点：相比于单一注意力机制可能会增加计算成本，同时增加了模型结构的复杂性，可能需要更多的调试和优化。

11、SAN-SAW: 域泛化语义分割模型

论文: Semantic-Aware Domain Generalized Segmentation

简述: 本文提出了一个新的框架, 其中包括两个模块: 语义感知标准化 (SAN) 和语义感知白化 (SAW)。SAN 模块用于在不同图像样式之间进行类别级的特征对齐; SAW 模块则用于对齐特征分布, 并增强类内紧凑性与类间分离性。通过在多个标准数据集上的实验验证, 这种方法各种主干网络上都表现出了明显优于现有技术的改进。



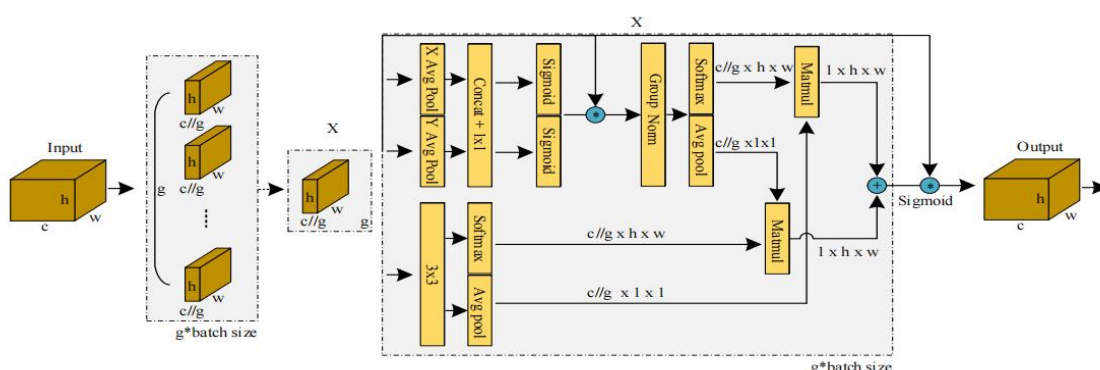
优点: 通过语义注意力和语义感知权重设计, 提高了模型在处理不同数据集时的泛化能力。这种模型能够更好地理解图像中的语义信息, 从而在域泛化语义分割任务中取得较好的性能。

缺点: 在实现语义注意力和语义感知权重时, 可能会增加模型的计算复杂度。此外, 模型的性能可能对超参数的选择比较敏感, 需要仔细调整以达到最佳效果。

12、EMA：高效多尺度注意力模块

论文：Efficient Multi-Scale Attention Module with Cross-Spatial Learning

简述：本文提出了一种叫做 **EMA** 的新型多尺度注意力模块，它能够同时保持信息有效性和降低计算成本，**EMA** 通过将通道分成子组和平滑特征组内的空间语义特征来实现这一目标。还设计了一个双分支结构，以编码全局信息，并通过跨维度的交互进一步加强特征。对 CIFAR100 图像分类和 MS COCO 与 VisDrone2019 数据集上的目标检测进行的实验验证了 **EMA** 的有效性，它在不增加网络复杂度的前提下，超越了多个现有的注意力机制。



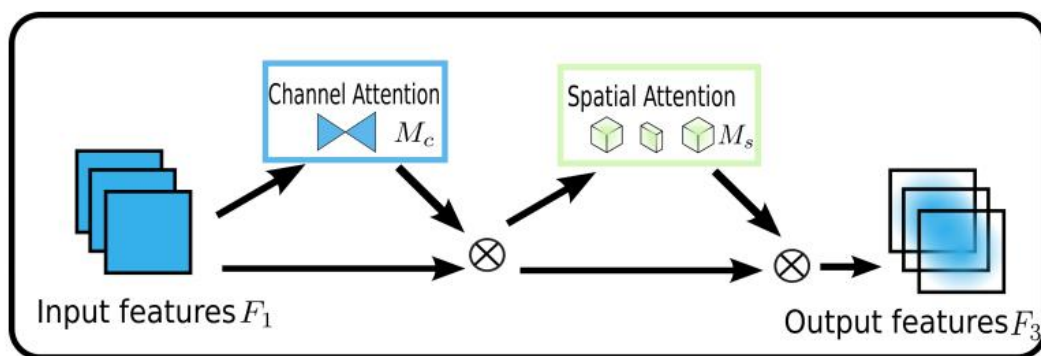
优点：能够有效地整合不同尺度的特征，增强模型对多尺度图像内容的理解，从而提高模型在处理复杂图像结构时的性能。同时，**EMA** 通常设计得较为高效，以平衡多尺度处理带来的额外计算成本。

缺点：在实现多尺度特征融合时，仍可能增加模型的参数量和计算复杂度。

13、GAM 注意力模块

论文: Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions

简介: 各种注意力机制被研究用以提高不同计算机视觉任务的性能。然而, 之前的方法忽略了保留通道和空间两个方面的信息以增强跨维度交互的重要性。因此, 论文提出一种全局注意力机制, 通过通道和空间双注意力减少信息损失, 增强全局特征交互, 从而提升视觉任务的性能。



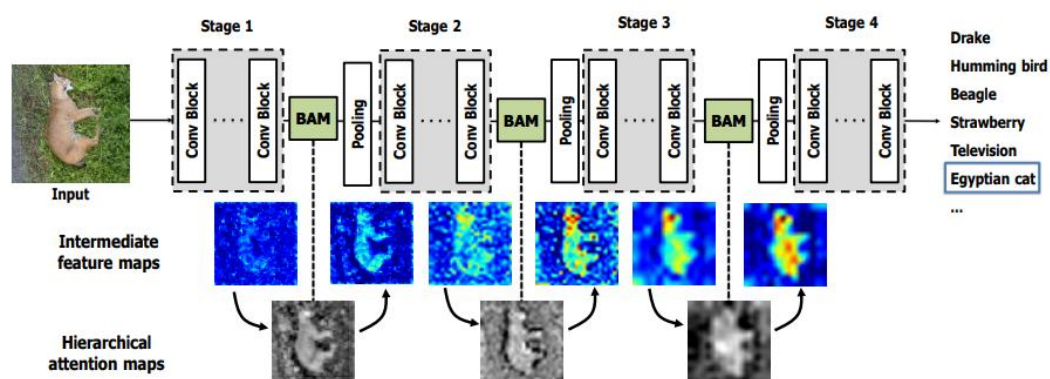
优点: 利用外部指导信息来引导注意力, 能够更精确地捕捉到重要的特征, 从而提高模型在图像分类、目标检测等视觉任务中的性能。同时, **GAM** 可以灵活地集成到不同的网络架构中, 提高了模型的适应性和鲁棒性。

缺点: **GAM** 依赖于外部指导信息, 这可能会增加模型的复杂性。

14、BAM 双边注意力模块

论文：BAM: Bottleneck Attention Module

简介：BAM 被提出用于提高深度神经网络的表示能力。BAM 模块通过两个独立的路径（通道和空间）来推断注意力图，可以与任何前馈卷积神经网络集成。这个模块在每个特征图的瓶颈处放置，即特征图下采样的位置。BAM 构建了一个具有参数的层次化注意力，可以与任何前向传播模型一起端到端地训练。



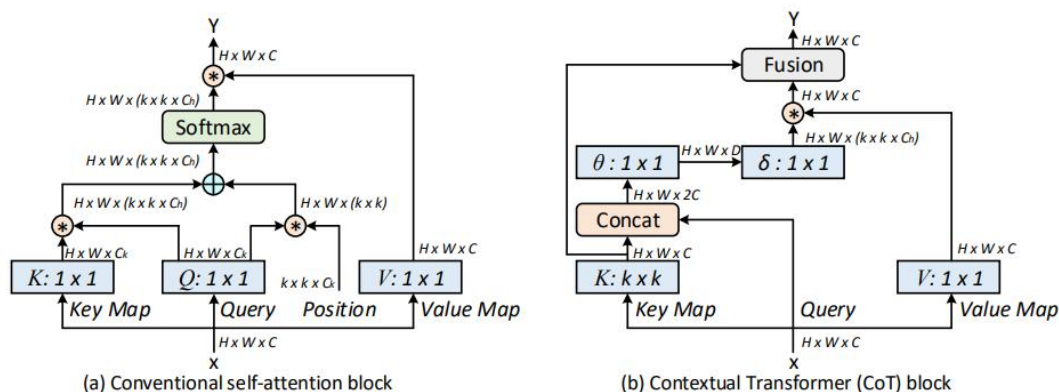
优点：能通过同时考虑通道和空间注意力来提高网络的表示能力，广泛适用于各种网络架构，并在多个基准测试中显示出性能提升。

缺点：包括增加计算复杂度，需要调整新的超参数，以及集成到复杂模型结构中可能较为困难。

15、CoTAttention

论文：Contextual Transformer Networks for Visual Recognition

简介：CoTAttention 是一种新型的 Transformer 模块，用于增强视觉识别能力。它在 YOLOv8 目标检测中得到了应用，并展示了其有效性。CoTAttention 模块通过 3×3 卷积编码上下文信息，并结合动态多头注意力矩阵，提高了视觉表示的能力。此外，它还提到了 Large Separable Kernel Attention (LSKA) 模块，它解决了大内核卷积的计算效率问题。CoTAttention 模块在 YOLOv8 中的应用，展示了其在目标检测任务中的优势。



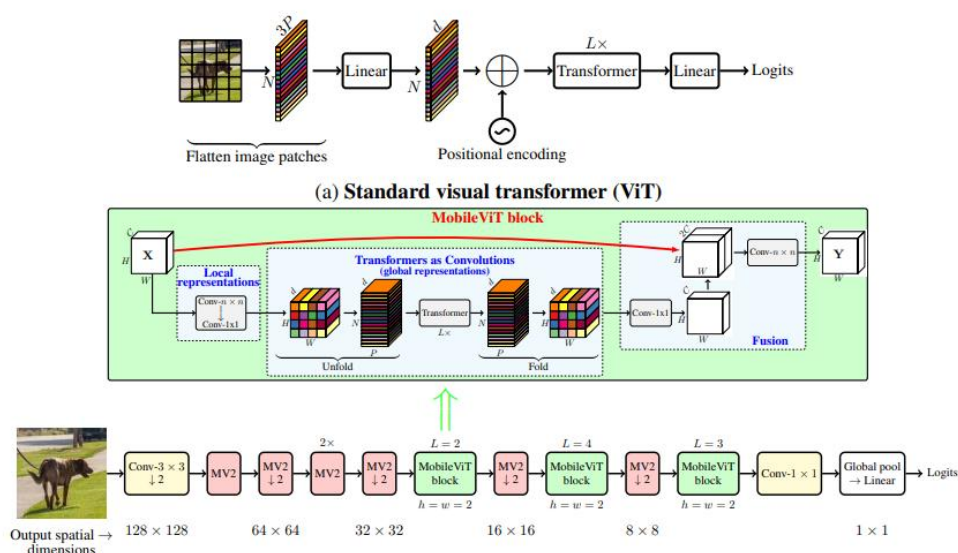
优点：通过结合上下文信息和动态多头注意力矩阵，显著提升了视觉识别的能力，尤其是在目标检测任务中。此外，它通过 LSKA 模块提高了大内核卷积的计算效率。

缺点：将 CoTAttention 集成到现有网络架构中可能需要一定的工程量，且可能需要调整超参数以达到最佳性能。

16、MobileViTAttention

论文： MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer

简介： MobileViTAttention 是苹果公司在 2021 年提出的一种轻量级、通用且适用于移动设备的视觉转换器注意力模块。这个模块结合了卷积神经网络（CNN）和视觉 Transformer（ViT）的优势，旨在为移动视觉任务构建一个轻量级、低延迟的网络。 MobileViTAttention 模块的主要特点包括其轻量级的结构和对数据增强的不敏感性。



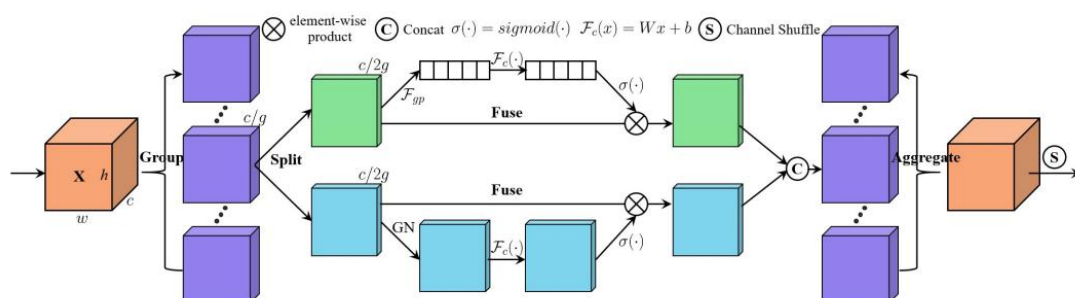
优点：轻量级和移动友好性，能够提供高效且适用于移动设备的视觉转换器注意力。它在 ImageNet-1k 数据集上达到了 78.4% 的 top-1 准确率，并在 MS-COCO 对象检测任务中表现出色。

缺点：可能需要更多的训练数据和更长的训练时间，以及对数据增强的高度敏感，这可能会增加训练难度。

17、ShuffleAttention

论文：SA-Net: Shuffle Attention for Deep Convolutional Neural Networks

简介：这篇论文提出了一种名为 Shuffle Attention (SA) 的模块，用于深度卷积神经网络（CNNs）。SA 模块旨在有效地结合空间注意力机制和通道注意力机制，以提高网络的性能。SA 模块通过将通道维度分组为多个子特征，并行处理这些子特征，然后采用“通道洗牌”操作符，以实现不同子特征之间的信息交流，这种设计使得 SA 模块既高效又有效。



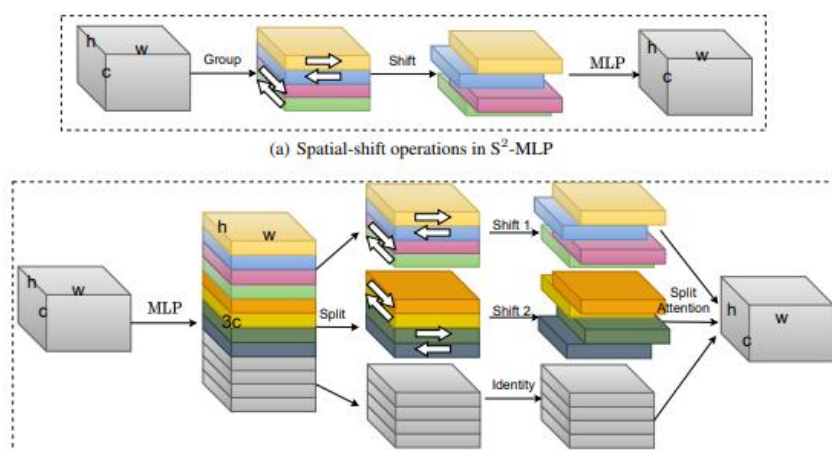
优点：能够有效地结合空间注意力机制和通道注意力机制，提高深度卷积神经网络的性能。SA 模块既高效又有效，在诸如 ImageNet-1k 分类、MS COCO 对象检测和实例分割等常用基准测试中，与当前最先进的方法相比，显著提高了准确性，同时保持了较低的模型复杂度。

缺点：需要调整超参数以达到最佳性能，以及在某些情况下，它可能会对模型性能产生一定影响。

18、S2Attention

论文：S2-MLPv2: Improved Spatial-Shift MLP Architecture for Vision

简介：这篇论文介绍了一种名为 **S2Attention** 的注意力机制模块，旨在解决基于 **Transformer** 的大型语言模型（LLMs）中软注意机制容易将上下文中的无关信息纳入其潜在表示中的问题。**S2Attention** 通过重新生成输入上下文，仅包括相关部分，然后关注重新生成的上下文以引出最终响应。这种方法在包含意见或无关信息的任务上，如问答、数学问题和长篇生成上，优于标准的基于注意力的 LLMs。



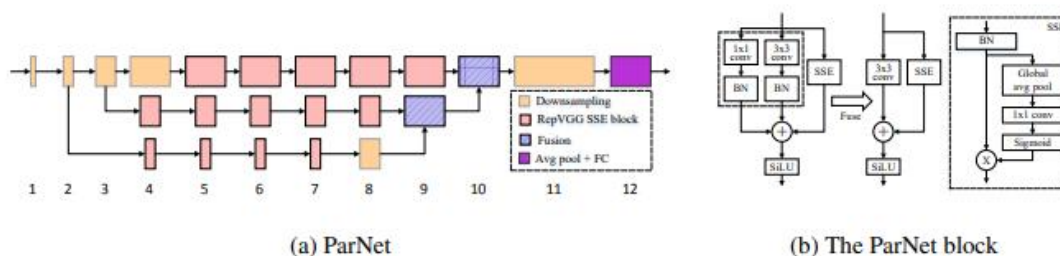
优点：能够有效地解决基于 **Transformer** 的大型语言模型（LLMs）中软注意机制容易将上下文中的无关信息纳入其潜在表示中的问题。增加了事实性和客观性，减少了虚假性。

缺点：潜在缺点可能包括需要调整超参数以达到最佳性能，以及在某些情况下，它可能会对模型性能产生一定影响。

19、ParNetAttention

论文：Non-deep Networks

简介：这篇论文介绍了 ParNetAttention（Parallel Networks Attention）模块，用于深度神经网络的注意力机制。ParNetAttention 通过将序列分割成多个子序列，并对每个子序列进行独立的注意力计算，从而降低了计算复杂度，使得网络在保持高性能的同时，有效地减少深度。



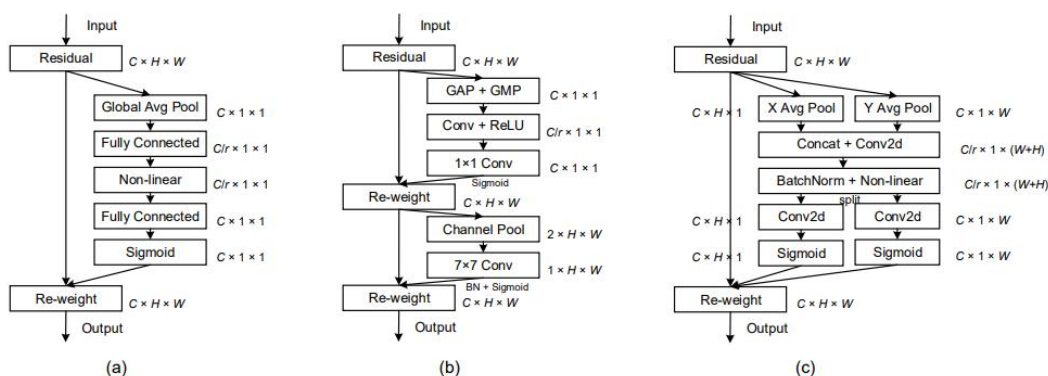
优点：通过将序列分割成多个子序列，并对每个子序列进行独立的注意力计算，从而降低了计算复杂度，使得网络在保持高性能的同时，有效地减少深度。

缺点：需要调整超参数以达到最佳性能，以及在某些情况下，它可能会对模型性能产生一定影响。

20、CoordAttention

论文：Coordinate Attention for Efficient Mobile Network Design

简介：这篇论文介绍了一种名为 CoordAttention 的注意力机制模块，用于轻量级移动网络。CoordAttention 通过嵌入位置信息到通道注意力中，使得网络能够关注大范围的区域，避免了空间位置信息的丢失。这种设计使得 CoordAttention 能够捕获跨通道的信息，还能捕获方向感知和位置感知的信息，帮助模型更精准地定位和识别感兴趣的目标。



优点：它能够通过嵌入位置信息到通道注意力中，使得网络能够关注大范围的区域，避免了空间位置信息的丢失。这种设计使得 CoordAttention 能够捕获跨通道的信息，还能捕获方向感知和位置感知的信息，帮助模型更精准地定位和识别感兴趣的目标。

缺点：需要调整超参数以达到最佳性能，以及在某些情况下，它可能会对模型性能产生一定影响。此外，集成 CoordAttention 模块到现有网络架构中可能需要一定的工程量，特别是对于复杂的模型结构。