

Expectation-Maximization Attention Networks for Semantic Segmentation

ICCV 2019 Oral

目录

CONTENTS

语义分割

PART ONE

语义分割中的
attention机制

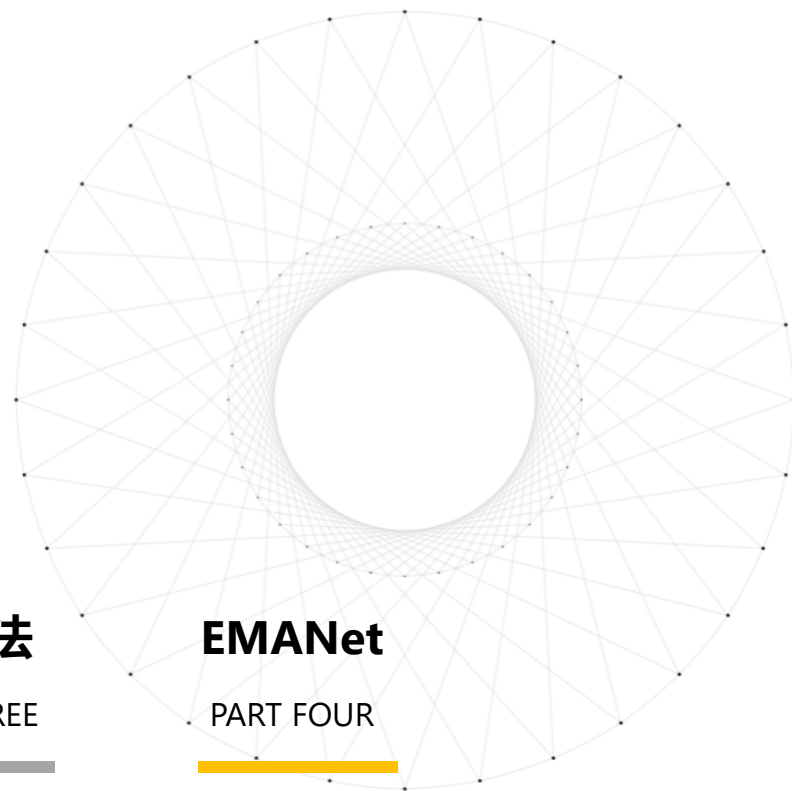
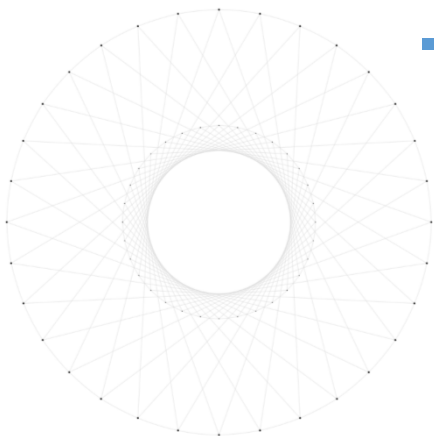
PART TWO

EM算法

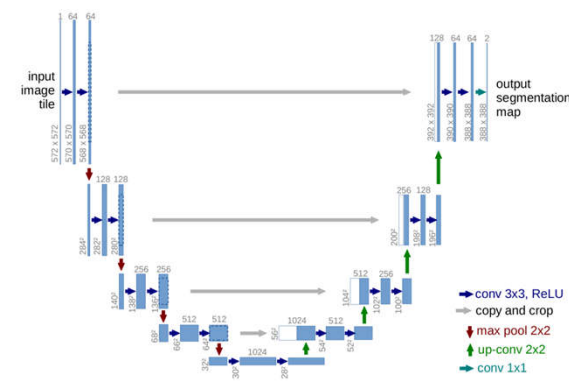
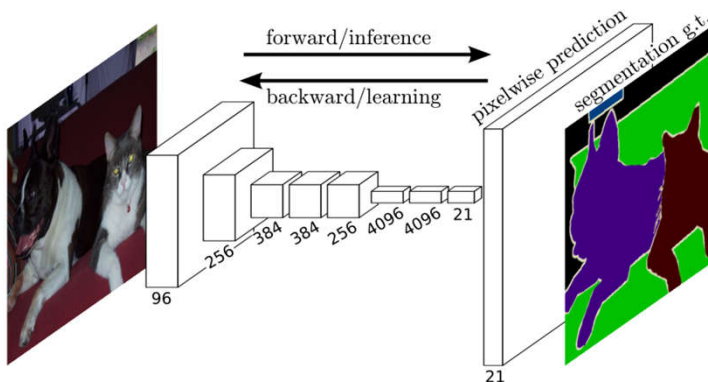
PART THREE

EMANet

PART FOUR

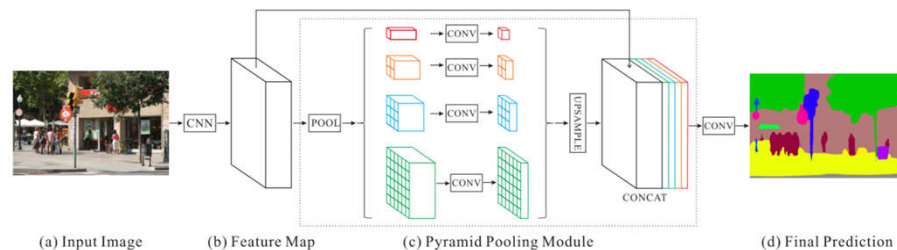
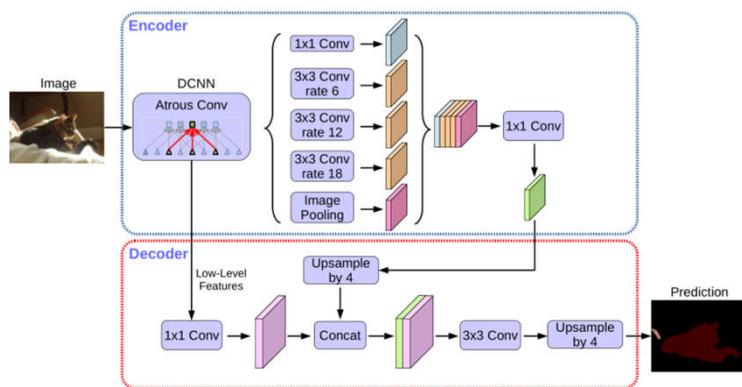


语义分割



1. 像素到语义节点映射;
2. 语义节点间推理;
3. 节点向像素反映射。

step 1 & 3 构成的对像素特征的低秩重建发挥了关键作用。



语义分割

映射的关键，在于寻找一组“最合适”的feature map，
具有如下性质：

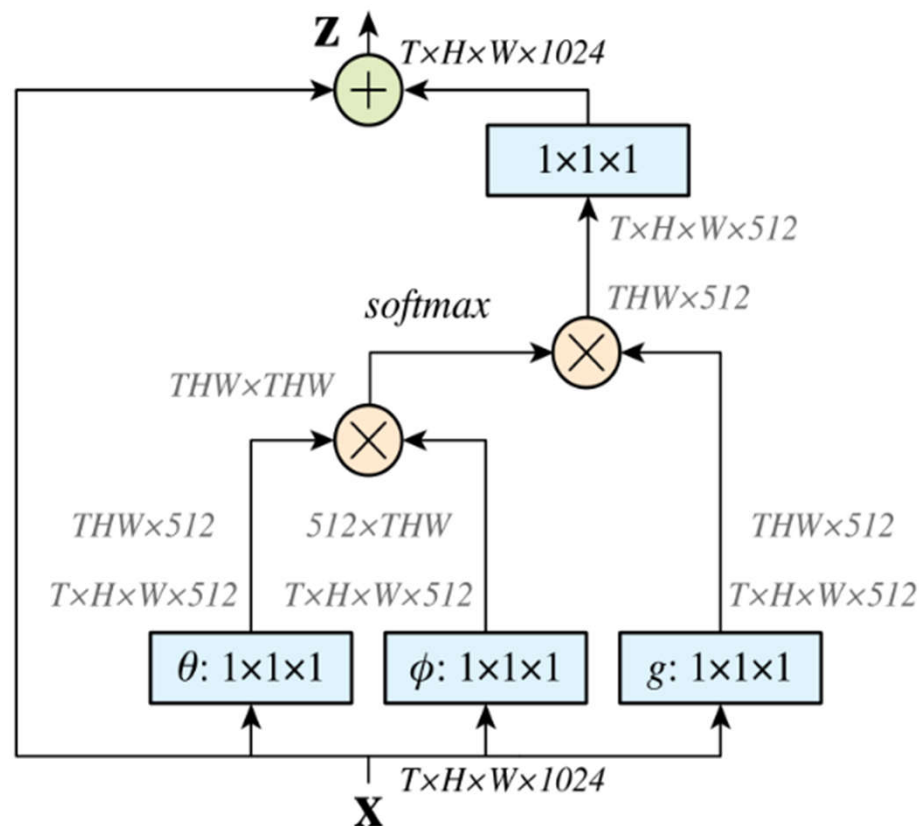
1. 具有代表性
2. 数量少
3. 互不相似

语义分割中的attention机制

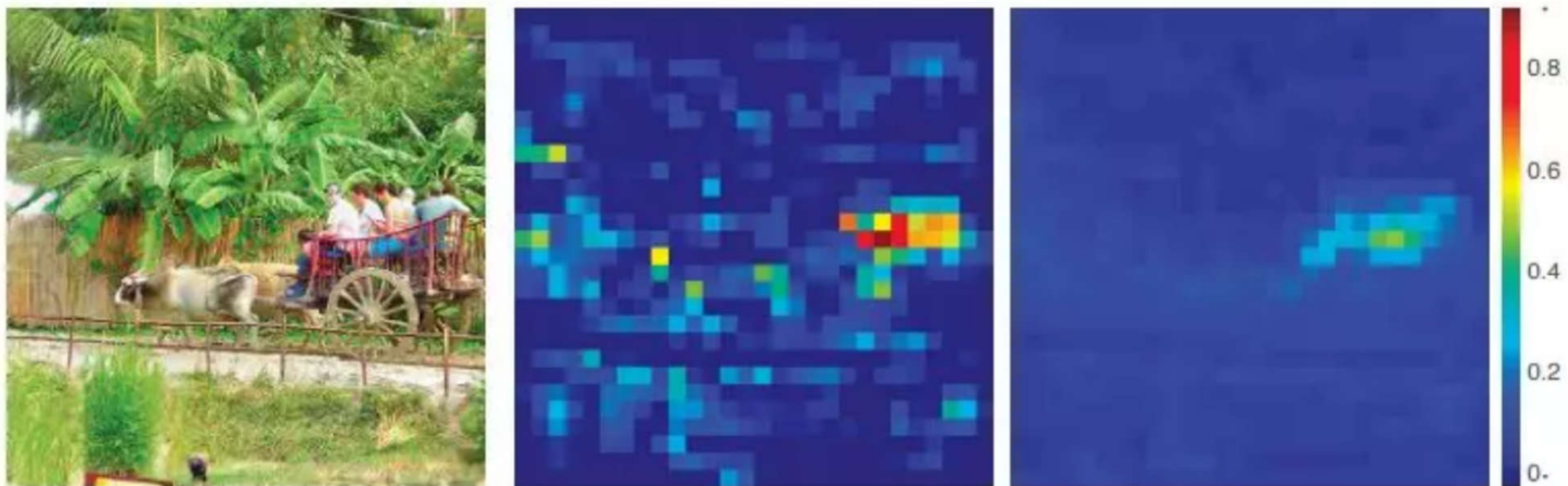
$$\mathbf{y}_i = \frac{1}{C(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j) \quad (1)$$

1. f 表示 计算 x_i 和 x_j 之间的相关度
2. g 对 x_j 进行变换
3. y_i 可以看作是 $g(x_i)$ 的加权平均, 输出的 y_i 是 x_i 的重构

$$\mathbf{y} = \text{softmax}(\mathbf{x}^T W_\theta^T W_\phi \mathbf{x}) (\mathbf{x}^T W_\sigma^T) \quad (2)$$



语义分割中的attention机制



a. 原图 b. Feature map c. Feature map after Nonlocal

即高维 feature map 里存在大量的冗余信息，
attention机制可以消除大量噪音。

[用Attention玩转CV，一文总览自注意力语义分割进展]

https://mp.weixin.qq.com/s?__biz=MzA3Mzl4MjgzMw==&mid=2650768770&idx=3&sn=aec7b055da21a94999adac0ce45dfe01&chksm=871ac41fcb06dc8ead45b8b99a7b9bc59aedc64373a45f781db00f528e029a7c861e95f8094c0&token=310258758&lang=zh_CN#rd

语义分割中的attention机制

Nonlocal 对于每个 x_i 的计算，都要在全图上进行，因此复杂度为 $O(N^2C)$ 。

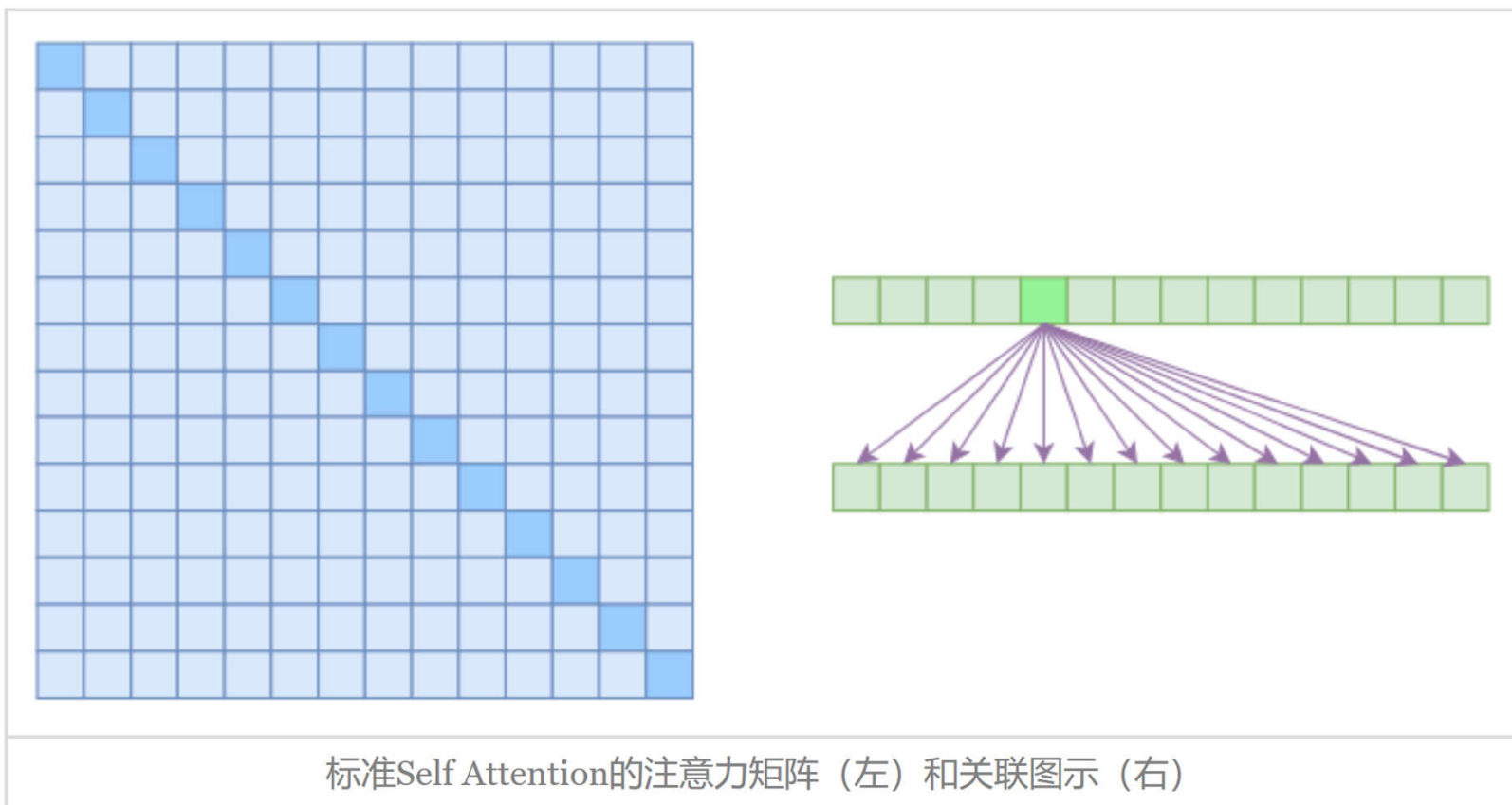
对于 attention 机制的扩展：

1. PSANet (双路attention)
2. DANet (双路attention)
3. OCNet

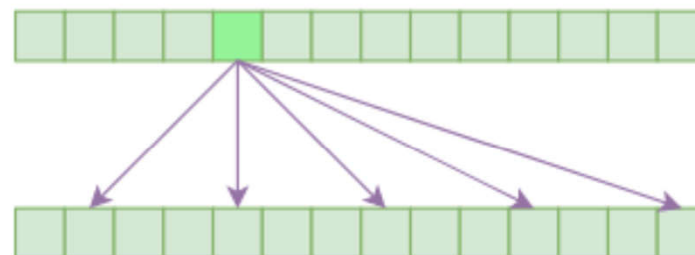
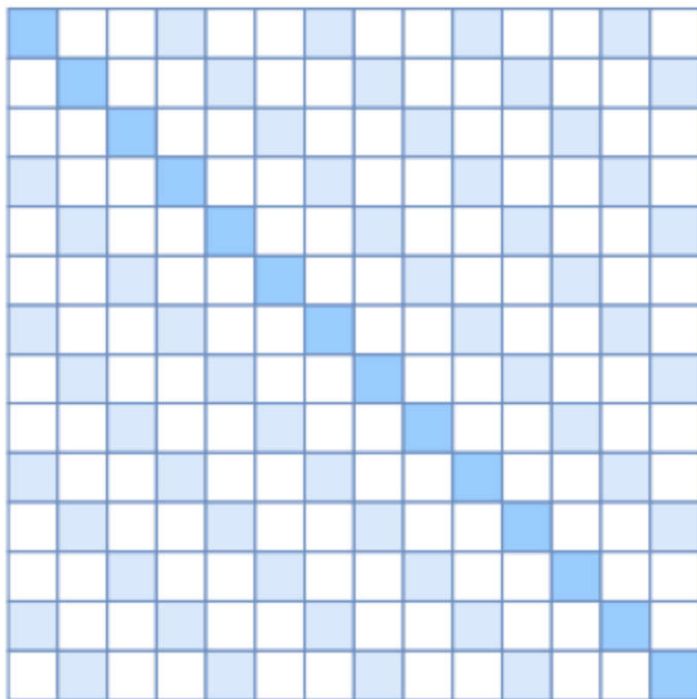
对于 Nonlocal 计算的优化：

- | | |
|------------------------------|------------------|
| 1. CCNet (运算过程分解) | 5. A2Net (乘法结合律) |
| 2. ISANet (运算过程分解) | ... |
| 3. DGMN (MC采样) | |
| 4. Local Relation Net (局部计算) | |

语义分割中的attention机制

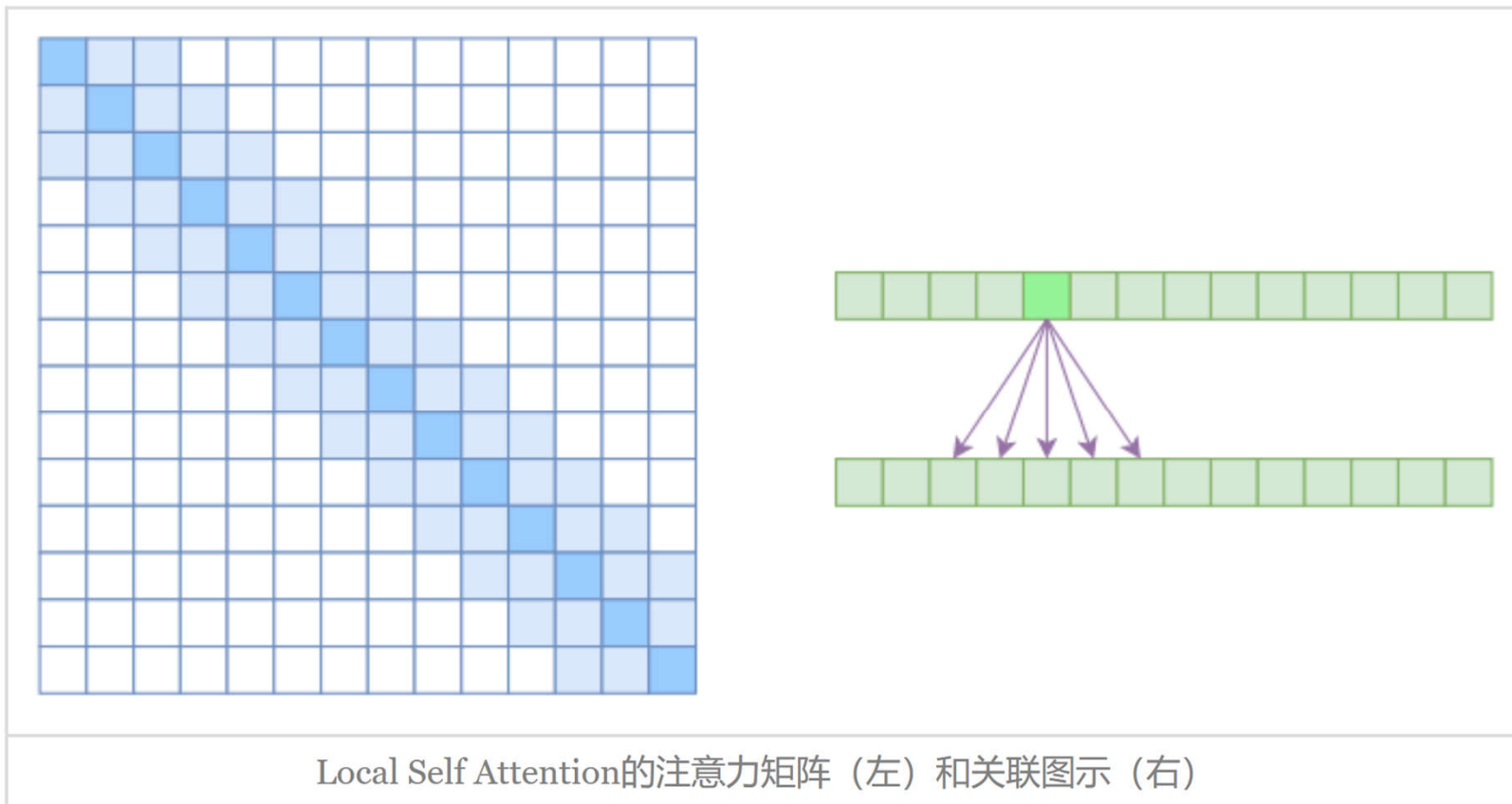


语义分割中的attention机制

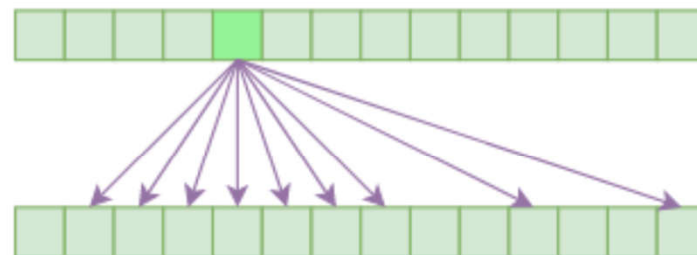
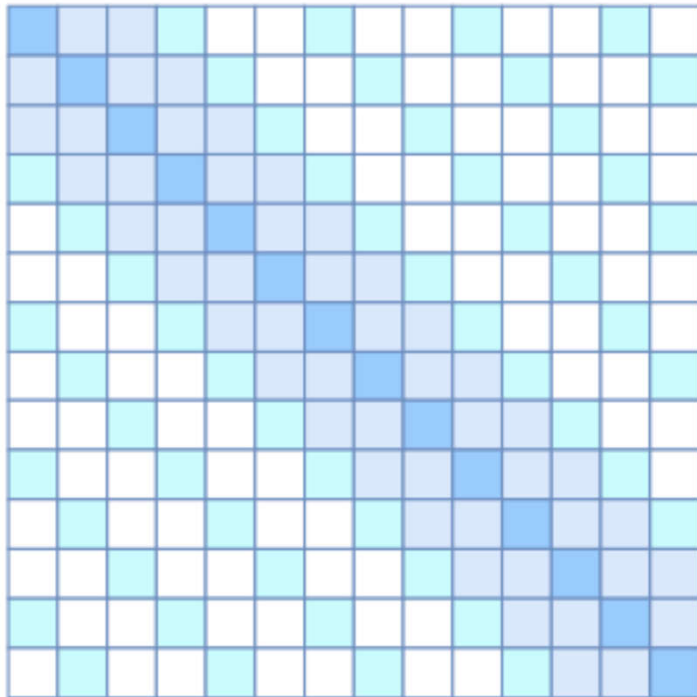


Atrous Self Attention的注意力矩阵（左）和关联图示（右）

语义分割中的attention机制



语义分割中的attention机制



Sparse Self Attention的注意力矩阵（左）和关联图示（右）

EM算法

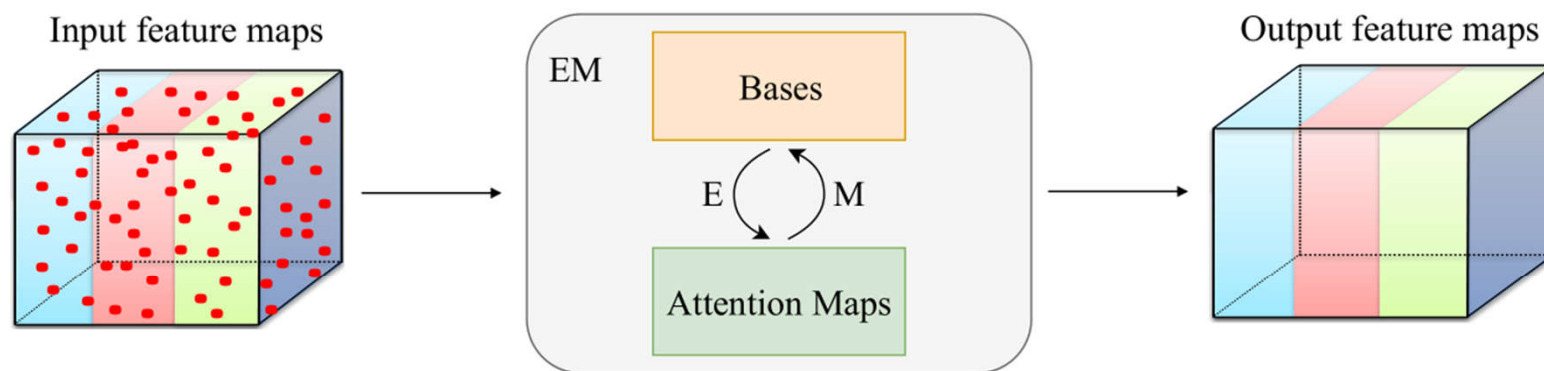


Figure 1: Pipeline of the proposed expectation-maximization attention method.

- 摒弃了在全图上计算注意力图的流程
- 转而通过EM算法迭代出一组紧凑的基，在这组基上运行注意力机制，降低了复杂度。

EM算法

期望最大化算法

期望最大化 (EM) 算法旨在为隐变量模型寻找最大似然解。对于观测数据 $X=\{x_1, x_2, \dots, x_N\}$, 每一个数据点 x_i 都对应隐变量 z_i 。我们把 $\{X, Z\}$ 称为完整数据, 其似然函数为 $\ln p(X, Z|\theta)$, θ 是模型的参数。

E 步根据当前参数 θ^{old} 计算隐变量 Z 的后验分布, 并以之寻找完整数据的似然 $Q(\theta, \theta^{old})$:

$$Q(\theta, \theta^{old}) = \sum_{\mathbf{z}} p(\mathbf{z}|\mathbf{X}, \theta^{old}) \ln p(\mathbf{X}, \mathbf{z}|\theta) \quad (1)$$

M 步通过最大化似然函数来更新参数得到 θ^{new} :

$$\theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{old}) \quad (2)$$

EM 算法被证明会收敛到局部最大值处, 且迭代过程完整数据似然值单调递增。

EMANet

Nonlocal:

$$\mathbf{y}_i = \frac{1}{C(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j) \quad (3)$$


其中 $f(\dots)$ 表示广义的核函数, $C(x)$ 是归一化系数。它将第 i 个像素的特征 x_i 更新为其他所有像素特征经过 g 变换之后的加权平均 y_i , 权重通过归一化后的核函数计算, 表征两个像素之间的相关度。这里 $1 < j < N$, 所以视为像素特征被一组完备的基进行了重构。这组基数目巨大, 且存在大量信息冗余。

EMANet

第一步 A_E 求期望, 估计隐变量 z

$$\mathbf{y}_i = \frac{1}{C(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j) \quad (3)$$

第 k 个基对第 n 个像素的权重
可以计算为


$$z_{nk} = \frac{\mathcal{K}(\mathbf{x}_n, \boldsymbol{\mu}_k)}{\sum_{j=1}^K \mathcal{K}(\mathbf{x}_n, \boldsymbol{\mu}_j)} \quad (4)$$

内核 \mathcal{K} 的一种选择

$$\mathbf{Z} = \text{softmax}(\lambda \mathbf{X} (\boldsymbol{\mu}^\top)) \quad (5)$$

其中, λ 作为超参数来控制 z 的分布。

EMANet

第二步 A_M 期望最大化更新基 μ

$$\mathbf{y}_i = \frac{1}{C(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) \overline{g(\mathbf{x}_j)} \quad (3)$$

μ_k 是 x_n 的加权平均

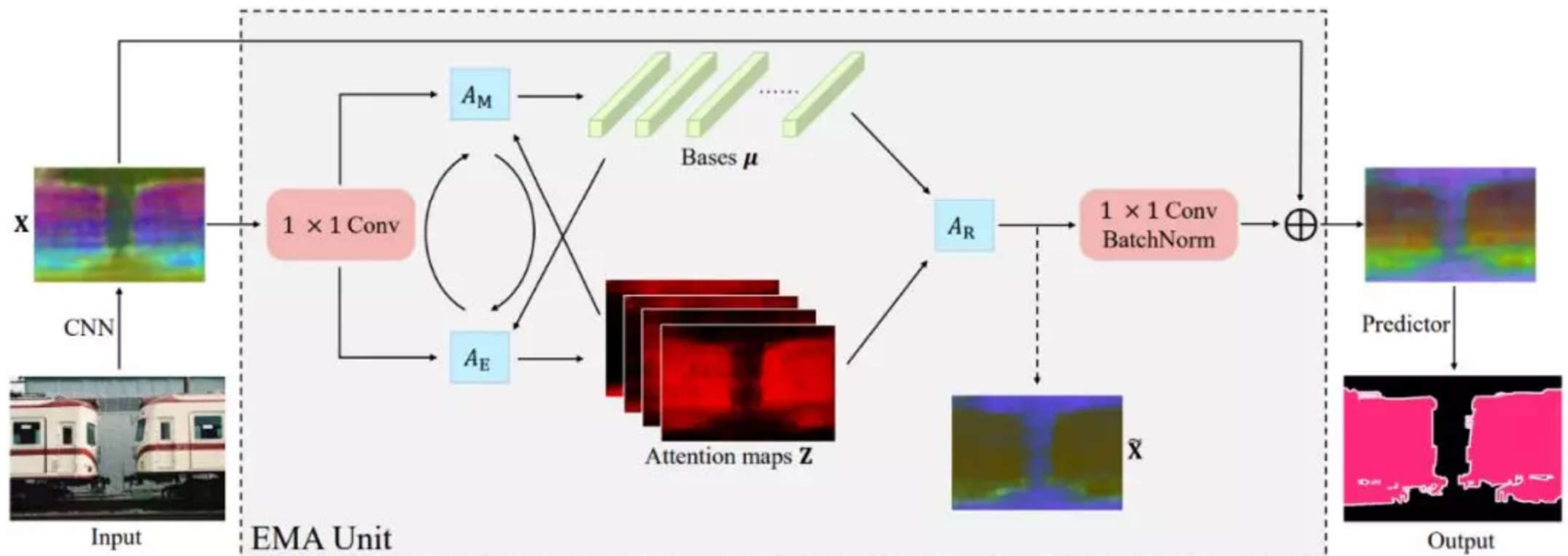
$$\mu_k = \frac{z_{nk} \mathbf{x}_n}{\sum_{m=1}^N z_{mk}} \quad (6)$$


第三步 A_R 期望最大化更新基 μ

A_E 和 A_M 交替执行 T 步。此后，近似收敛的 μ 和 Z 便可以被用来对 X 进行重估计得 \tilde{X}

$$\tilde{X} = Z\mu \quad (7)$$

EMANet



EMA Unit

EMANet

		Evaluation Iterations (mIoU %)							
		1	2	3	4	5	6	7	8
Training Iterations	1	77.34	77.52	77.60	77.59	77.59	77.59	77.59	77.59
	2		77.75	78.04	78.15	78.15	78.12	78.12	78.17
	3			78.52	78.80	78.86	78.88	78.89	78.88
	4				78.14	78.25	78.27	78.28	78.27
	5					77.70	77.76	77.82	77.86
	6						77.85	77.91	77.92
	7							77.11	77.14
	8								77.24

EMANet

Method	SS	MS+Flip	FLOPs	Memory	Params
ResNet-101	-	-	190.6G	2.603G	42.6M
DeeplabV3 [4]	78.51	79.77	+63.4G	+66.0M	+15.5M
DeeplabV3+ [5]	79.35	80.57	+84.1G	+99.3M	+16.3M
PSANet [38]	78.51	79.77	+56.3G	+59.4M	+18.5M
EMANet (256)	<u>79.73</u>	<u>80.94</u>	+21.1G	+12.3M	+4.87M
EMANet (512)	80.05	81.32	<u>+43.1G</u>	<u>+22.1M</u>	<u>+10.0M</u>

EMANet

Table 3: Comparisons with state-of-the-art on the PASCAL Context test set. ‘+’ means pretrained on COCO Stuff.

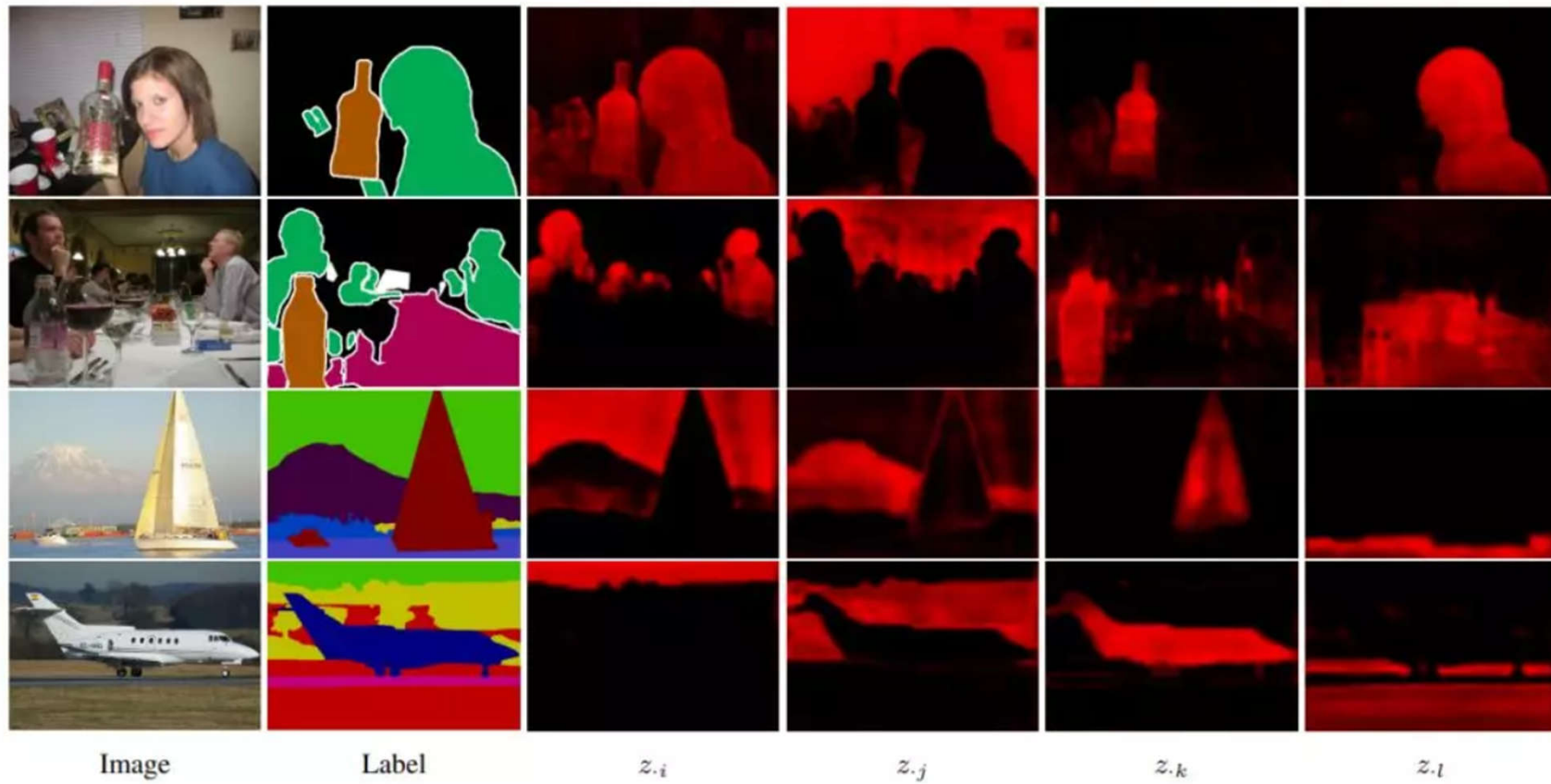
Method	Backbone	mIoU (%)
PSPNet [37]	ResNet-101	47.8
DANet [11]	ResNet-50	50.1
MSCI [20]	ResNet-152	50.3
EMANet	ResNet-50	<u>50.5</u>
SGR [18]	ResNet-101	50.8
CCL [8]	ResNet-101	51.6
EncNet [35]	ResNet-101	51.7
SGR+ [18]	ResNet-101	52.5
DANet [11]	ResNet-101	52.6
EMANet	ResNet-101	53.1

EMANet

Table 4: Comparisons on the COCO Stuff test set.

Method	Backbone	mIoU (%)
RefineNet [21]	ResNet-101	33.6
CCL [8]	ResNet-101	35.7
DANet [11]	ResNet-50	37.2
DSSPN [19]	ResNet-101	37.3
EMANet	ResNet-50	<u>37.6</u>
SGR [18]	ResNet-101	39.1
DANet [11]	ResNet-101	39.7
EMANet	ResNet-101	39.9

EMANet



EMANet

Expectation-Maximization Attention Networks for Semantic Segmentation

<https://arxiv.org/abs/1907.13426>

为节约而生：从标准Attention到稀疏Attention

<https://spaces.ac.cn/archives/6853>

用Attention玩转CV，一文总览自注意力语义分割进展

https://mp.weixin.qq.com/s?__biz=MzA3Mzl4MjgzMw==&mid=2650768770&idx=3&sn=aec7b055da21a94999adac0ce45dfe01&chksm=871a41fcb06dc8ead45b8b99a7b9bc59aedc64373a45f781db00f528e029a7c861e95f8094c0&token=310258758&lang=zh_CN#rd

ICCV 2019 | 解读北大提出的期望最大化注意力网络EMANet

https://mp.weixin.qq.com/s?__biz=MzA3Mzl4MjgzMw==&mid=2650768486&idx=4&sn=8dd39c05a69021007f8f2d9ccae5ffb6&chksm=871a4018b06dc90e5ef9320dc9a032a92e7a609a34765ea37f6eacd7382b2f93b23d3f51f717&scene=21#wechat_redirect

EM算法的九层境界：Hinton和Jordan理解的EM算法

<https://mp.weixin.qq.com/s/NbM4sY93kaG5qshzgZzZIQ?>