人工智能原理与技术

深度学习前沿技术

深度强化学习

强化学习：相关概念

环境（environment）

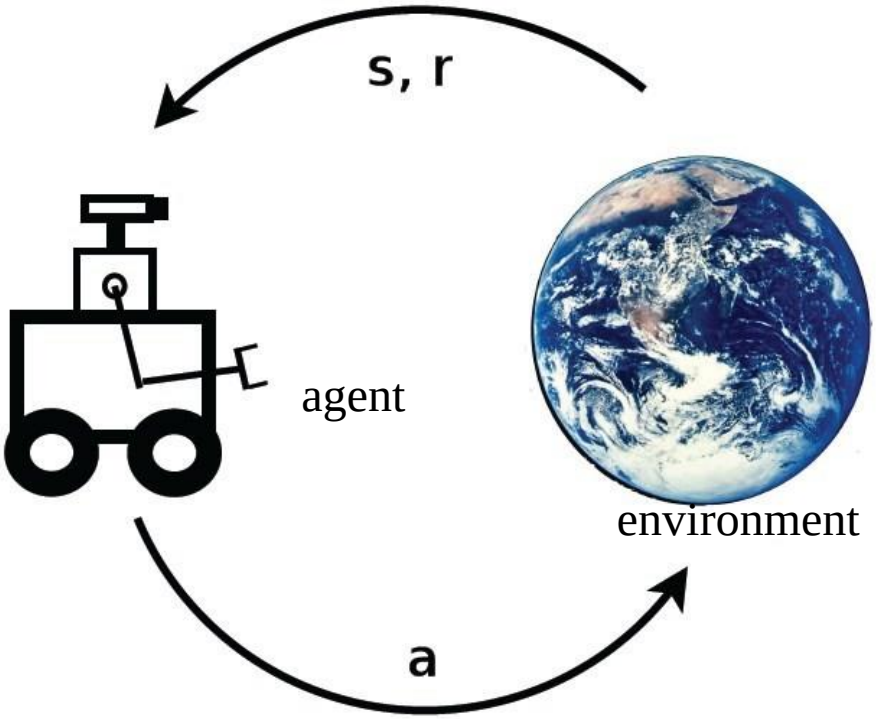
操作者或智能体（ agent ）

状态（state）：s

动作（action）：a

奖励（reward）：r

策略（policy）：π

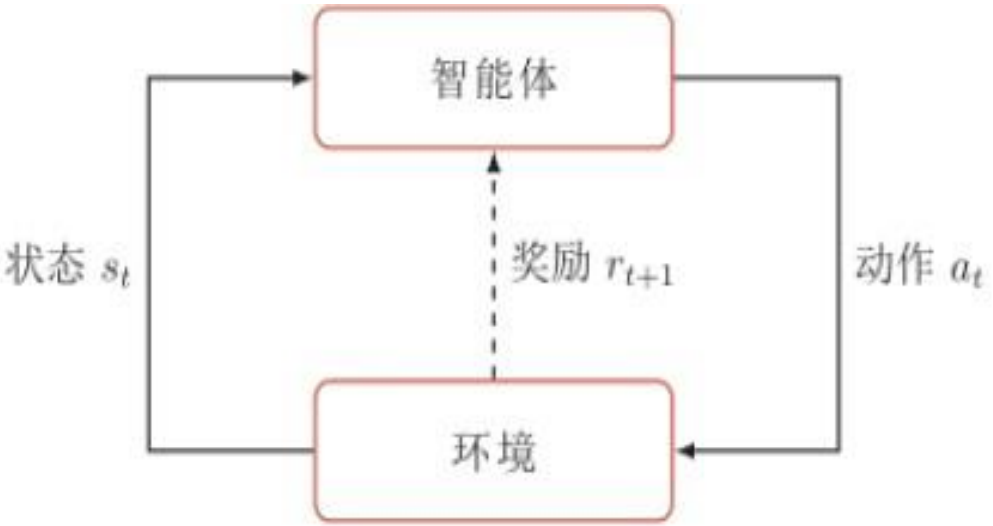


$$\pi : S \rightarrow A$$

$$\max_{\pi} R = \sum_{t=0}^{\infty} \gamma^t r_t$$

强化学习：定义

- 强化学习的目标可以描述为一个智能体从与环境的交互中不断学习以完成特定目标（比如取得最大奖励值）
- 强化学习就是智能体与不断与环境进行交互，并根据经验调整其策略来最大化所有奖励的累积值



强化学习 vs. 机器学习

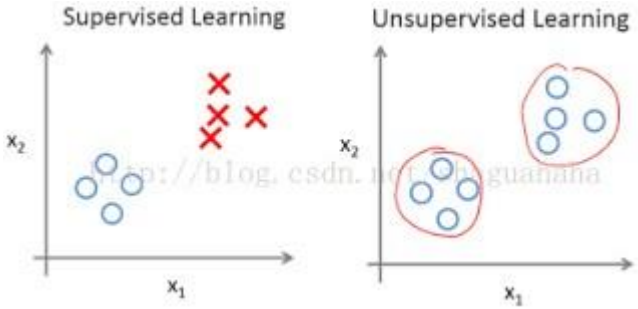
学习类型	学习靠谁？	有没有老师指导	特点
无监督学习	自己	无	全然无师自通，主要是聚类
监督学习	老师	有	学习效率高，但对数据要求高
强化学习	自己	有	算法较复杂，但对数据要求低

监督学习 用标记过的数据来训练模型。

无监督学习 使用的数据是没有标记过的，寻找数据的模型和规律

强化学习 也是使用未标记的数据，但可以通过某种方法知道离正确答案越来越近还是越来越远（即**奖惩函数**）。

⇨ **根据奖惩学习决策模型**



hotter or colder？

强化学习：实例

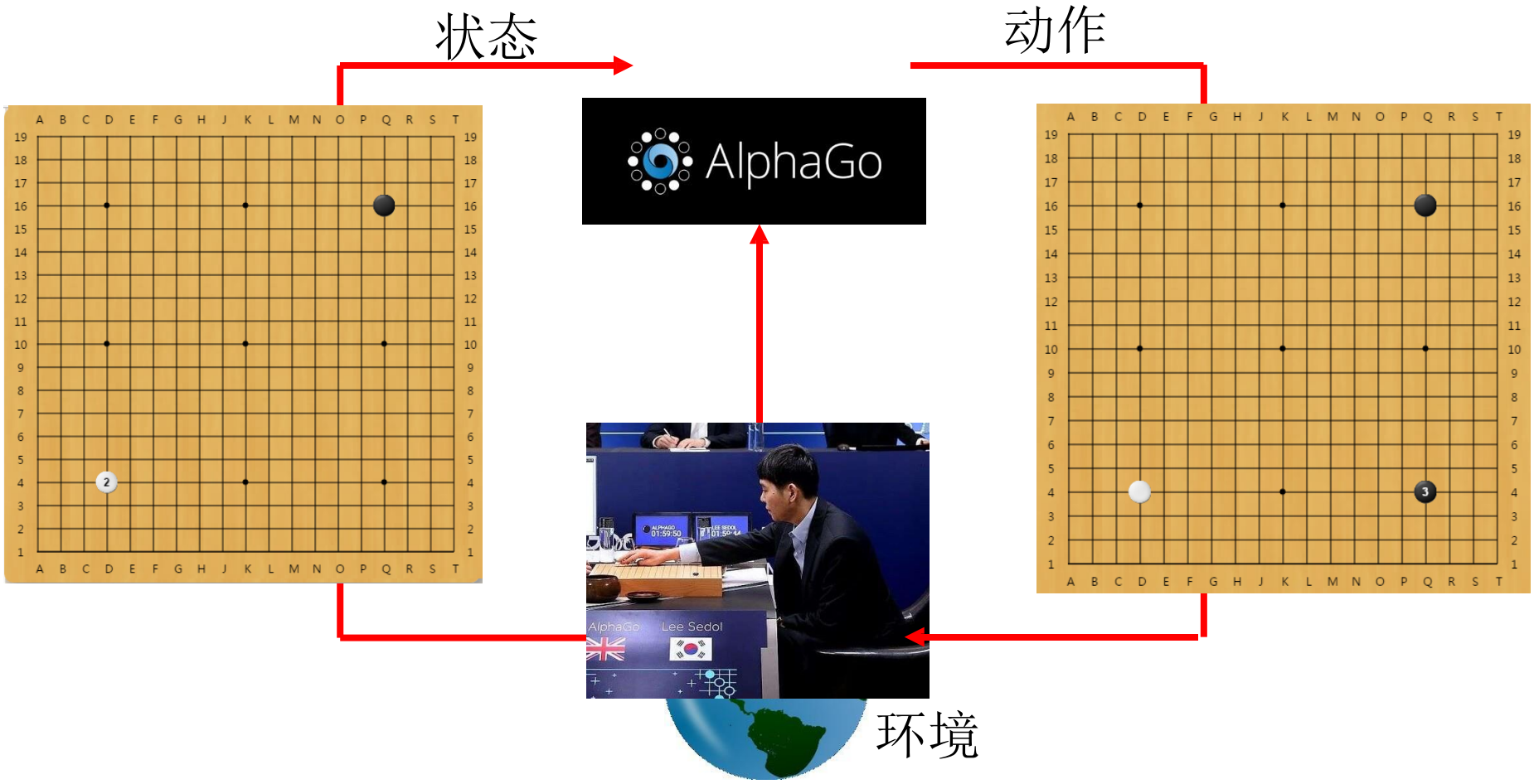
Atari Breakout 游戏

✓三种动作：向左，向右以及开火（把球射出）
✓状态：所处的位置，屏幕上方状况等
✓奖励：得分增减

传统监督学习 vs 强化学习

（有稀疏并延时的标签—奖励）

强化学习：实例



@李宏毅

强化学习：实例


状态
动作
AlphaGo
奖励
reward = 0 in most cases
If win, reward = 1
If loss, reward = -1
AlphaGo
Lee Sedol
环境

@李宏毅

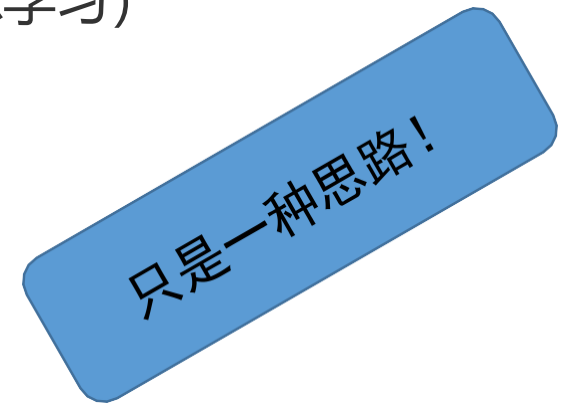
深度强化学习：AI=DL+RL

- 强化学习（RL）是一个进行决策制定（decision-making）的通用框架：
  - RL是要学习一个智能体（agent）用来执行一些动作（action）
  - 每一个action都会影响智能体的未来状态
  - 用一个标量奖励值（reward）来衡量成功与否
  - 目标：选择一系列actions来最大化未来奖励

- 深度学习（RL）是一个进行表示学习（representation-making）的通用框架：
  - 给定一个目标函数
  - 学习特征表示以达到目标
  - 直接从输入中学习
  - 需要很少的领域知识

深度强化学习：AI=DL+RL

- 我们寻求一个智能体能够解决通用的人类任务：
  - RL定义目标
  - DL提供方法机制（特征表示学习）
  - 通用人工智能=RL+DL

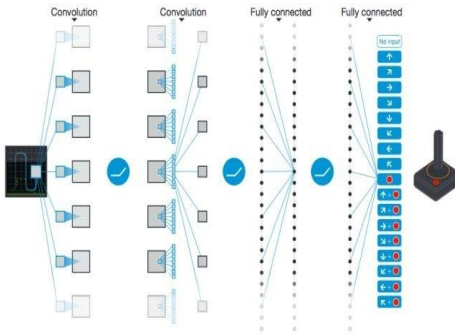


深度强化学习：里程碑

- NIPS 2013, DeepMind, Playing Atari with Deep Reinforcement Learning, https://arxiv.org/abs/1312.5602

- Nature cover paper 2015, DeepMind, Human-level control through deep reinforcement learning, www.nature.com/articles/nature14236

- Nature cover paper 2016, DeepMind, Mastering the game of Go with deep neural networks and tree search, www.nature.com/articles/nature16961



DeepMind

"Human-level control through deep reinforcement learning"

letter

Deep Q-Learning

深度强化学习：应用

机器人控制

- **状态State**：摄像头图像，机械臂关节的角度
- **动作Actions**：电机输出扭矩，机械臂关节扭矩
- **反馈Rewards**：由具体任务定义，比如保持平衡，移动到特定位置，或者某种服务等

深度强化学习：应用

商业决策

库存管理。观察当前的库存情况，选择补充每一种货物的数量，反馈就是最后的收益。

投资分配。观察当前的资金储备和市场，选择资金分配策略，反馈就是收益。

物流运输。滴滴的专车调配；观察当前不同地区的顾客需求，选择专车调配策略，反馈是满足顾客需求的程度。

深度强化学习：应用

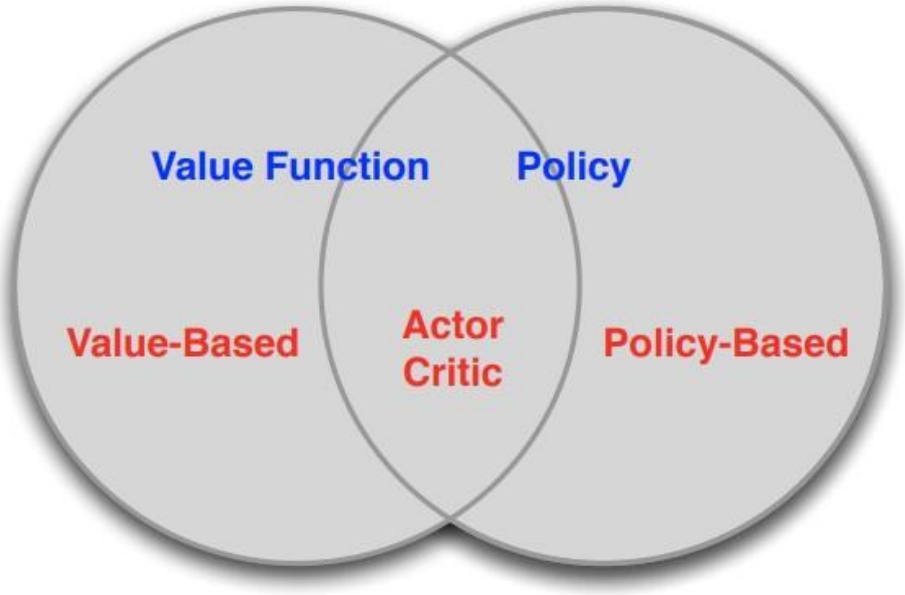
游戏

AI research in the real-time strategy game StarCraft II & DOTA 2

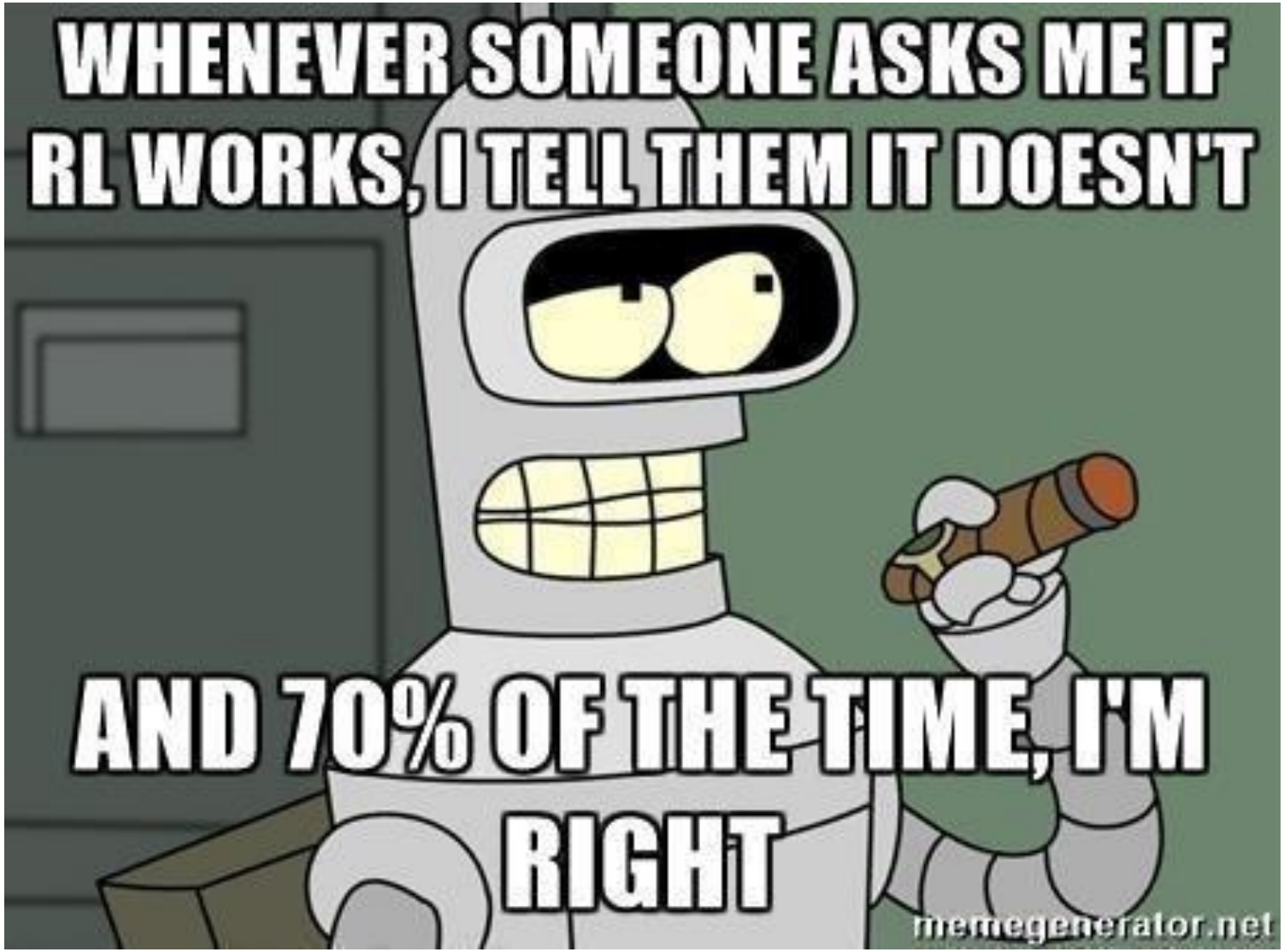
Image courtesy: (L) SC2LE - an RL environment based on StarCraft II from DeepMind & Blizzard and (R) A bot which beats the world's top professionals at 1v1 matches of Dota 2 under standard tournament rules

深度强化学习：三类方法

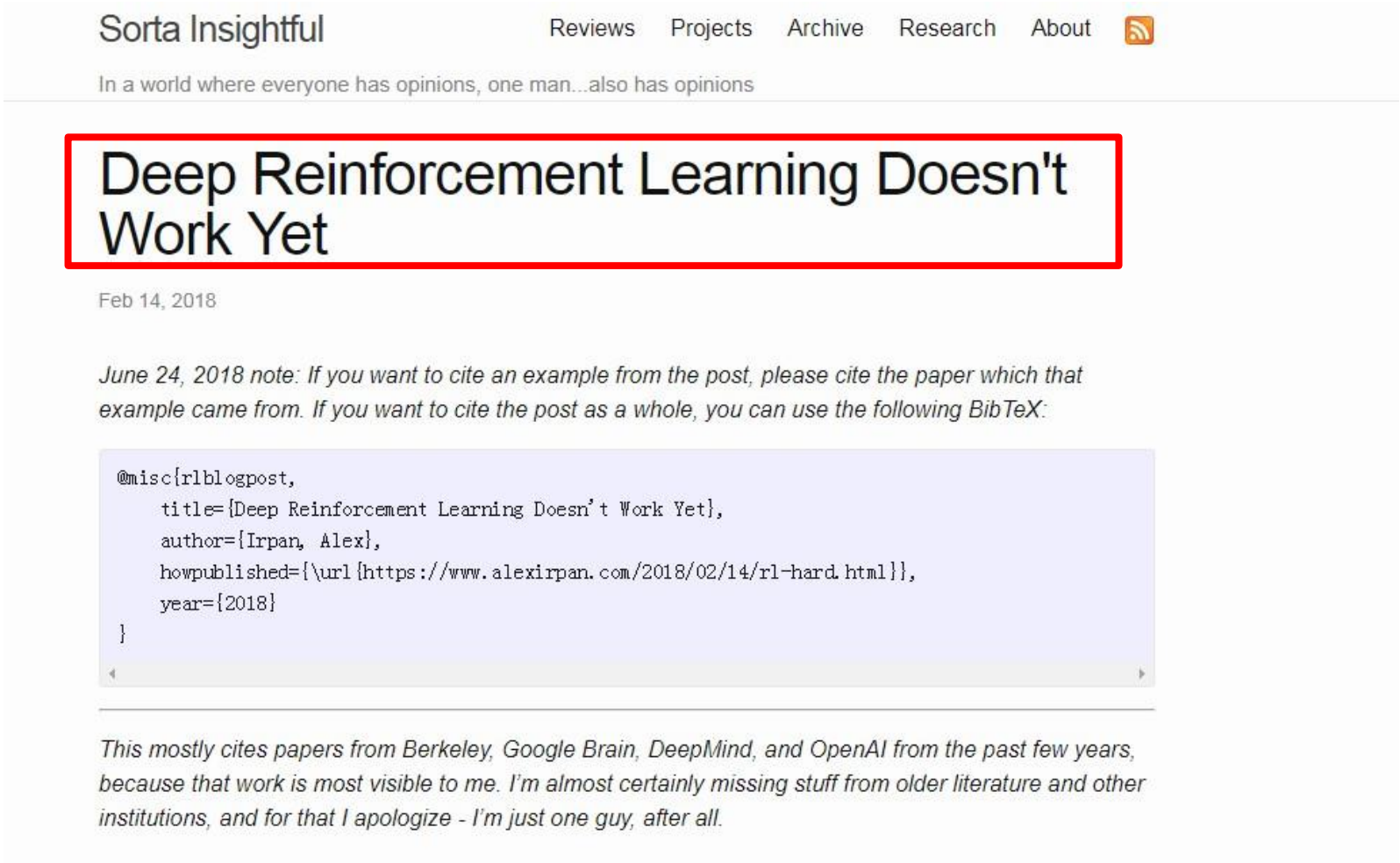
- 基于策略的方法（Policy-based）：
  - 没有值函数
  - 学习策略
- 基于值的方法（Value-based）：
  - 学习评估值（类似于reward）的函数
  - 隐式的策略方法
    - Actor-Critic：
      - 学习值函数
  - 学习策略



深度强化学习劝退?


WHENEVER SOMEONE ASKS ME IF
RL WORKS, I TELL THEM IT DOESN'T
AND 70% OF THE TIME, I'M
RIGHT
memegenerator.net

深度强化学习劝退?



Sorta Insightful

Reviews Projects Archive Research About

In a world where everyone has opinions, one man...also has opinions

Deep Reinforcement Learning Doesn't Work Yet

Feb 14, 2018

June 24, 2018 note: If you want to cite an example from the post, please cite the paper which that example came from. If you want to cite the post as a whole, you can use the following BibTeX:

```
@misc{rlblogpost,
    title={Deep Reinforcement Learning Doesn't Work Yet},
    author={Irpan, Alex},
    howpublished={\url{https://www.alexirpan.com/2018/02/14/rl-hard.html}},
    year={2018}
}
```

This mostly cites papers from Berkeley, Google Brain, DeepMind, and OpenAI from the past few years, because that work is most visible to me. I'm almost certainly missing stuff from older literature and other institutions, and for that I apologize - I'm just one guy, after all.

https://www.alexirpan.com/2018/02/14/rl-hard.html

深度强化学习劝退：理由

- 它的样本利用率非常低。
- 最终表现很多时候不够好。
- DRL成功的关键离不开一个好的奖励函数（reward function），然而这种奖励函数往往很难设计。
- 局部最优/探索和剥削（exploration vs. exploitation）的不当应用。
- 对环境的过拟合。
- 不稳定性。

深度强化学习劝退？No! Keep Working!

- 局部最优或许已经足够好
- 硬件为王
- 人为添加一些监督信号
- 更多融合基于模型的学习从而提高样本使用率
- 仅仅把DRL用于fine-tuning
- 自动学习奖励函数
- 迁移学习和强化学习的进一步结合
- 好的先验
- 有的时候复杂的任务反而更容易学习

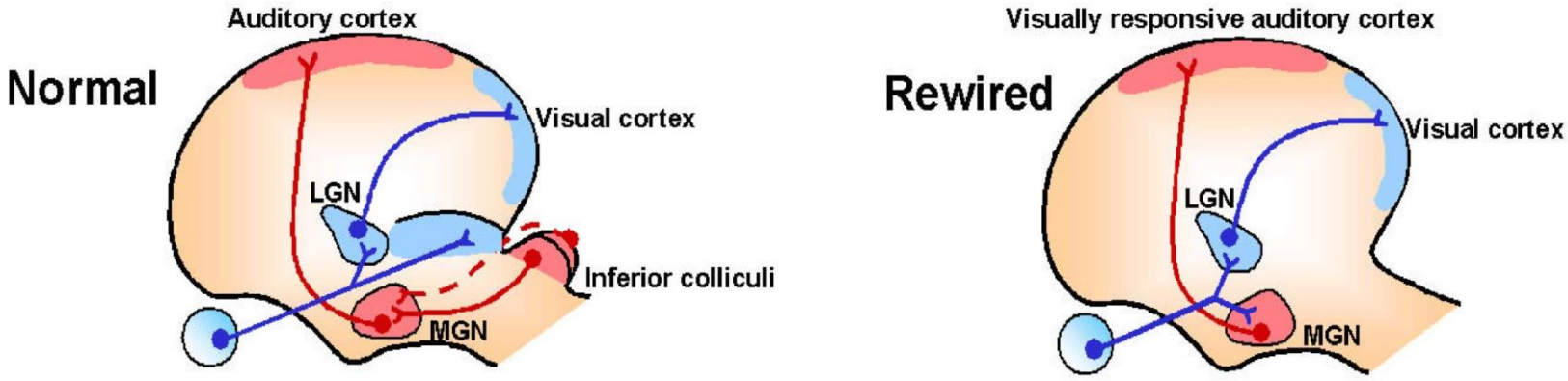
深度强化学习资源

- **http://www.zhuanzhi.ai/**
- **李宏毅深度强化学习2018：https://www.bilibili.com/video/av24724071/**
- https://github.com/jgvictores/awesome-deep-reinforcement-learning
- https://github.com/tigerneil/awesome-deep-rl
- https://github.com/aikorea/awesome-rl
- https://github.com/wwxFromTju/awesome-reinforcement-learning-zh
- https://deepmind.com/blog/deep-reinforcement-learning/
- Deep Reinforcement Learning: An Overview
  (https://arxiv.org/abs/1701.07274)
- https://zhuanlan.zhihu.com/p/25239682
- https://simoninithomas.github.io/Deep_reinforcement_learning_Course/
- https://icml.cc/2016/tutorials/deep_rl_tutorial.pdf
- http://rail.eecs.berkeley.edu/deeprlcourse/

自注意力机制

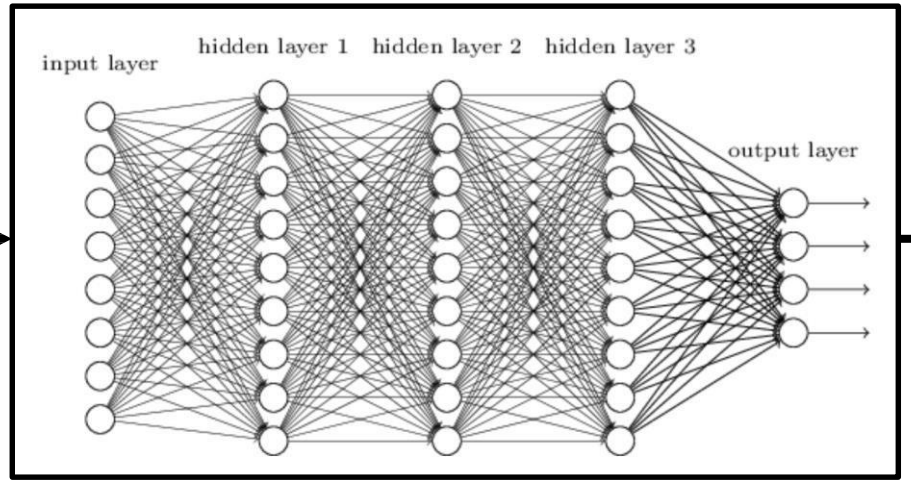
Human Brain

- Human cortex can universally perceive different senses



Intelligent Machines

- A **universal** learning pipeline



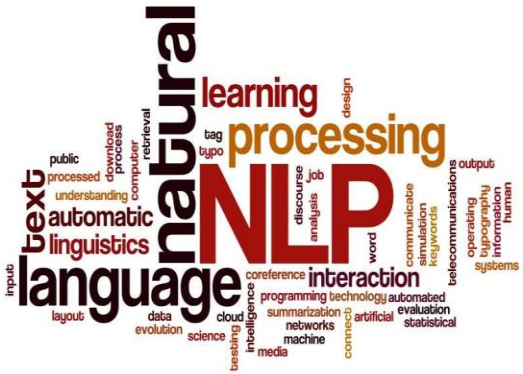
data → prediction

Loss(prediction, label) → back-propagation

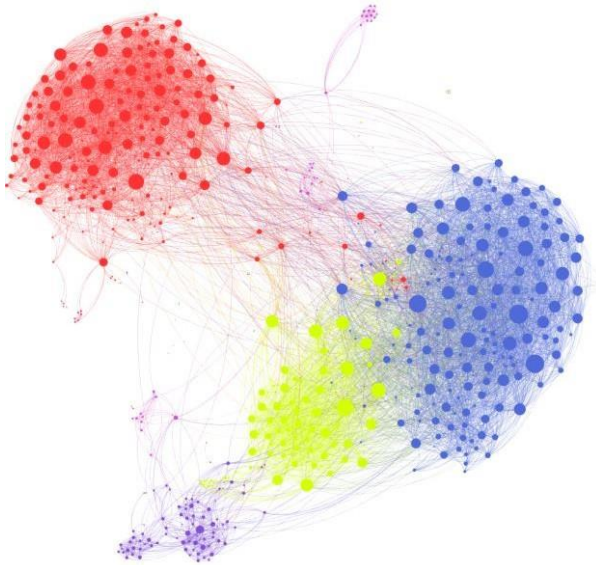
Intelligent Machines

- **Particular** basic model for different task/data

convolution



LSTM, GRU, convolution, self-attention, …

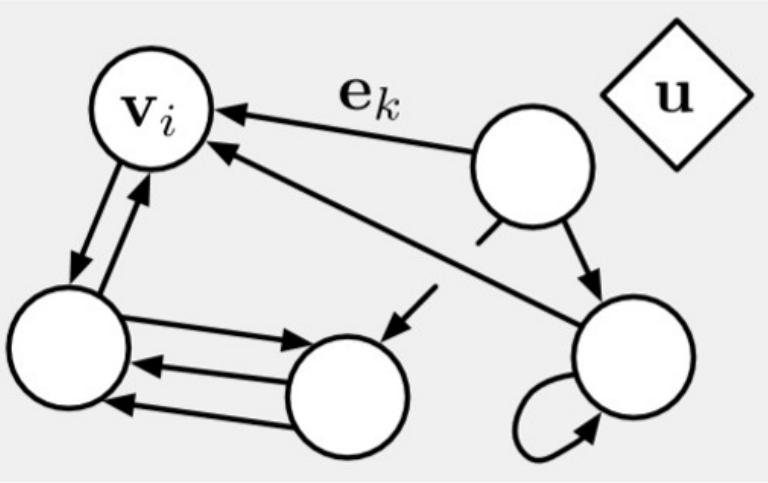


graph networks

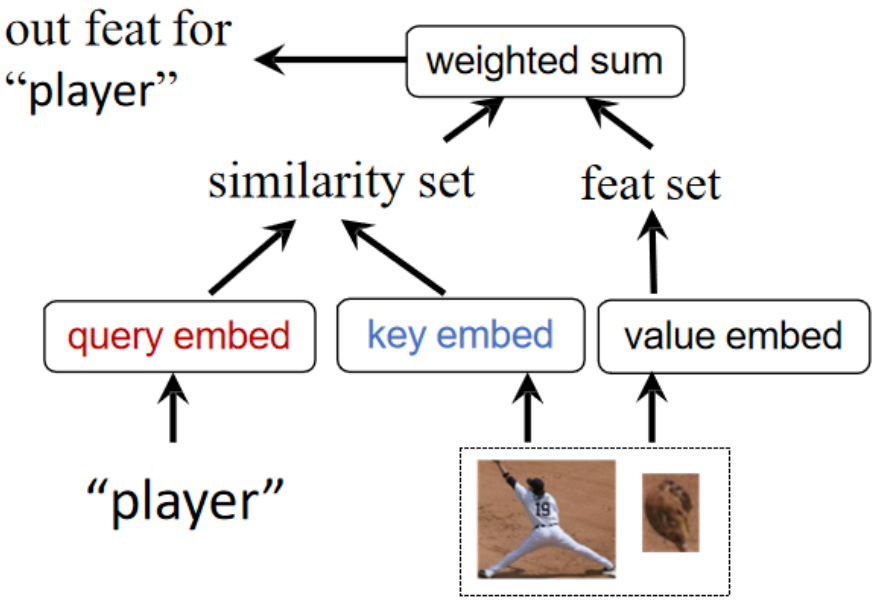
Universal Basic Models for Intelligent Machines?

Relation Networks: Towards Universal Basic Models

similar things: ***graph neural networks***, *self-attention*, …

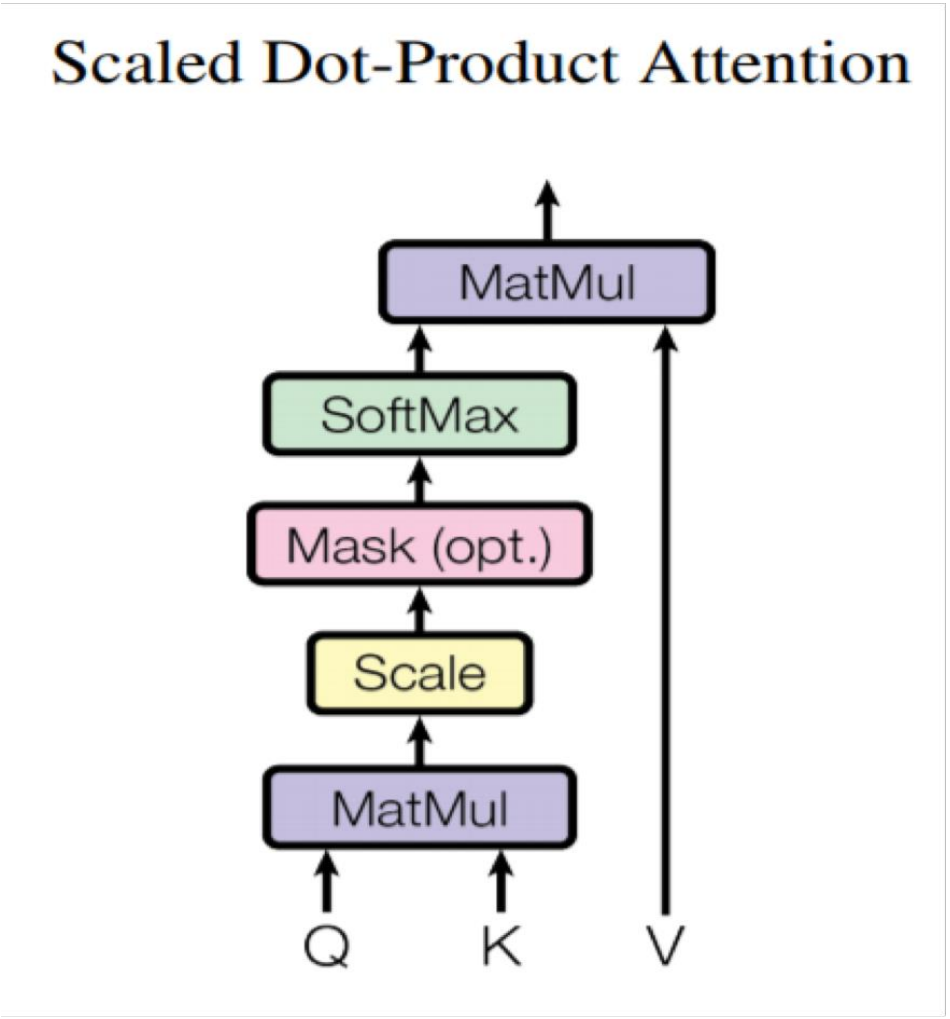


graph neural networks



(self)-attention

Attention is All You Need

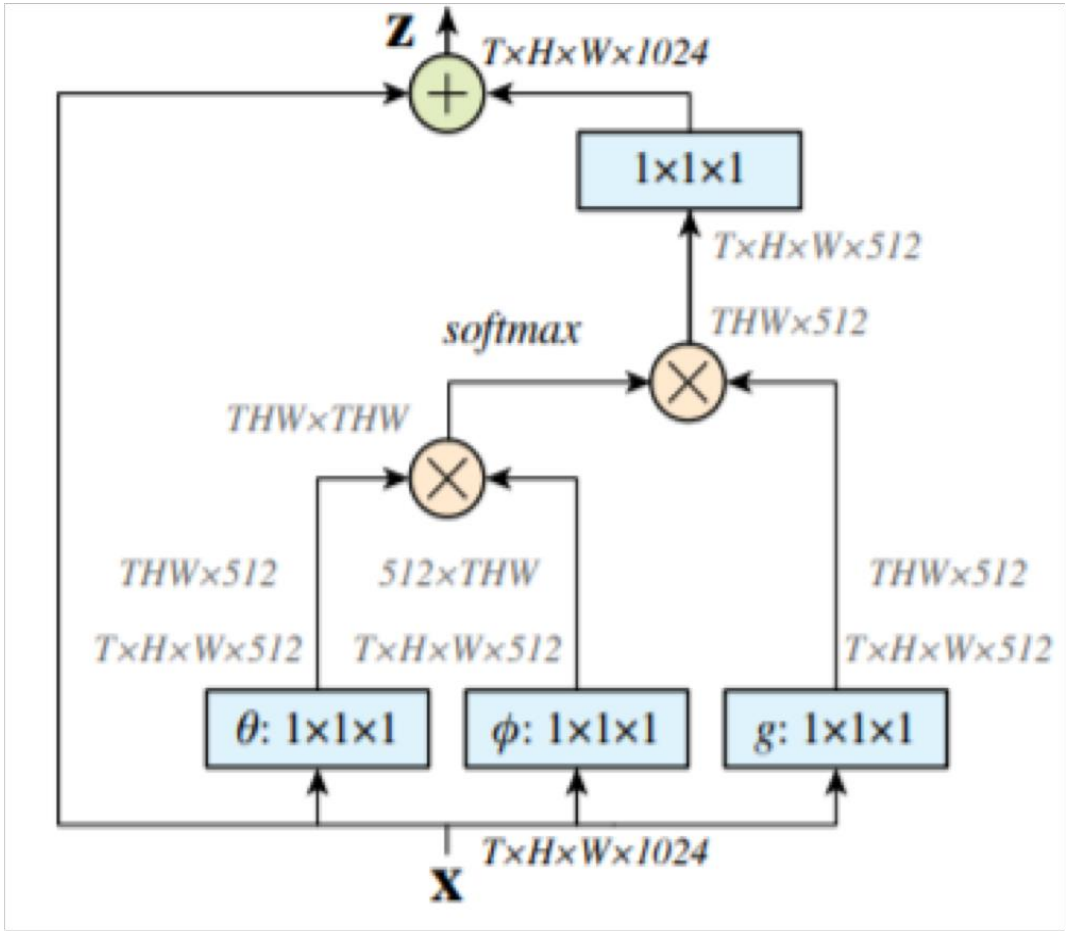


- (query, key, vale) 三元组可以捕捉长距离依赖
- key和query通过点乘获得权重，最后把权重和value做点乘得到输出

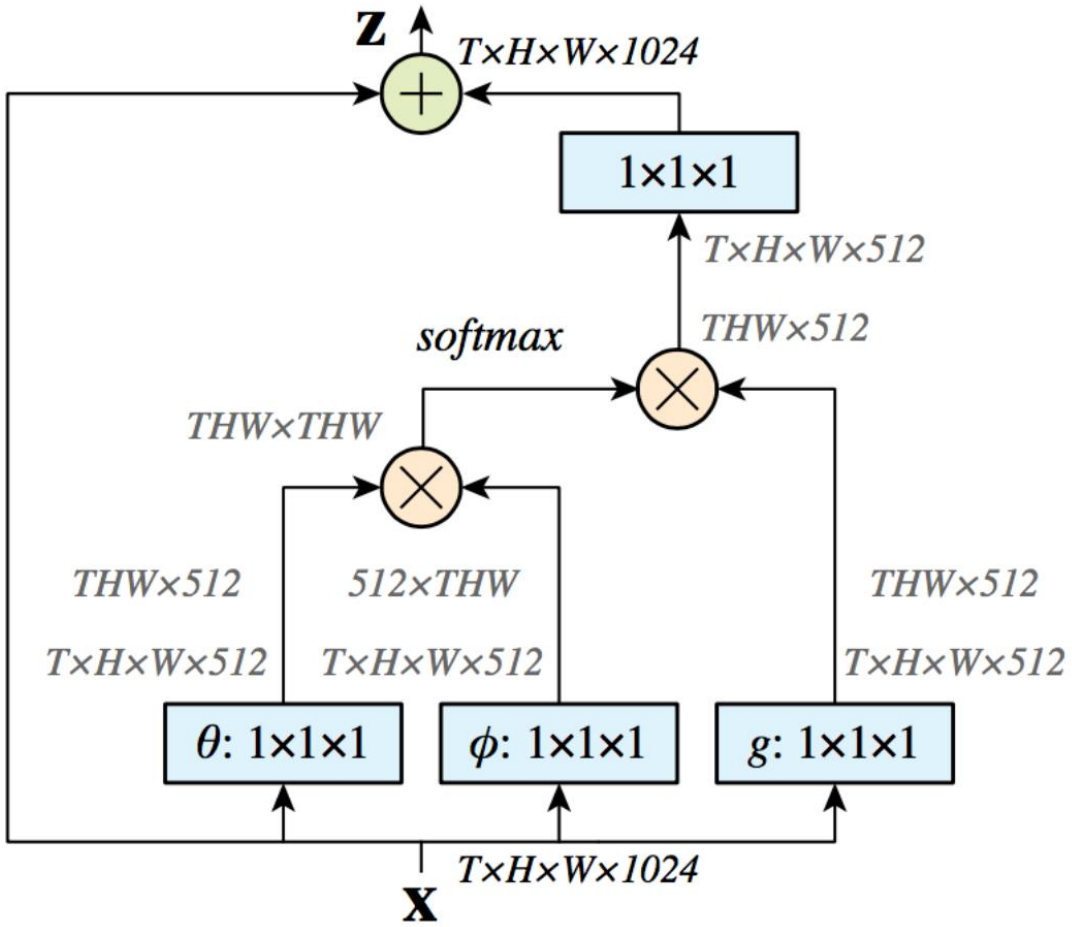
Attention Is All You Need. ***NIPS 2017***

Non-local Neural Networks

Non-local Module



[1] Xiaolong Wang, Ross Girshick, Abhinav Gupta, Kaiming He .
Non-local Neural Networks. CVPR 2018
[2] https://zhuanlan.zhihu.com/p/33345791



自注意力：强调重要特征，抑制噪声

Key, Query, Value 为输入的线性映射

计算 Key 和 Query 间的相似性，softmax进行归一化

结合 Value 得到权重矩阵

主要问题：每个点都要计算相似性，非常烧显存

Non-local Neural Network, *CVPR 2018*

Dual Attention Networks for Scene Segmentation

CVPR 2018

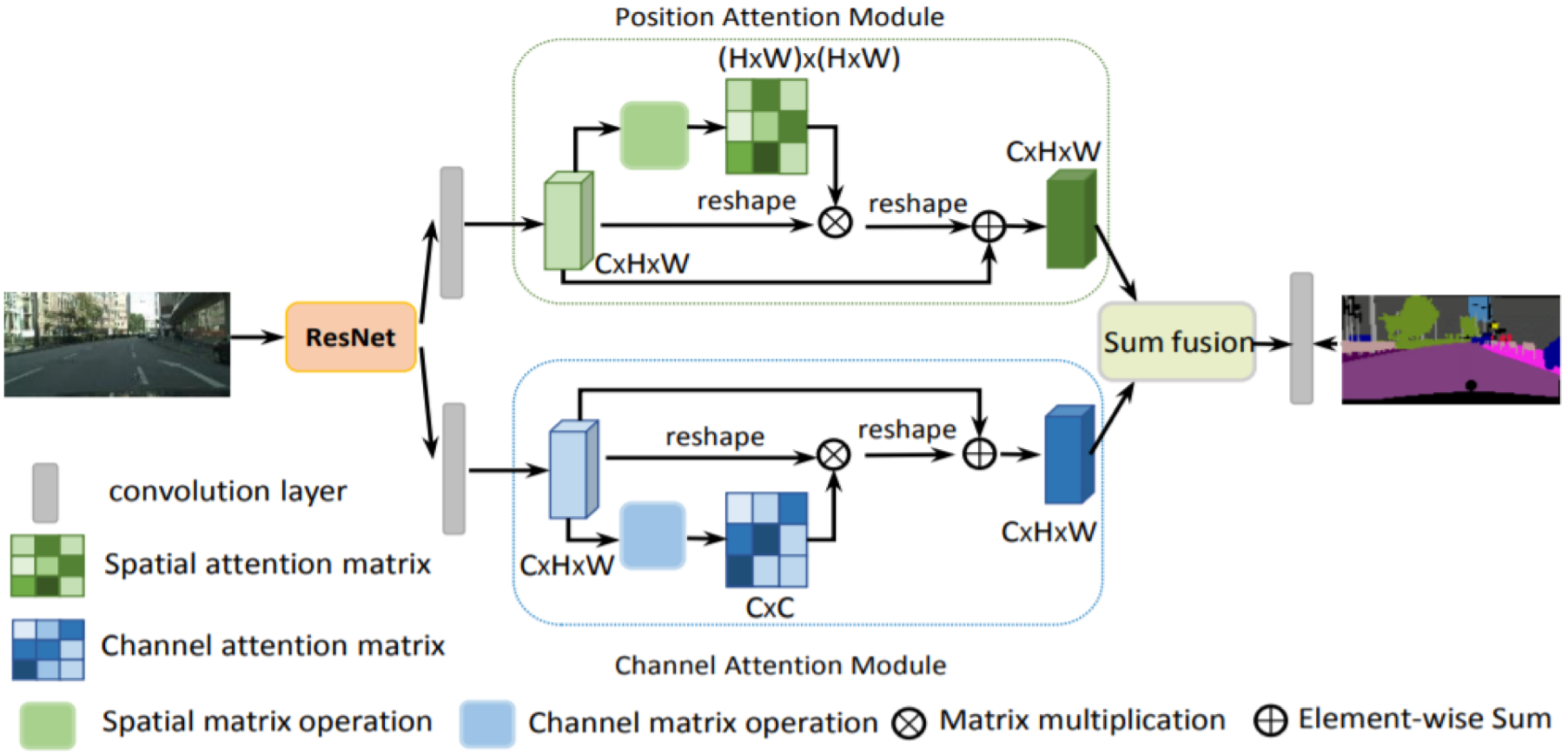
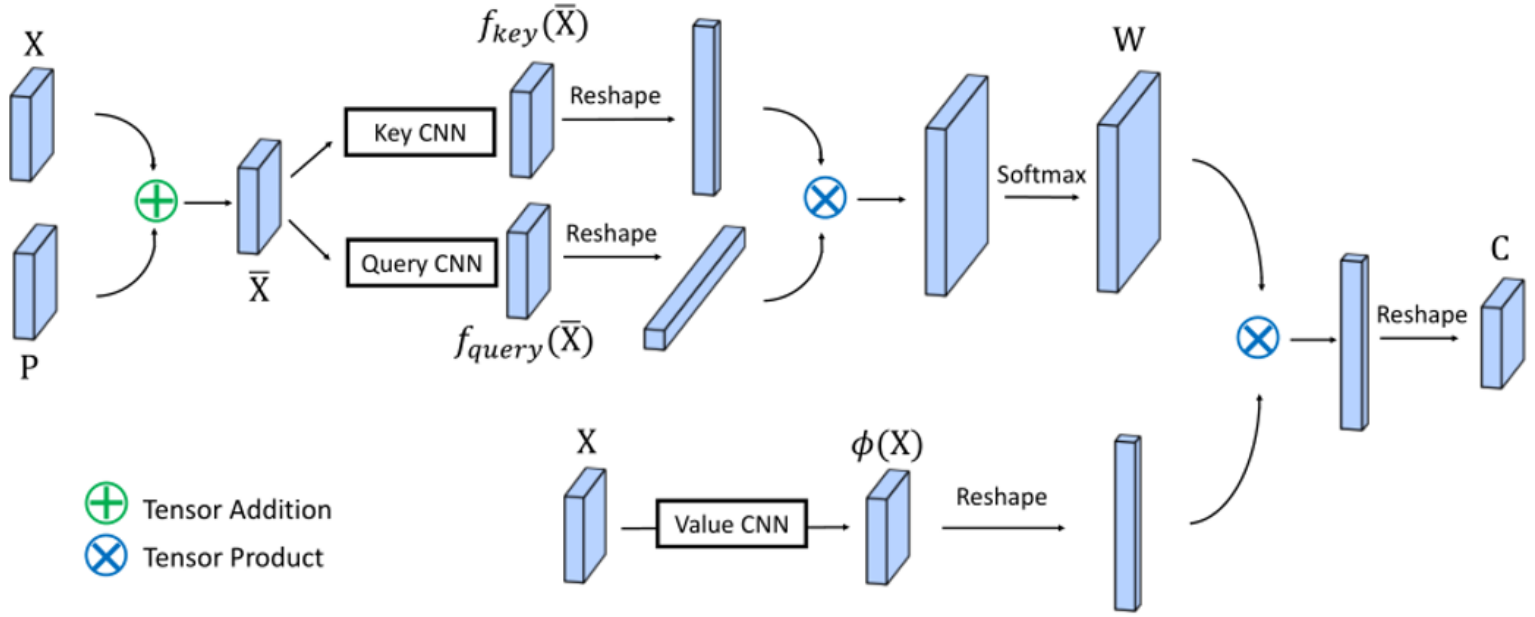


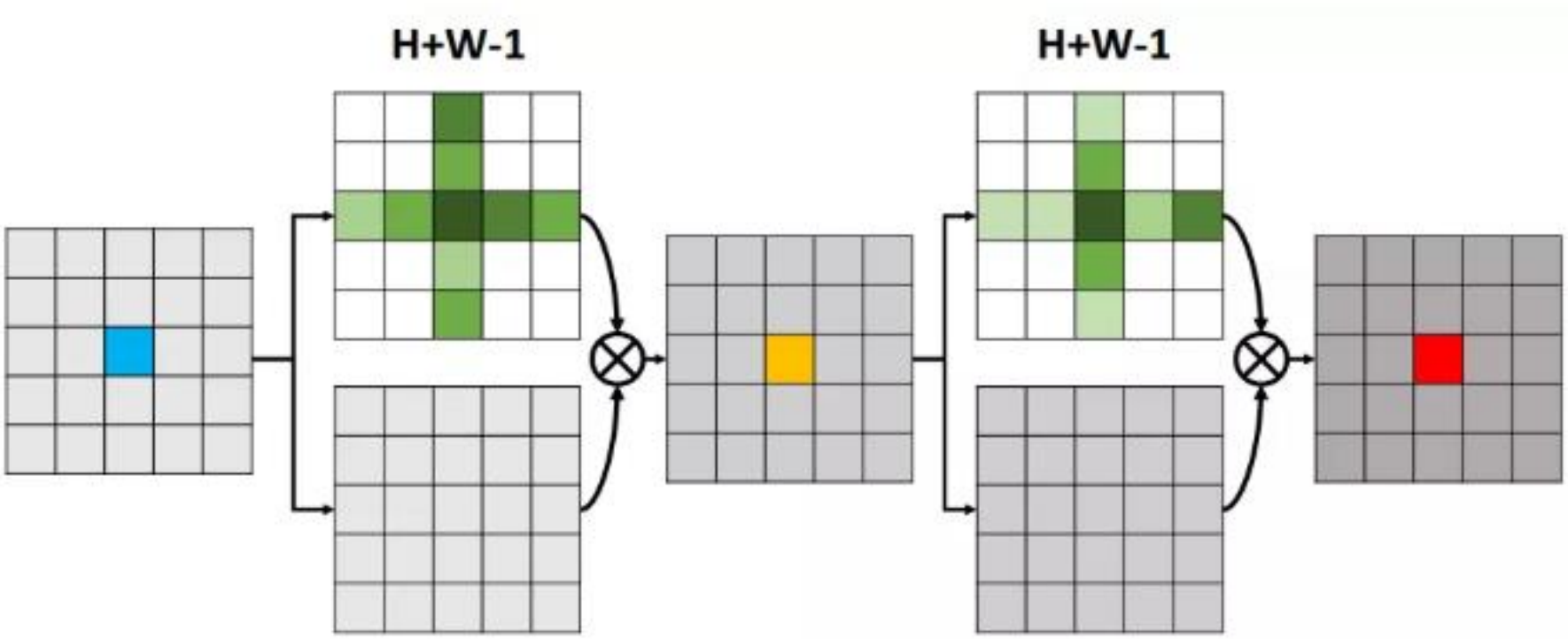
Figure 2: An overview of the Dual Attention Network. (Best viewed in color)

Object Context Network for Scene Parsing， 2018

和DANet同套用non-local来做语义分割，思想完全一样。但这个工作只利用了spatial信息，并且添加了多尺度处理



CCNet: Criss-Cross Attention for Semantic Segmentation 2018

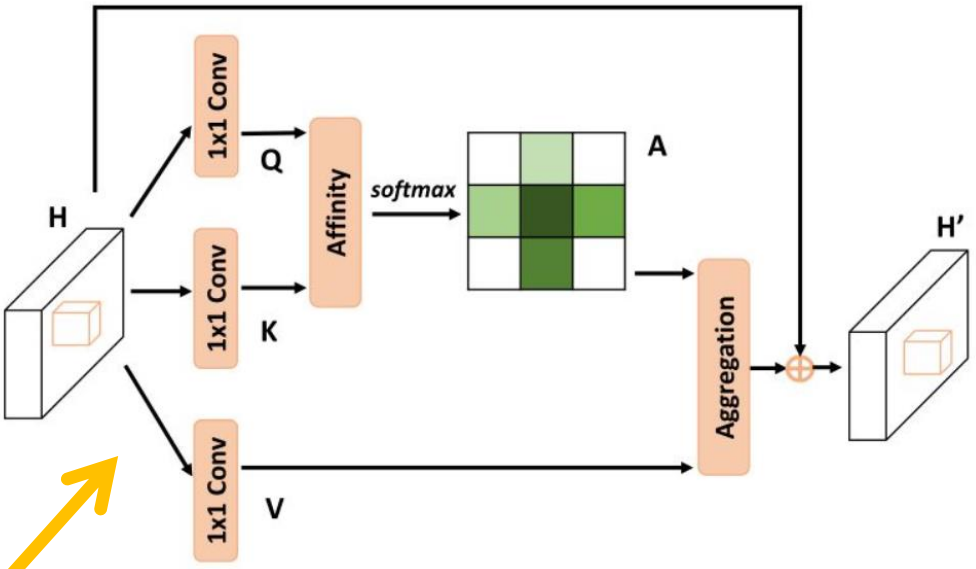


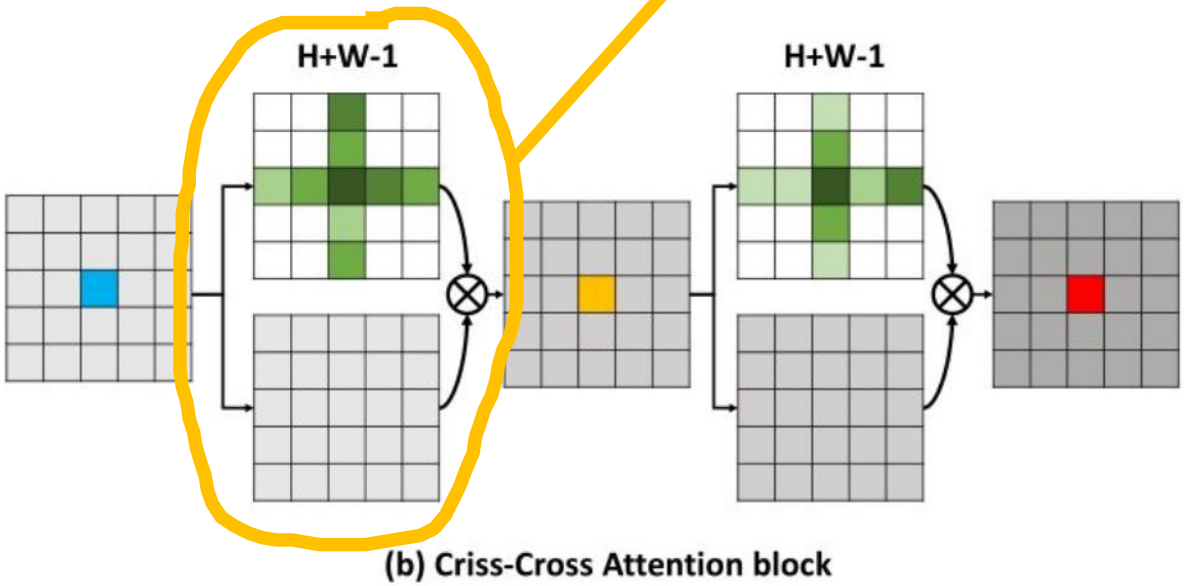
(b) Criss-Cross Attention block

CCNet 只需要计算每个像素点所在的行与列中的像素点之间的关系

同学们可能有问题：没有获取全局信息啊?

注意：第二次计算时，因为行列里的像素包含了其所在行列像素的信息，相当于捕获了全局的context

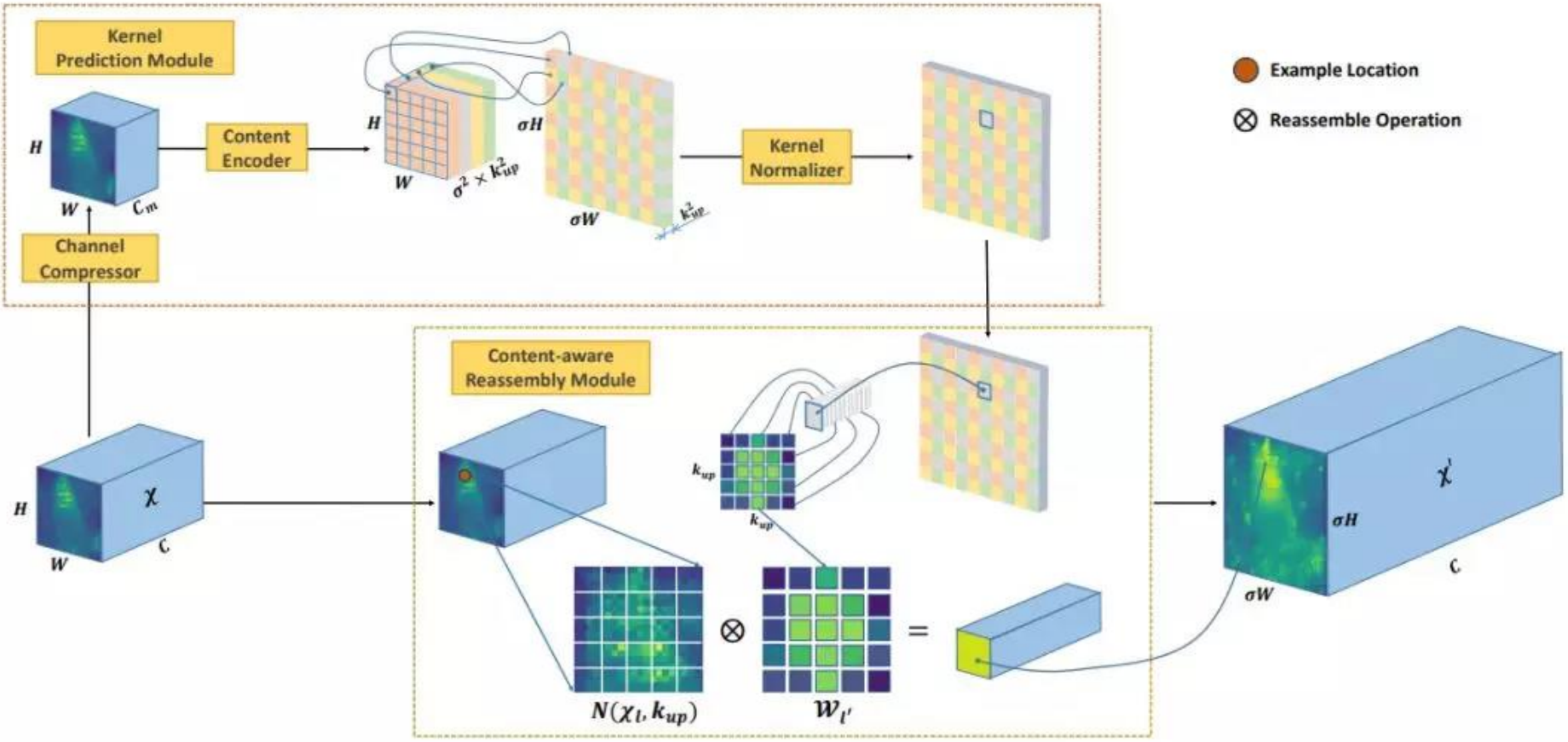




(b) Criss-Cross Attention block

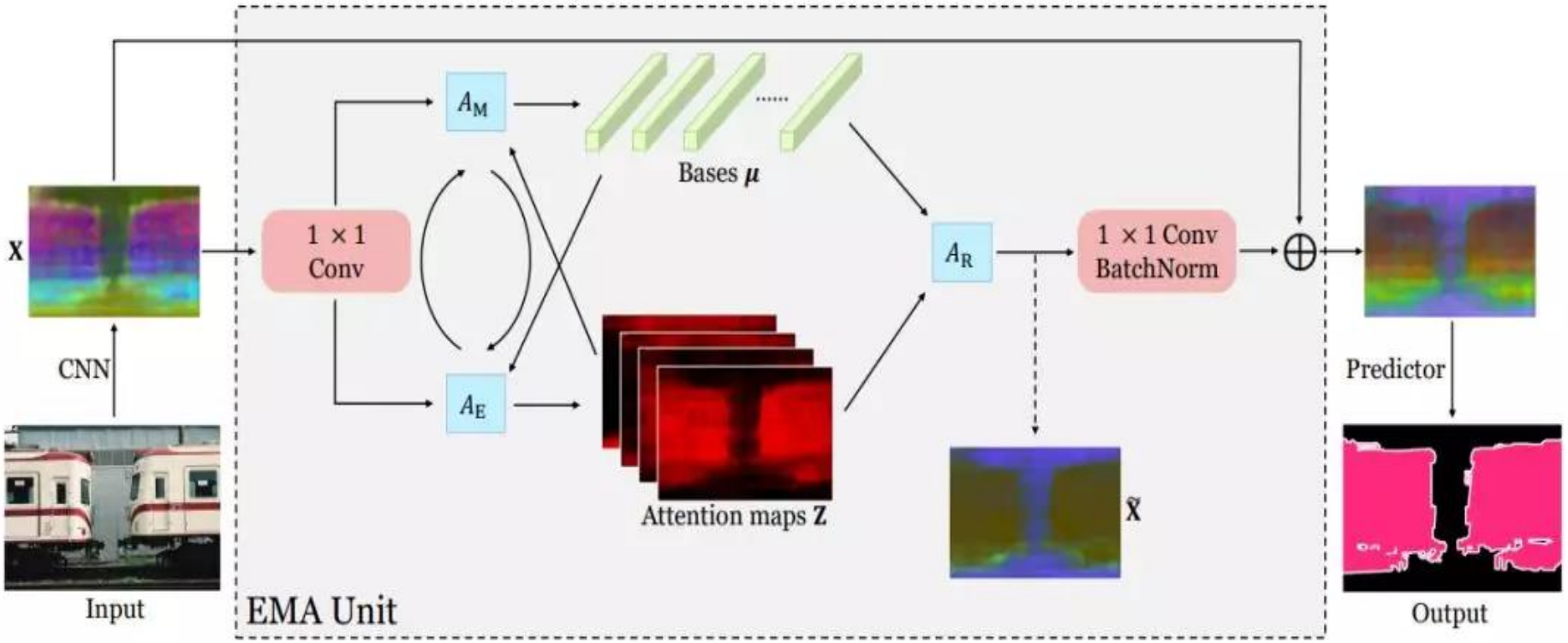
CARAFE: Content-Aware ReAssembly of FEatures， 2019

MMLab 的 CARAFE，用来进行特征上采样。其计算方式也是用窗口内像素特征的特征加权平均。其特殊之处在于，用于加权的权重是学习出来的，通过对特征变换、pixelshuffle上采样和通道归一化得到。



Expectation Maximization Attention Networks for Semantic Segmentation， ICCV 2019 Oral

EMNet摒弃了在全图上计算注意力图的流程，转而通过期望最大化（EM）算法迭代出一组紧凑的基，在这组基上运行注意力机制，从而大大降低了复杂度。其中，E 步更新注意力图，M 步更新这组基。E、M 交替执行，收敛之后用来重建特征图。本文把这一机制嵌入网络中，构造出轻量且易实现的 EMA Unit。在多个语义分割数据集上取得了较高的精度





Jie Hu[1] , Li Shen[2] , Gang Sun[1]

[1] Momenta [2] University of Oxford

CVPR 2017

胡杰的SENet取得冠军

Momenta 是一家自动驾驶公司

SENet，启发我们，通道上的信息非常重要

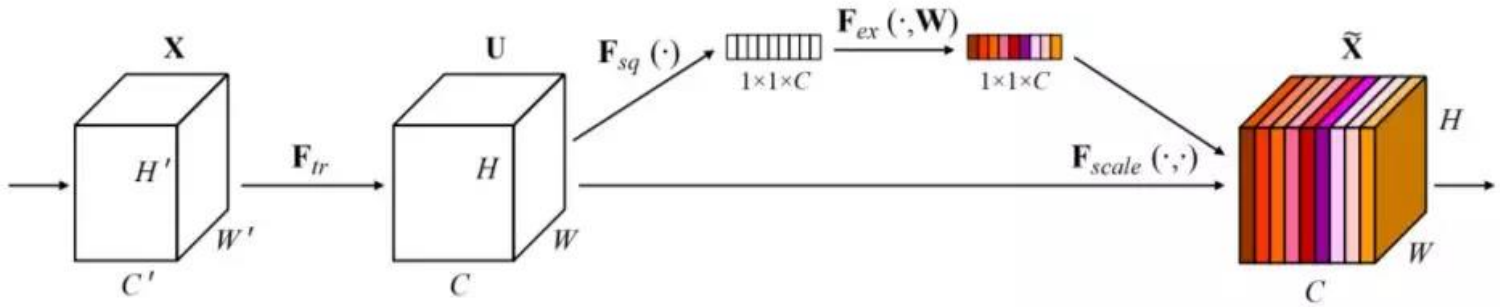
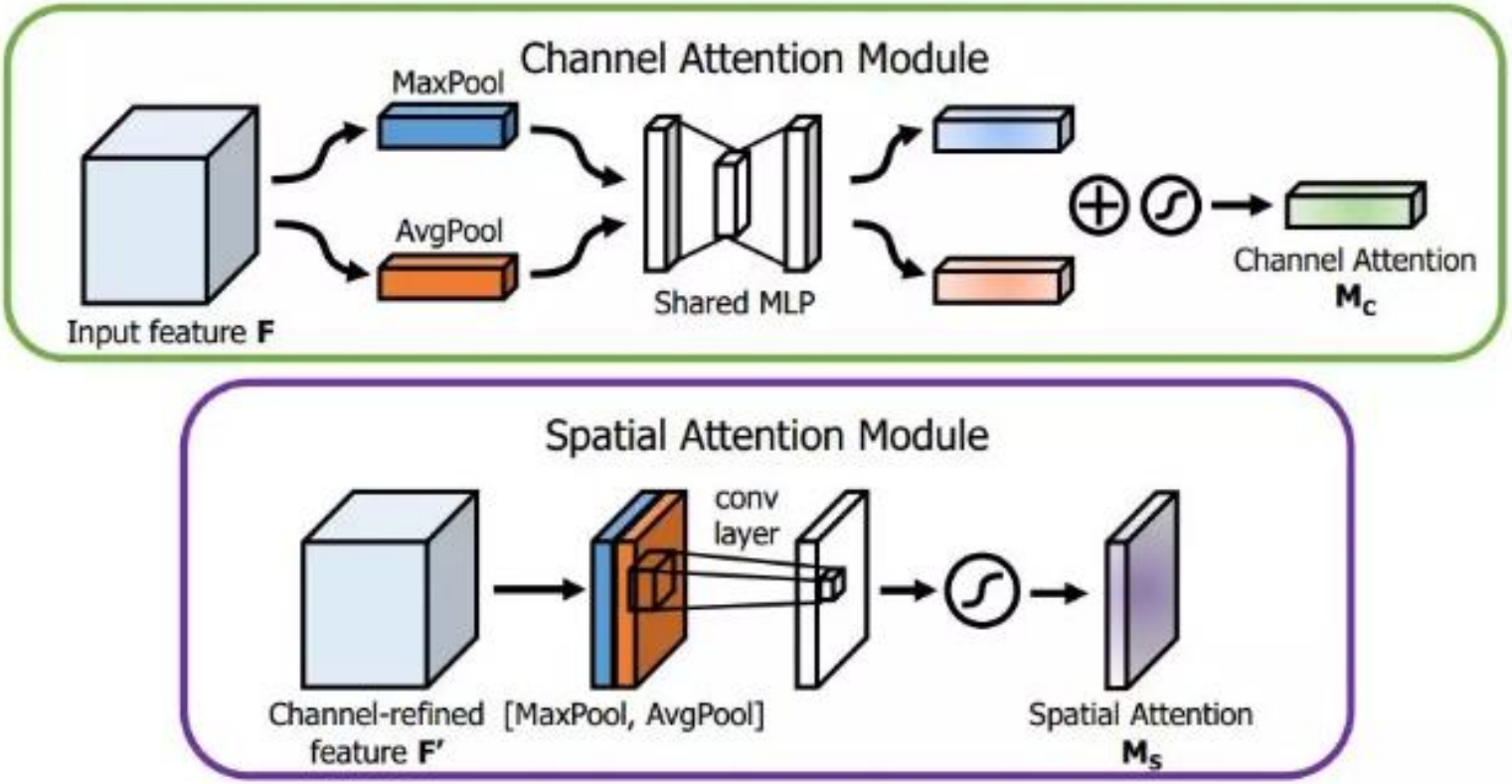
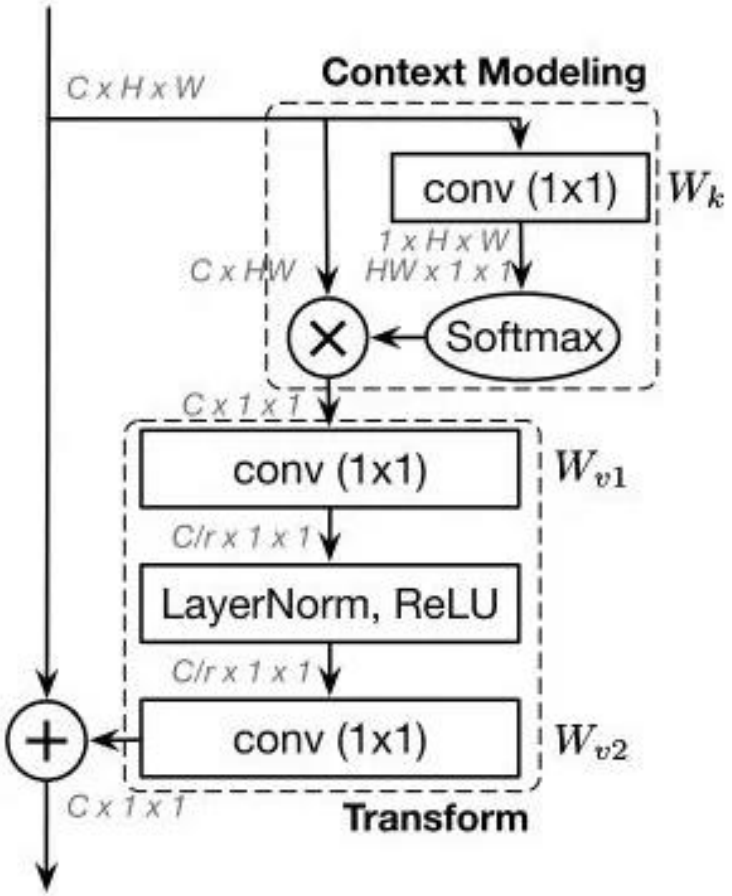


Fig. 1. A Squeeze-and-Excitation block.

CBAM： Convolutional Block Attention Module, *ECCV 2018*



GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond



Cross-modal 里的注意力机制

首先回顾一下，什么是 VQA?

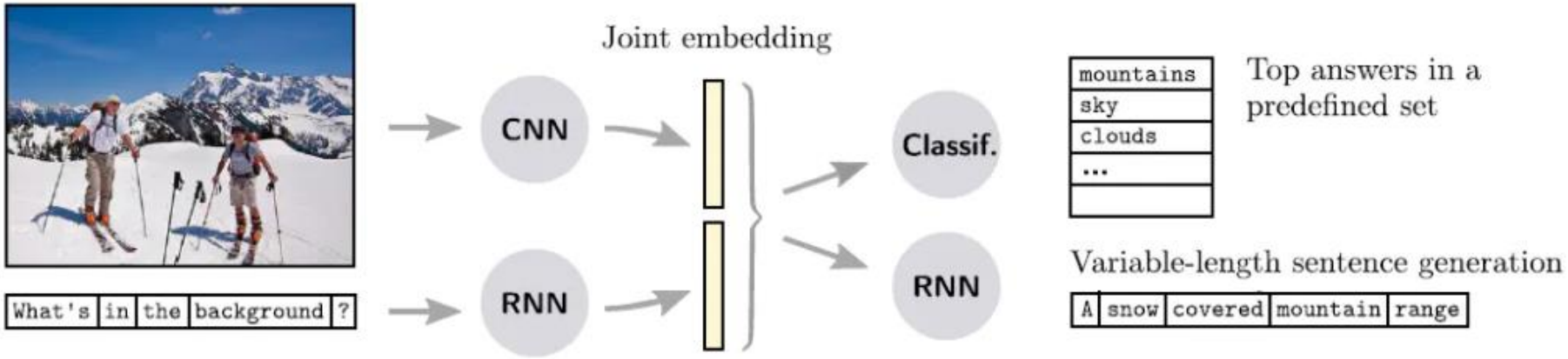


Image Captioning

a group of zebras standing in a field [Vinyals, CVPR15]
a group of zebras grazing on grass [You, CVPR16]
a group of zebras grazing in a field [Yao, ICCV17]
a group of zebras and a rainbow in the sky [Peter, CVPR18]
a group of zebras grazing in a field with a rainbow in the sky [Yao, ECCV18]

给定一张图，生成右边的语言描述
基于深度学习的方法主要是从15年开始
18年开始关注region之间的关系，可以生成更加丰富的描述

Image Captioning

同样是从2018年开始，人们开始关注 attention，挖掘图像、文本间的关系

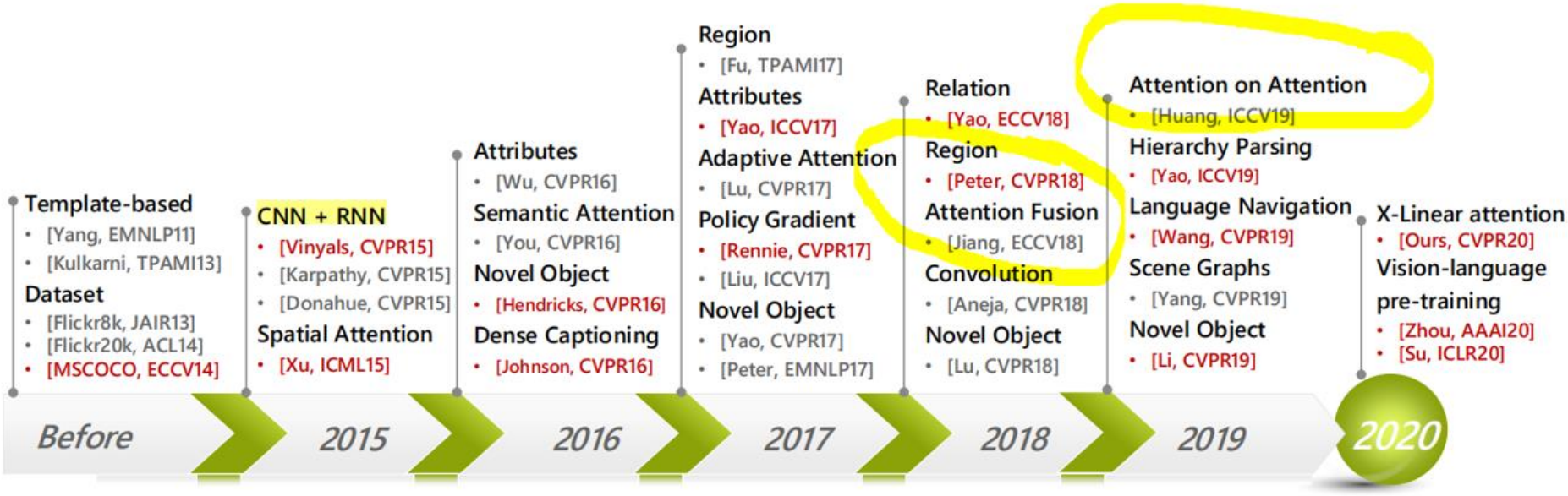
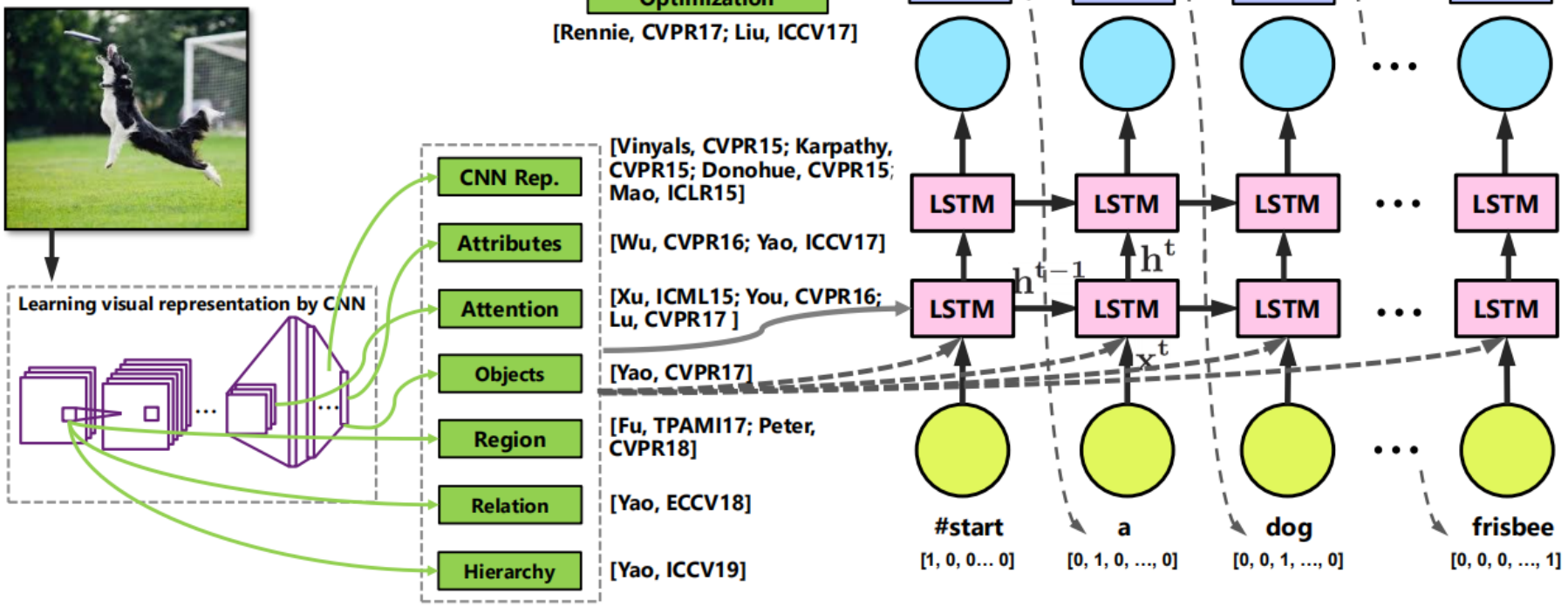
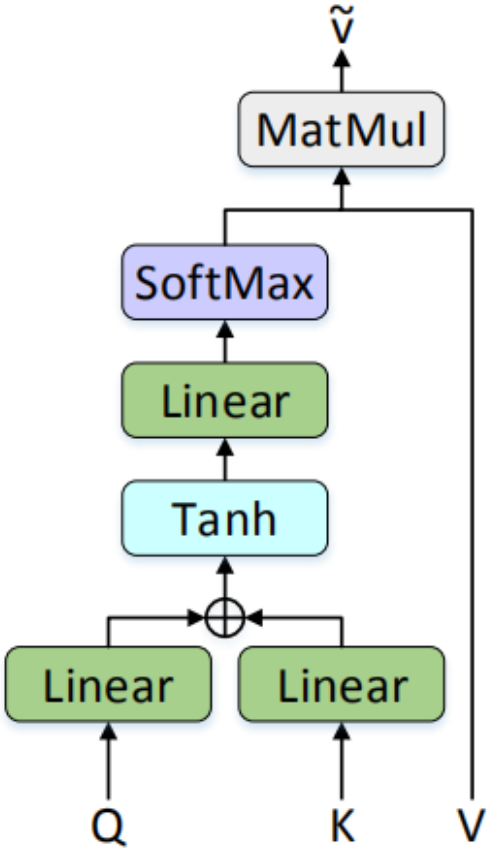


Image Captioning





Query: 来自于文本decoder

Key = Value : 图像的区域级表达（一般是用faster rcnn提取几个区域，然后每个区域的表达）

Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering,

CVPR 2018 ***2018年度VQA竞赛的冠军***

杭电余宙的 co-attention，2019 VQA竞赛冠军

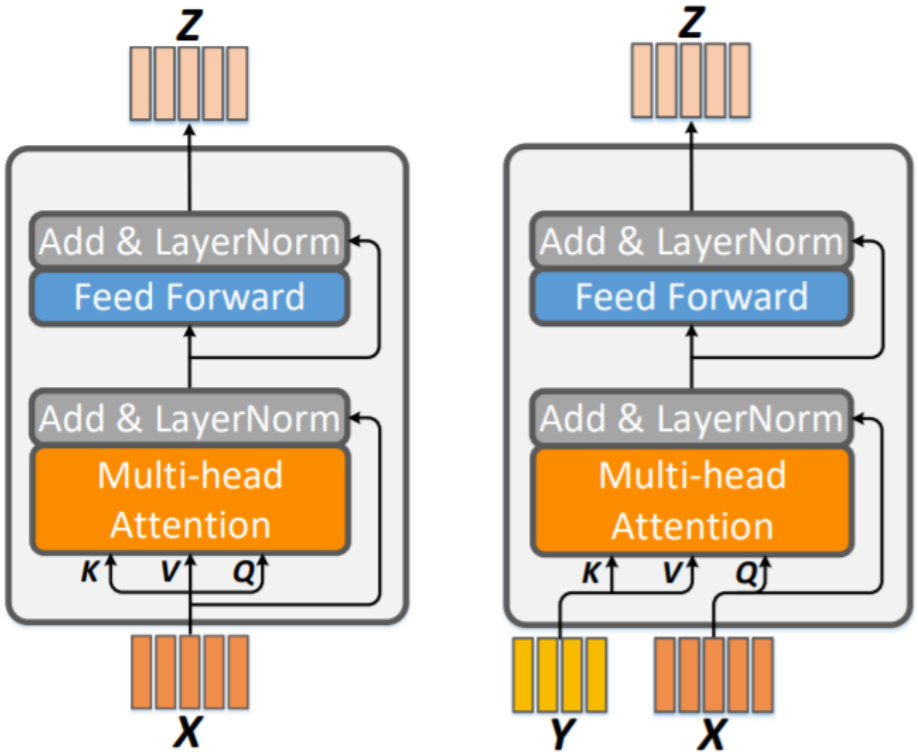
MCAN: Deep Modular Co-Attention Networks for Visual Question Answering

Team **MIL@HDU** with members
Zhou Yu, Jun Yu, Yuhao Cui, Jing Li

Media Intelligence Lab (MIL),
Hangzhou Dianzi University, P.R. China

这篇论文主要使用了 Transformer 的思路，设计了两个 attention 模块：

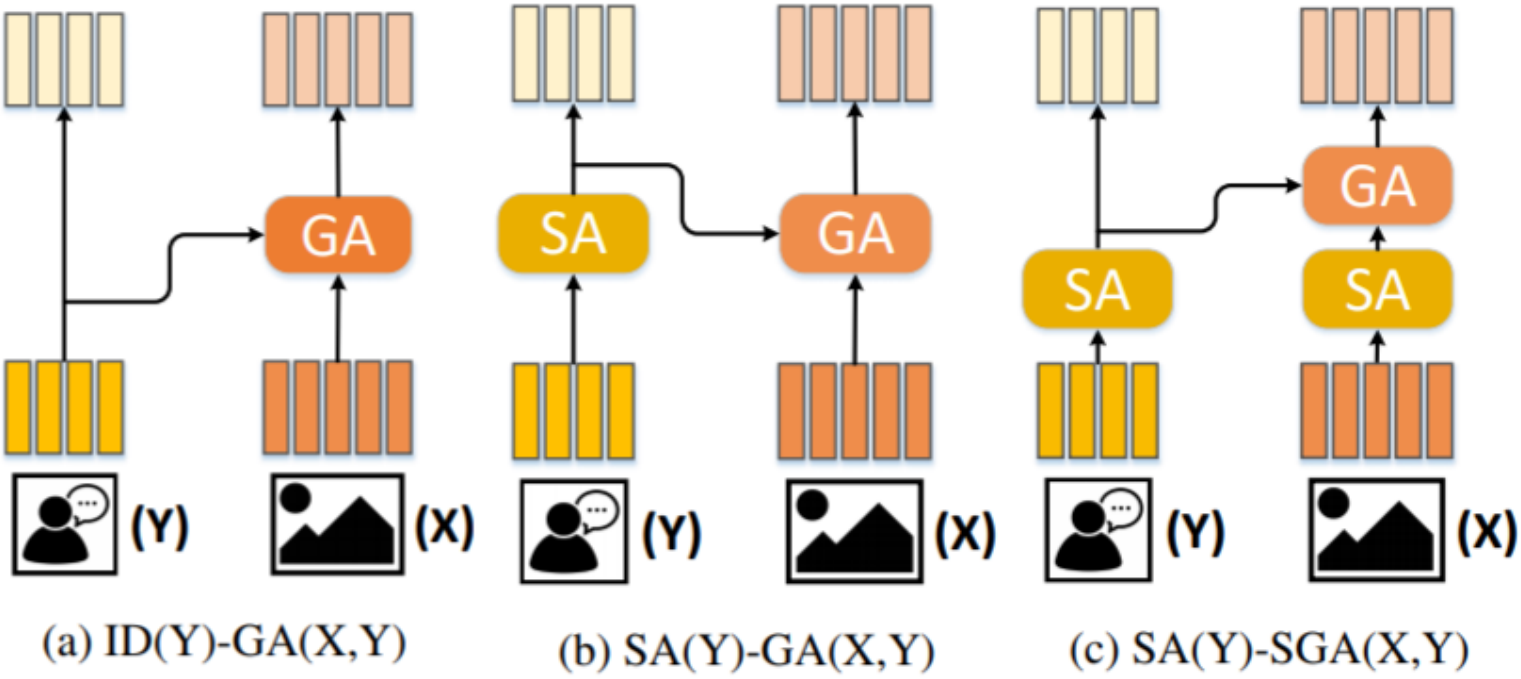
一个 Self-Attention (SA)，另一个是 Guided-Attention (GA)



(a) Self-Attention (SA)　(b) Guided-Attention (GA)

其中，X 是图像特征，
Y 是文本特征

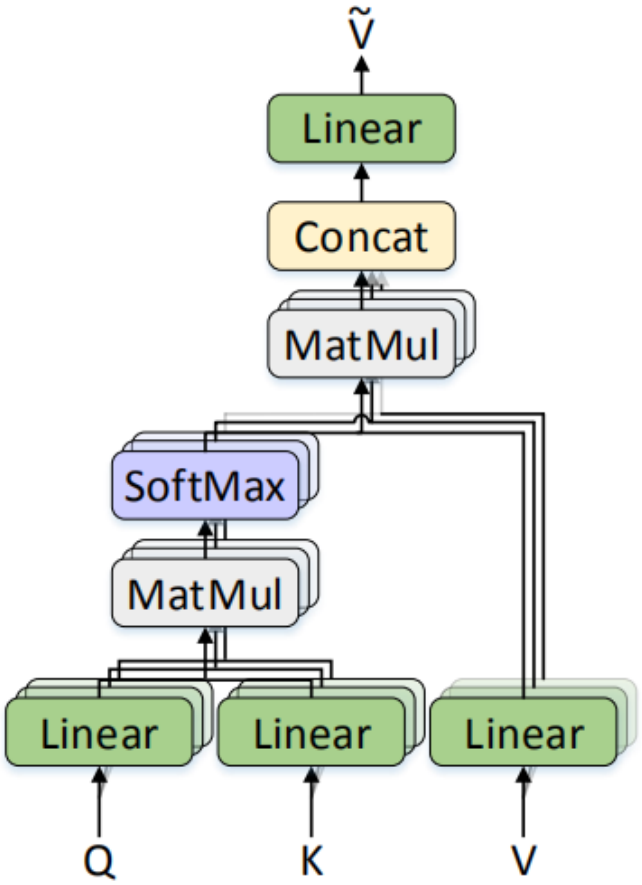
如何利用 SA 和 GA 呢？ 有三种方式
通过实验，余宙最终选择了第三种方式



(a) ID(Y)-GA(X,Y)　(b) SA(Y)-GA(X,Y)　(c) SA(Y)-SGA(X,Y)

为什么要 co-attention 呢？多模态融合需要同时理解图像、文本的含义，所以要同时利用两个模态的注意力机制

ACL 2018的一个工作，开始引入了 transformer

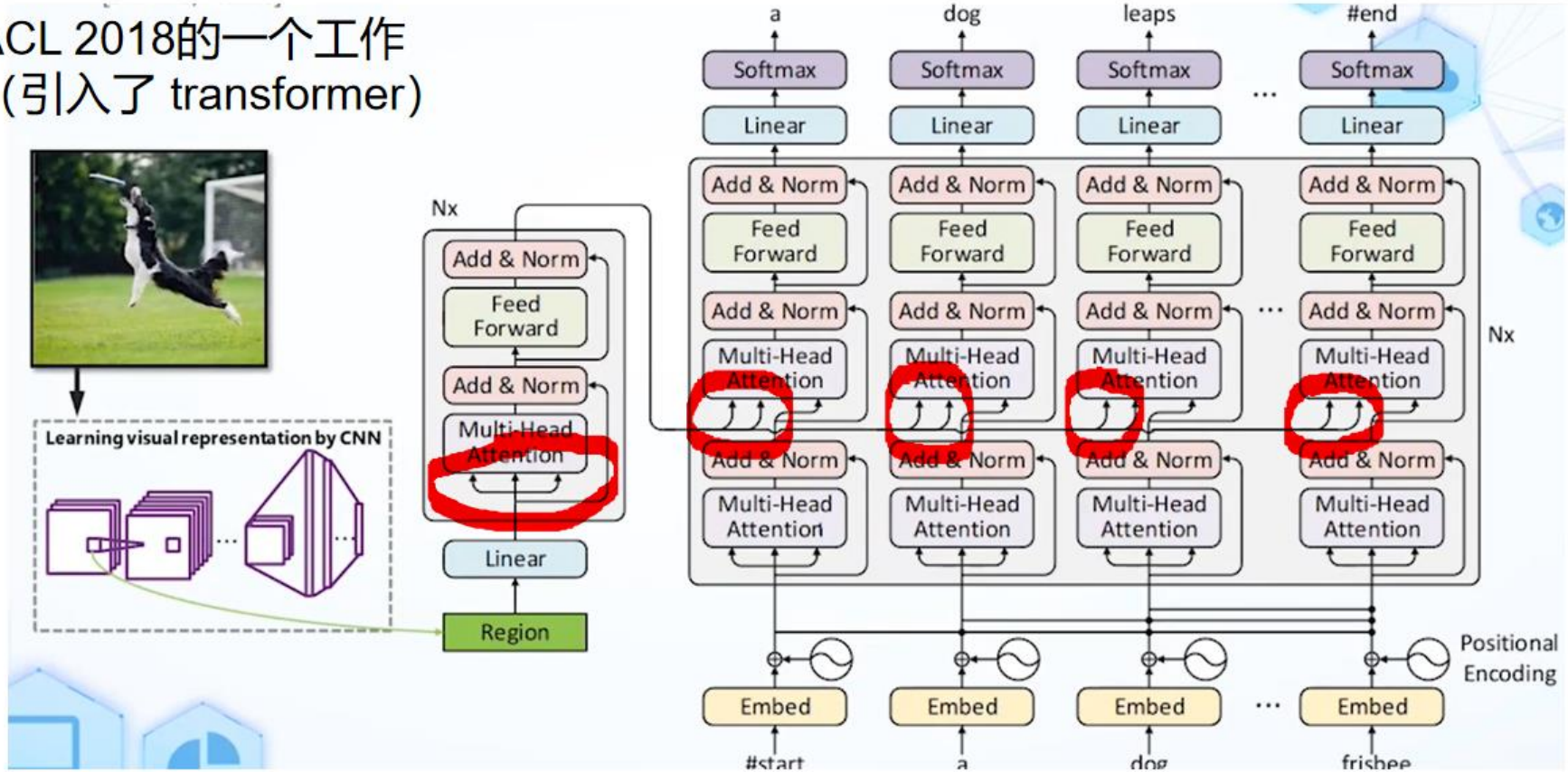


Multi-head Attention

Sharma, ACL 2018

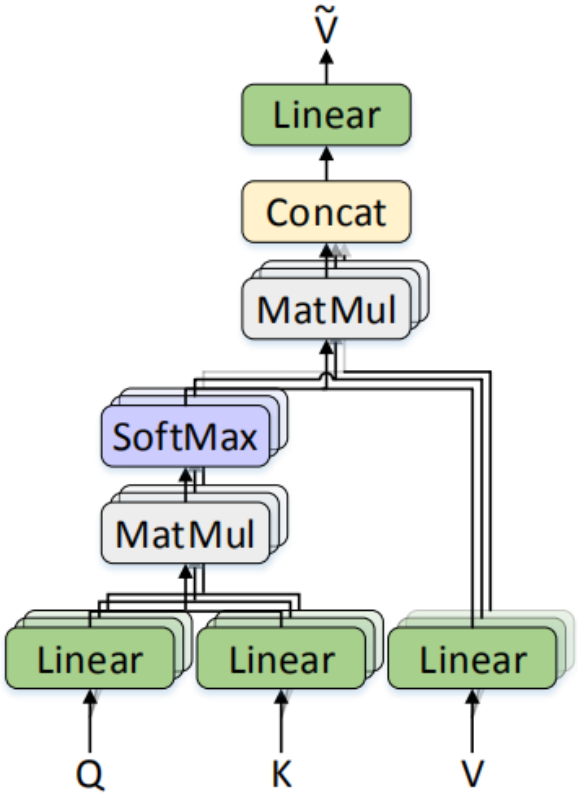
Image Captioning

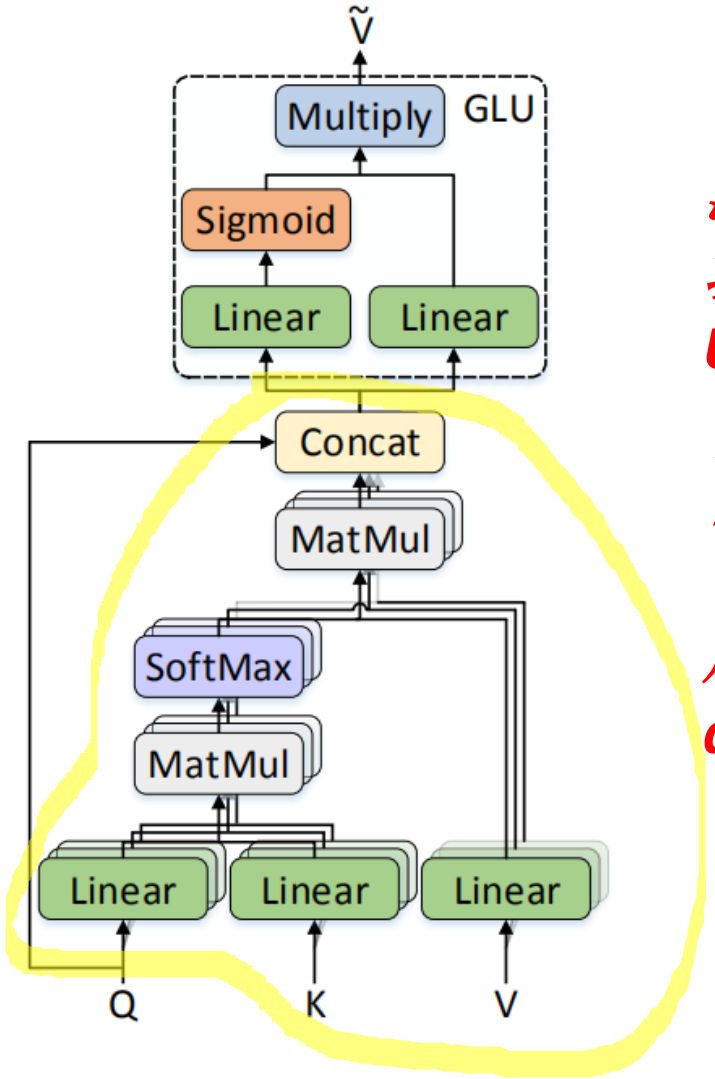
ACL 2018的一个工作
(引入了 transformer)



ICCV 2019 Attention on Attention for Image Captioning

ACL 2018



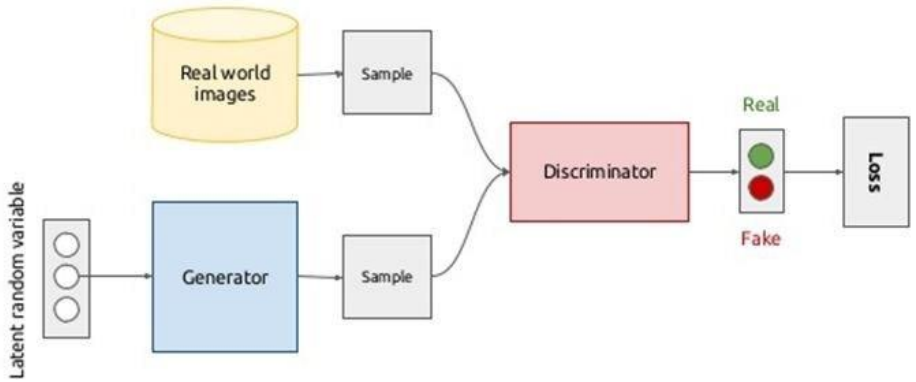


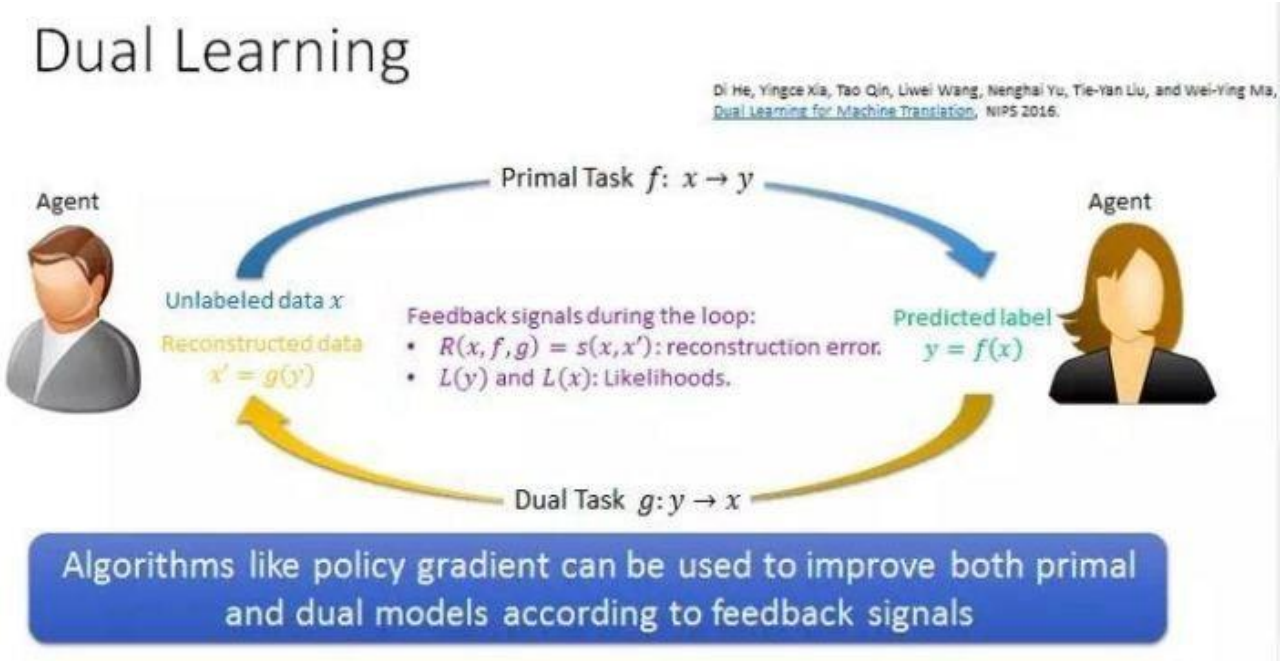
*添加一个**GLU** 模块（**Gated Linear Units**，在自然语言处理中应用非常多）*

*所以相当于在 **attention** 上又加了一个 **attention***

深度学习的未来：趋势

挑战1：标注数据代价昂贵
趋势1：从无标注的数据里学习





深度学习的未来：趋势

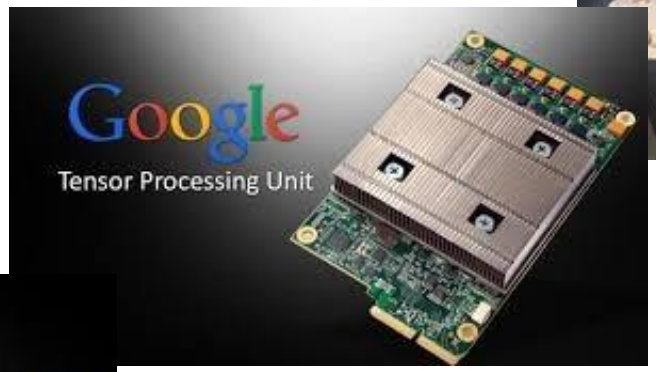
挑战2：大模型不方便在移动设备上使用
前沿2：降低模型大小

剪枝　　权值共享　　量化（二进制网络）

Network	Original Size	Compressed Size	Compression Ratio	Original Accuracy	Compressed Accuracy
LeNet-300	1070KB	27KB	**40x**	98.36%	98.42%
LeNet-5	1720KB	44KB	**39x**	99.20%	99.26%
AlexNet	240MB	6.9MB	**35x**	80.27%	80.30%
VGGNet	550MB	11.3MB	**49x**	88.68%	89.09%
GoogleNet	28MB	2.8MB	**10x**	88.90%	88.92%
SqueezeNet	4.8MB	0.47MB	**10x**	80.32%	80.35%

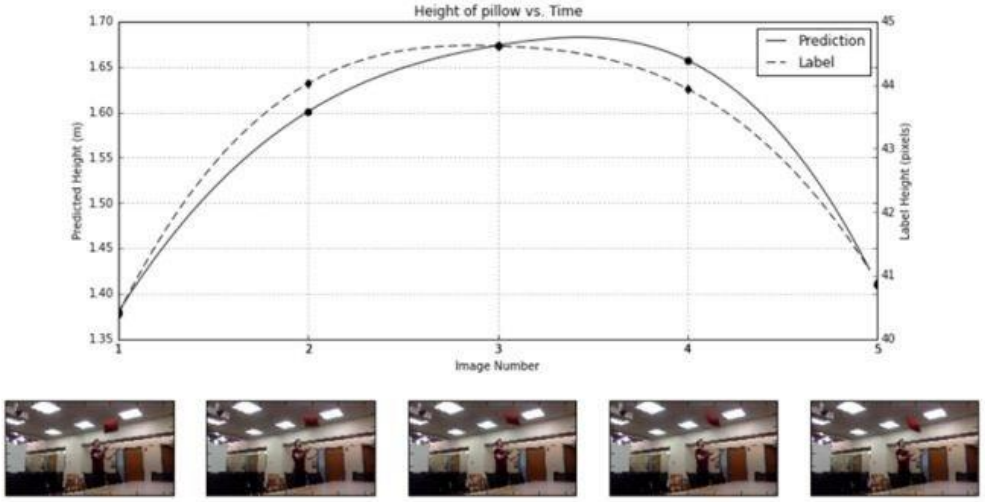
深度学习的未来：趋势

挑战3：大计算需要昂贵的物质、时间成本
趋势3：全新的硬件设计、算法设计、系统设计



深度学习的未来：趋势

挑战4：如何像人一样从小样本进行有效学习?
趋势4：数据+知识，深度学习与知识图谱、逻辑推理、符号学习相结合



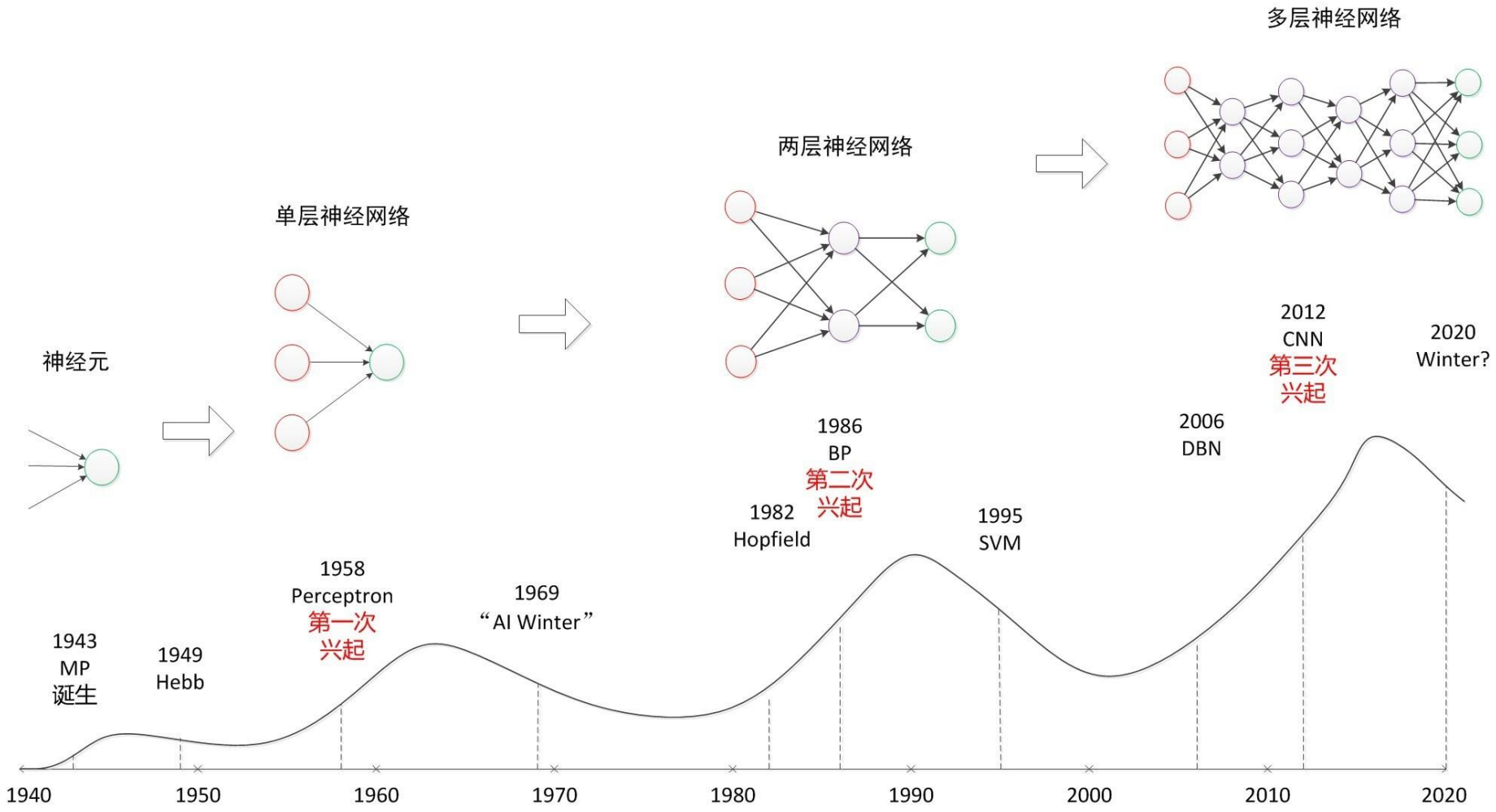
Stewart, Russell, and Stefano Ermon. "Label-free supervision of neural networks with physics and domain knowledge." AAAI2017.

无监督信息，利用物理知识（抛物线）跟踪
(AAAI 2017最佳论文)

深度学习的未来：趋势

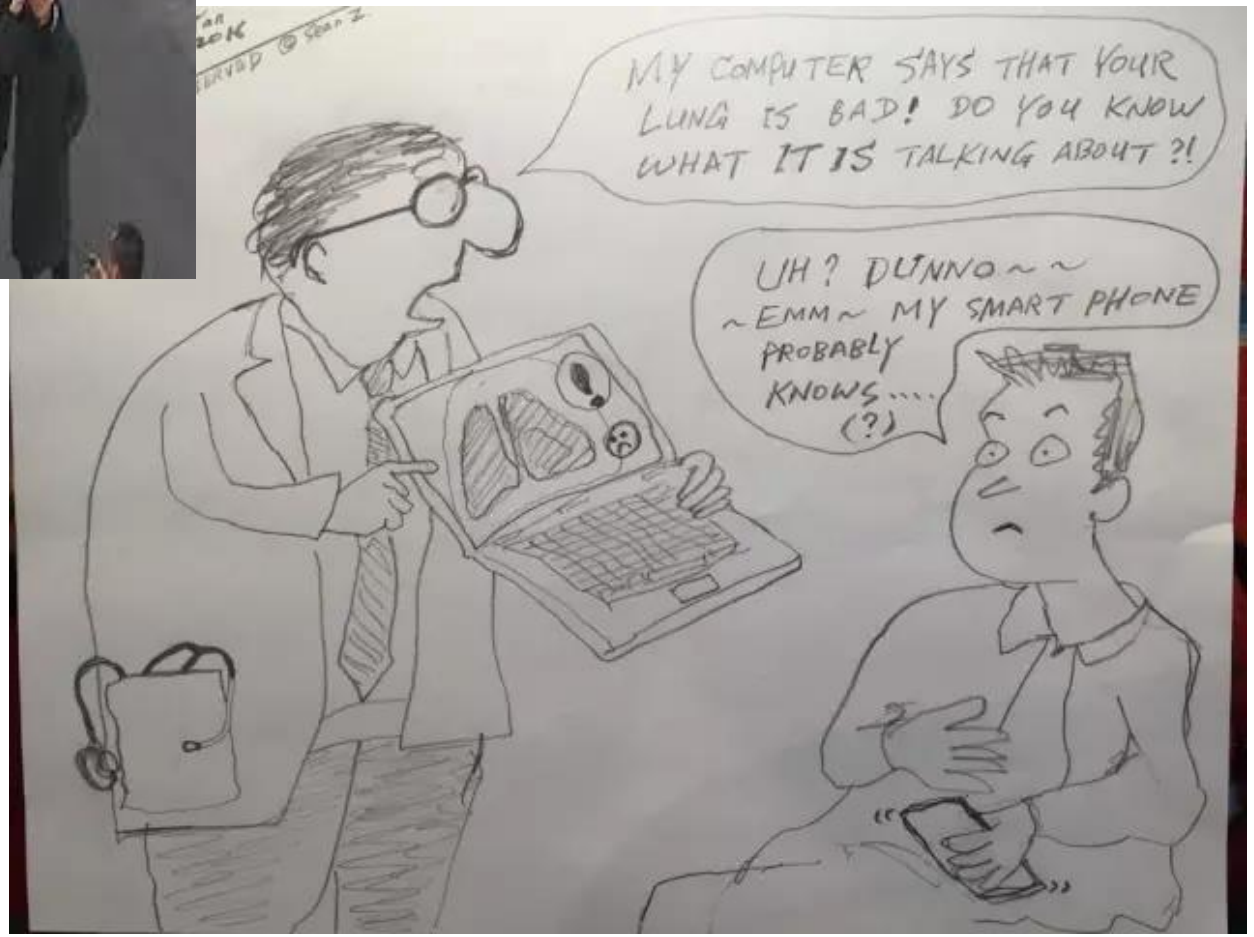
挑战5：如何从感知认知性的任务扩展到决策性任务？
趋势5：博弈机器学习

深度学习的未来：趋势


多层神经网络
两层神经网络
单层神经网络
神经元
2012
CNN
第三次
兴起
2020
Winter?
2006
DBN
1986
BP
第二次
兴起
1982
Hopfield
1995
SVM
1958
Perceptron
第一次
兴起
1969
“AI Winter”
1943
MP
诞生
1949
Hebb
1940
1950
1960
1970
1980
1990
2000
2010
2020

深度学习的未来：趋势


我们
敬礼！


MY COMPUTER SAYS THAT YOUR LUNG IS BAD! DO YOU KNOW WHAT IT IS TALKING ABOUT?!
UH? DUNNO~~ ~EMM~ MY SMART PHONE PROBABLY KNOWS.... (?)

THANKS