OpenMMLab

通用视觉框架OpenMMLab

第1讲 计算机视觉与OpenMMLab概述

林达华 教授

课程介绍

内容：

- 深度学习下的计算机视觉，各个应用方向的算法和原理
- OpenMMLab 开源视觉工具包

先修要求：

- （建议）Python编程
- （建议）机器学习
- （建议）基础的图像处理
- （不必须）PyTorch深度学习框架
- （不必须）深度学习
- （不必须）计算机视觉

OpenMMLab 项目概述

现状

1 架构

- 所有项目基于一致架构开发

15+ 研究方向

- 涵盖多个研究热点方向，算法覆盖完善

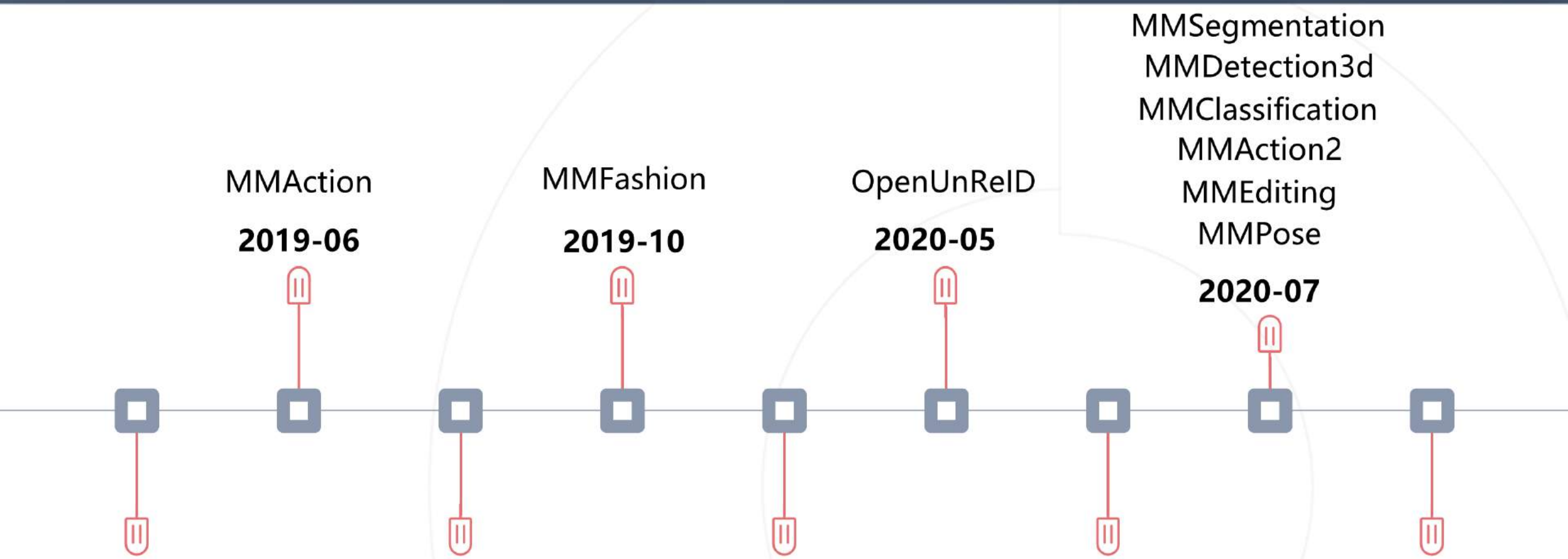
140+ 算法

- 包括 140+ 先进算法，性能领先

1200+ 预训练模型

- 拥有超过 1200 个预训练模型，真正实现开箱即用

发展历程


2018-08
MMCV
MMDetection
MMAction
2019-06
2019-08
MMSR
MMSkeleton
MMFashion
2019-10
2020-03
OpenPCDet
OpenUnReID
2020-05
2020-06
OpenSelfSup
MMSegmentation
MMDetection3d
MMClassification
MMAction2
MMEditing
MMPose
2020-07
2021-01
MMTracking

总体架构概览

OpenMMLab

算法框架和数据集

15+ 算法库

- MMClassification 图像分类
- MMSegmentation 语义分割
- MMSkeleton 基于骨骼分析
- MMDetection 目标检测
- MMEditing 图像和视频编辑
- MMTracking 目标跟踪
- MMDetection3D 3D目标检测
- MMAction2 视频动作分析
- MMPose 人体关键点检测

40+ 数据集

- WIDER
- FineGym
- MovieNet
- Placepedia

视觉基础库

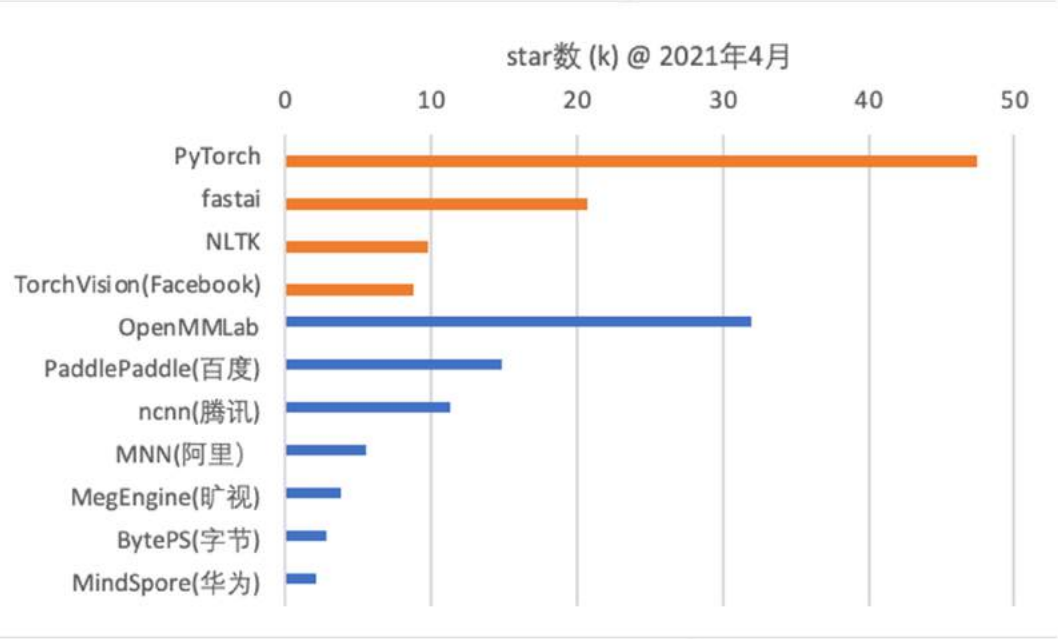
- 公用底层模块
- 抽象训练接口

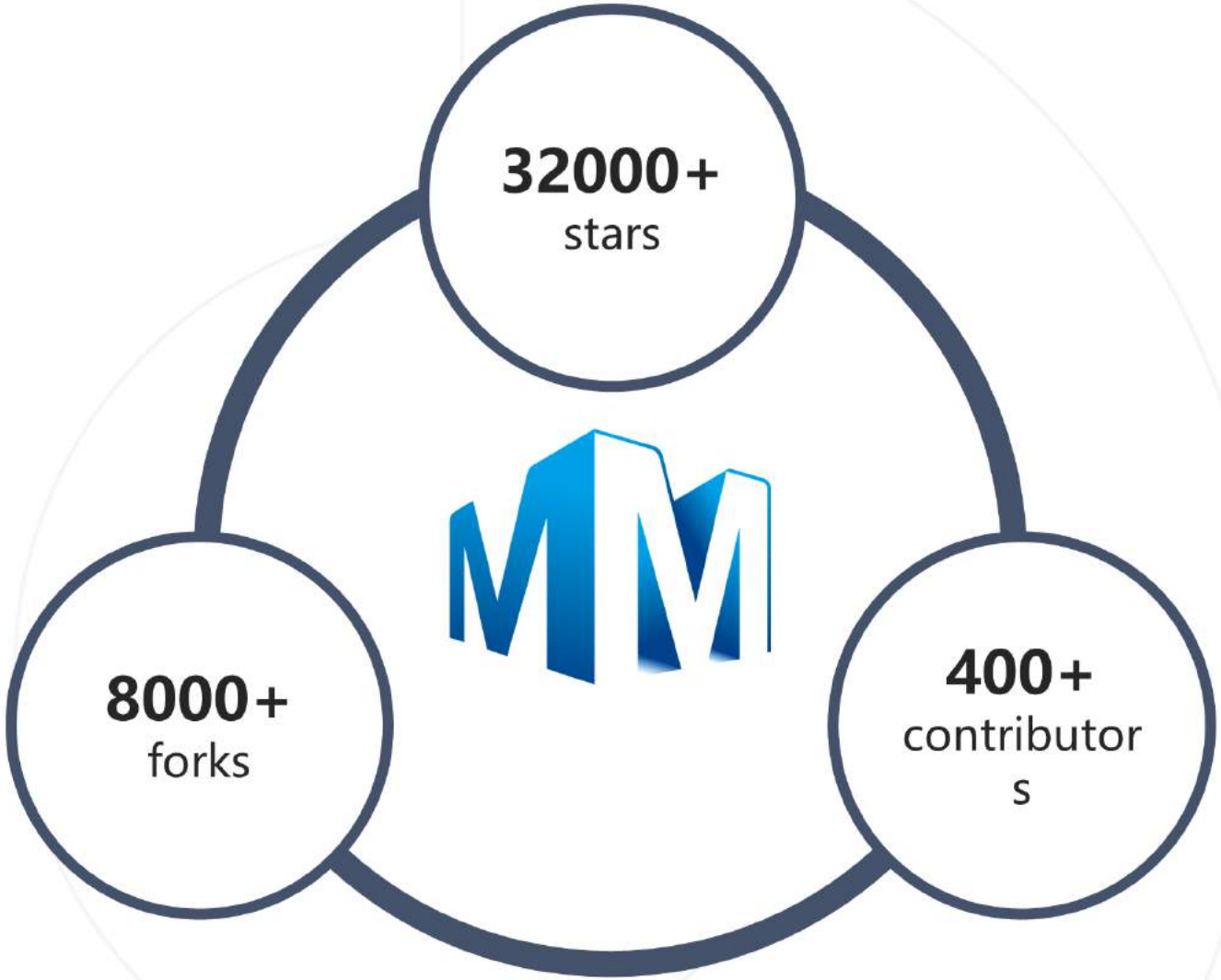
训练框架

- SenseParrots
- PyTorch

项目意义及社区影响力

- 提供高质量代码框架，减少算法复现难度
- 提供完善的科研平台，加速产出
- 缩短算法落地链条，促进产学研打通

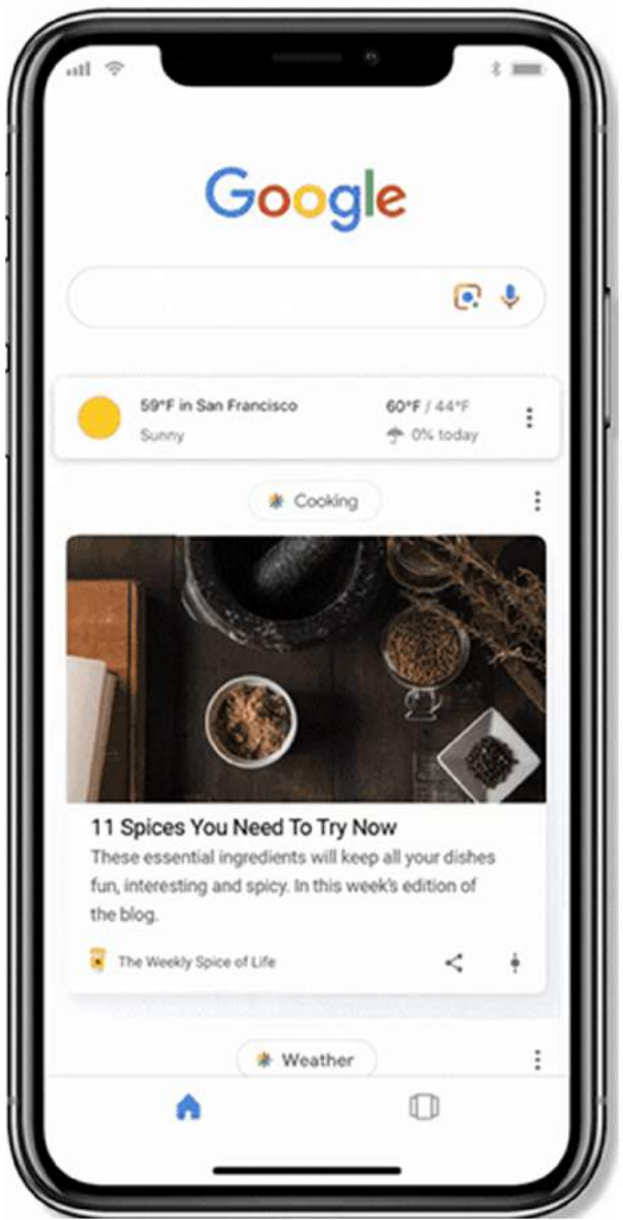


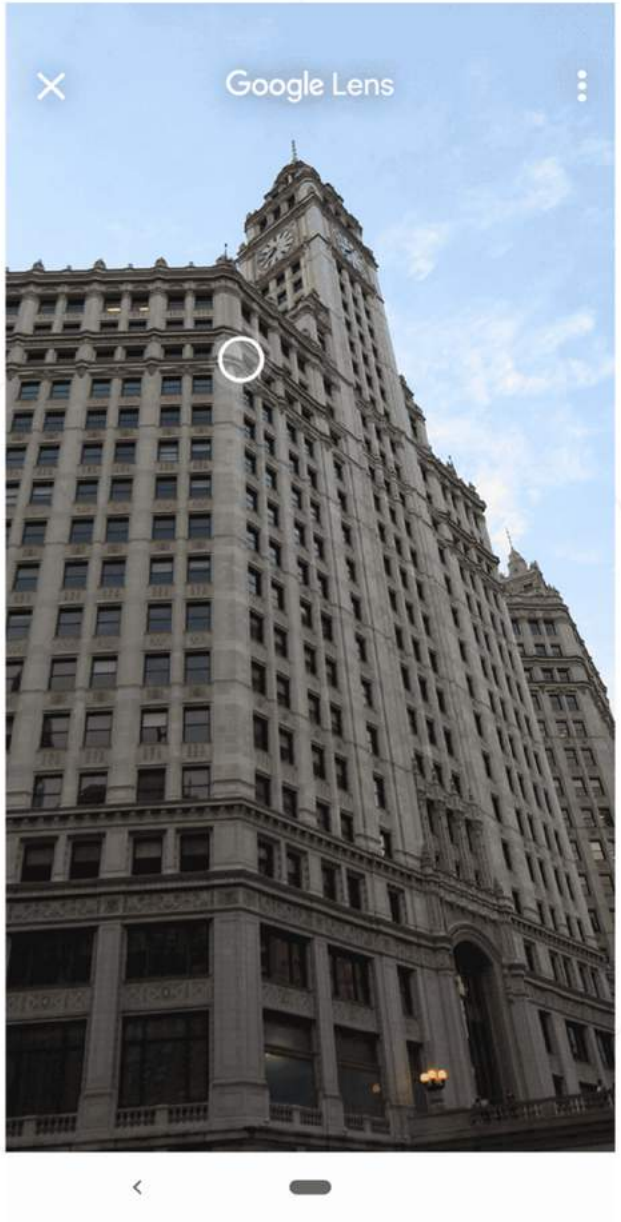


丰富多彩的计算机视觉世界

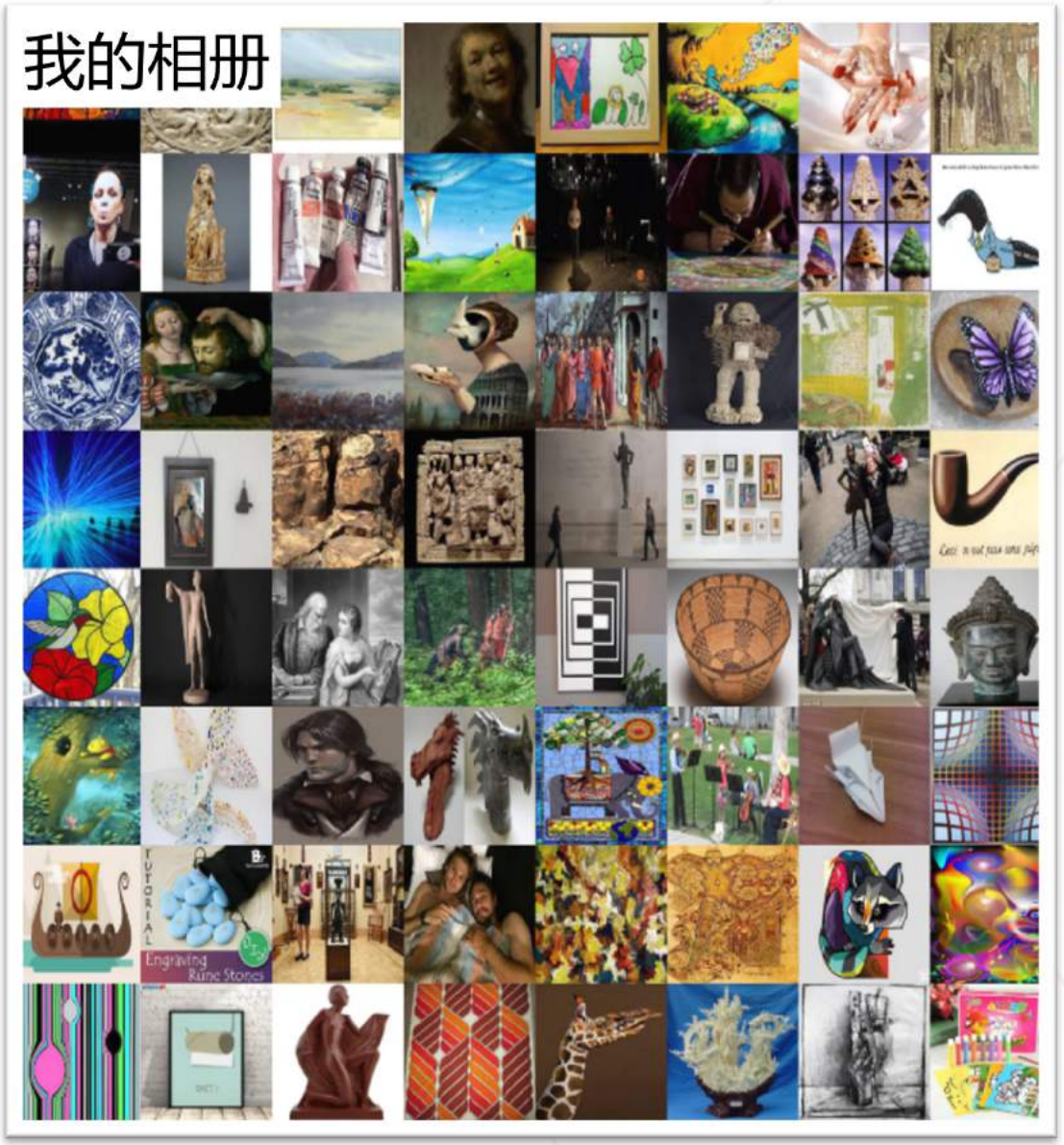
这是什么？我在哪里？

这是什么？我在哪里？


Google
59°F in San Francisco
Sunny
60°F / 44°F
0% today
Cooking
11 Spices You Need To Try Now
These essential ingredients will keep all your dishes fun, interesting and spicy. In this week's edition of the blog.
The Weekly Spice of Life
Weather


Google Lens

找不到照片?



基本任务：识别图像中的物体是什么

……

猫　　……

狗　　柴犬

汽车　　牧羊犬

建筑物　　……

……

图像标签

图像检索

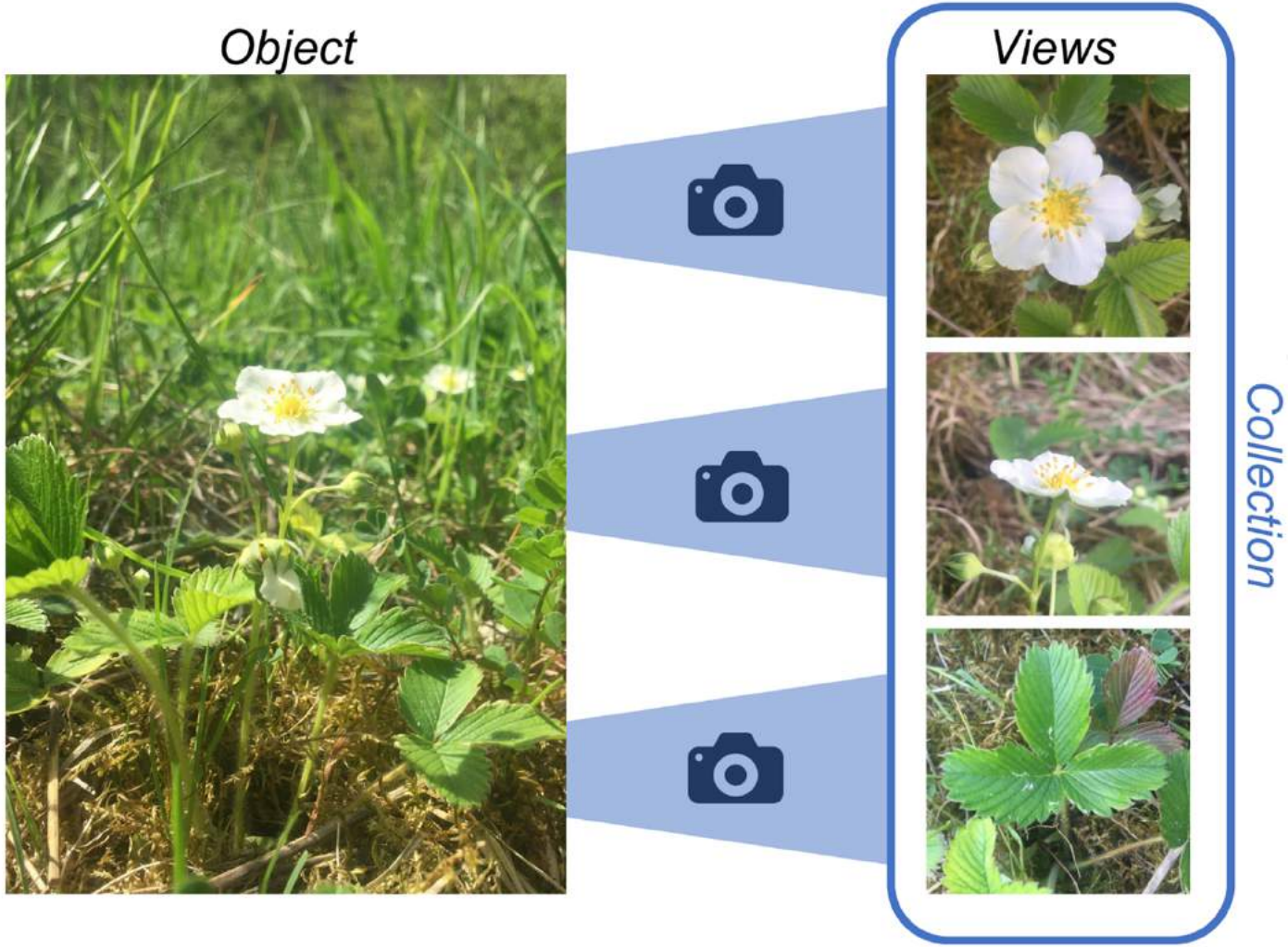
困难1：同类物体外观差异巨大

外观各异的蘑菇

困难2：不同物体外观相似

吉娃娃还是松饼?

困难3：拍摄环境的影响



从数据中学习

收集数据

狗

狗

……

猫

定义模型

$y = f(x;\Theta)$

x：图像像素
y：类别标签
Θ：模型参数

例子：

$y = \Theta^T x$

训练

通过优化算法找到让模型在数据集上获得最大正确率的参数 Θ^*

预测

$\hat{y} = f(\hat{x};\Theta^*)$

$\hat{x}$：新的图像
$\hat{y}$：模型的预测
Θ^*：最佳参数

图像特征与分类器 (1990s~2012)

图像

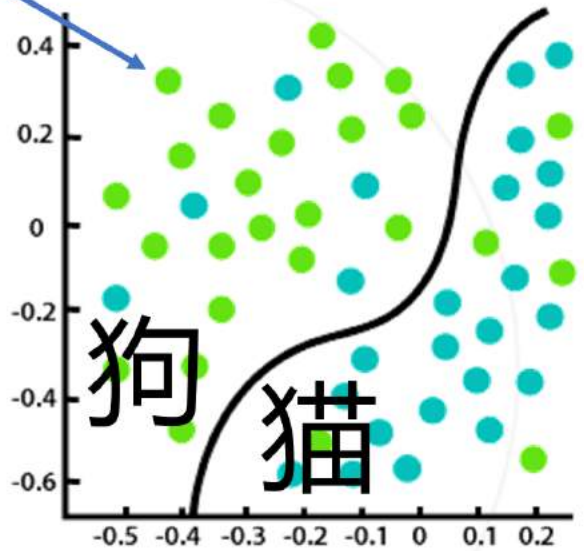
特征算法

```
[[0.06813484],
 [0.00789936],
 [0.00509447],
 ...,
 [0.02194785],
 [0.0908481 ],
 [0.05443394]]
```

图像特征

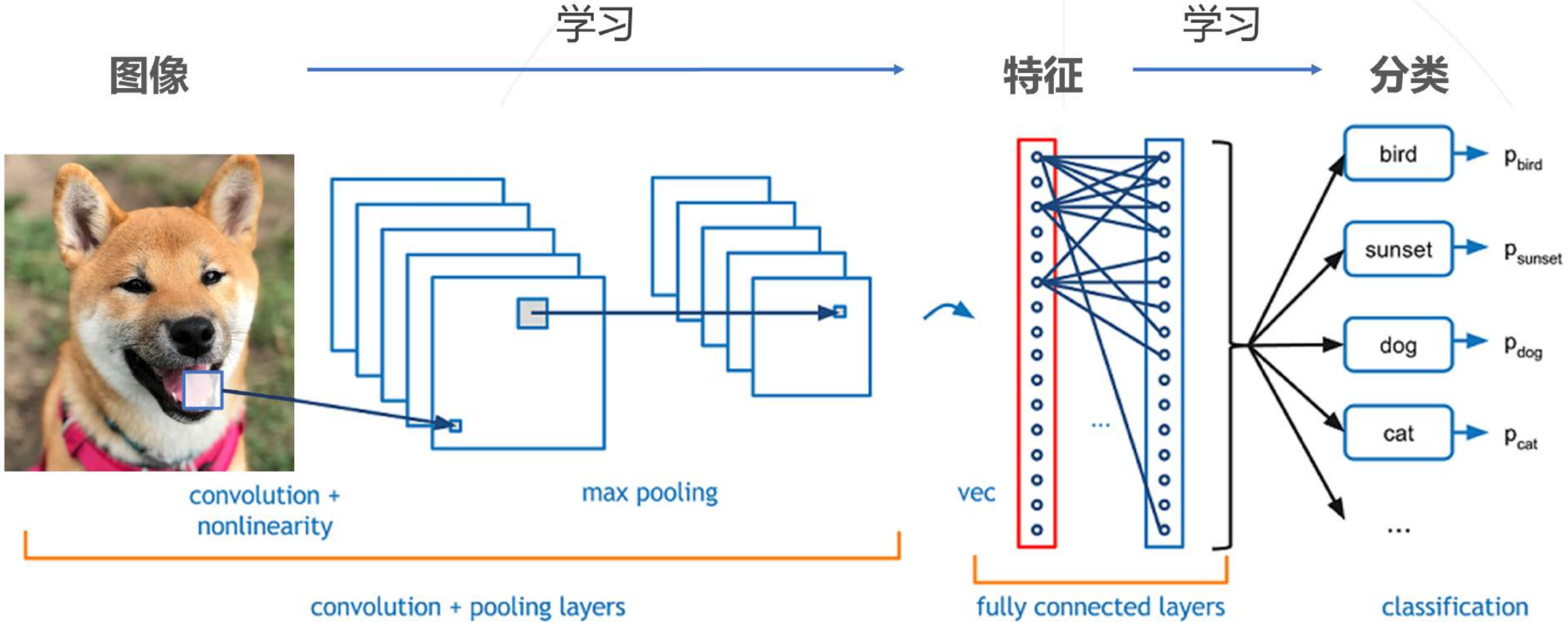
学习

高维空间中的样本点



分类

特征与分类联合学习


学习
学习
图像
特征
分类
bird
sunset
dog
cat
...
p_{bird}
p_{sunset}
p_{dog}
p_{cat}
convolution + nonlinearity
max pooling
vec
convolution + pooling layers
fully connected layers
classification

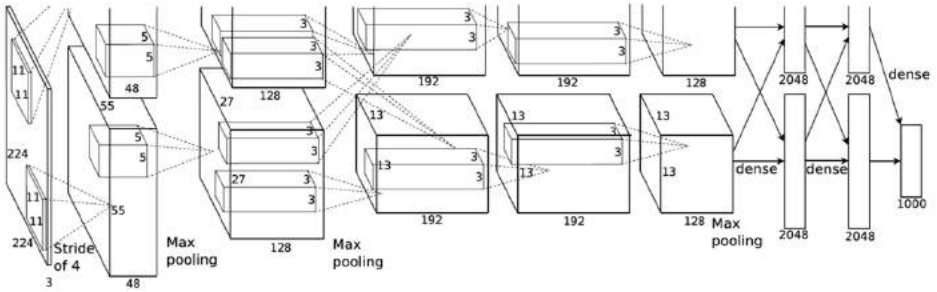
ImageNet 数据集 与 ILSVRC 竞赛

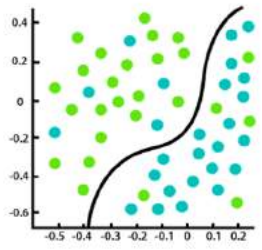
ImageNet 是一个图像数据集，创建于 2006 年，目前包含多达 2 万类，共计约 1500 万张图片

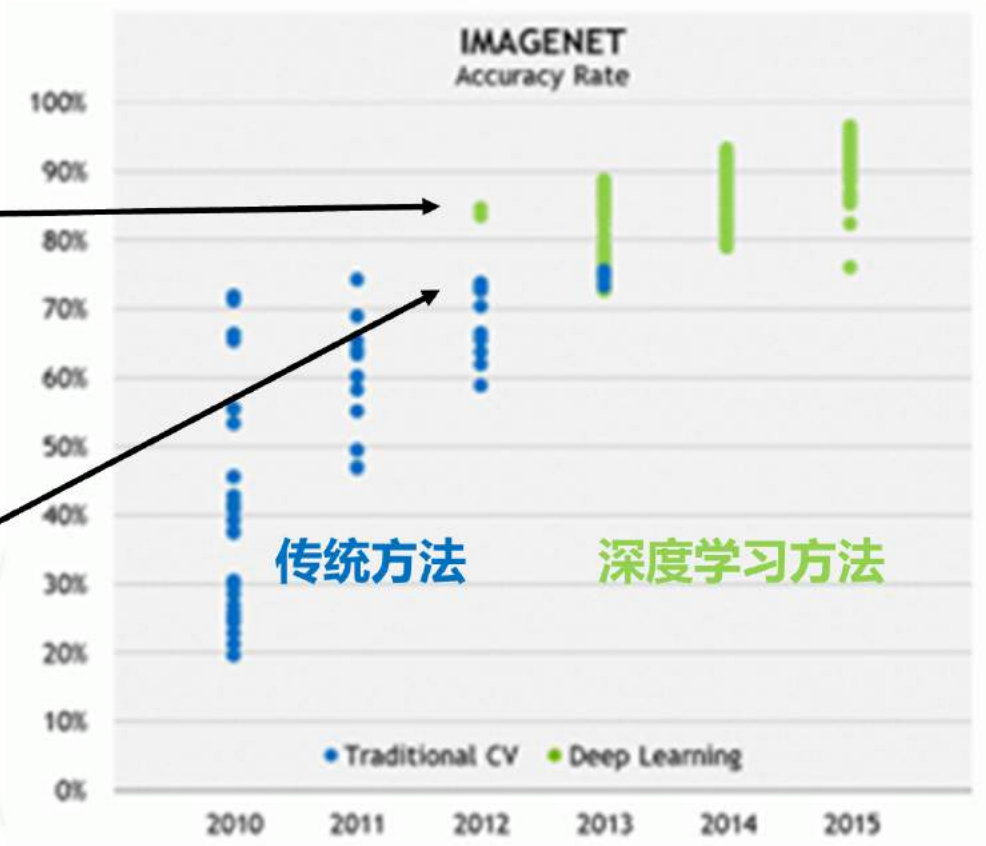
自 2010 年起 ImageNet 举办了一年一度的大规模视觉识别挑战赛 ILSVRC

AlexNet 的诞生 (2012)

AlexNet 模型

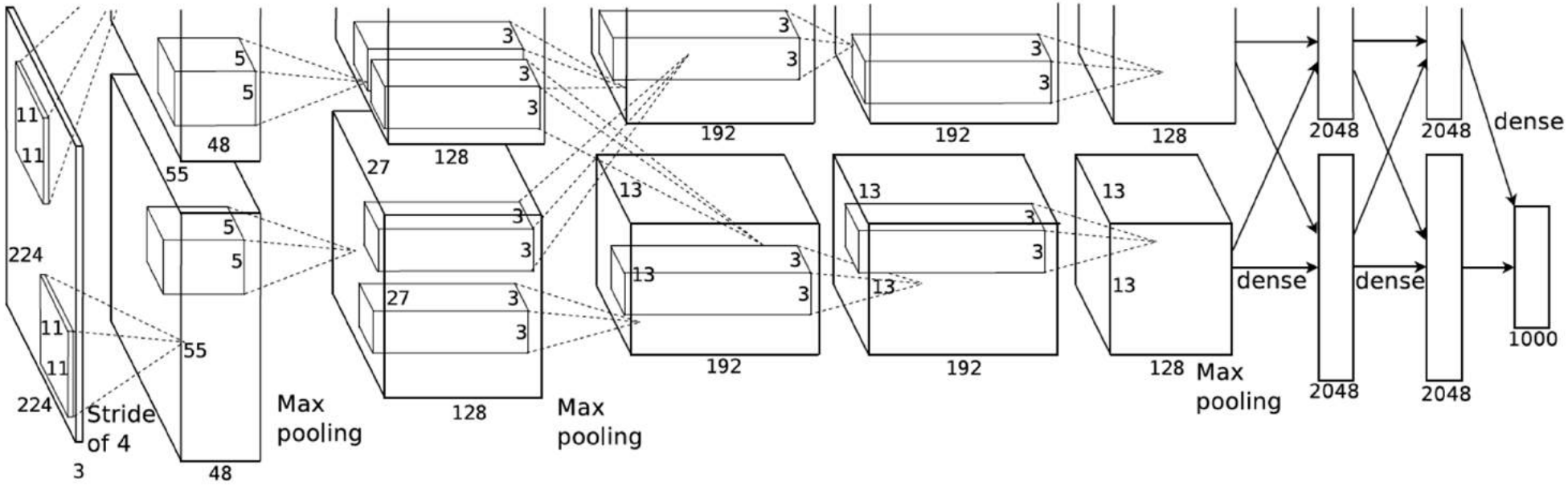






2012 年之前，参赛团队大多使用传统视觉方法。2012 年，深度学习模型 AlexNet 一举领先传统方法 10 个百分点，让人们意识到了深度学习的巨大潜力。相比之下，传统方法已经达到性能瓶颈

AlexNet 的诞生 (2012)



AlexNet 是一个 8 层的卷积神经网络，包含 60M 个参数

在 2 块 GPU 上训练了一周

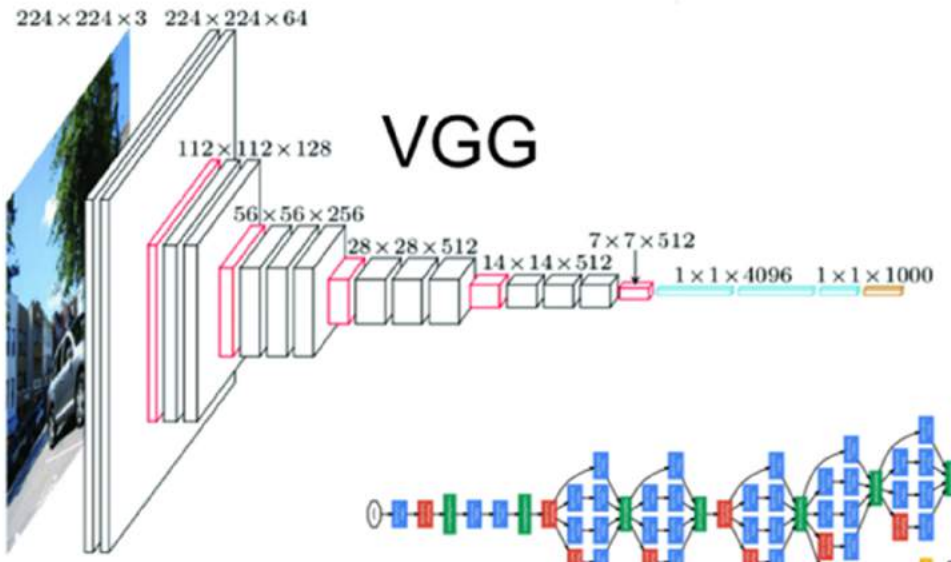
网络结构的演进 (2012~2017+)

更深的网络

AlexNet (2012)：8 层

VGGNet (2014)：11~19 层

GoogLeNet (2014)：22 层
使用不同尺度的卷积核

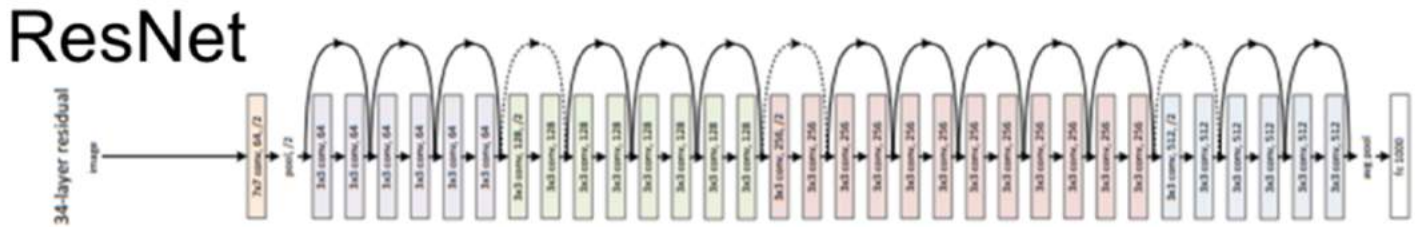


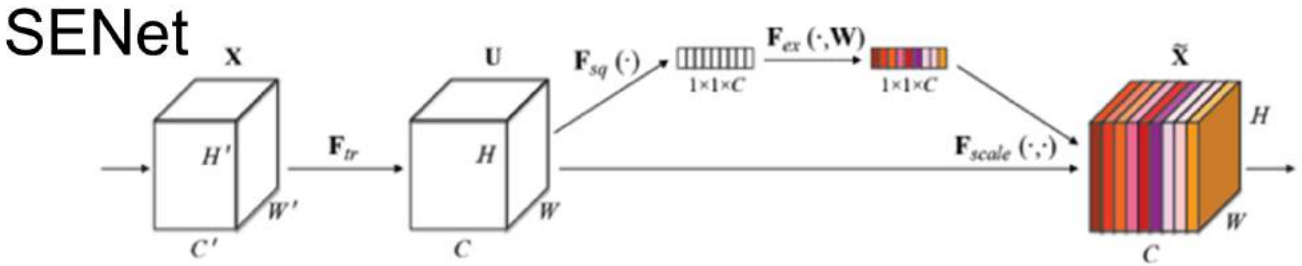
更复杂的连接方式
更多样化的模块设计

ResNet (2015)：18~152 层
引入跨层连接，实现超深网络

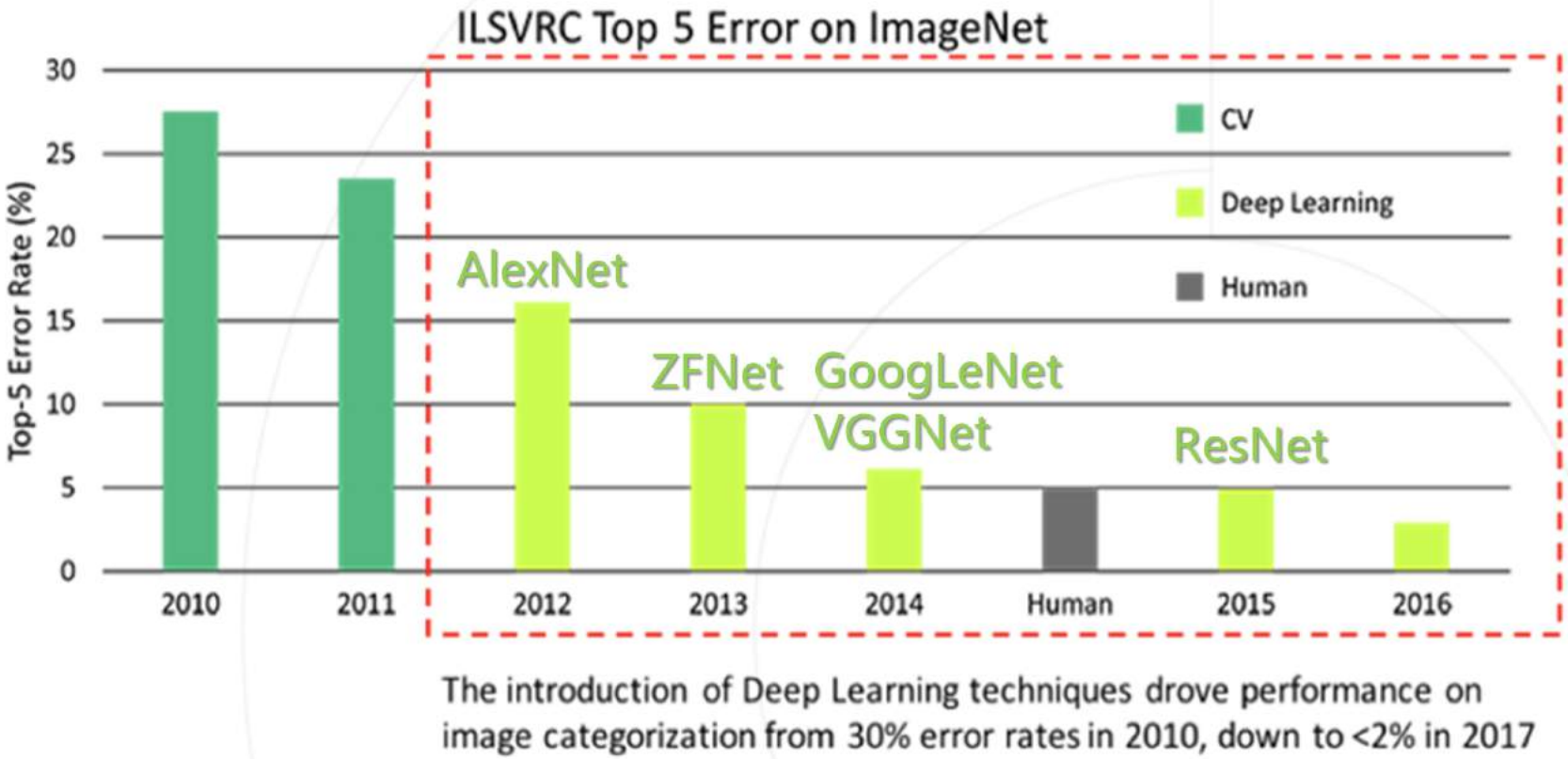
DenseNet (2017)
更复杂的跨层连接

SENet (2018)
加入注意力机制



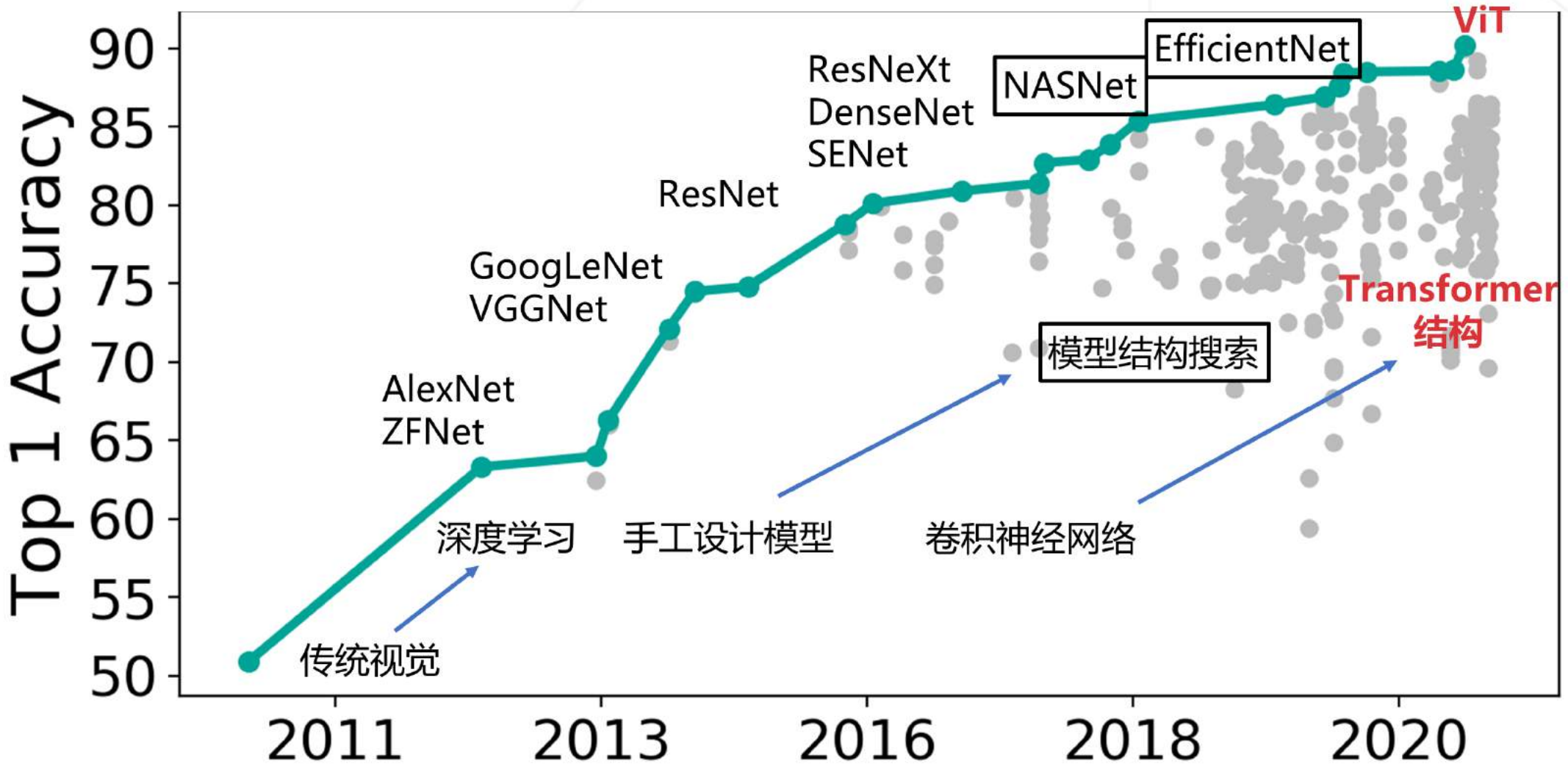


网络结构的演进 (2012~2017+)



随着网络结构不断演进，ImageNet 的分类精度最终超过人类识别精度，到达 2% 以下。最终，ILSVRC 分类比赛于 2017 年停办。

新的趋势 (2017~)


Top 1 Accuracy
90
85
80
75
70
65
60
55
50
2011
2013
2016
2018
2020
AlexNet
ZFNet
GoogLeNet
VGGNet
ResNet
ResNeXt
DenseNet
SENet
NASNet
EfficientNet
ViT
Transformer
结构
模型结构搜索
传统视觉
深度学习
手工设计模型
卷积神经网络

一站式图像分类工具包 MMClassification

OpenMMLab

MMClassification

丰富的模型

VGG VGG-BN	ResNet ResNet V1D	ResNeXt ShuffleNet	SE-ResNet ShuffleNet V2 MobileNet V2	ResNeSt	ViT
2014	2015	2017	2018	2020	2021

丰富的数据集支持

- CIFAR10
- ImageNet
- ……

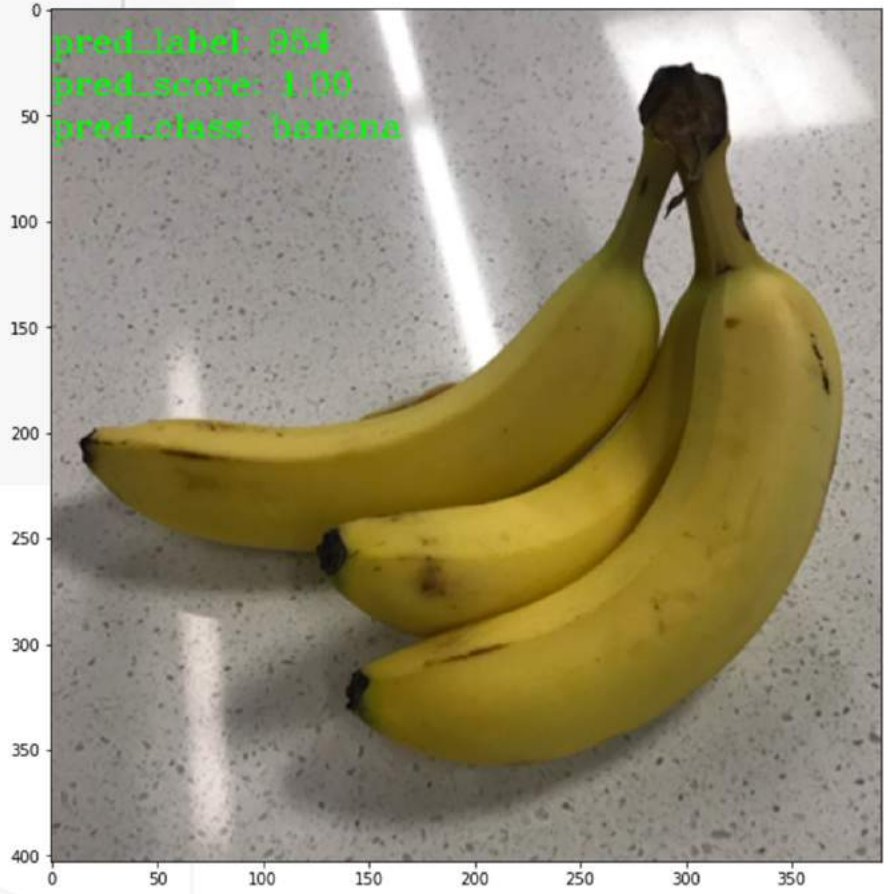
丰富的训练技巧与策略

- 数据增强策略
- 优化器和学习率策略

便捷的 API

```python
from mmcls.apis import inference_model, init_model, show_result_pyplot

# Specify the path to config file and checkpoint file
config_file = 'configs/mobilenet_v2/mobilenet_v2_b32x8_imagenet.py'
checkpoint_file = 'checkpoints/mobilenet_v2_batch256_imagenet_20200708-3b2dc3af.pth'
# Specify the device. You may also use cpu by `device='cpu'`.
device = 'cuda:0'
# Build the model from a config file and a checkpoint file
model = init_model(config_file, checkpoint_file, device=device)
# Test a single image
img = 'demo/banana.png'
result = inference_model(model, img)
# Show the results
show_result_pyplot(model, img, result)
```




```
{'pred_class': 'banana',
 'pred_label': 954,
 'pred_score': 0.9999284744262695}
```

统一的模型与算法配置

配置模型

配置数据集

配置训练算法

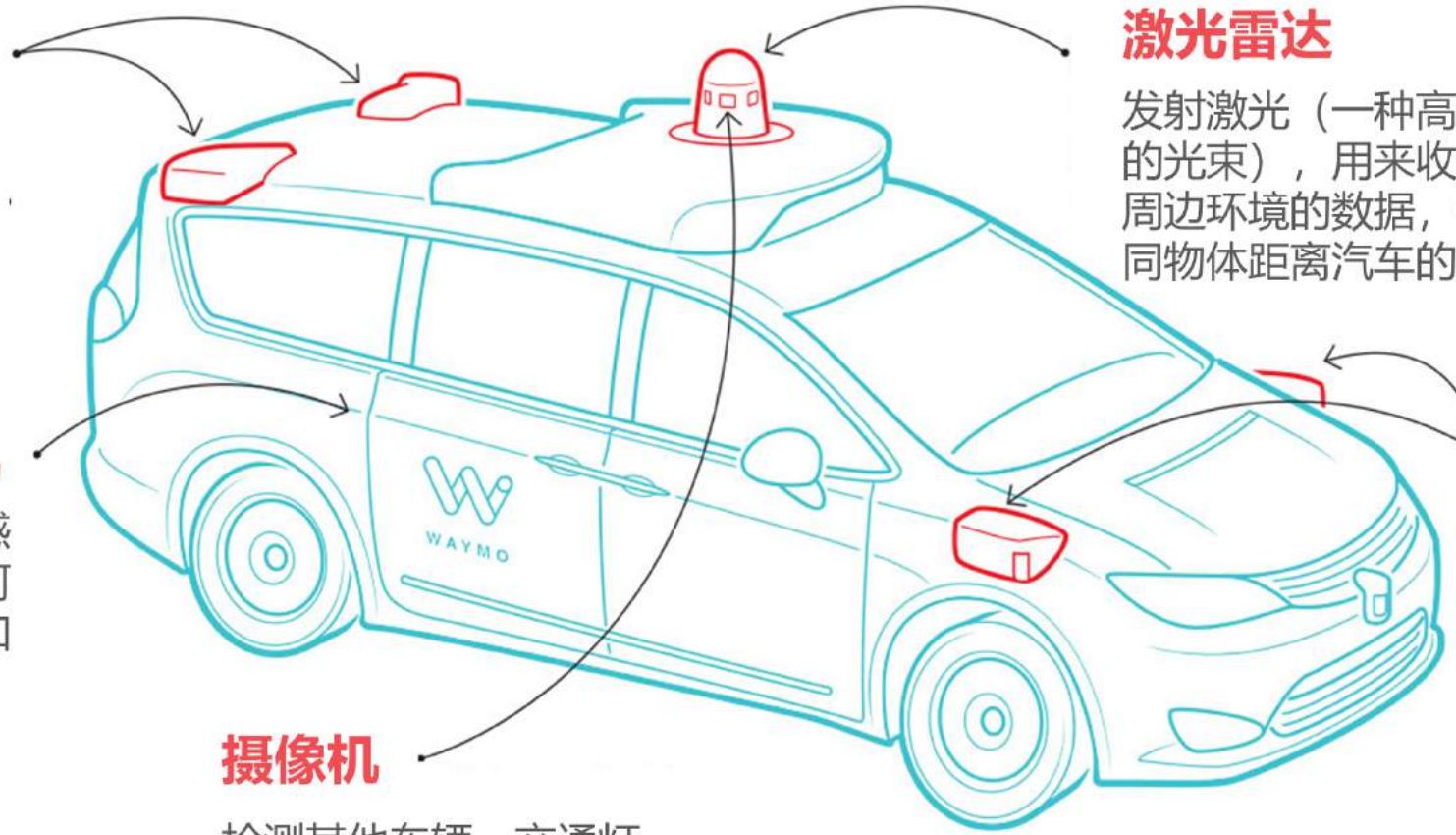
```python
# model settings
model = dict(
    type='ImageClassifier',
    backbone=dict(type='LeNet5', num_classes=10),
    neck=None,
    head=dict(
        type='ClsHead',
        loss=dict(type='CrossEntropyLoss', loss_weight=1.0),
    ))
# dataset settings
dataset_type = 'MNIST'
data = dict(
    samples_per_gpu=128,
    workers_per_gpu=2,
    train=dict(
        type=dataset_type, data_prefix='data/mnist'),
    val=dict(
        type=dataset_type, data_prefix='data/mnist'))
# optimizer
optimizer = dict(type='SGD', lr=0.01, momentum=0.9)
runner = dict(type='EpochBasedRunner', max_epochs=5)
```

雷达

发射无线电波并接收周围物体反射的回波，判断物体与汽车的距离远近；

计算机

位于车内；分析来自传感器的数据，以此判断作何回应，例如转向、变速和刹车；



激光雷达

发射激光（一种高度密集的光束），用来收集汽车周边环境的数据，例如不同物体距离汽车的距离；

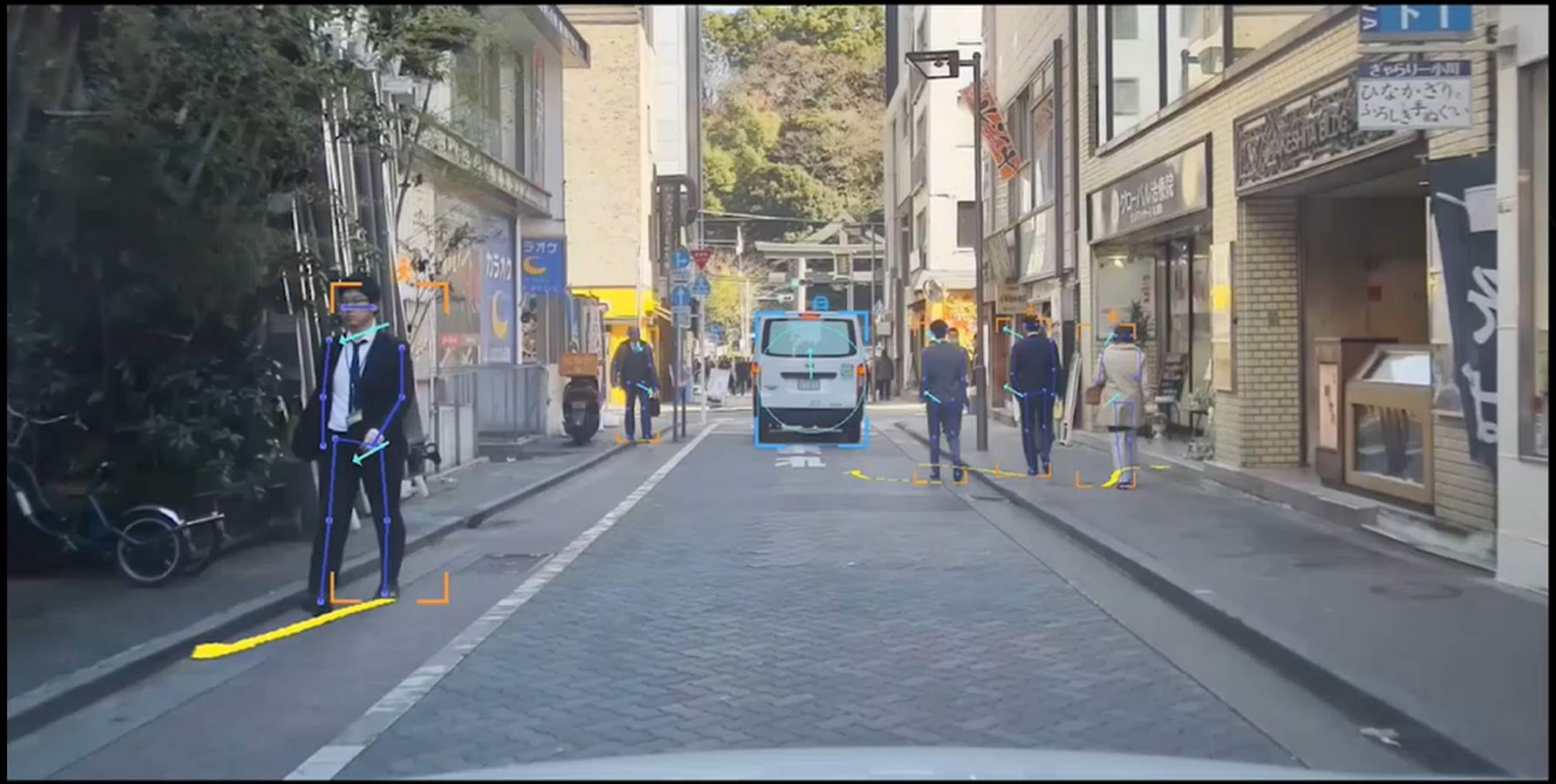
雷达以及激光雷达

摄像机

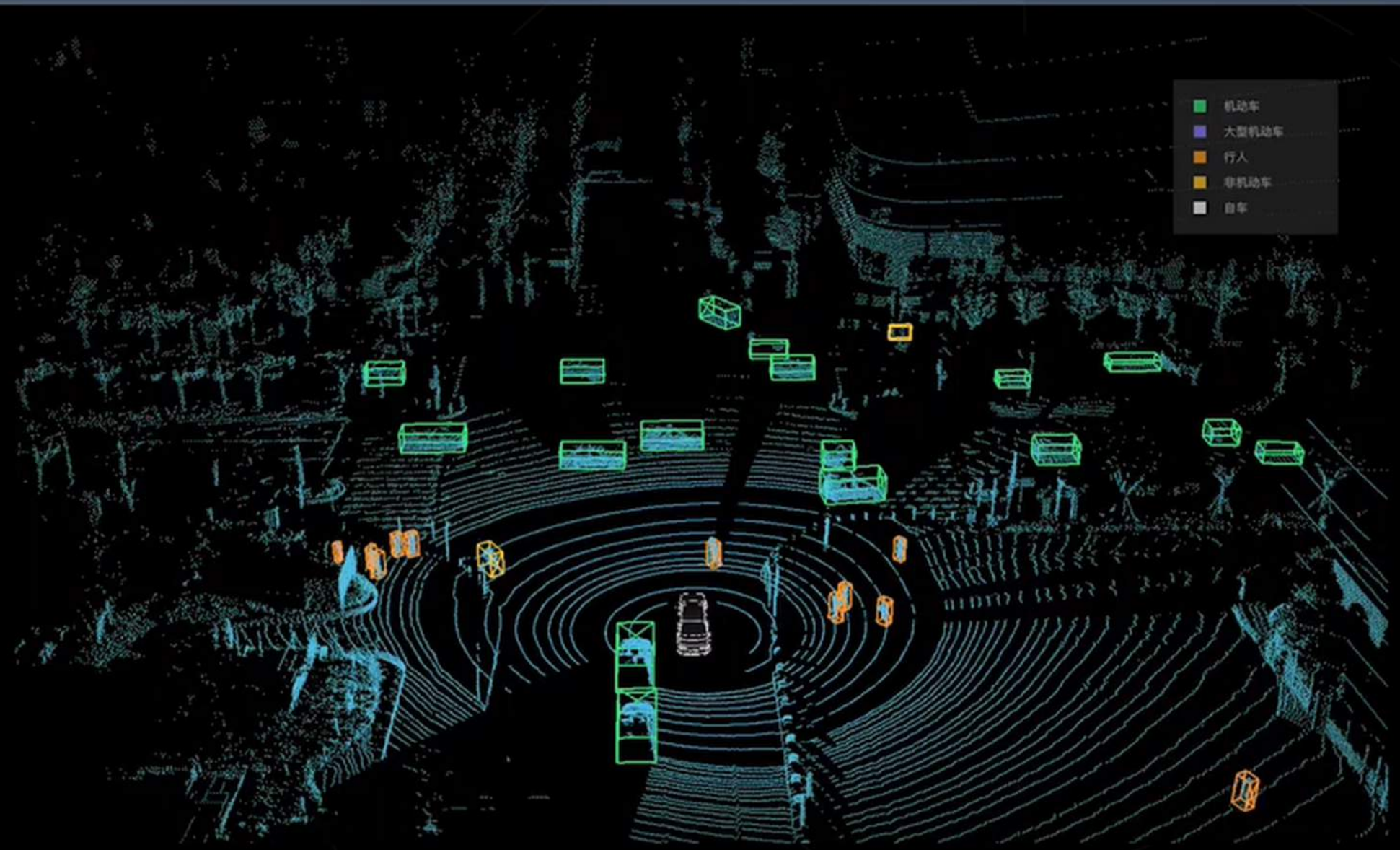
检测其他车辆、交通灯、指示牌、行人以及自行车；

现代的汽车配置了大量的传感器，以感知周围的环境

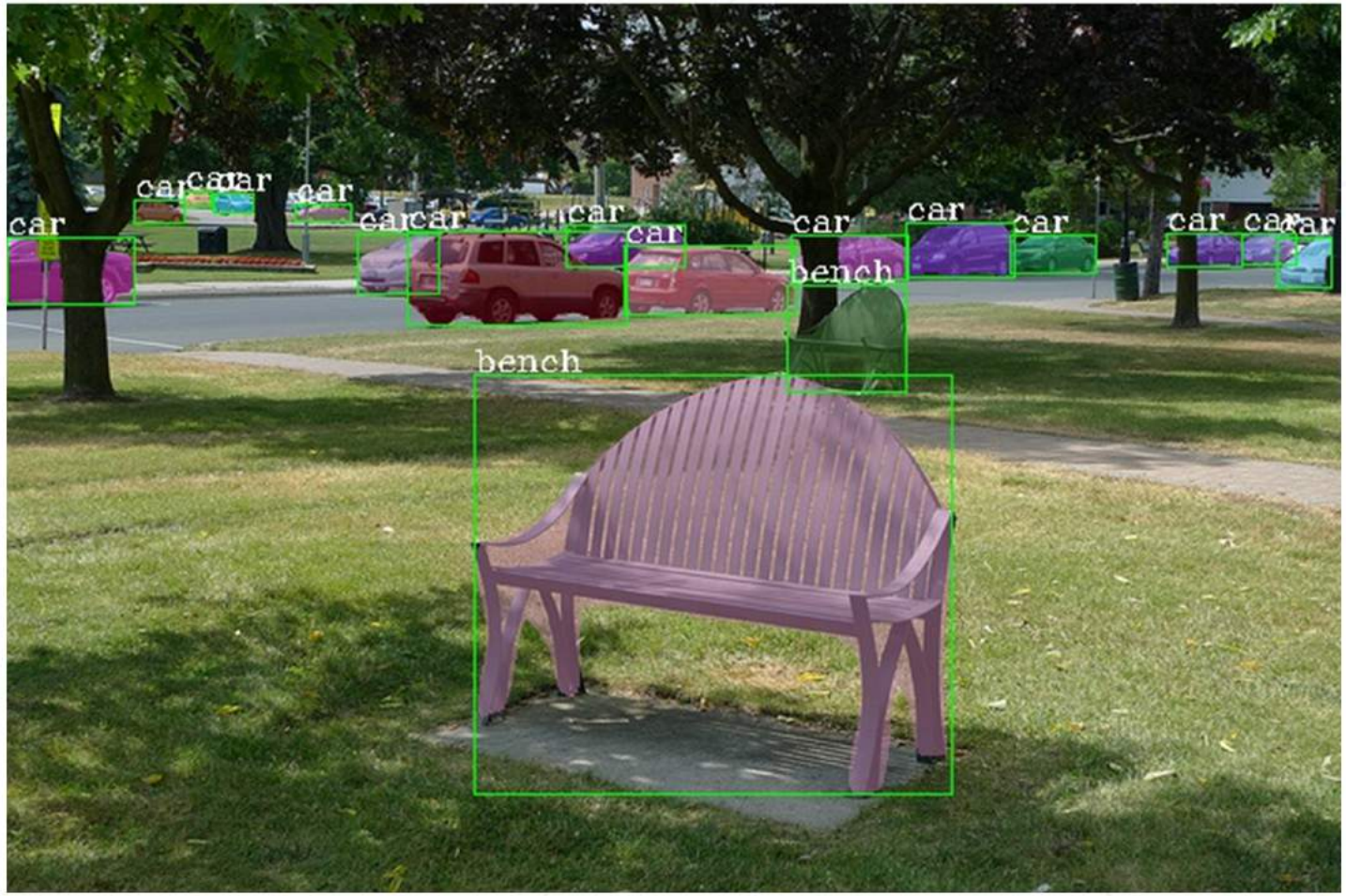
视觉感知



激光雷达感知



目标检测：What & Where



任务：

识别图像中出现的物体，分类并定位；

实例分割：

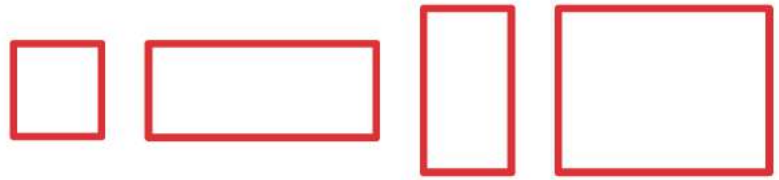
用矩形框标出物体；

目标检测：

精确定位物体占据的每个像素。

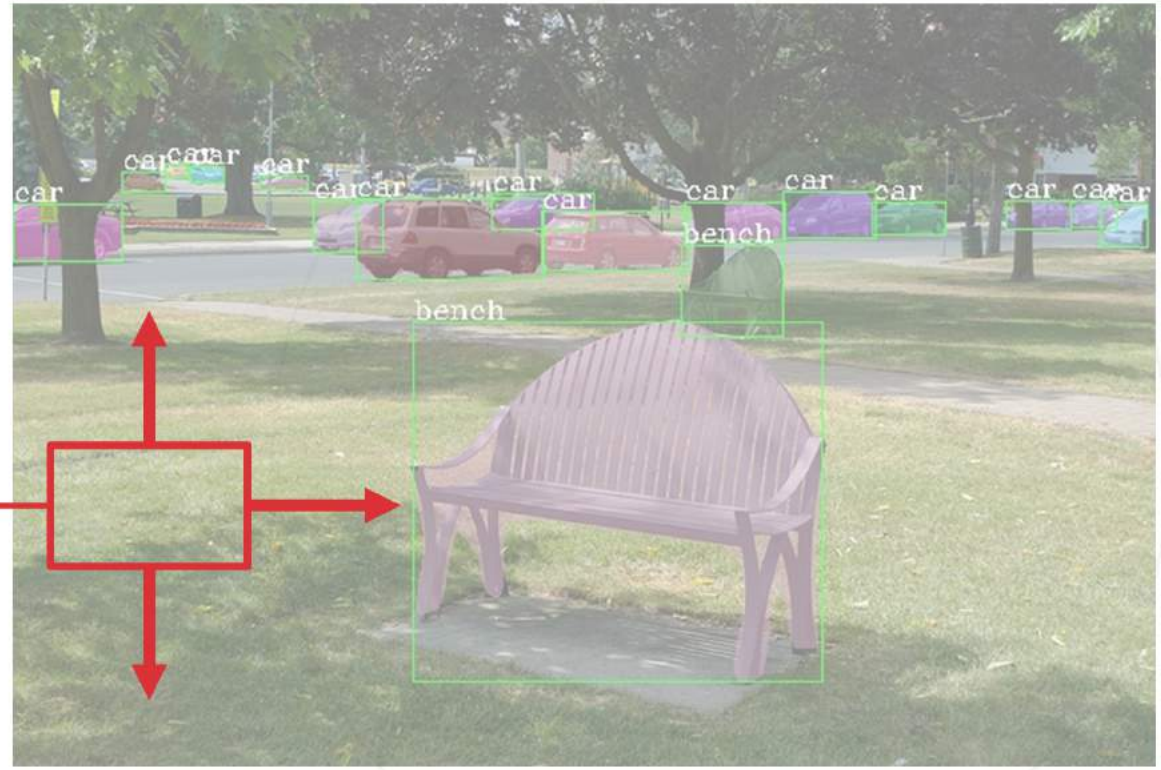
基本思想：滑窗检测 (sliding windows)

1. 设定一些列大小形状不同的窗口



2. 用这些窗口扫过整张图片

3. 对于每一个区域用图像分类模型识别其中的物体



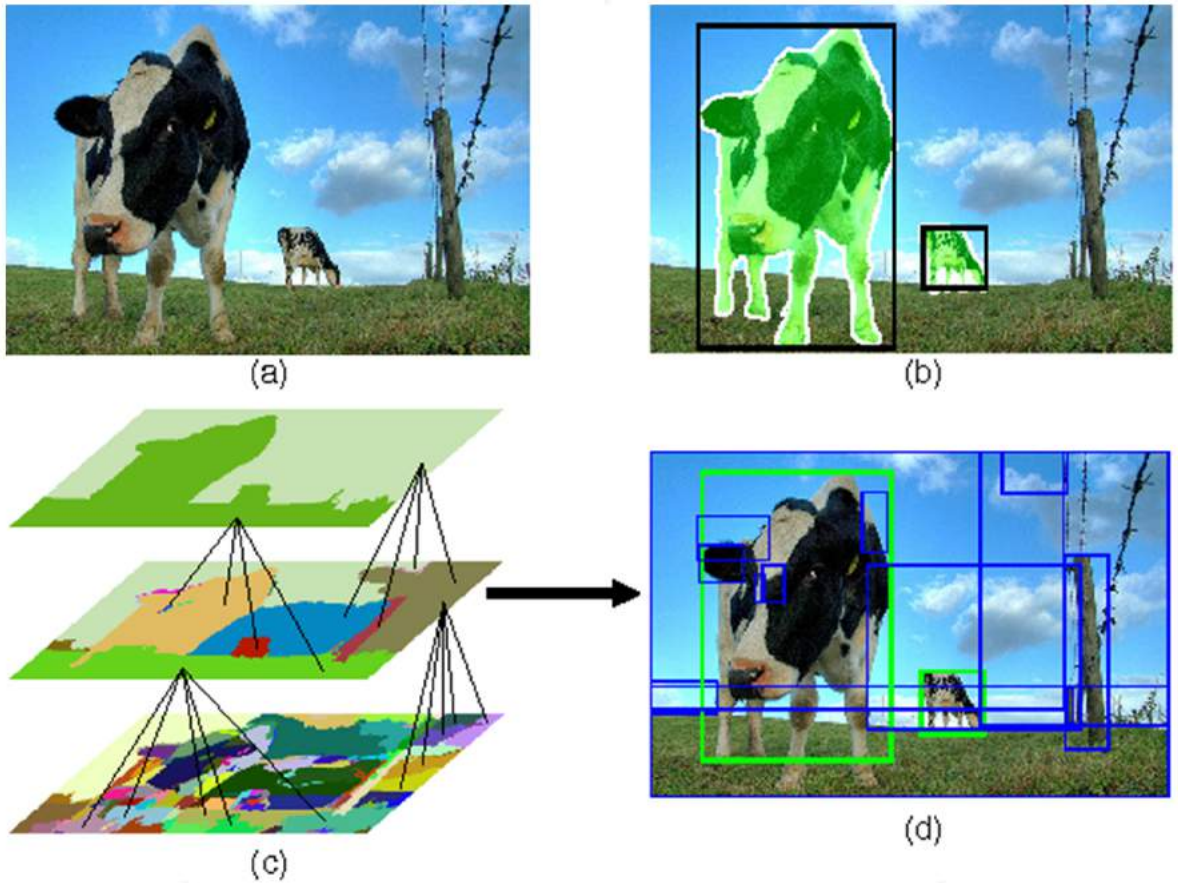
优点：准确不易遗漏

缺点：窗口数目巨大，计算量巨大

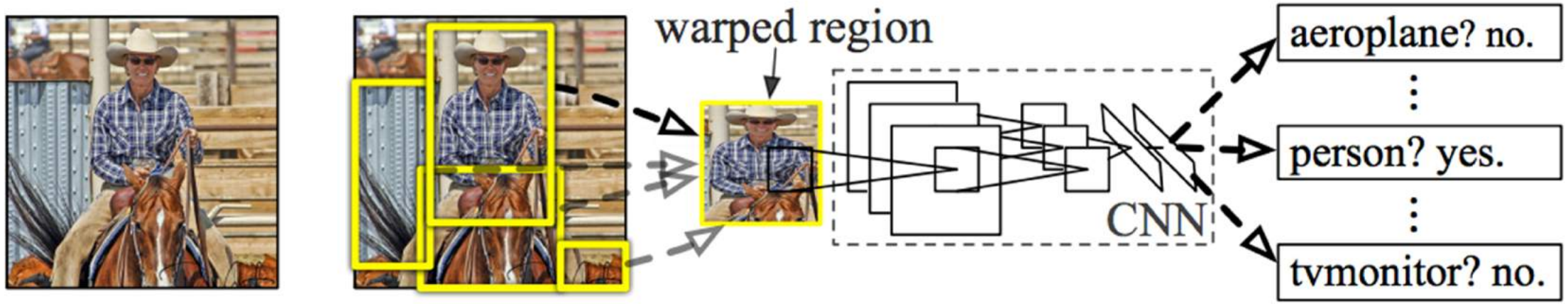
基本思想：区域提议 (region proposal)

基于局部颜色或图像特征，一次性找出所有可能含有物体的区域。

- 优点：快
- 缺点：有可能遗漏



基于区域的方法 RCNN (2013)

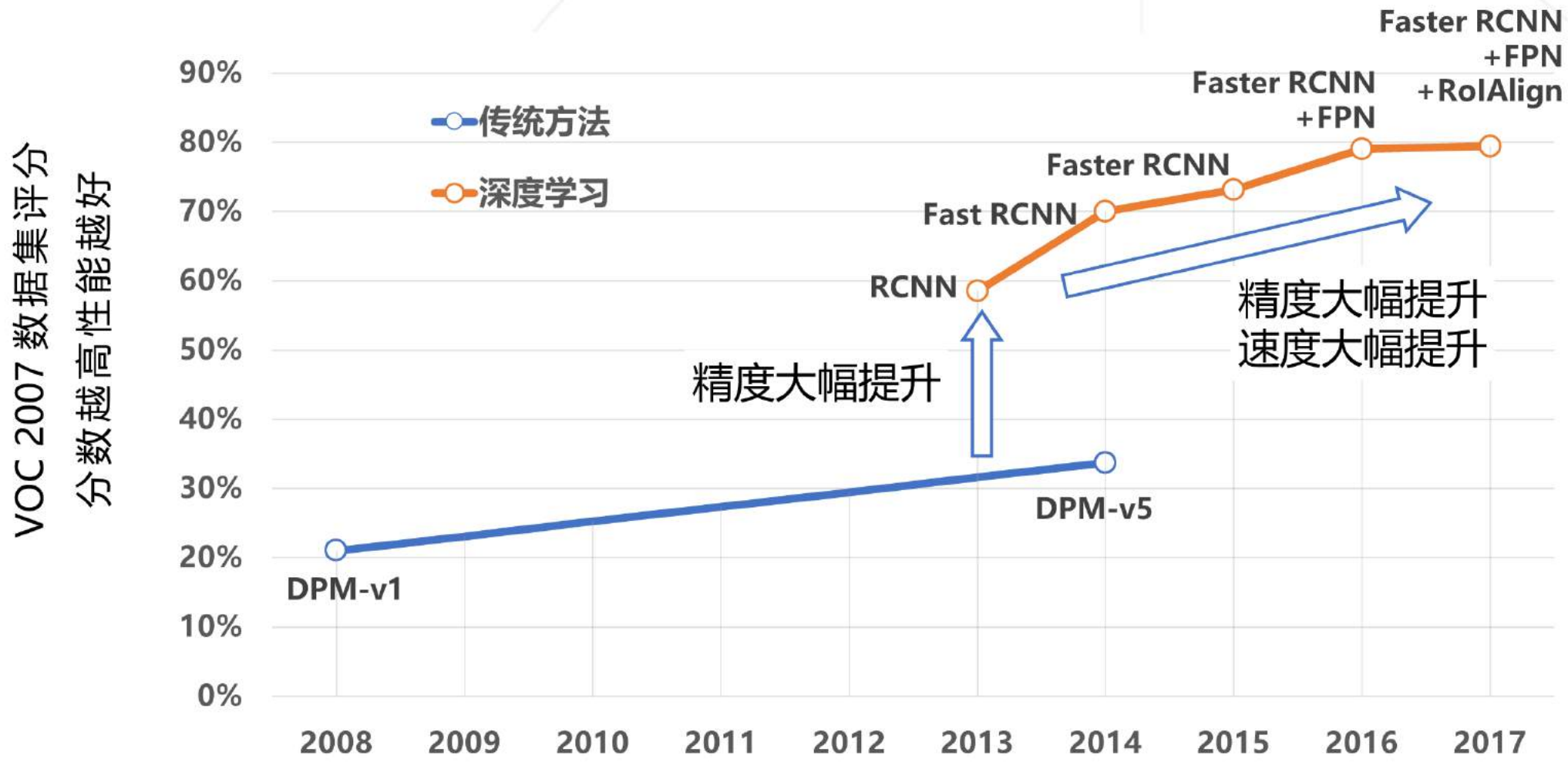


输入图片　　区域提议　　对于每个区域，使用卷积网络预测物体类别

快但是**不准确**　　**慢**但是**准确**

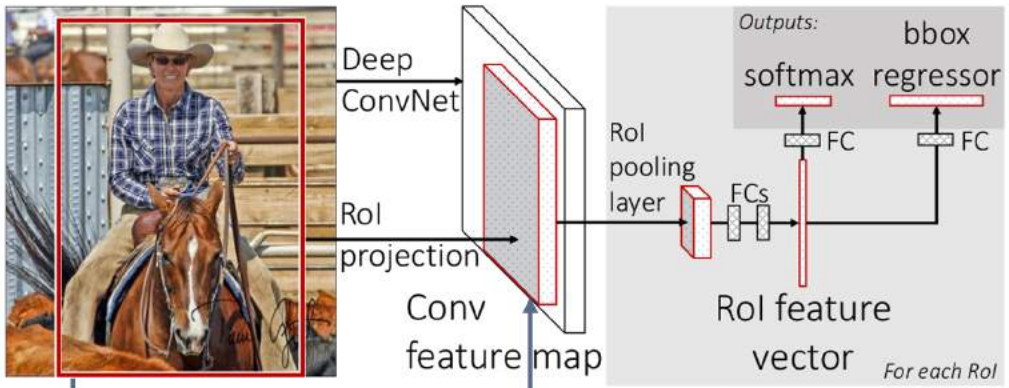
基于区域的方法包含两个步骤，因此也称为两阶段方法

两阶段方法的演进 (2014 - 2017)



两阶段方法的演进 (2014 - 2017)

Fast RCNN 2014

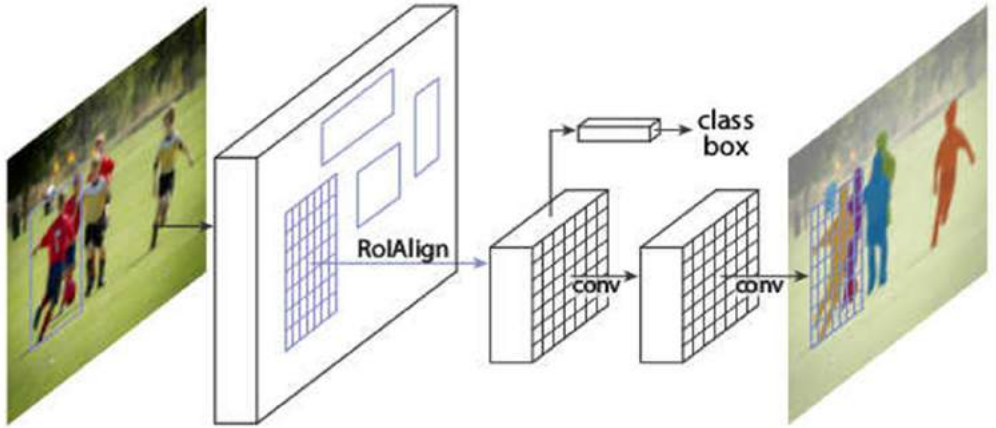


提出了 RoIPooling 方法，把区域从图像移动到特征图上，大幅降低了计算量。

Faster RCNN 2015

提出了 RPN 网络，用于替换传统方法，产生区域提议，进一步提高效率。

Mask RCNN 2017



提出了 ROIAlign 算法升级替换 ROIPooling。

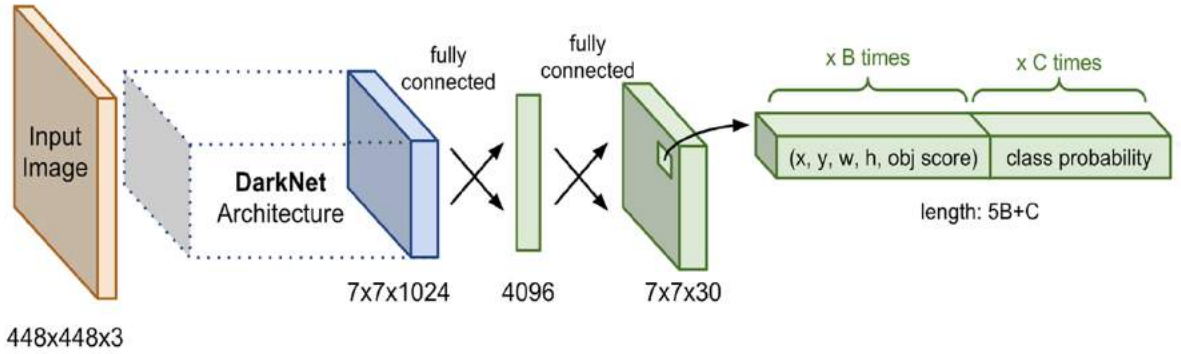
加入 FCN 分支用于实例分割。

一阶段方法 (2015~)

在两阶段算法发展的同时，有另一类算法被发明出来。这类算法不依赖区域提议，直接基于特征图进行预测。因为不需要两阶段方法中的区域提议，这类算法被称为一阶段算法。

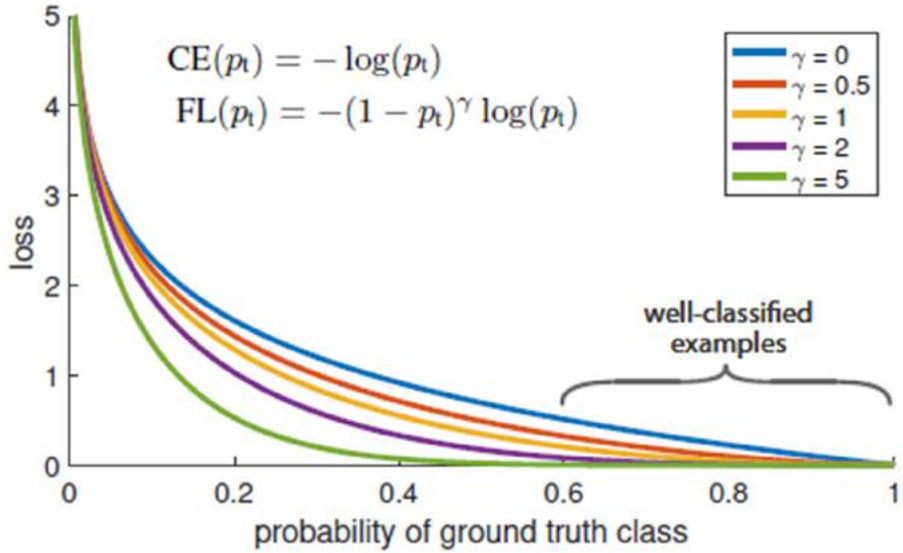
YOLO=
You Only Look Once

YOLO 系列
2015 ~ 2020



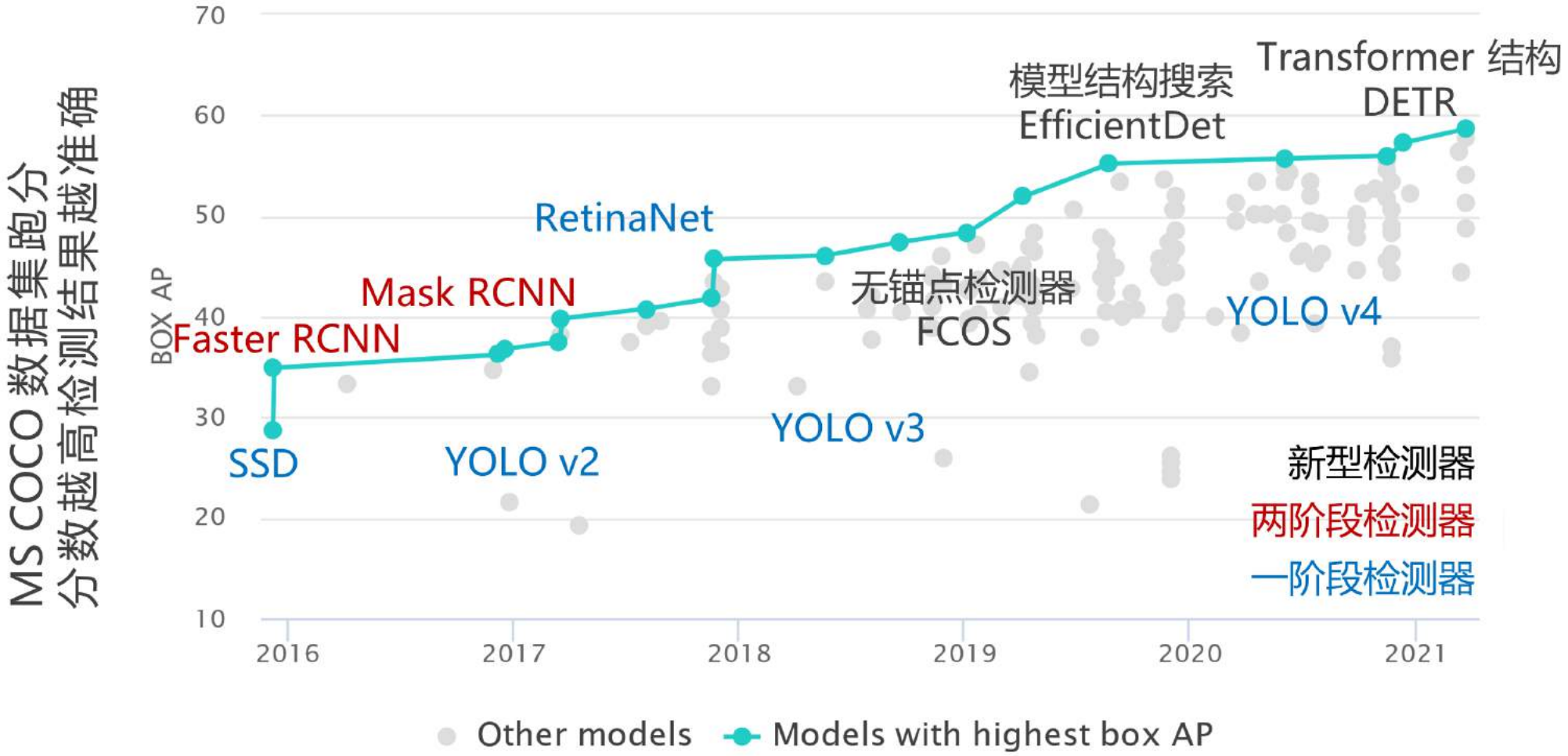
简单而快速，在工业界有广泛的应用，持续改进到 v4 版本。

RetinaNet
2017



提出了Focal Loss，解决了一阶段检测器的难题——正负样本不均。

新的趋势 (2018~)


MS COCO 数据集跑分
分数越高检测结果越准确
BOX AP
70
60
50
40
30
20
10
2016
2017
2018
2019
2020
2021
Faster RCNN
Mask RCNN
RetinaNet
SSD
YOLO v2
YOLO v3
无锚点检测器
FCOS
模型结构搜索
EfficientDet
YOLO v4
Transformer 结构
DETR
新型检测器
两阶段检测器
一阶段检测器
Other models
Models with highest box AP

一站式目标检测工具包 MMDetection

OpenMMLab

- 2018-10 发布
- 2019-07 v1.0
- 2020-05 v2.0

任务支持

- 目标检测
- 实例分割

覆盖广泛

- 375 个预训练模型
- 58 篇论文复现
- 常用学术数据集

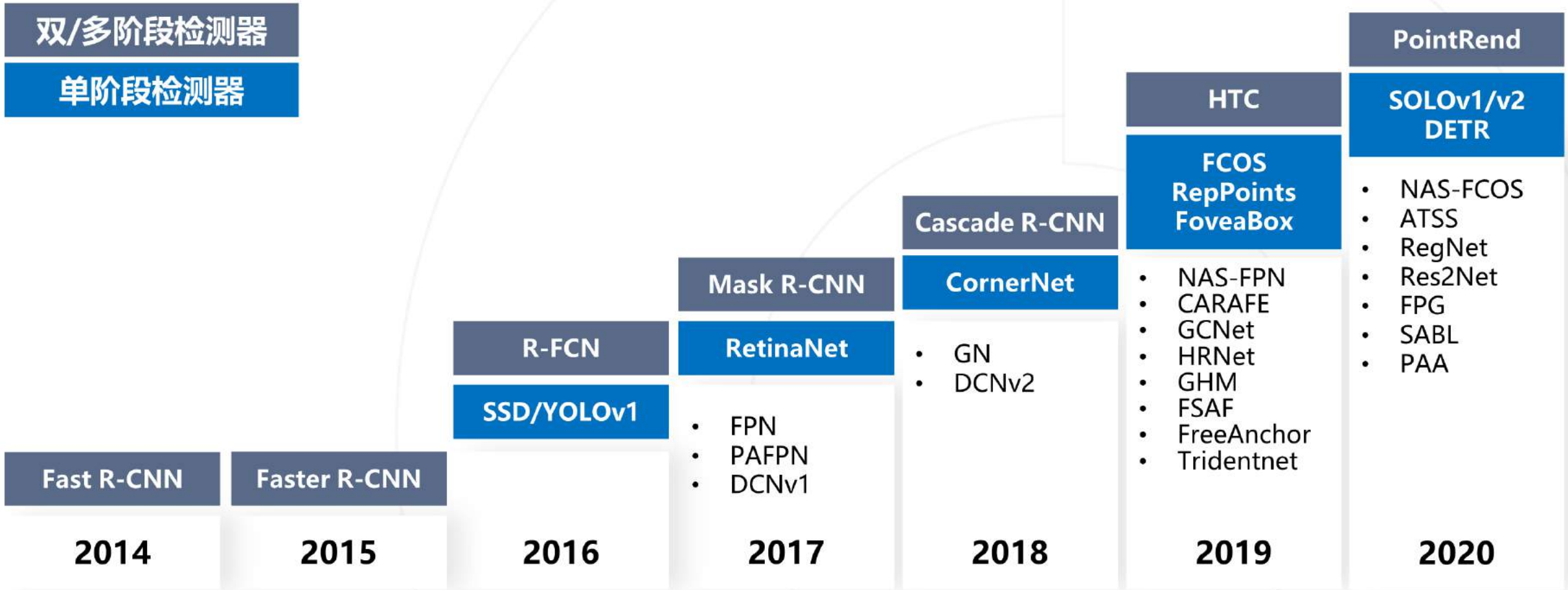
算法丰富

- 两阶段检测器
- 一阶段检测器
- 级联检测器
- 无锚点检测器
- Transformer

使用方便

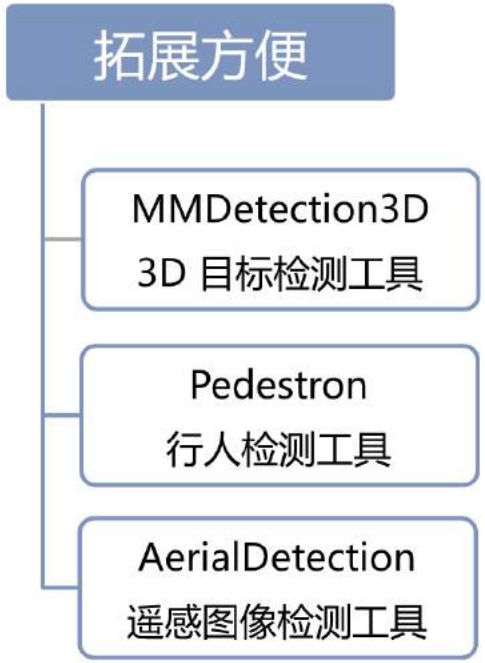
- 训练工具
- 测试工具
- 推理 API

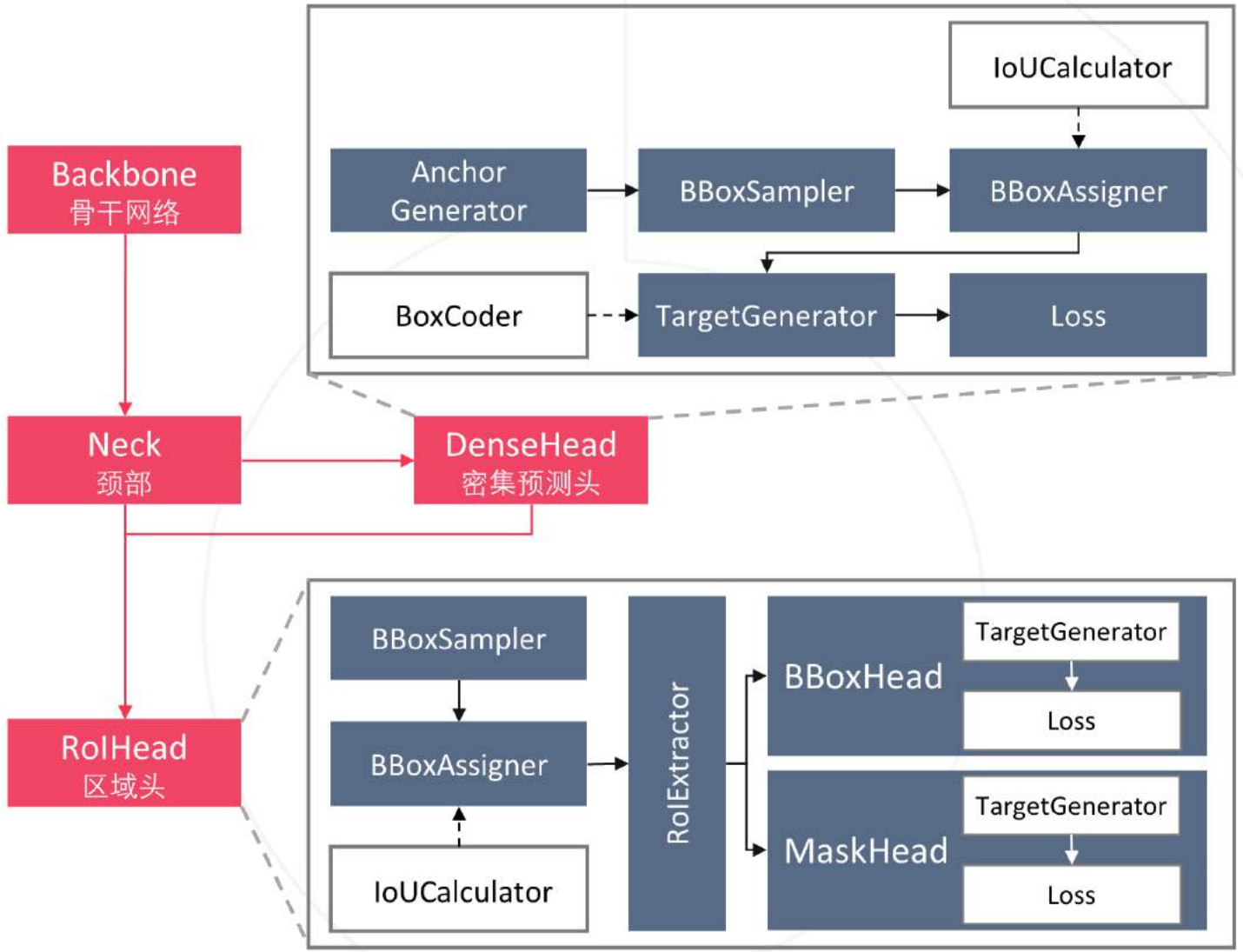
算法丰富


双/多阶段检测器
单阶段检测器
Fast R-CNN
Faster R-CNN
R-FCN
SSD/YOLOv1
Mask R-CNN
RetinaNet
• FPN
• PAFPN
• DCNv1
Cascade R-CNN
CornerNet
• GN
• DCNv2
HTC
FCOS
RepPoints
FoveaBox
• NAS-FPN
• CARAFE
• GCNet
• HRNet
• GHM
• FSAF
• FreeAnchor
• Tridentnet
PointRend
SOLOv1/v2
DETR
• NAS-FCOS
• ATSS
• RegNet
• Res2Net
• FPG
• SABL
• PAA
2014
2015
2016
2017
2018
2019
2020

模块化设计

模块化设计是 OpenMMLab 的灵魂。在 MMDetection 中，我们将不同模型按照功能模块进行分解，方便用户自由组合和拓展。





社区活跃

OpenMMLab

- 170 contributors
- 1300 merged PRs
- 3200 closed issues
- 4900 forks
- 14100 stars

20+学术机构和企业











广泛应用

OpenMMLab

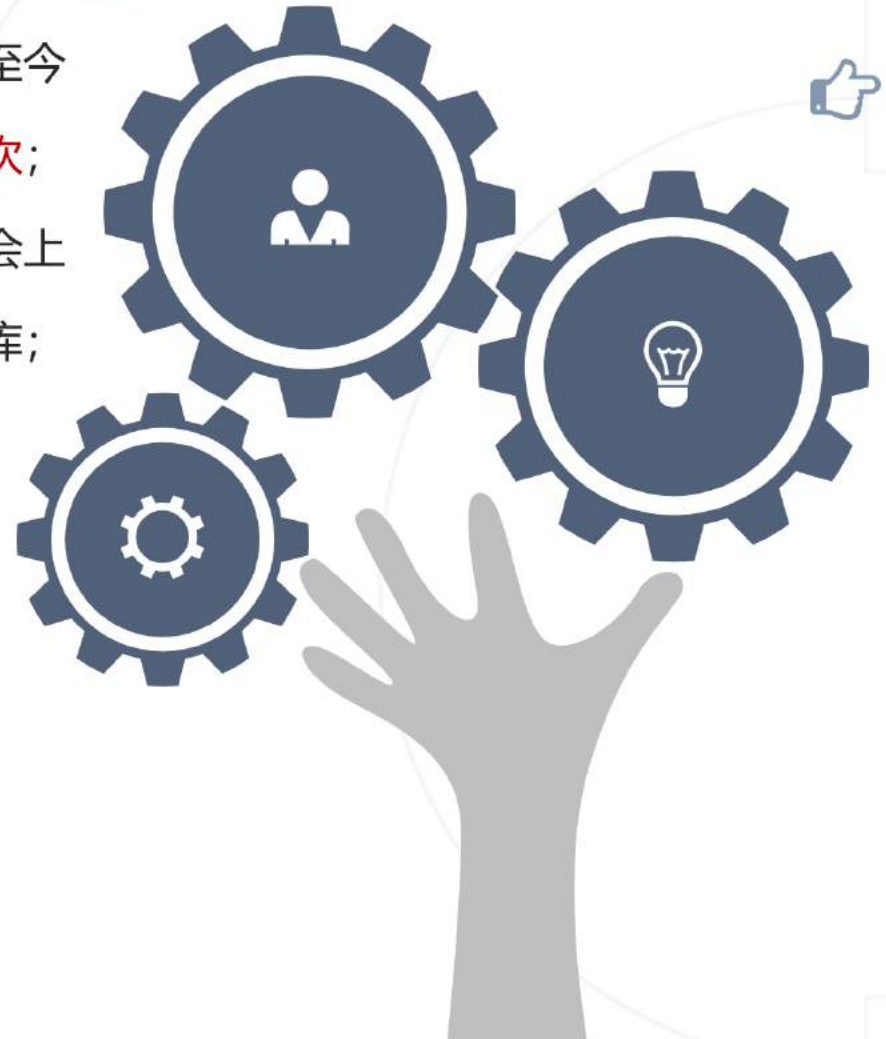
科研论文

2019 年 6 月至今

谷歌学术引用 超过 370 次;

仅计算机视觉三大顶会上

被 超过 50 篇论文 作为基础代码库;



学术比赛

COCO 2018 实例分割**冠军**

COCO 2019 实例分割**冠军**

Open Images 2019 物体检测**冠军**

Global Wheat Detection**冠军**

Crowd Human 人体检测**冠军**

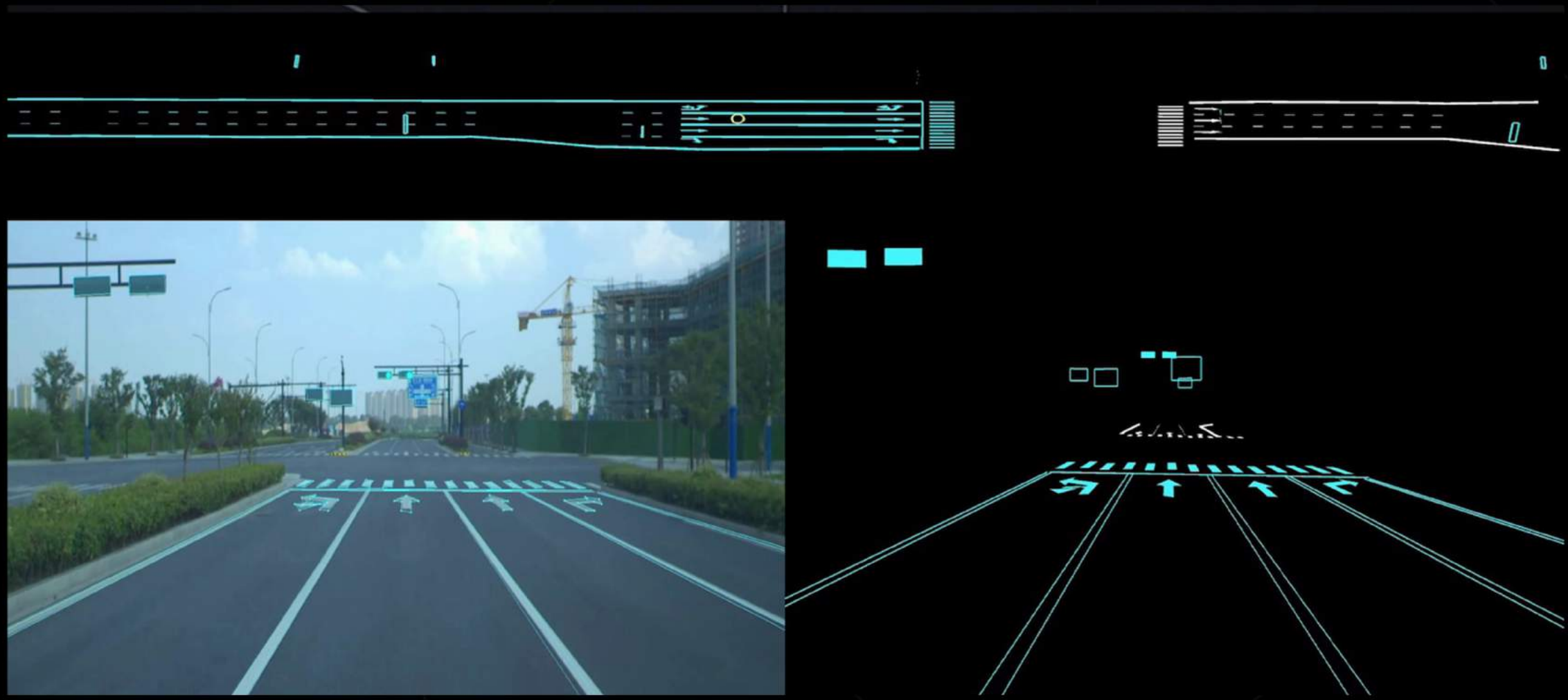
Materialist(FGVC6) 2019**冠军**

工业落地

商汤、腾讯、阿里、华为、

国内外初创公司，……

车道线定位

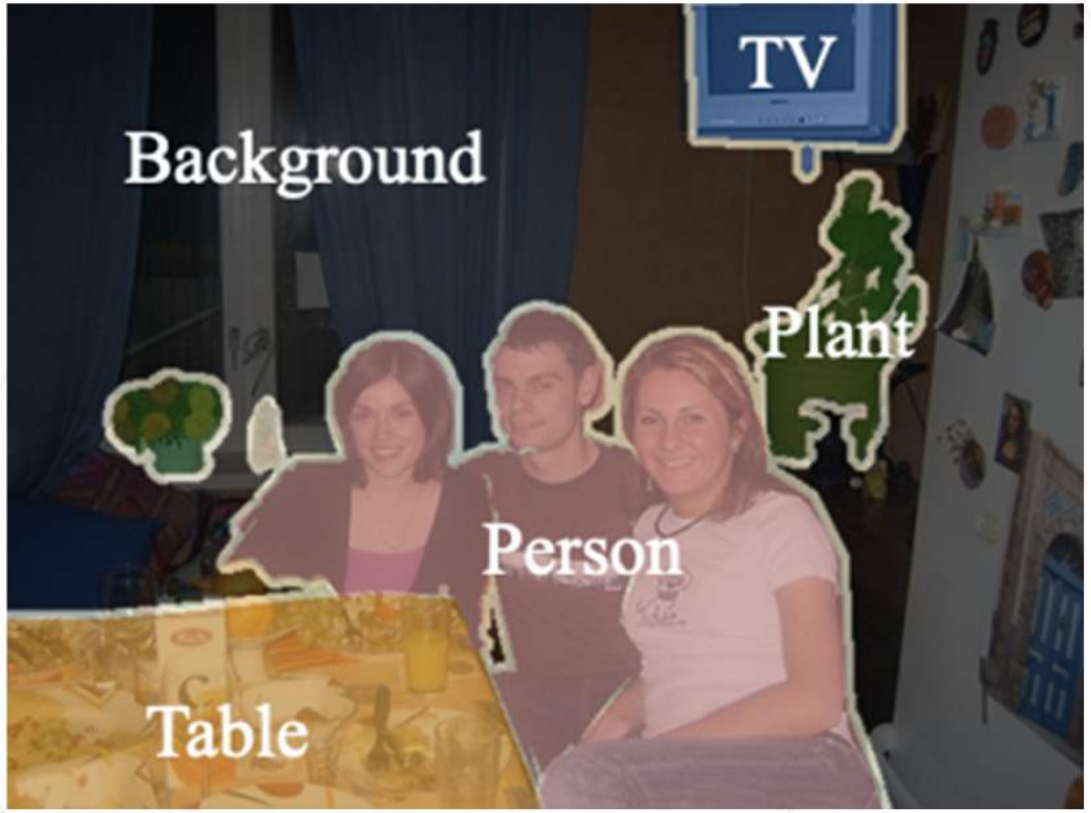


任务

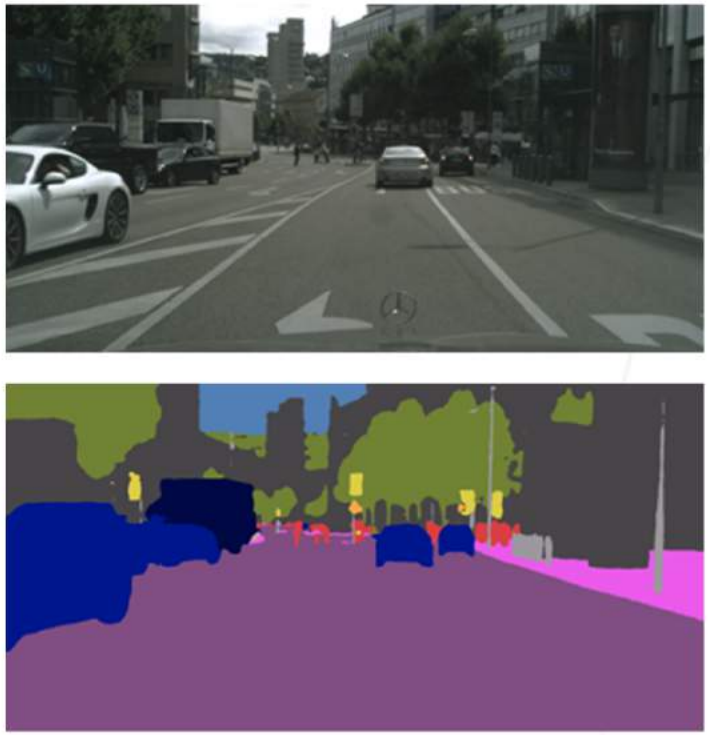
依据图像中出现的物体，将图像分割成不同的区域

等价的理解

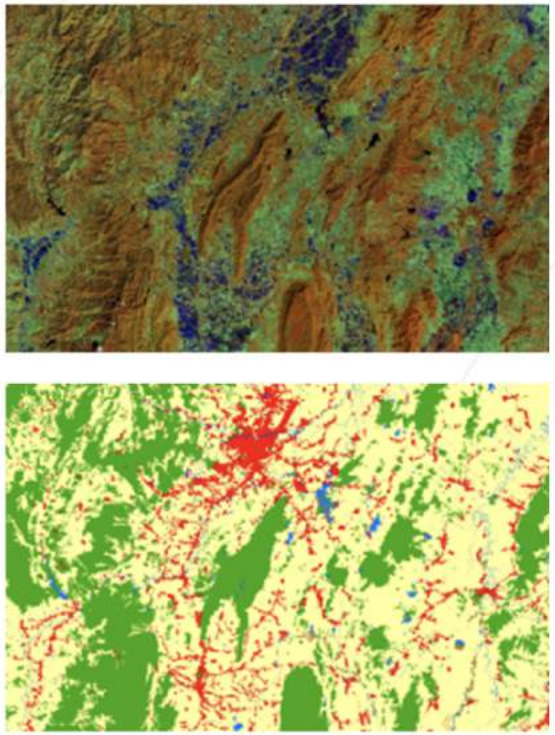
每一个像素进行分类



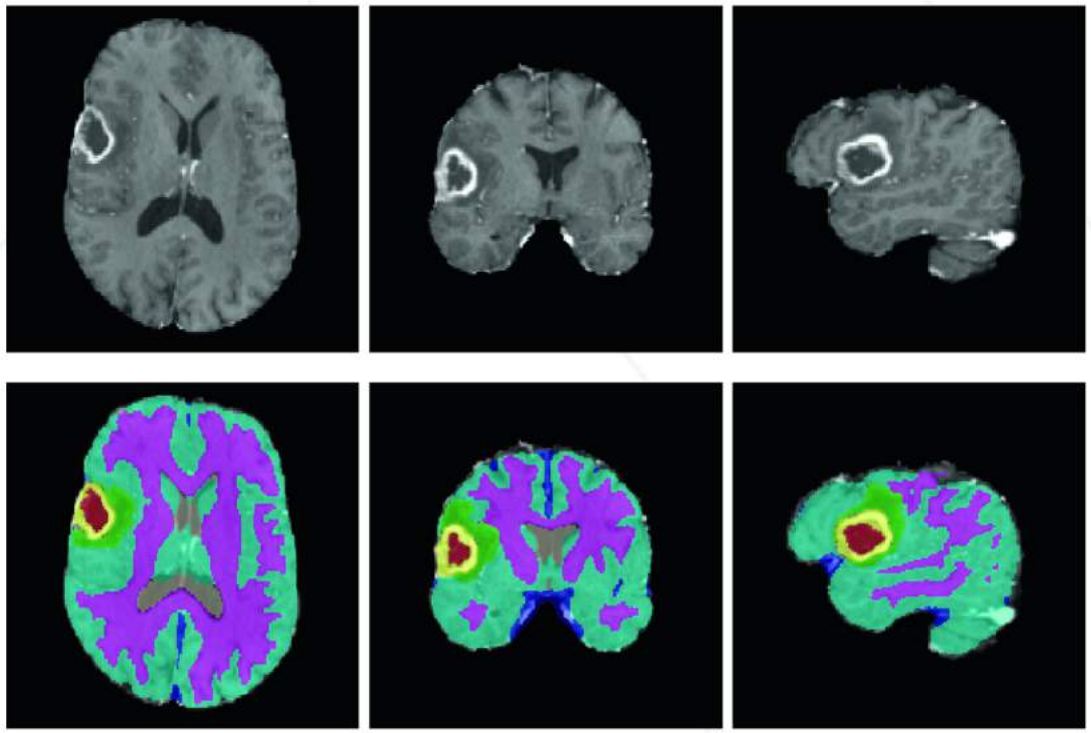
语义分割的应用



无人驾驶汽车



遥感

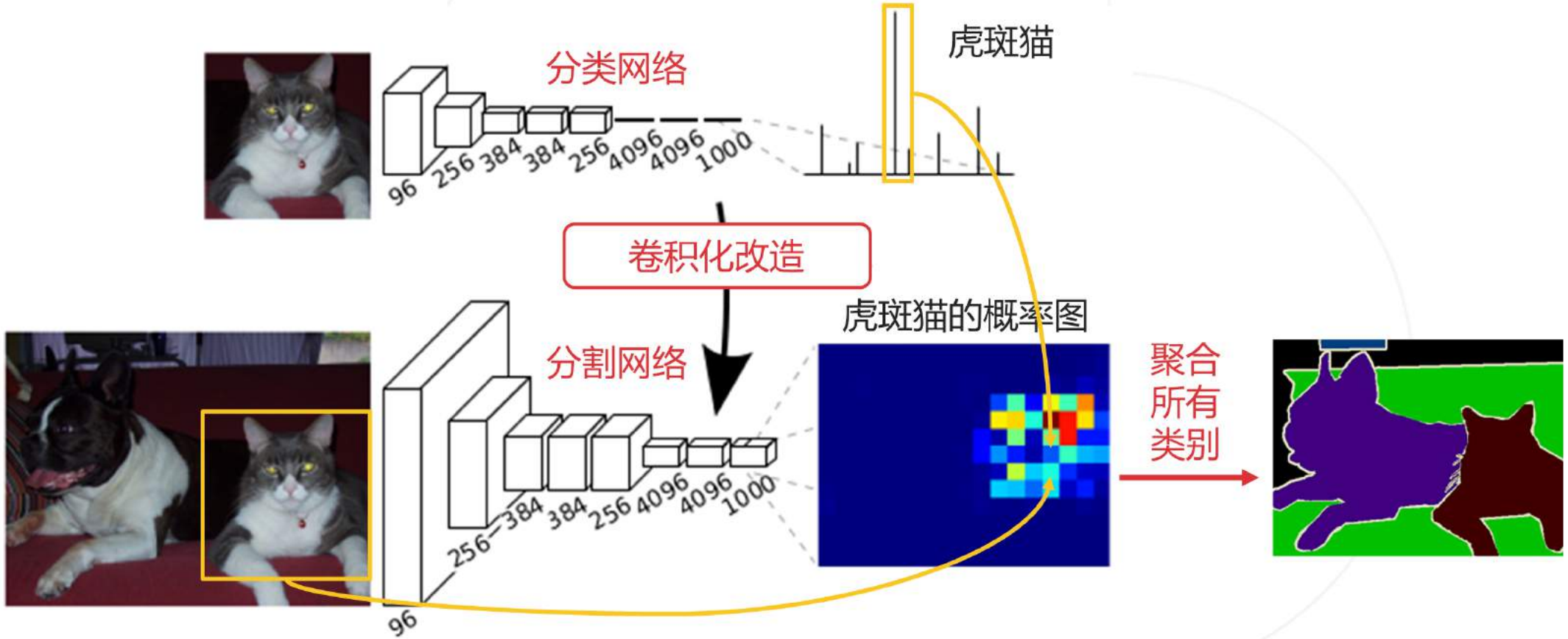


医疗影像分析

深度学习的突破：全卷积网络 FCN (2015)

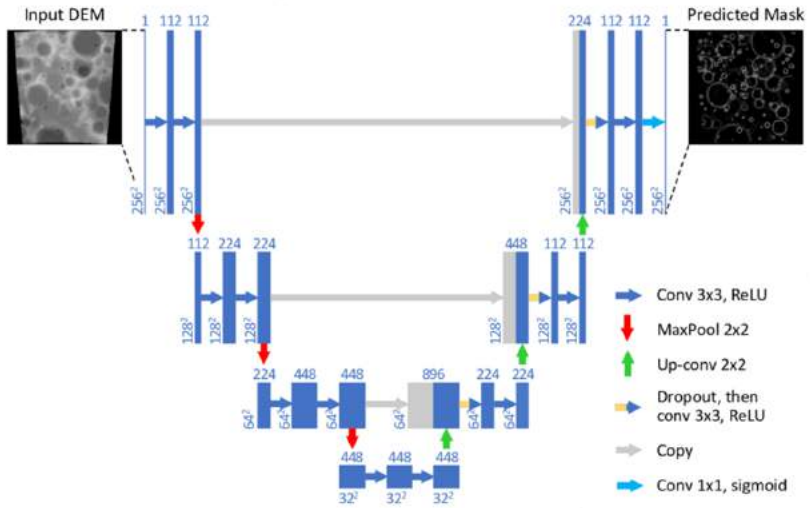
同样是分类问题，充分利用已有的分类网络是一个明智的策略。可以基于滑窗，但需要解决计算量问题。

FCN – 改造分类网络，实现卷积计算在滑窗之间的共享。



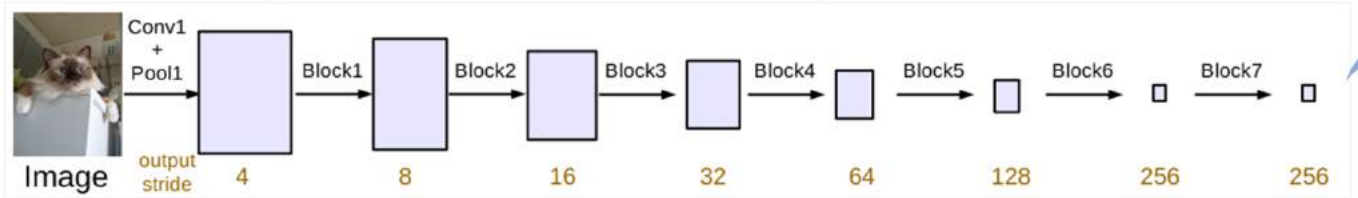
分割模型的进化 (2015~)

UNet 2015

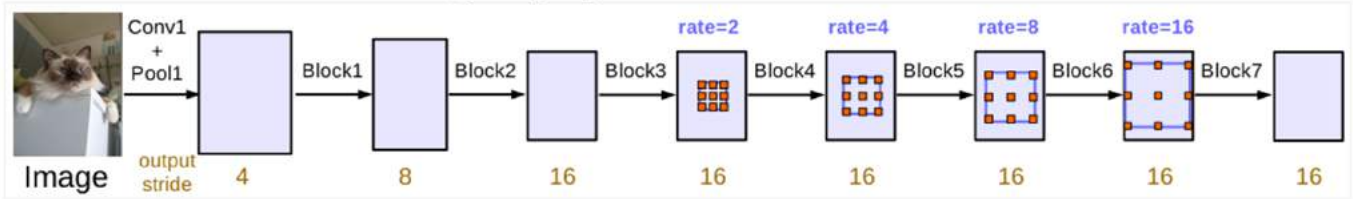


将 FCN 改造为**编码器-解码器**结构。在医疗领域应用广泛。因网络形状像 U 形得名。

DeepLab 2015



(a) Going deeper without atrous convolution.



(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

开创性地引入了空洞卷积结构，解决了语义分割中的分辨率与感受野的问题。

有歧义的区域

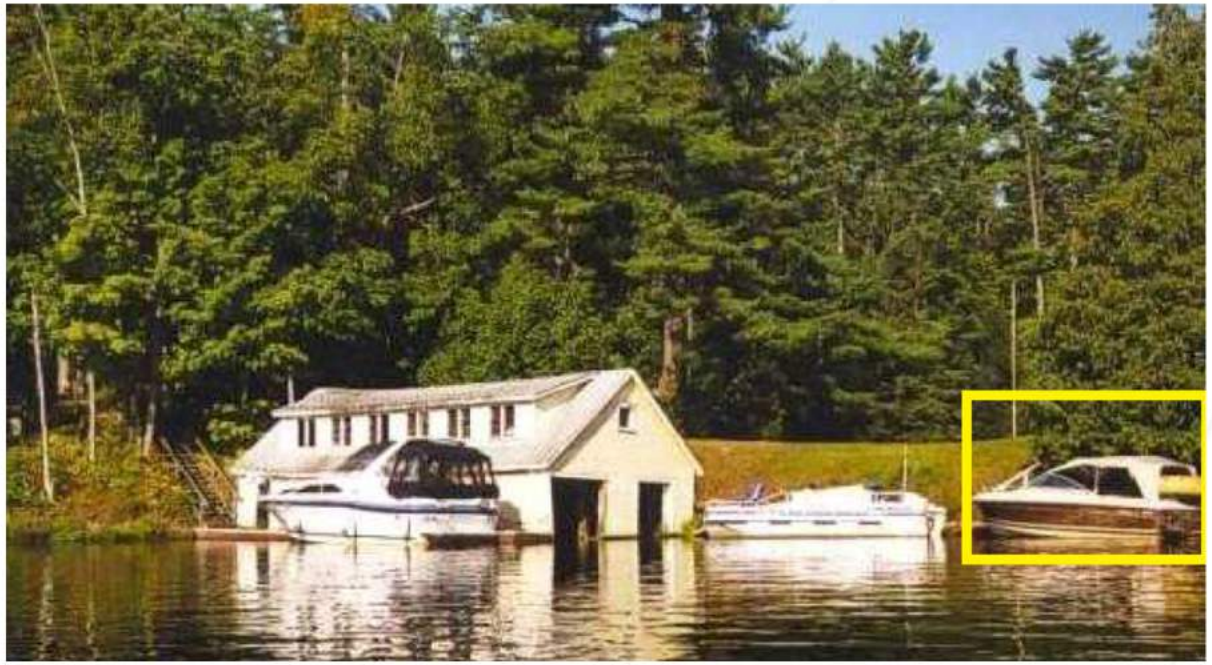
船?

汽车?

枕头?

被子?

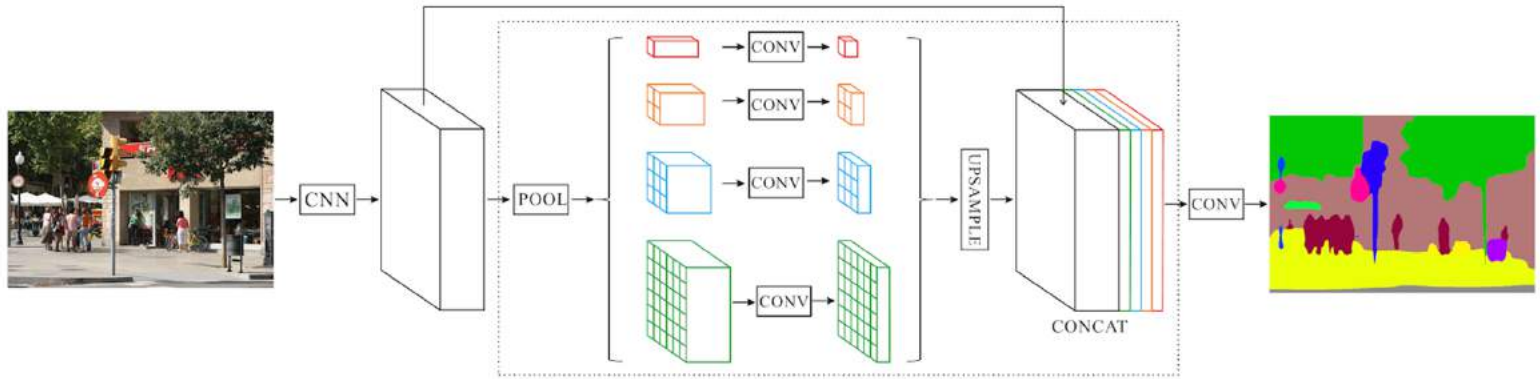
上下文的重要性



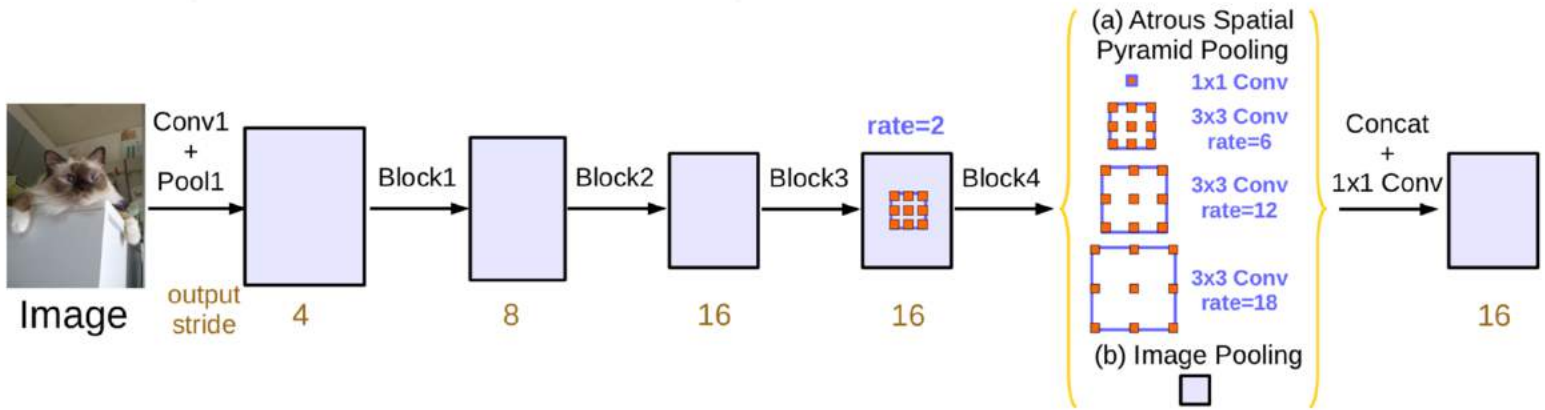
图像周围的内容（也称上下文）可以帮助我们做出更准确的判断。

在模型中捕捉上下文信息 (2016~)

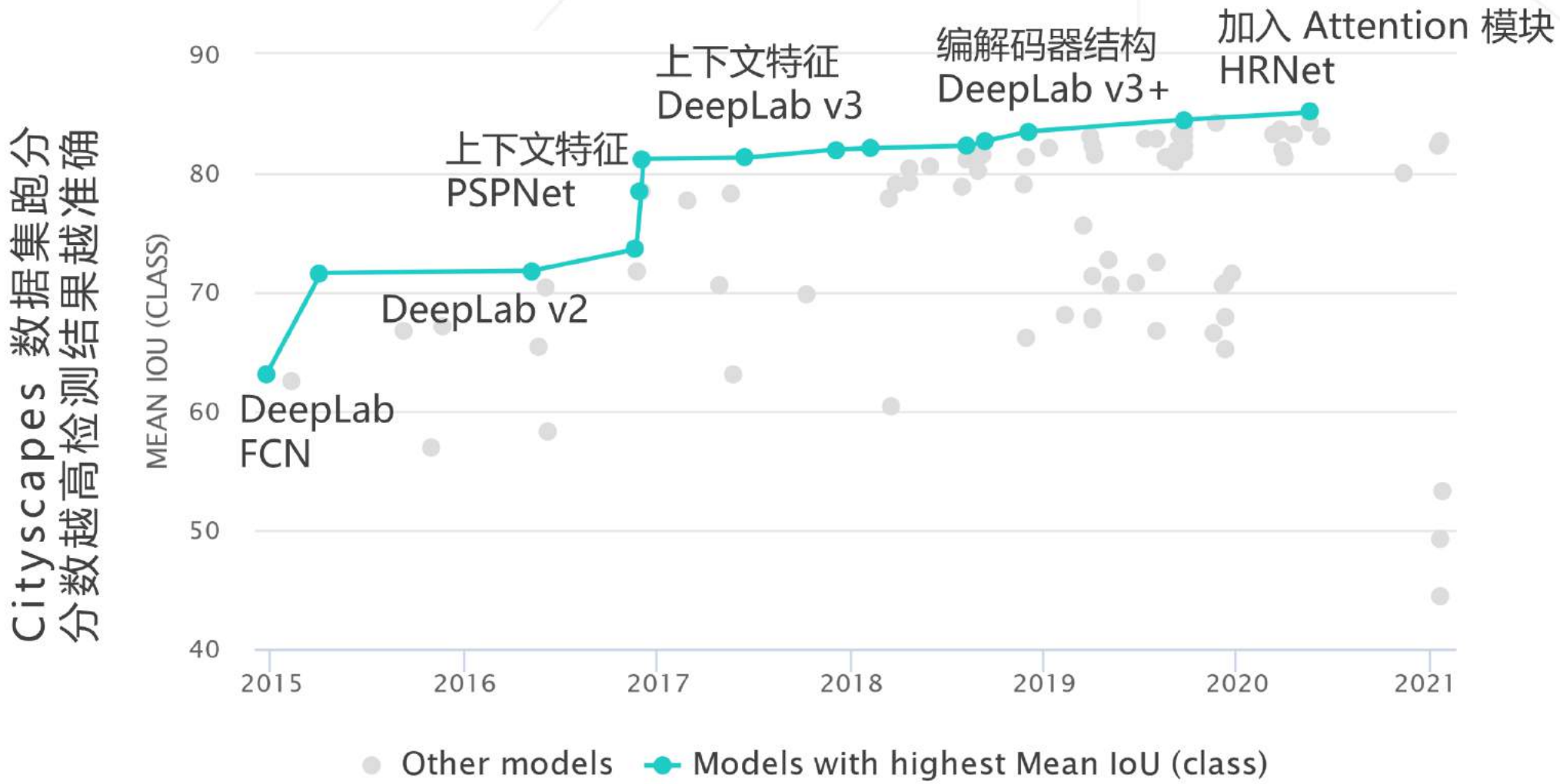
PSPNet (2016) 使用不同尺度的池化获取上下文信息



DeepLab v3 (2017) 使用不同尺度的空洞卷积获取上下文信息



语义分割的发展


Cityscapes 数据集跑分
分数越高检测结果越准确
MEAN IOU (CLASS)
90
80
70
60
50
40
2015
2016
2017
2018
2019
2020
2021
DeepLab
FCN
DeepLab v2
上下文特征
PSPNet
上下文特征
DeepLab v3
编解码器结构
DeepLab v3+
加入 Attention 模块
HRNet
Other models
Models with highest Mean IoU (class)

一站式语义分割工具包 MMSegmentation

MMSegmentation

算法丰富

- 378 个预训练模型
- 27 篇论文复现

模块化设计

- 配置简便
- 容易拓展

统一超参

- 大量消融实验
- 支持公平对比

使用方便

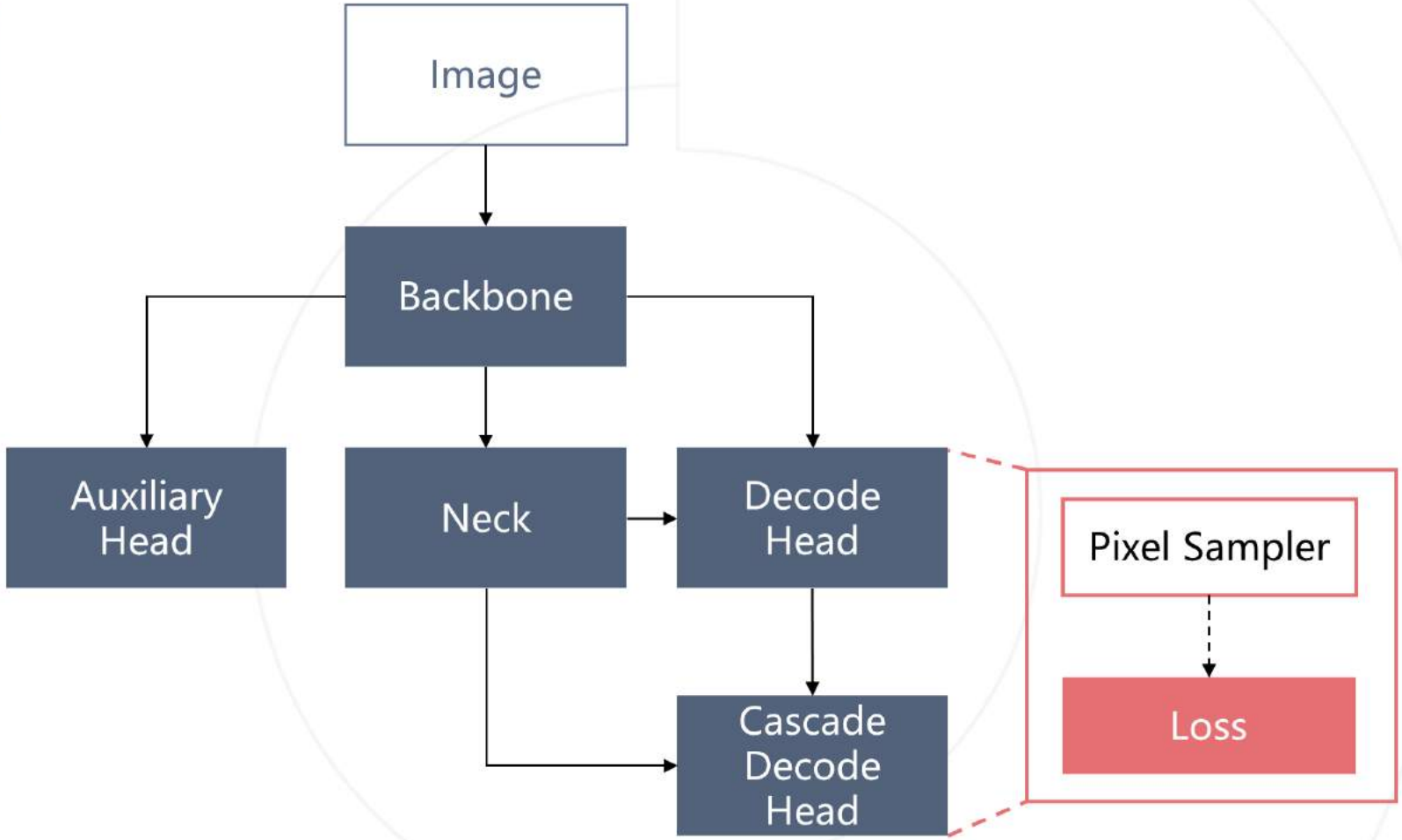
- 训练工具
- 测试工具
- 推理 API

模块化设计

我们将不同的分割模型统一拆解为：

骨干网络 **颈部** **解码头**

等的模块，方便用户根据自己的需求进行组装和扩展，构建自己的模型。

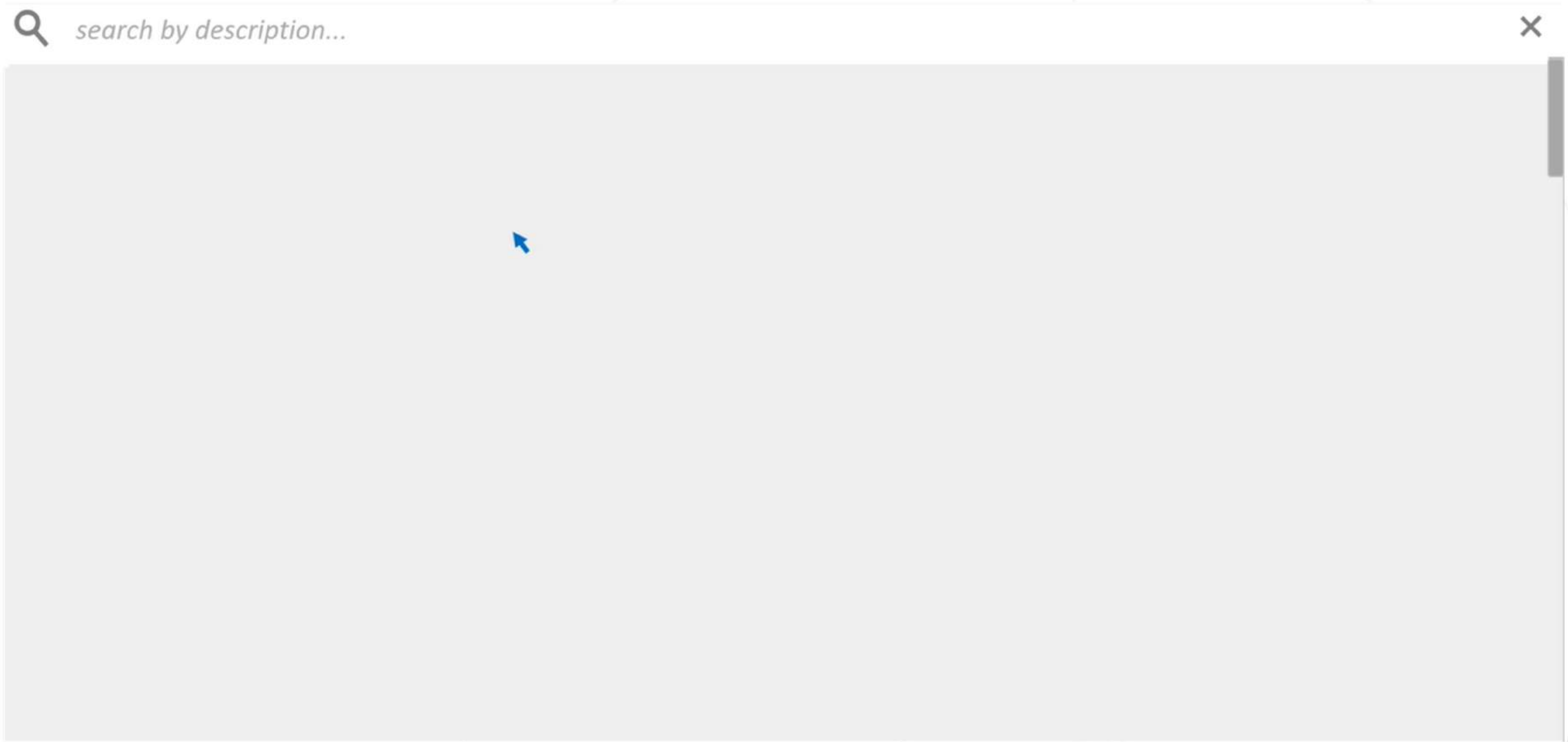


下节课继续

识别视频中的体育动作

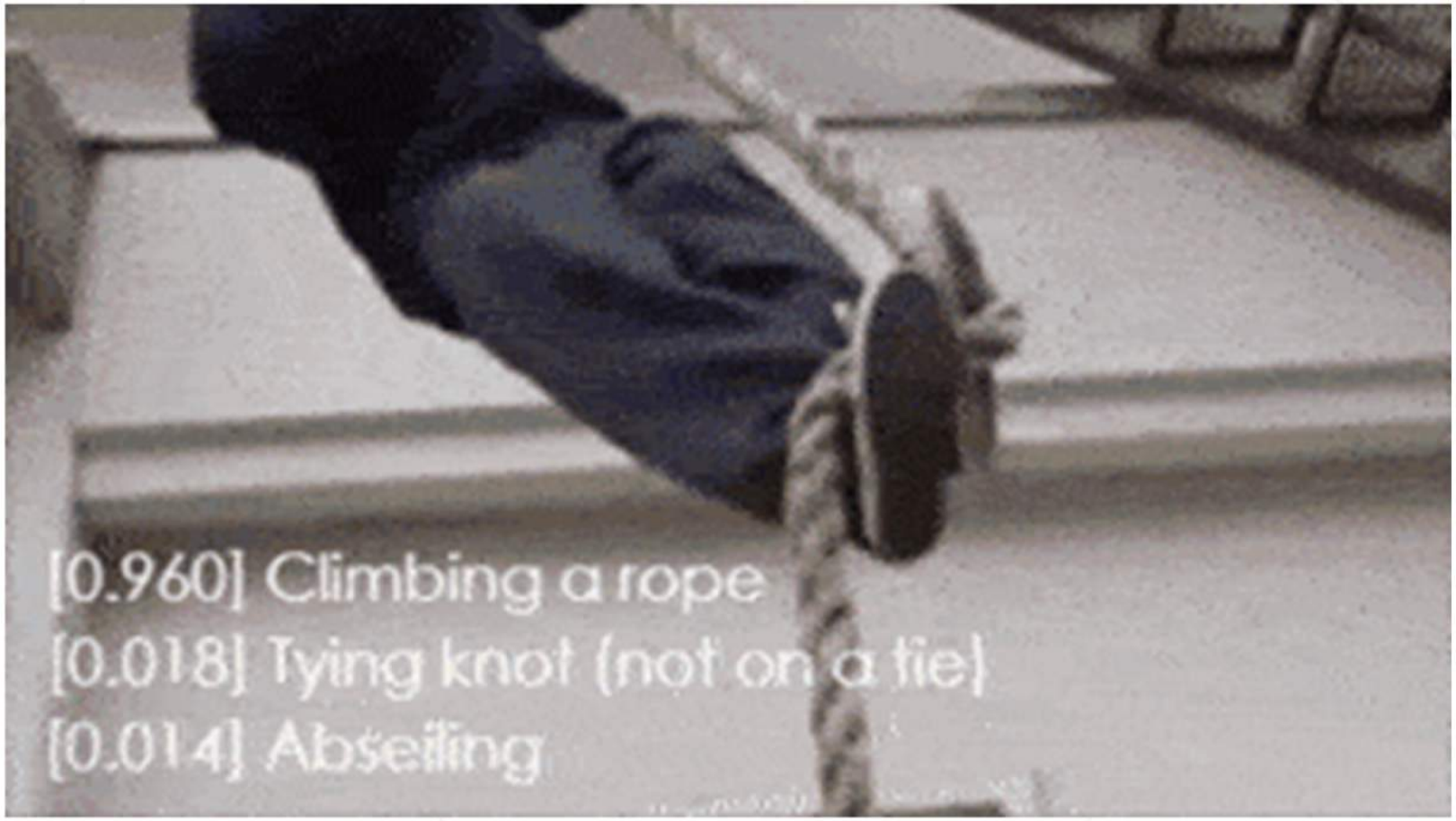
Rio Olympics 2016 – Man's 3M Spring Board Final

检索视频中的片段



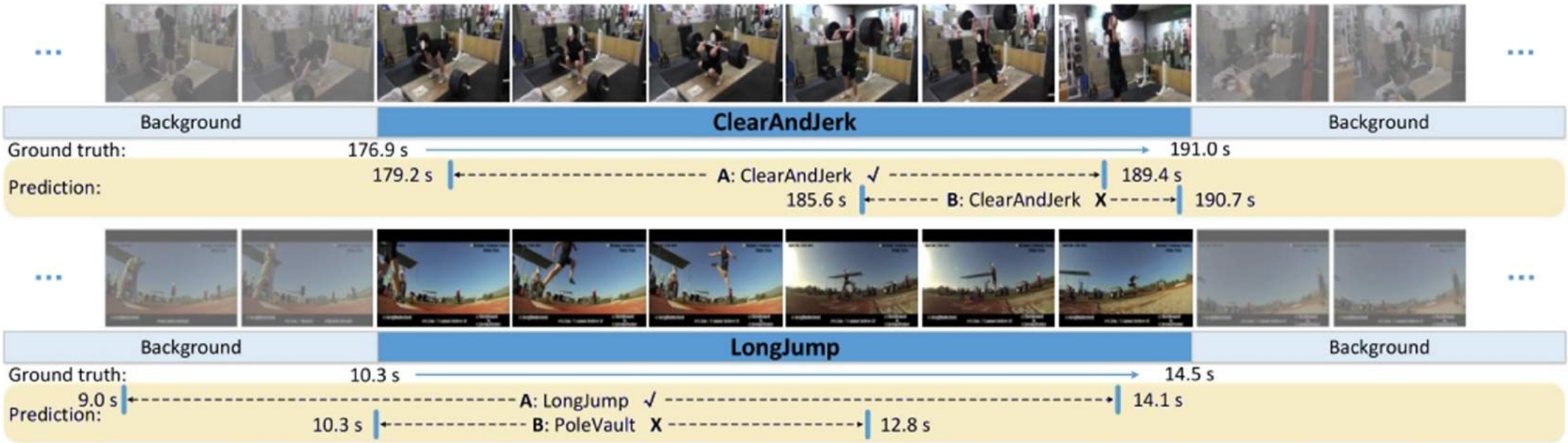
行为识别：What

- 识别视频片段中出现的动作
- 视频的分类问题



时序动作检测：When + What

在长视频中定位特定动作出现的时间段，并对动作进行分类



时空动作检测：Where + When + What

- 识别并定位视频中出现的人和动作
- 视频的检测问题

Left: Sit, Talk to, Watch; Right: Crouch/Kneel, Listen to, Watch

Left: Stand, Carry/Hold, Listen to; Middle: Stand, Carry/Hold, Talk to; Right: Sit, Write

Left: Sit, Ride, Talk to; Right: Sit, Drive, Listen to

Left: Stand, Watch; Middle: Stand, Play instrument; Right: Sit, Play instrument

视频 = 空间 + 时间 = 外观 + 运动

=

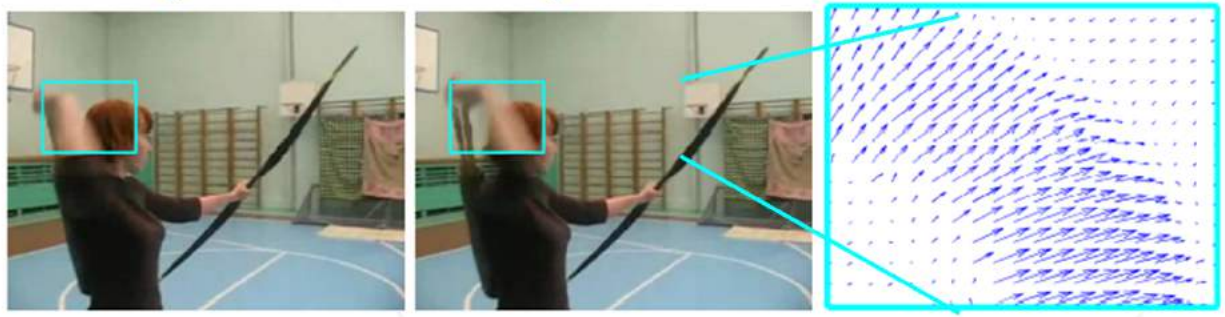
时间

外观

运动

视频中的动作

帧间运动

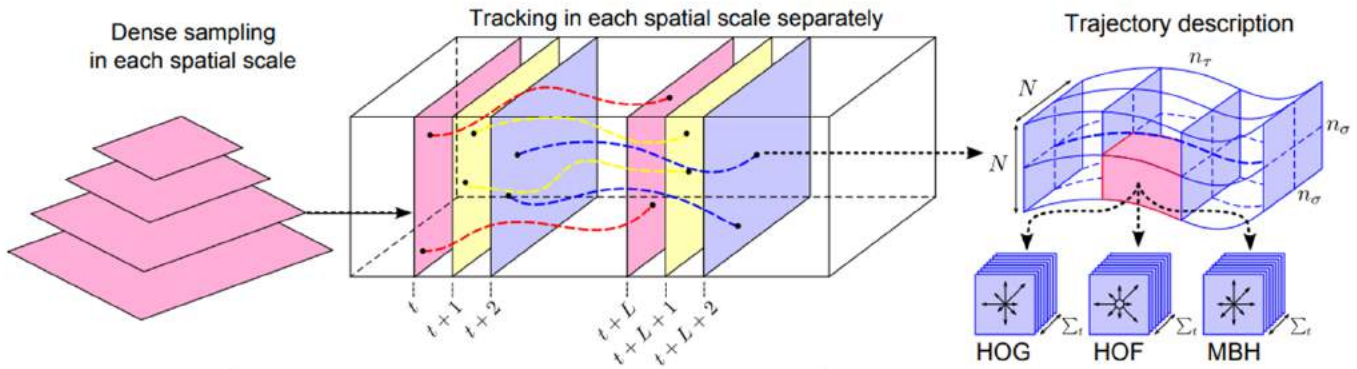


光流

帧间运动的描述

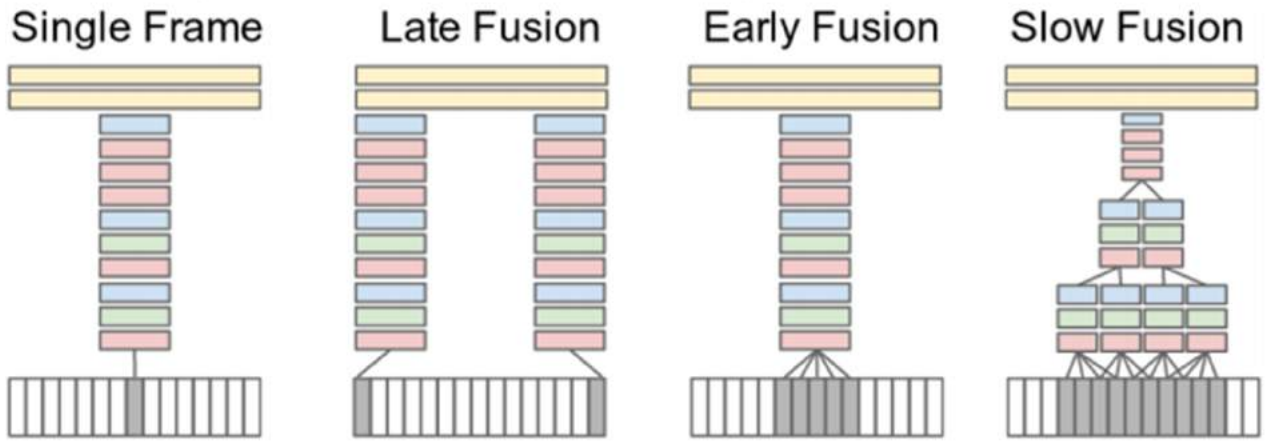
从手工特征到卷积网络 (~2014)

Dense Trajectories 2011 & Improved DT 2013



基于传统视觉方法构建运动特征。没有使用深度学习技术。

DeepVideo 2014



使用卷积网络，基于图像帧进行动作预测，并考虑融合帧间信息。没有显式利用运动特征，一些数据集的评测不如 IDT。

从单流到双流 (2014~)

双流网络 Two Stream Networks 2014

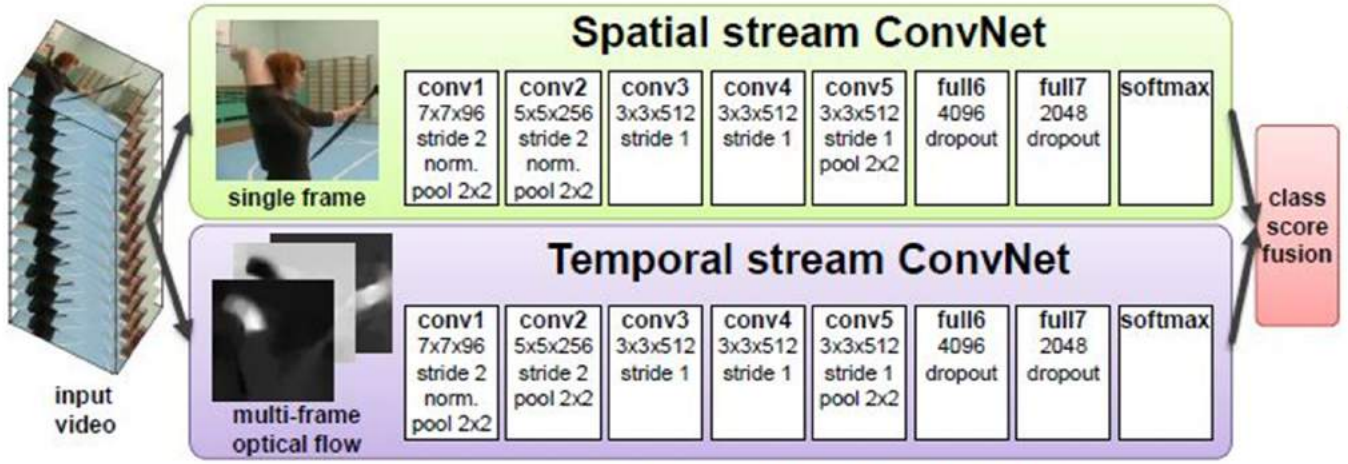
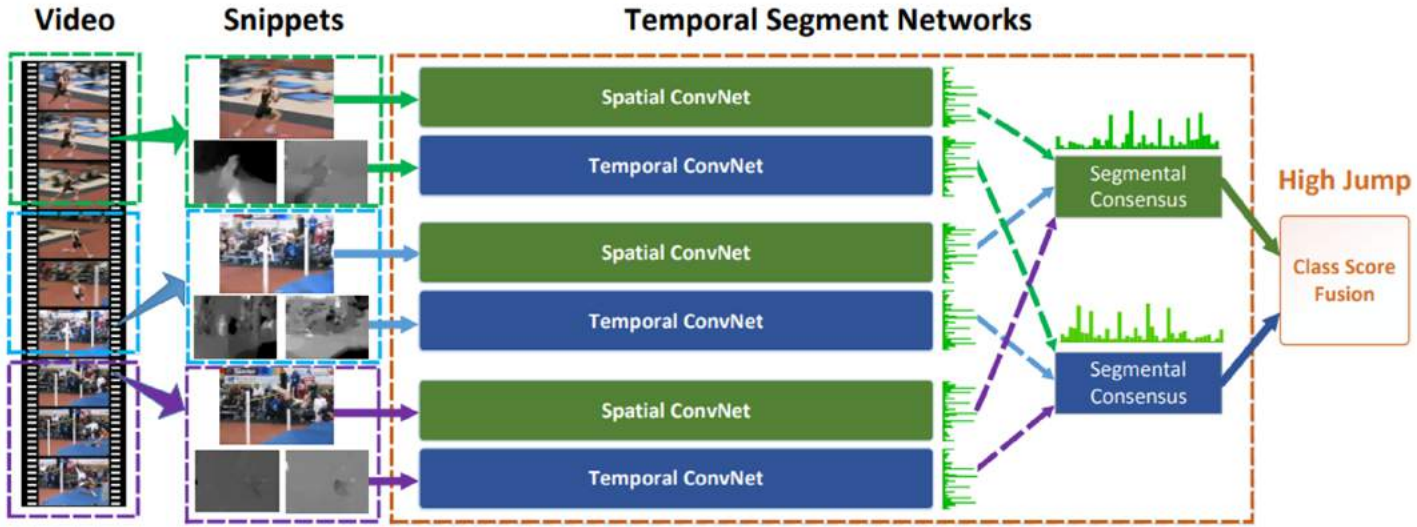


Figure 1: Two-stream architecture for video classification.

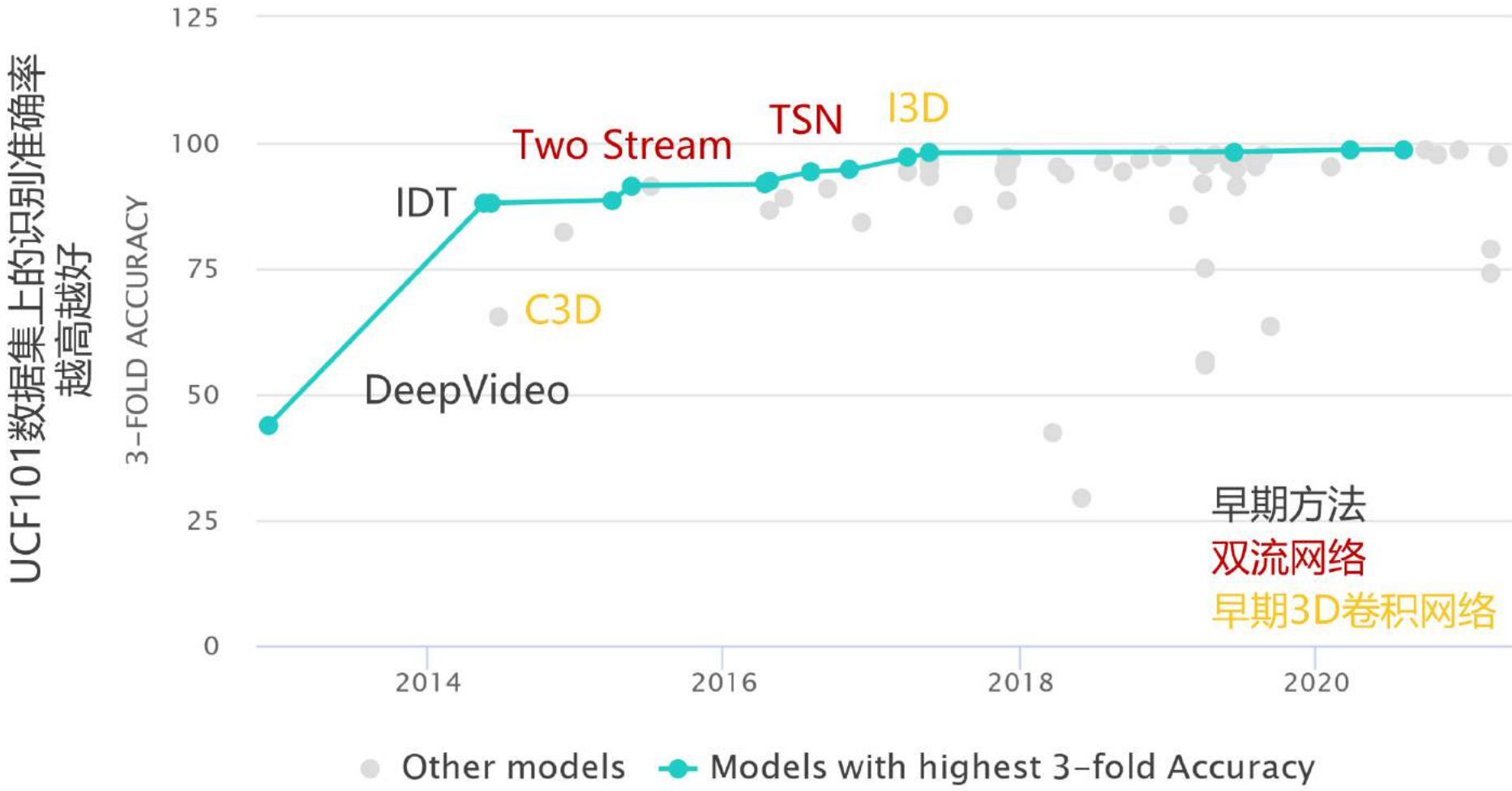
使用两个网络，分别基于图像和光流进行预测。
实验表明，使用光流对提高识别精度非常有效。

时序分段网络 TSN 2016



将视频切割成多个片段，分别使用双流网络预测再融合结果。
显式建模长时间帧间关系。

行为识别的技术发展

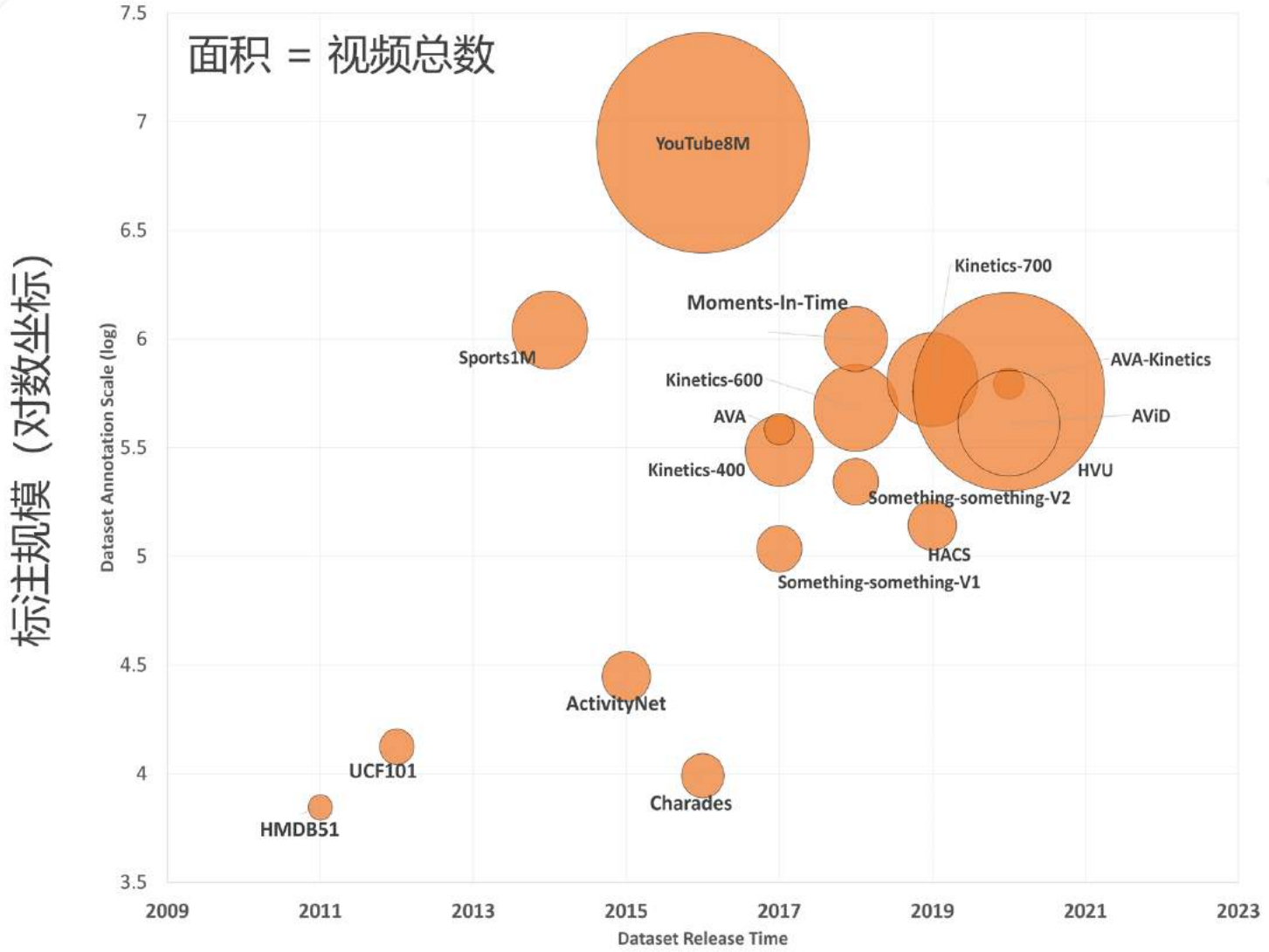

UCF101数据集上的识别准确率
越高越好
3-FOLD ACCURACY
125
100
75
50
25
0
2014
2016
2018
2020
IDT
Two Stream
TSN
I3D
C3D
DeepVideo
早期方法
双流网络
早期3D卷积网络
Other models
Models with highest 3-fold Accuracy

大规模视频数据集的出现 (2017~)

视频分析是数据密集型任务。在算法进步的同时，人们也意识到数据的重要性，相继提出更大规模的视频数据集。

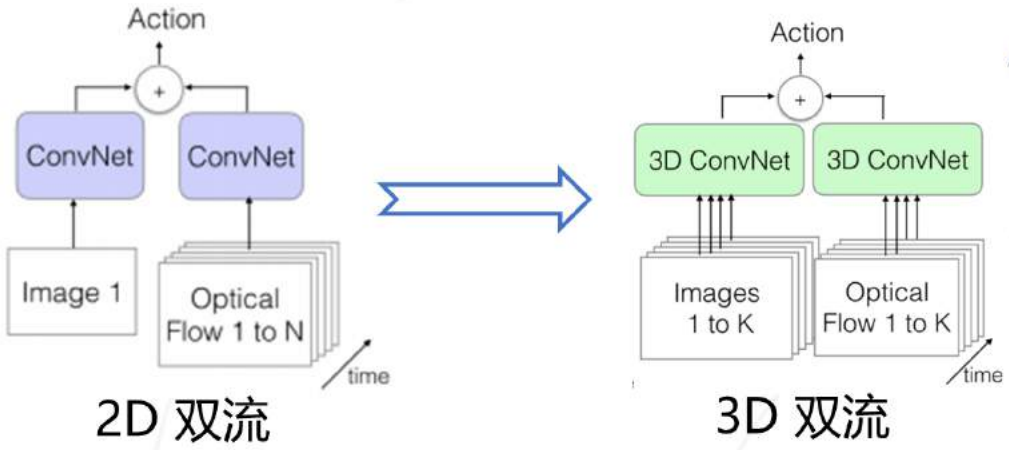
在数据的推动下，基于 3D 卷积网络的方法逐渐取代基于 2D 卷积网络的方法成为了主流。

另一方面，长视频数据集（如 Activity Net，HACS）逐渐出现，也推动了时序检测、时空检测等技术的发展。



从 2D 卷积到 3D 卷积 (2017~)

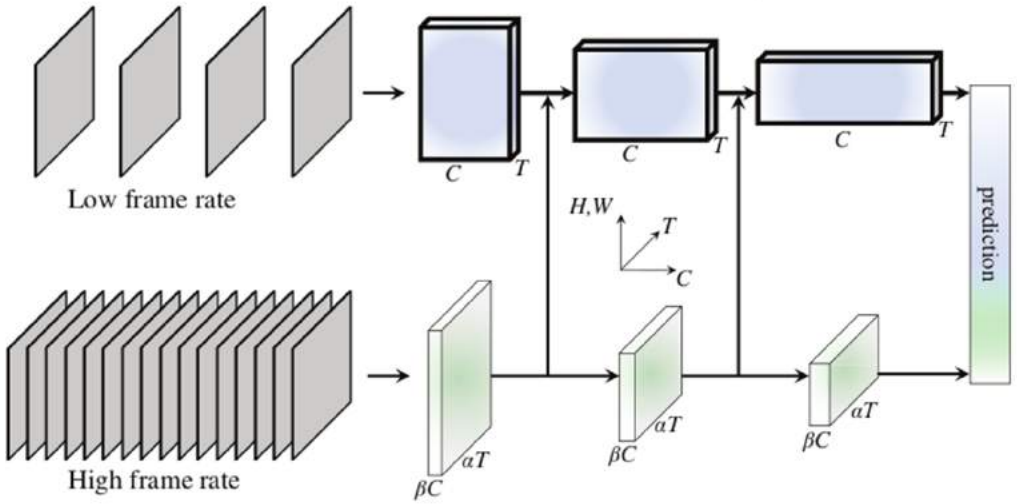
I3D 2017



2D 双流　　3D 双流

将 2D 卷积膨胀至 3D 卷积，基于图像序列个视频序列进行双流预测。
基于同时提出的 Kinetics-400 数据训练，取得了超越 2D 方法的精度。

SlowFast 2019



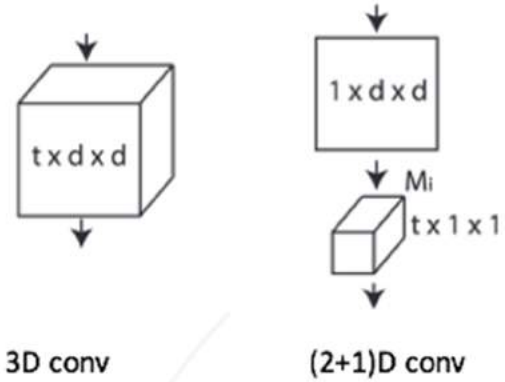
基于低速采样 (slow) 和高速采样 (fast) 的图像帧流进行预测。经过精细的参数调整，可以在不使用光流的情况下获得当时的最好精度。

压缩 3D 网络 (2018~)

3D 卷积网络可以取得更高得精度，但也带来了巨大的计算开销。随后一系列工作开始考虑压缩 3D 网络，降低计算开销的同时尽量不影响精度。

S3D 2018

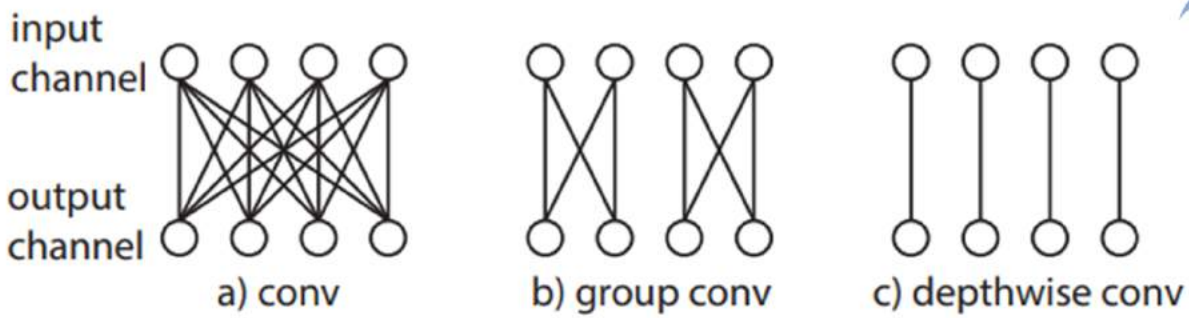
R2+1D 2018



分解卷积核：
将 3D 卷积核拆解为空间 2D 卷积核和时间 1D 卷积核

CSN 2019

X3D 2020



减少通道之间的连接：
在通道维度引入分组卷积或分层卷积。

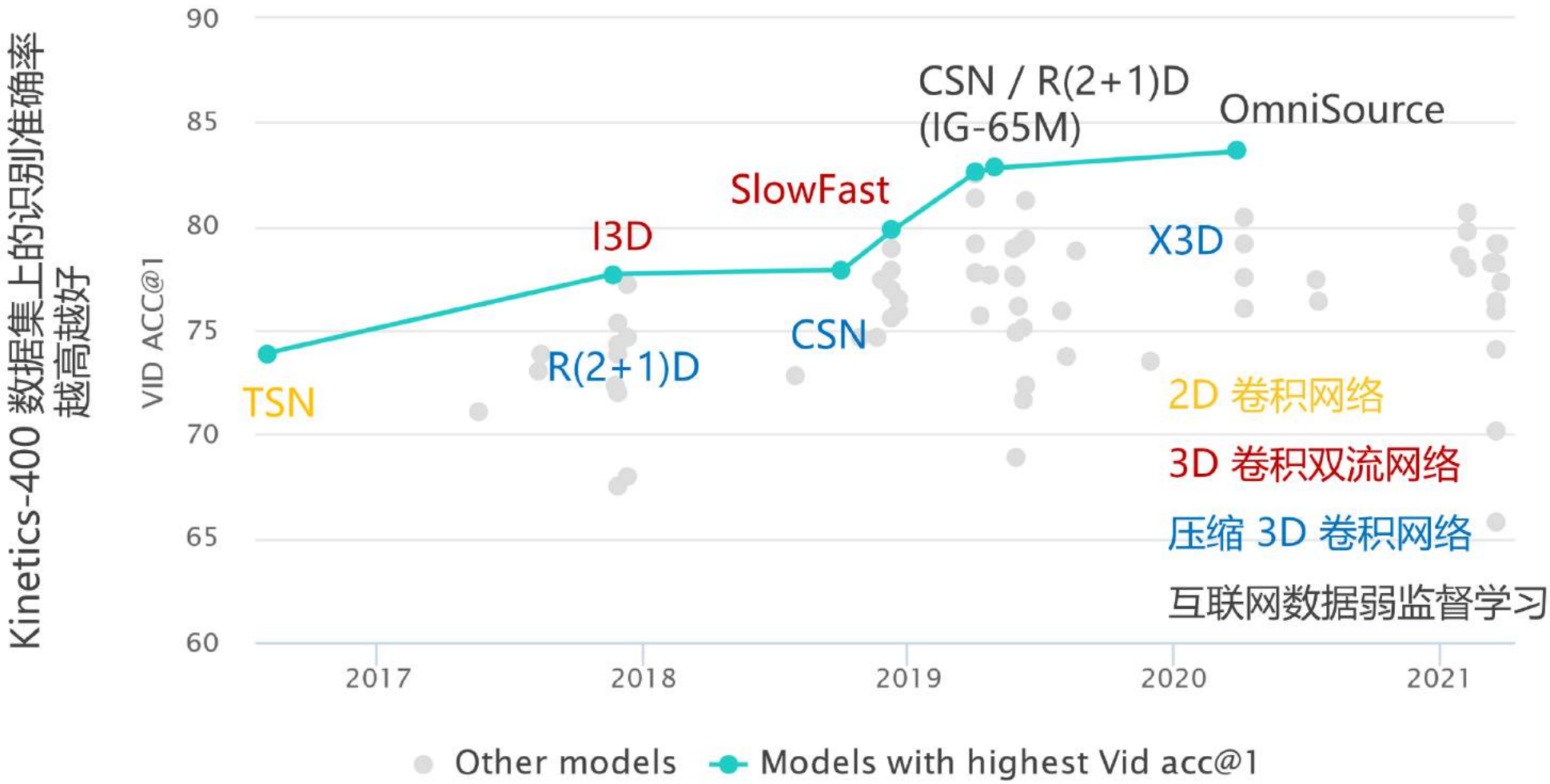
更大的数据 (2019~)

- 标注视频数据集的人力成本非常高昂。
- 互联网上有海量无标注的视频，能否帮助我们训练更好的模型？

弱监督学习 (2019~)

- 在机器学习中，使用标注不全或不准的数据进行学习的问题称为**弱监督学习**（weakly-supervised learning）。
- 近年来，一些工作开始探索这个方向，代表工作包括来自 Facebook 的 IG-65M 以及来自香港中文大学的 OmniSource 等。
- 实验表明，使用百倍于学术数据集的弱监督数据训练出的模型，可以很好地迁移到学术数据集上，取得领先的监督学习的精度。

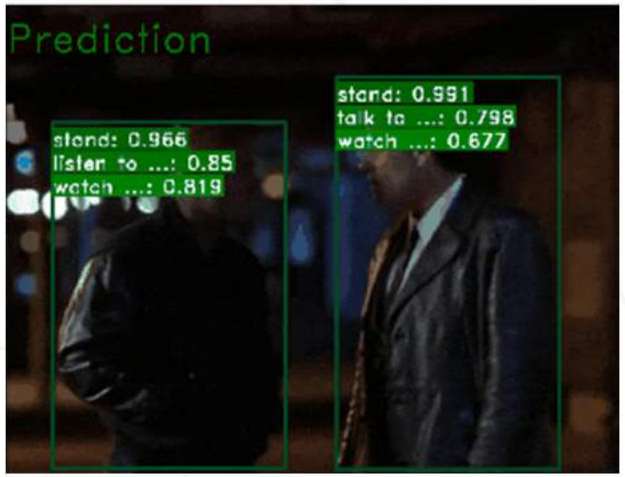
行为识别的技术发展


Kinetics-400 数据集上的识别准确率
越高越好
VID ACC@1
90
85
80
75
70
65
60
2017
2018
2019
2020
2021
TSN
I3D
R(2+1)D
SlowFast
CSN
CSN / R(2+1)D
(IG-65M)
OmniSource
X3D
2D 卷积网络
3D 卷积双流网络
压缩 3D 卷积网络
互联网数据弱监督学习
Other models
Models with highest Vid acc@1

一站式视频理解工具包 MMAction2

OpenMMLab

MMAction2



全面支持

- 动作识别
- 时序检测
- 时空检测

算法丰富

- 157 个预训练模型
- 21 篇论文复现

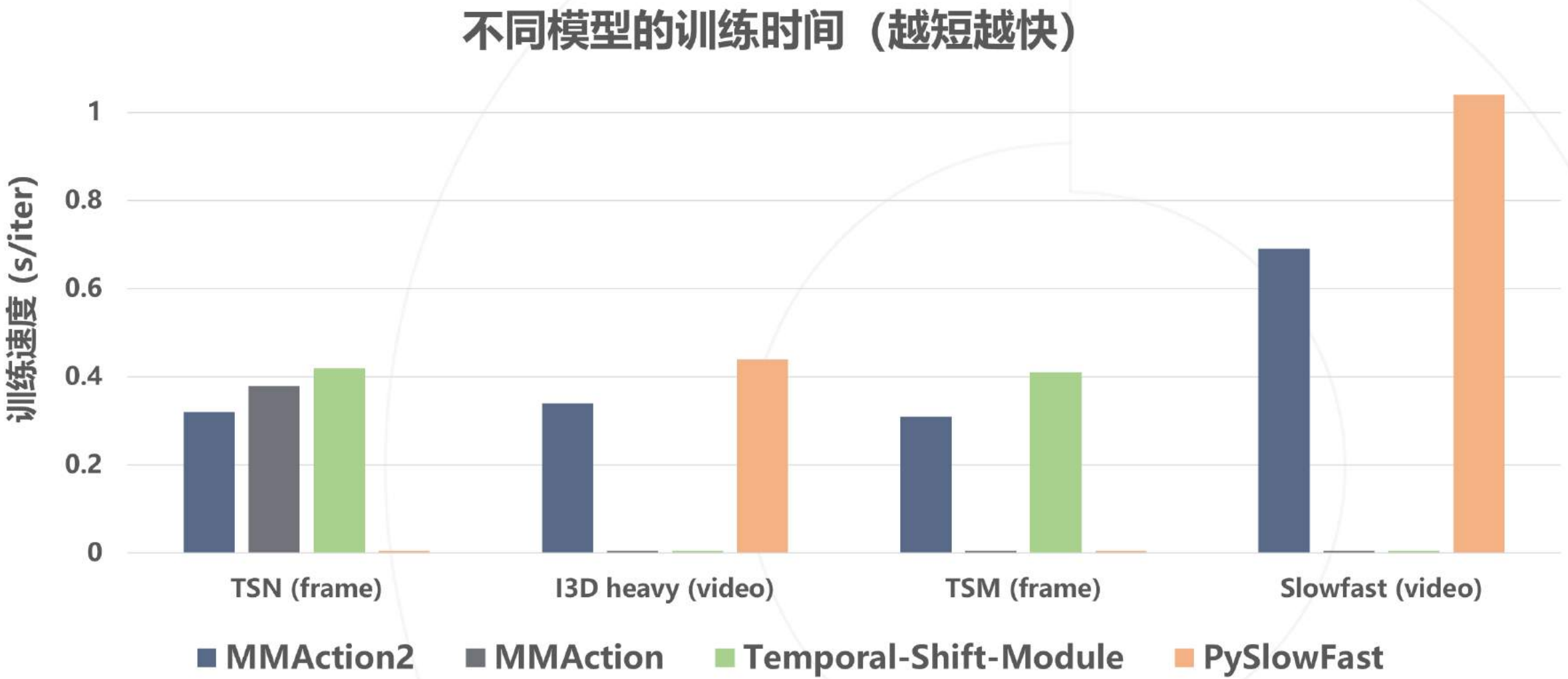
更优更快

- 训练速度
- 模型精度

使用方便

- 训练工具
- 测试工具
- 推理 API
- 多 IO 后端支持

飞一般的训练速度

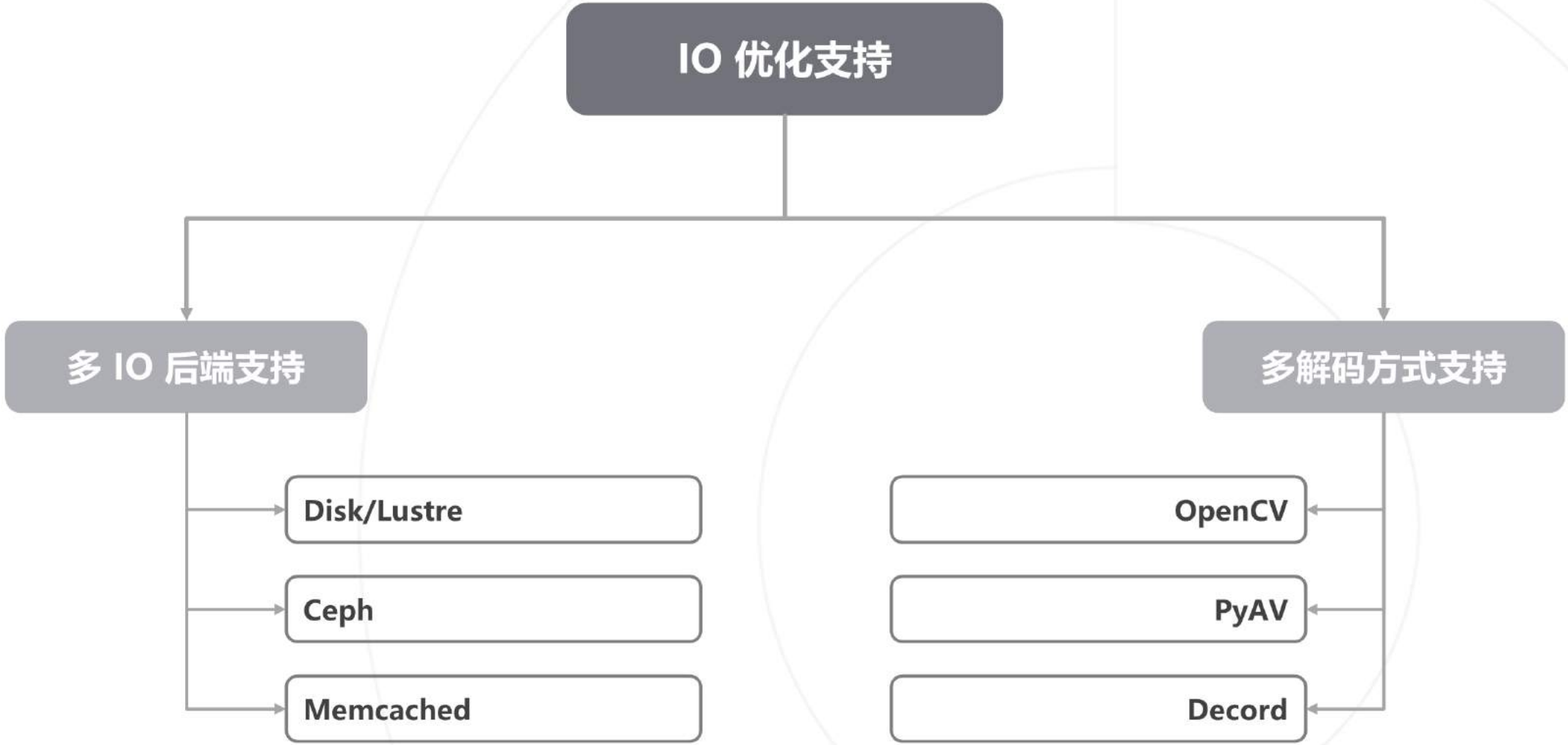

不同模型的训练时间（越短越快）
训练速度 (s/iter)
1
0.8
0.6
0.4
0.2
0
TSN (frame)
I3D heavy (video)
TSM (frame)
Slowfast (video)
MMAction2
MMAction
Temporal-Shift-Module
PySlowFast

更高的复现精度

MMAction2 的模型精度 普遍优于 其他代码框架

模型	数据集	分辨率	MMAction2	其他代码
TSN	Kinetics400	Short-side 320	**70.91**	**70.6**
I3D	Kinetics400	Short-side 256	**73.27**	**72.6**
TPN	Kinetics400	Short-side 320	**76.20**	**75.49**
CSN	Kinetics400	Short-side 320	**80.76**	**80.6**
TSM	Sth-v1	Height 100	**49.28**	**48.61**
TIN	Sth-v2	Height 240	**56.70**	**56.38**
TSM	Sth-v2	Height 240	**62.04**	**60.98**

丰富的 IO 后端支持


IO 优化支持
多 IO 后端支持
多解码方式支持
Disk/Lustre
Ceph
Memcached
OpenCV
PyAV
Decord

哪些是真实的人脸?

这些都是深度学习模型生成的假人脸!

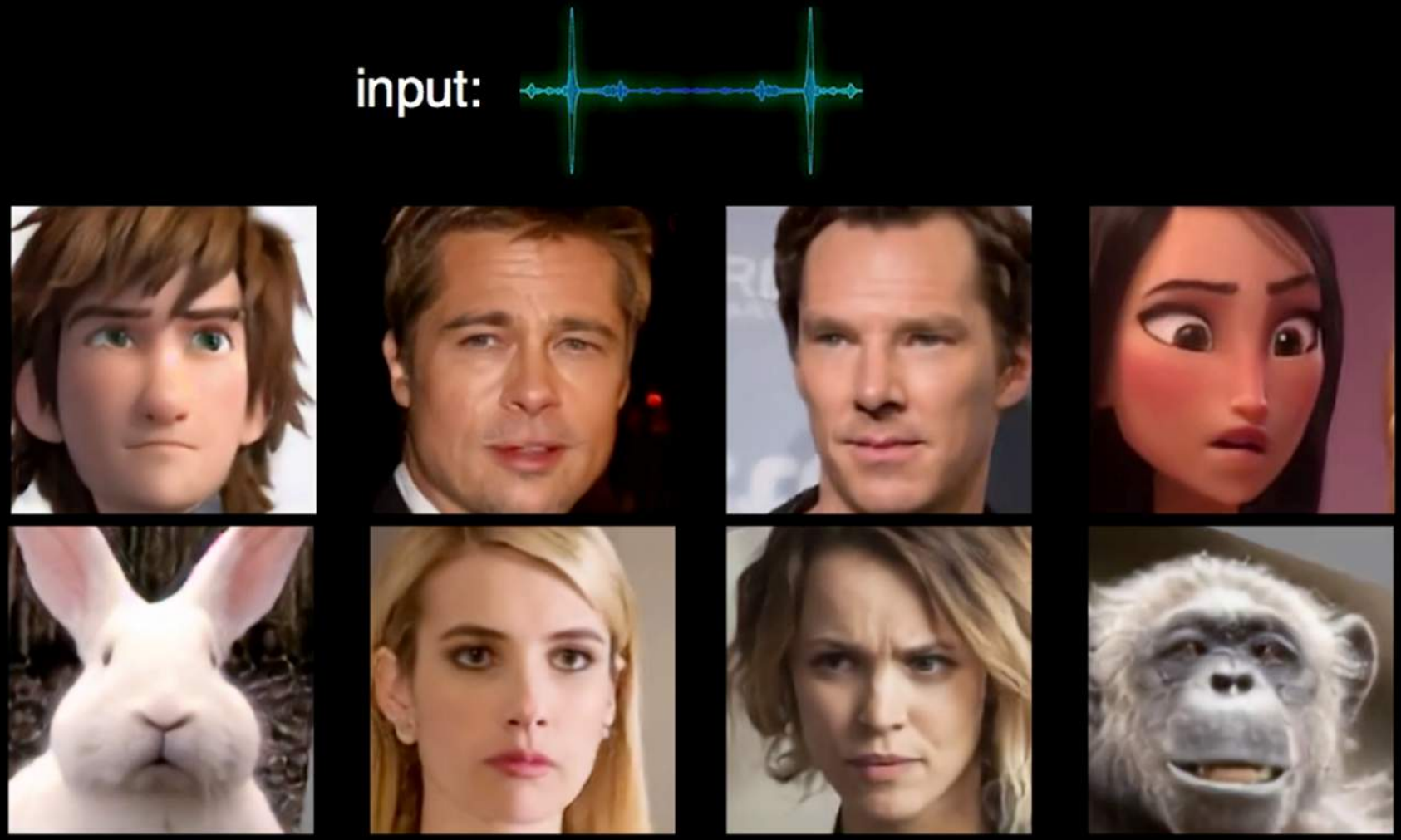
https://thispersondoesnotexist.com/





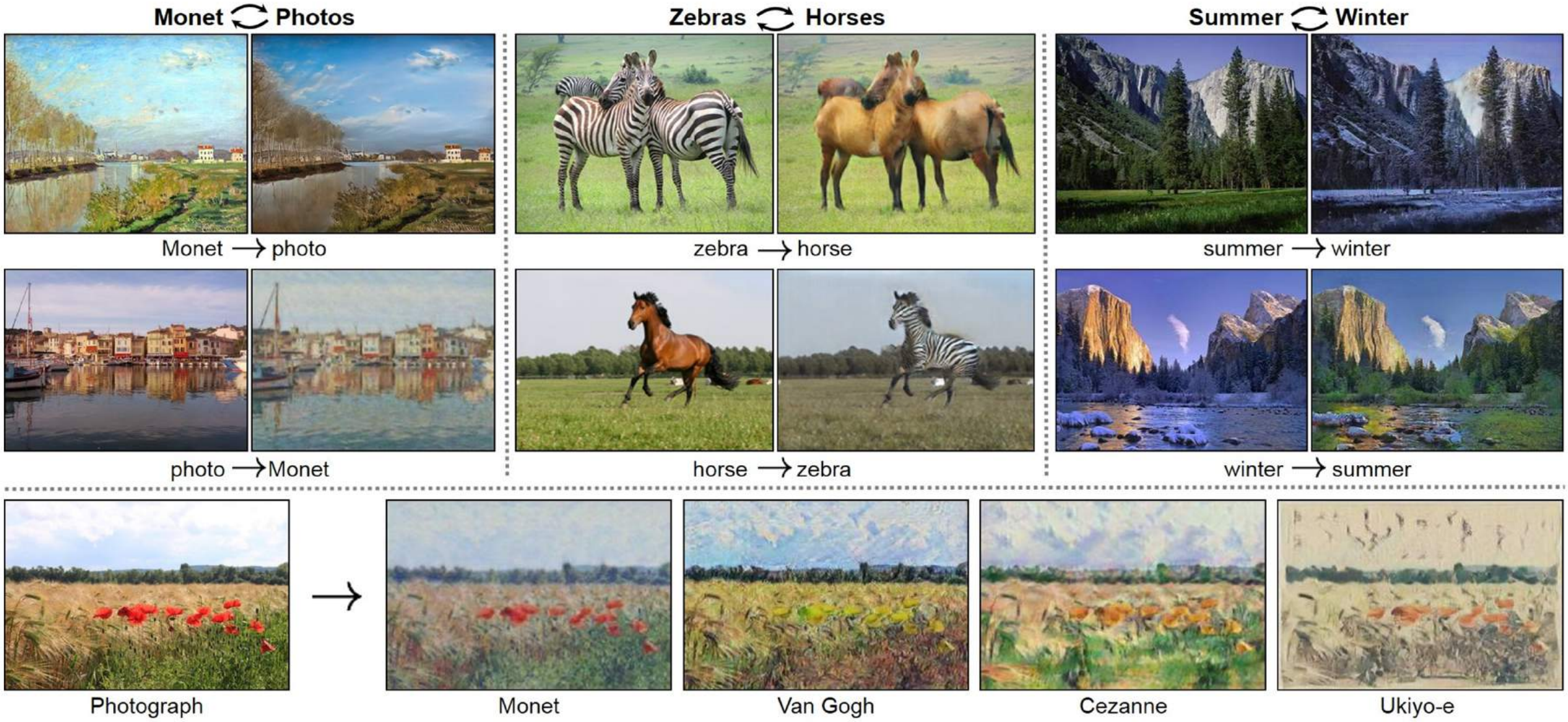



会说话的人脸



Talking Face Generation by Adversarially Disentangled Audio-Visual Representation

图像转换



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

动漫特效

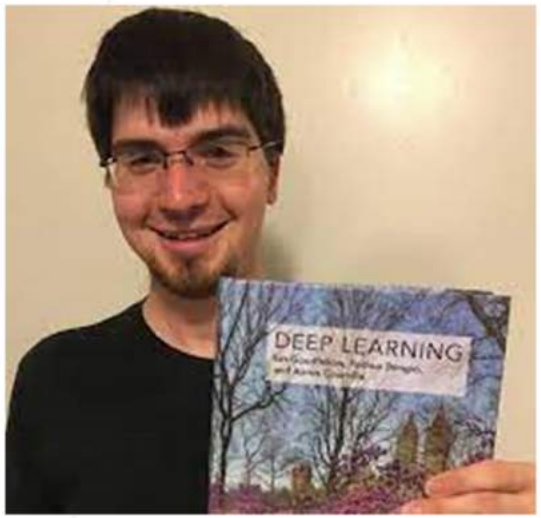
Landmark Assisted CycleGAN for Cartoon Face Generation

From 抖音

无中生有——生成对抗模型

"[GANs are] the most interesting idea in the last 10 years in Machine Learning" -- Yann LeCun

"生成对抗模型是机器学习领域近十年来最有趣的想法" -- 杨立昆



Ian Goodfellow
GAN 的发明者
《Deep Learning》 教材作者

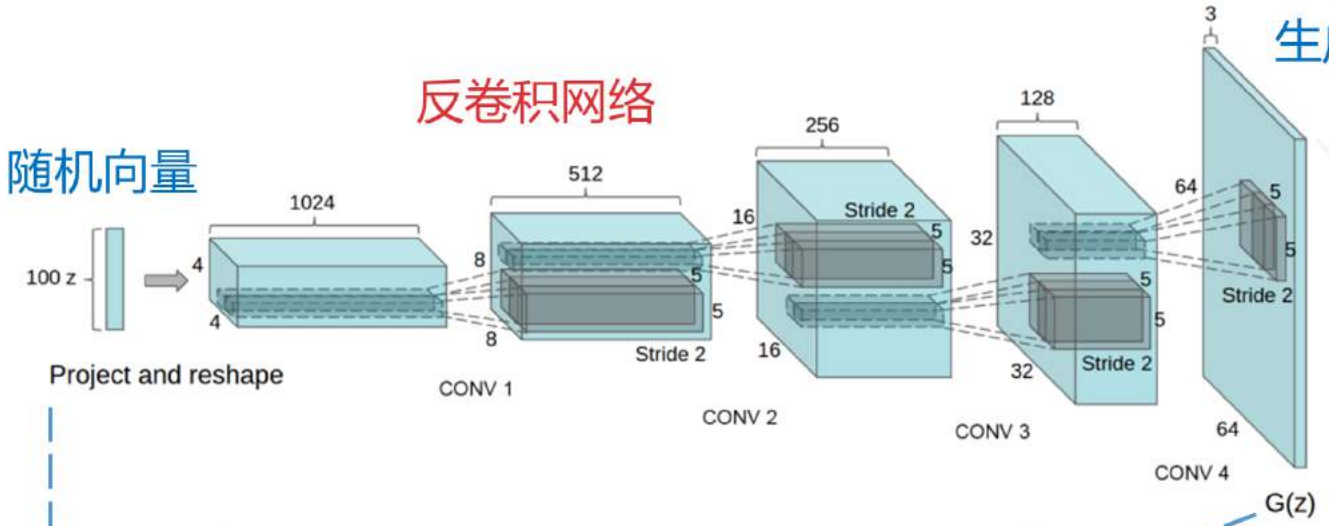
Yann LeCun
卷积网络之父

生成对抗模型 Generative Adversarial Network

Generative 模型可以**生成**图像

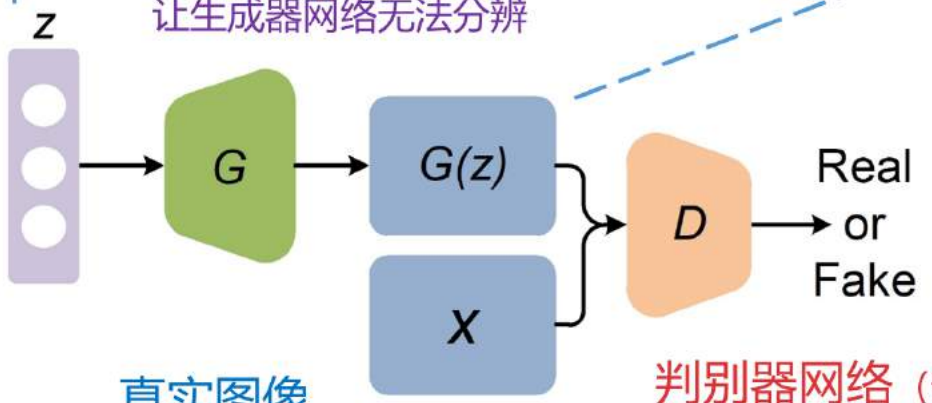
Adversarial 训练以**对抗**的方式进行

Network 神经网络结构



生成器网络

尽力生成接近真实的图像，
让生成器网络无法分辨



真实图像

来自数据集

判别器网络（普通的分类网络）

尽力分辨真实图像和生成图像

生成对抗模型的诞生

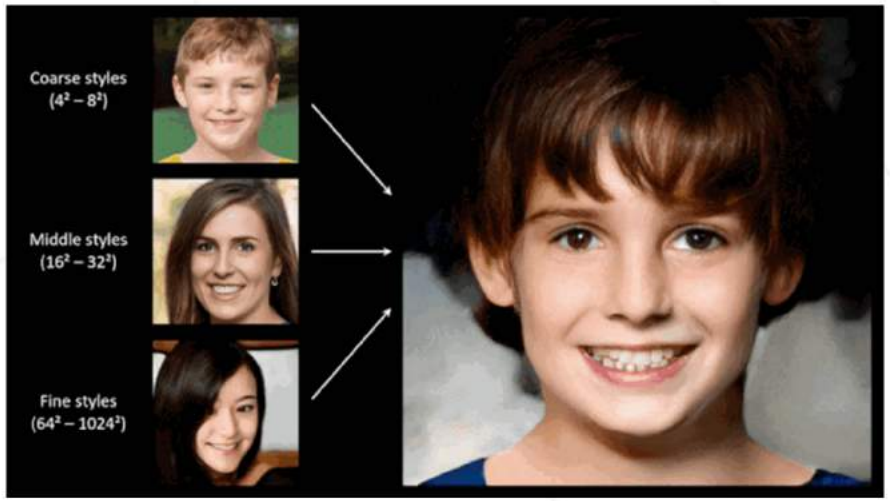
- 2014 Ian Goodfellow 首次提出了 GAN 模型，基于多层感知机模型
- 2015 基于卷积和反卷积的 DCGAN 生成了效果较好的图像
- 2017 WGAN 模型提出，改善了传统 GAN 难以训练的问题

Face Super-Resolution Through Wasserstein GANs
DCGAN Tutorial from PyTorch

生成对抗模型的发展

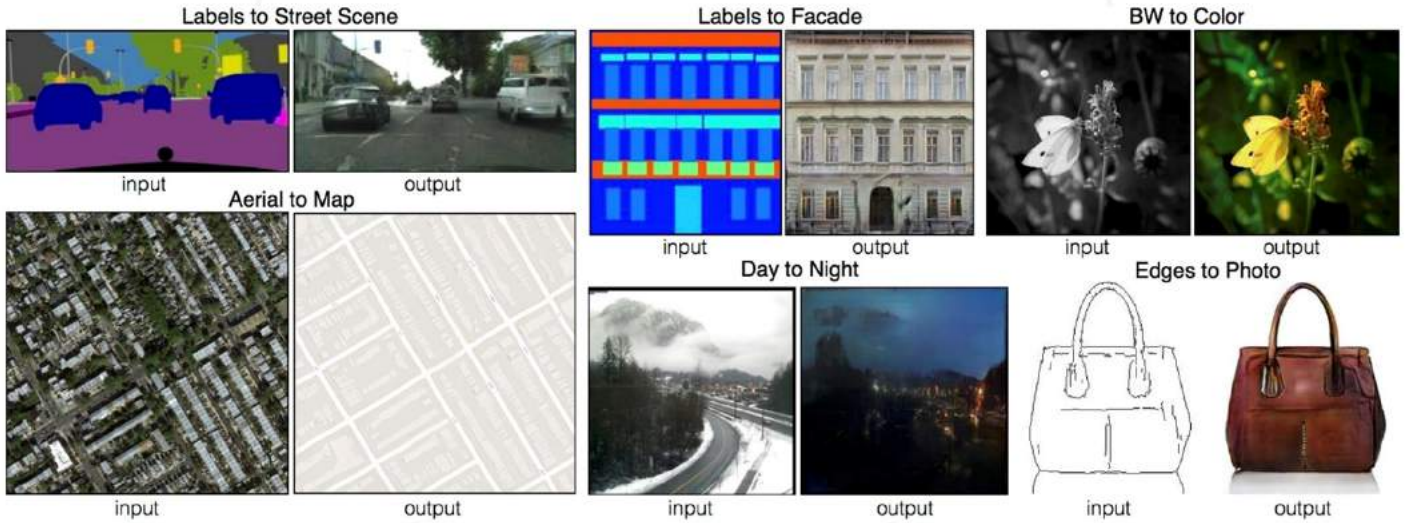
更强更大的 GAN 模型

2017 BigGAN (DeepMind)
2018 StyleGAN (NVIDIA)
大模型，大数据集
更清晰，更逼真，



用 GAN 做图像转译

2017 CycleGAN
2017 pix2pix



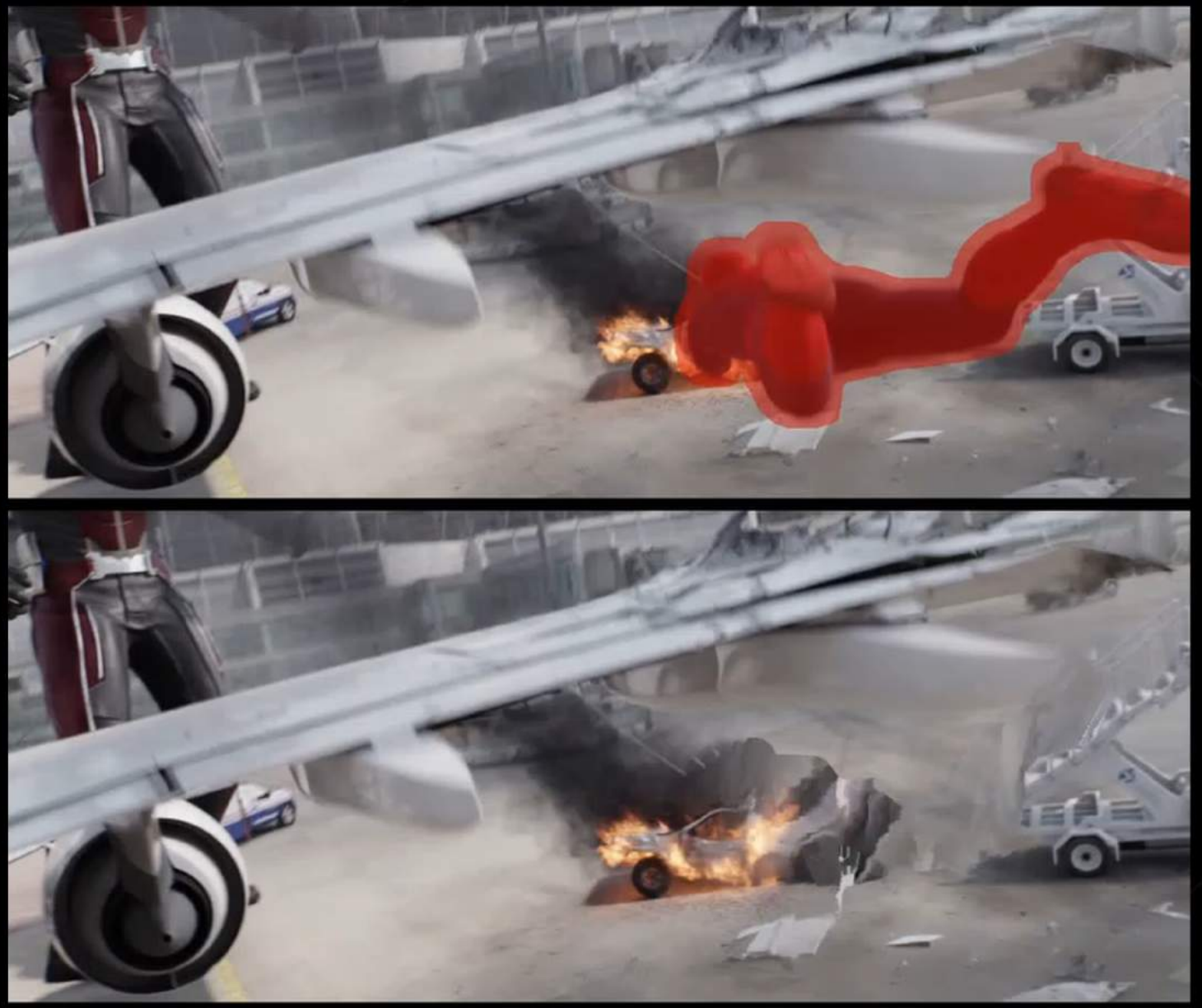
生成对抗模型的发展

可解释的 GAN 模型

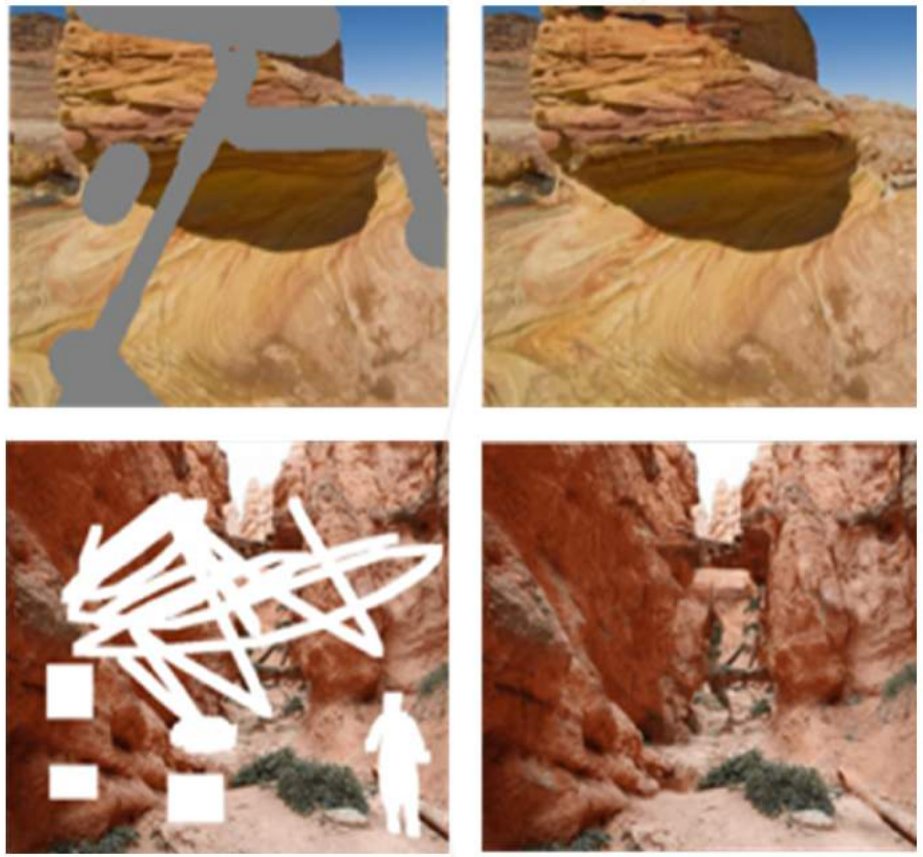
2020 SeFa

在隐空间直接控制各种属性

抹掉视频中的人



图像修复



目标

修复图像中的受损区域

应用

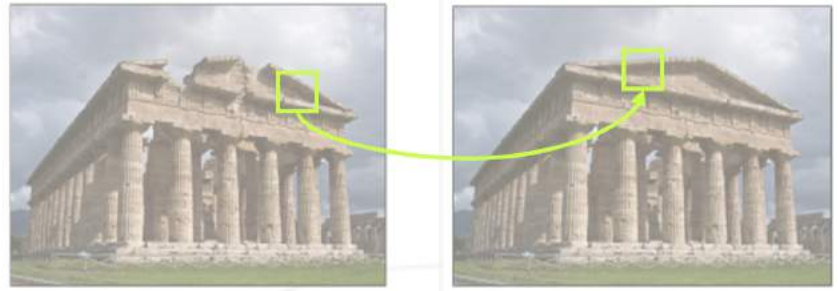
去除水印、消除人像、视频修复

图像修复的技术演进

传统视觉

PatchMatch (2009)

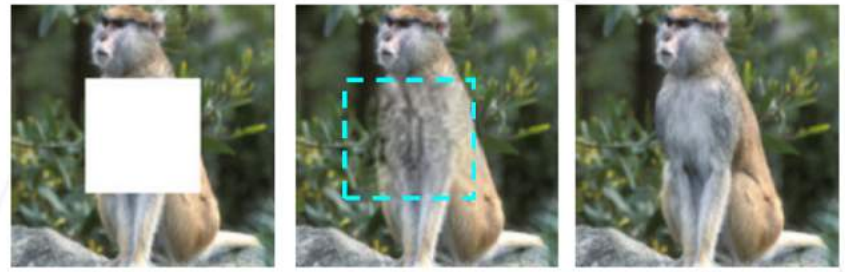
基于区块匹配，从原图上匹配相似的区域进行补全



深度学习早期探索

Context Encoder (2016)

使用**编码解码器**结构和**对抗训练**机制，对抗训练只考虑全局，局部恢复效果不佳



(a) Input (b) CE (c) GT

更多样的深度模型更好的恢复效果

Global & Local (2017)

在 CE 的基础上加入局部的对抗训练，获得较好效果

DeepFill (2018) v2 (2019)

Pconv (2018)

加入 Attention 机制

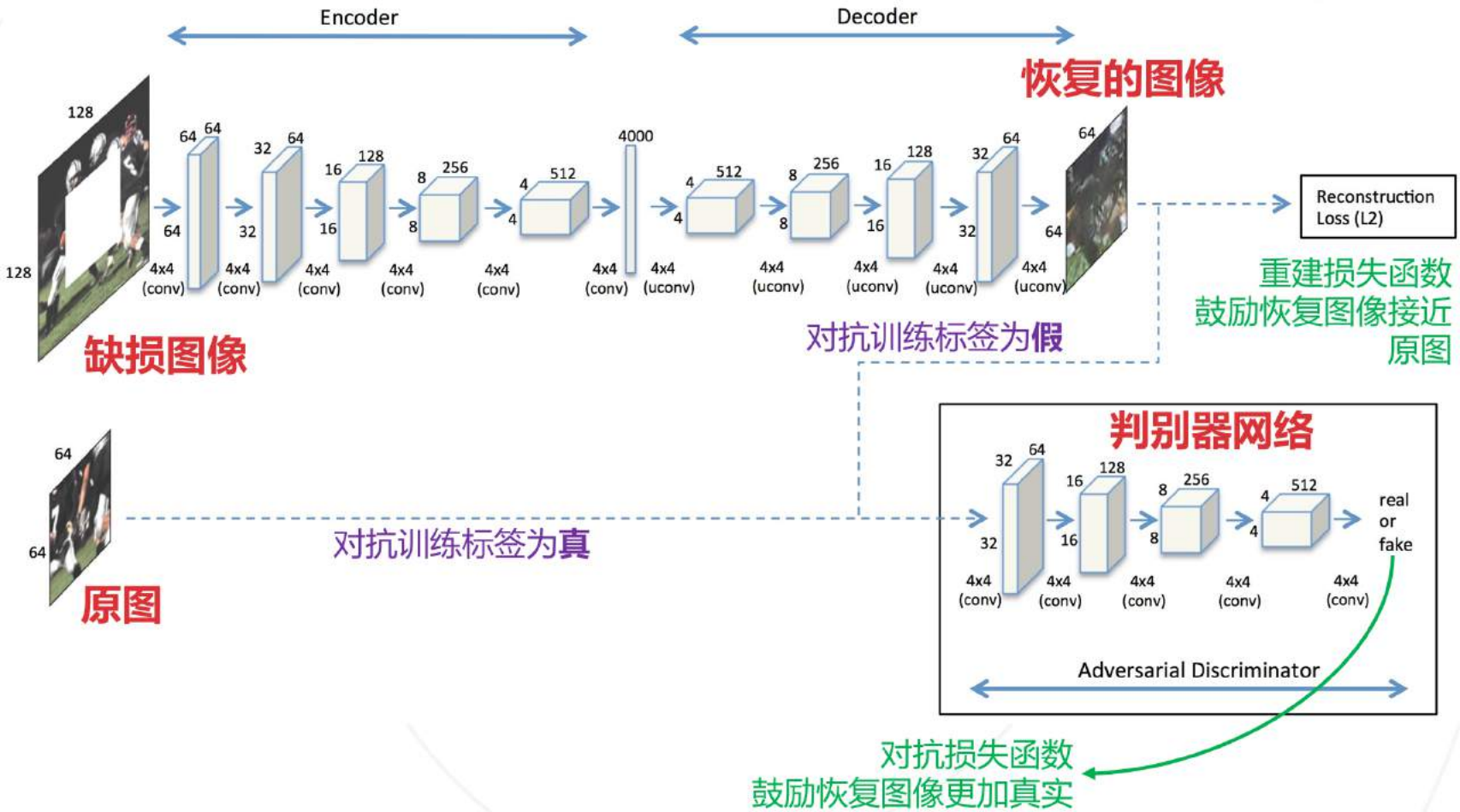
单阶段➔双阶段

什么是对抗训练

图像恢复要求尽量**真实**，但如何鼓励网络生成更真实的回复图像？

我们可以引入一个判别分类器，将完整图片分类为真，恢复图片判别为假。为了“**迷惑**”判别器，降低损失函数，图像恢复网络就必须生成更为真实的图像。

与此同时，判别器也要区分真实和回复图像之间越来越细微的差别。二者在训练过程中相互对抗，同步增强。

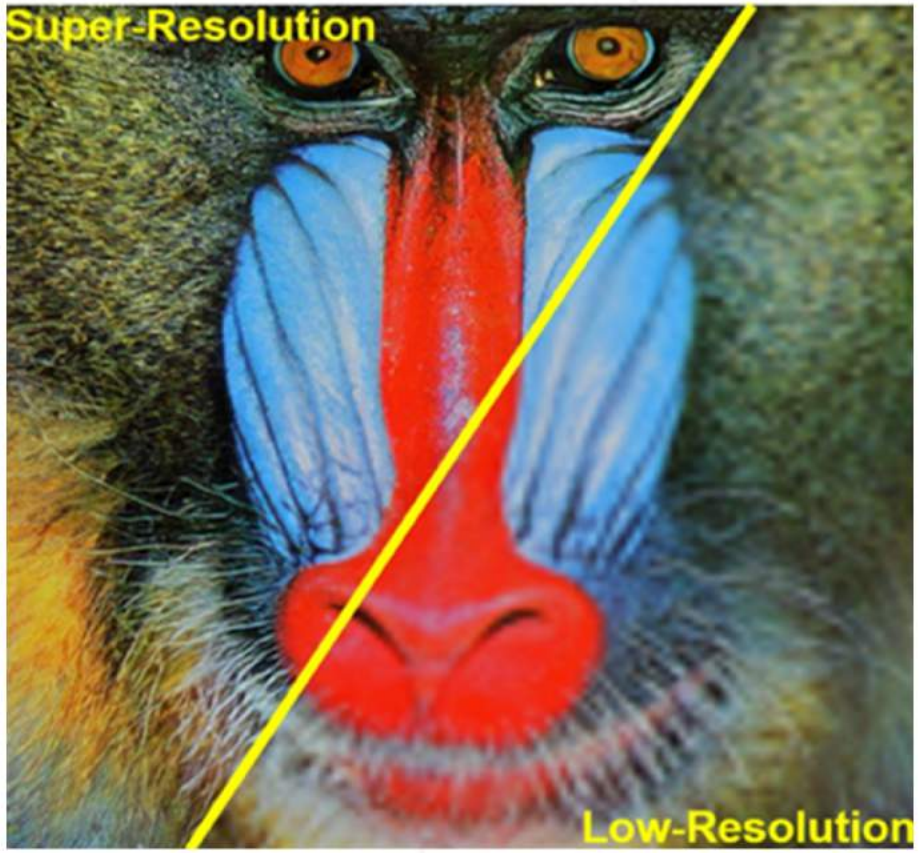


从低清视频到高清视频

Input
(320p)

After Super Resolution
(1080p)

图像超分辨率



超分辨率
Super-Resolution

目标

从低分辨率到高分辨率的图像和视频生成

应用

图像编辑，低功耗图像传感器

难点

可能存在不同的高分辨率图像，是个不定问题。

图像超分辨率的技术演进

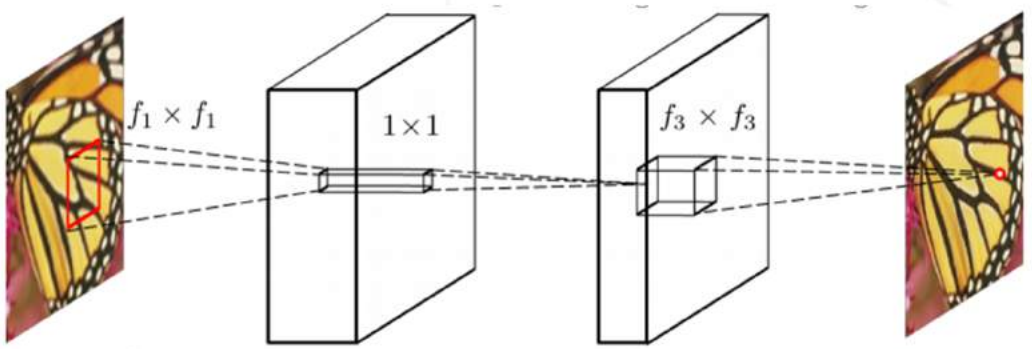
传统视觉

基于相似匹配和字典学习

卷积网络

SRCNN (2014)

深度学习超分辨率的开山之作，端到端生成高分辨率图像



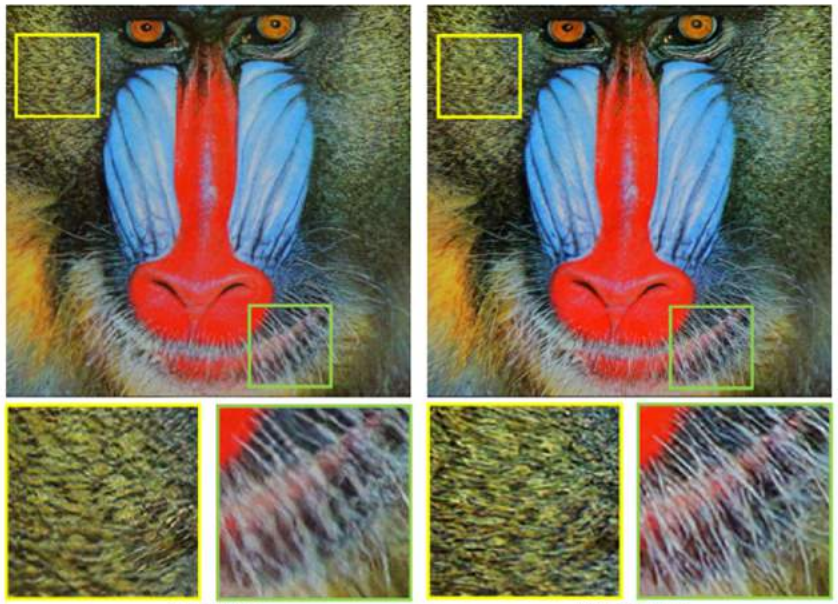
生成对抗网络

SRGAN (2016)

首次将**生成对抗网络**引入超分辨率，实现 4 倍清晰放大

ESRGAN (2018)

SRGAN的全面提升



新的方向

MetaSR (2019)

基于元学习技术实现任意倍率放大

视频超分辨率的兴起 (2015~)

视频超分辨率考虑从低分辨率视频重建高分辨率视频。相较于图像超分辨率，视频超分辨率还需要考虑图像帧之间的连贯性。

EDVR (2019)
使用可变形卷积完成帧间对齐

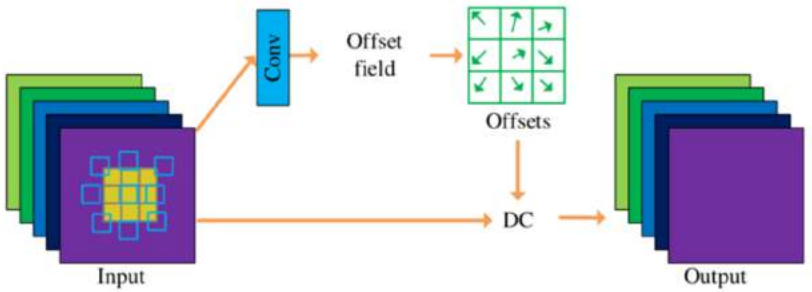
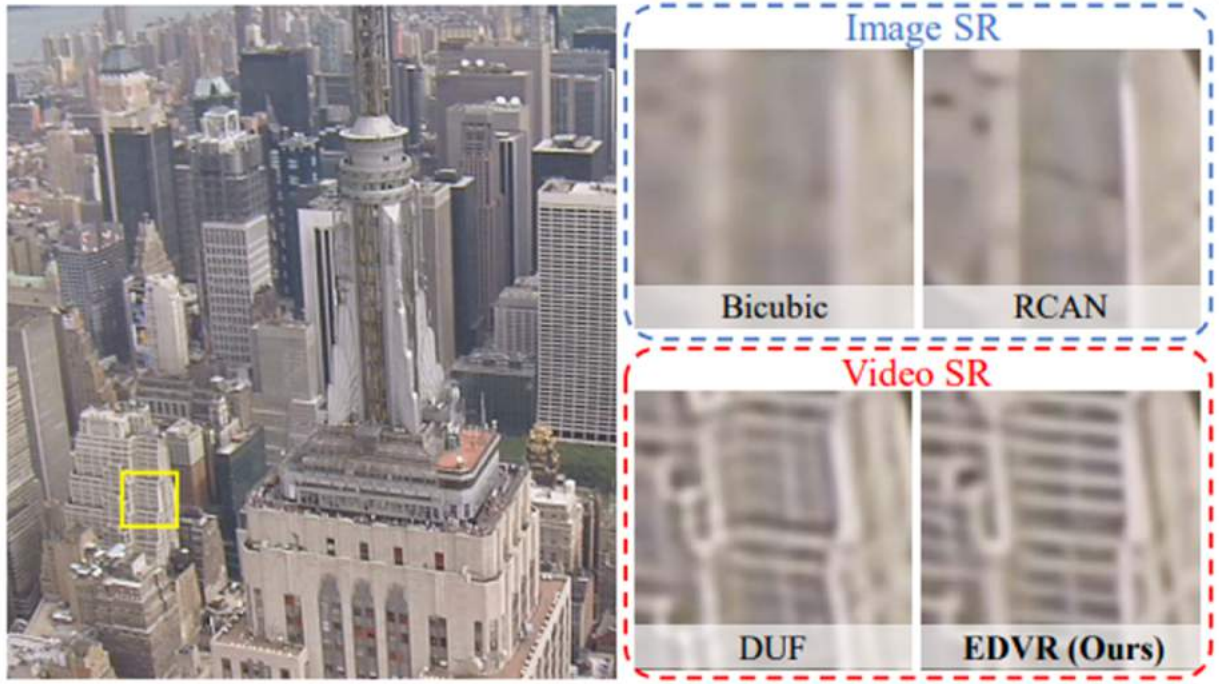


Fig. 19: A deformable convolution network.



抠图

抠图

图像抠图 Matting

目标

对图片中已知区域做精细化分割

建模

$\boldsymbol{C}_i = \alpha_i \boldsymbol{F}_i + (1 - \alpha_i)\boldsymbol{B}_i$

已知每点像素值 $\boldsymbol{C}_i$ ，预测透明度 α_i

应用

人物抠像，背景替换，影视制作

抠图的技术演进

传统视觉

Closed-Form Matting (2006)

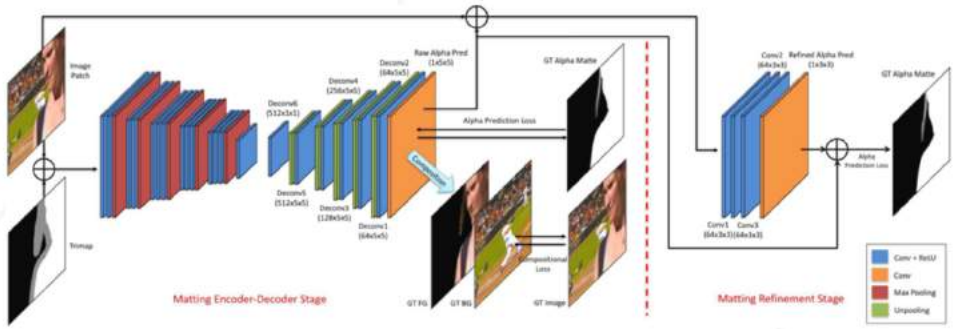
建模成全局优化问题并求解

$$J(\alpha,a,b)=\sum_{j\in I}\left(\sum_{i\in w_j}(\alpha_i-a_jI_i-b_j)^2+\varepsilon a_j^2\right)$$

深度学习方法

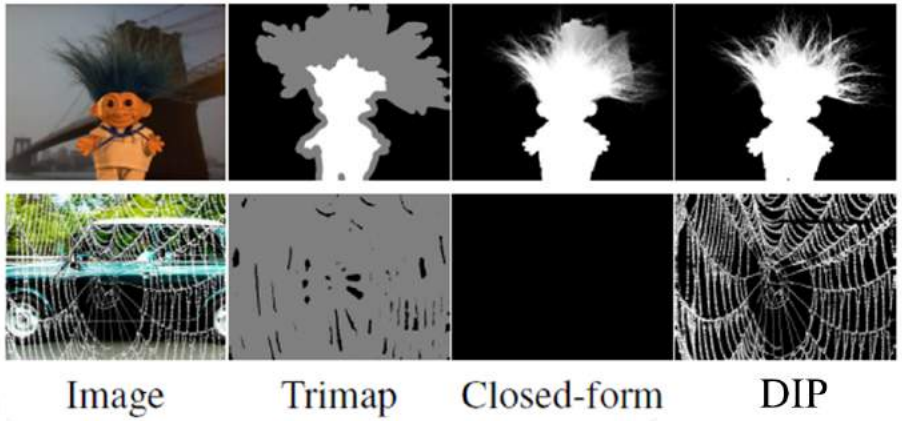
Deep Image Matting (2017)

首次使用深度学习，分两个阶段产生透明度图



FBA Matting (2020)

同时预测透明度、前景和背景



一站式图像编辑工具包 MMEditing

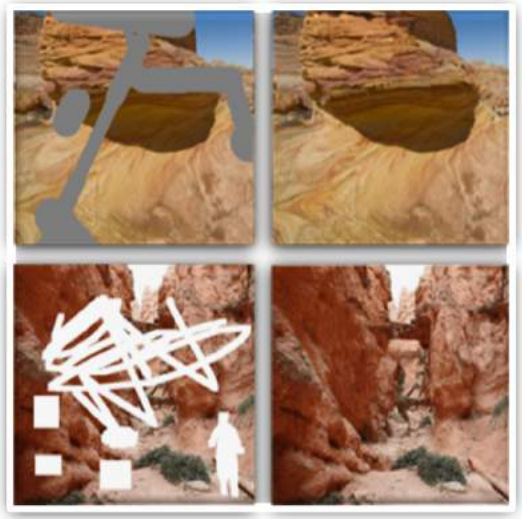
OpenMMLab

MMEditing

15 篇论文复现

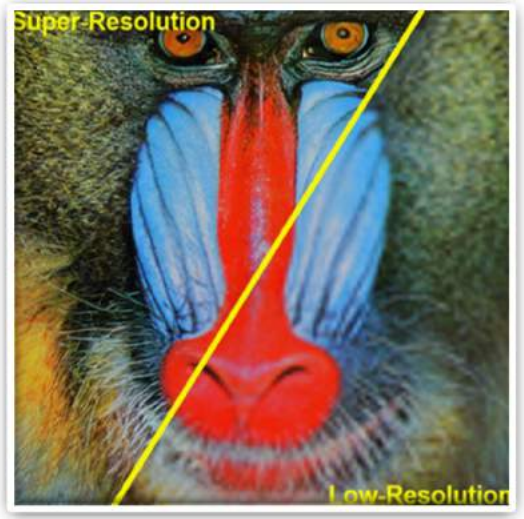
30 个预训练模型

图像修复

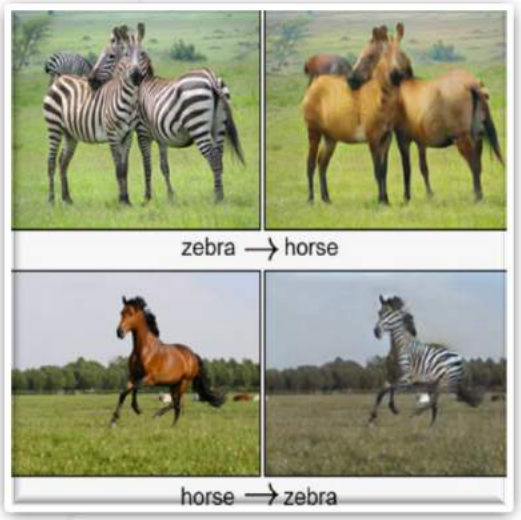


图像抠图

超分辨率



图像生成



开源社区联络站点

提供意见

交流讨论

消息渠道

OpenMMLab 社区

群号：144762544

扫一扫二维码，加入群聊。

QQ

技术文档

新闻速递

周边解读