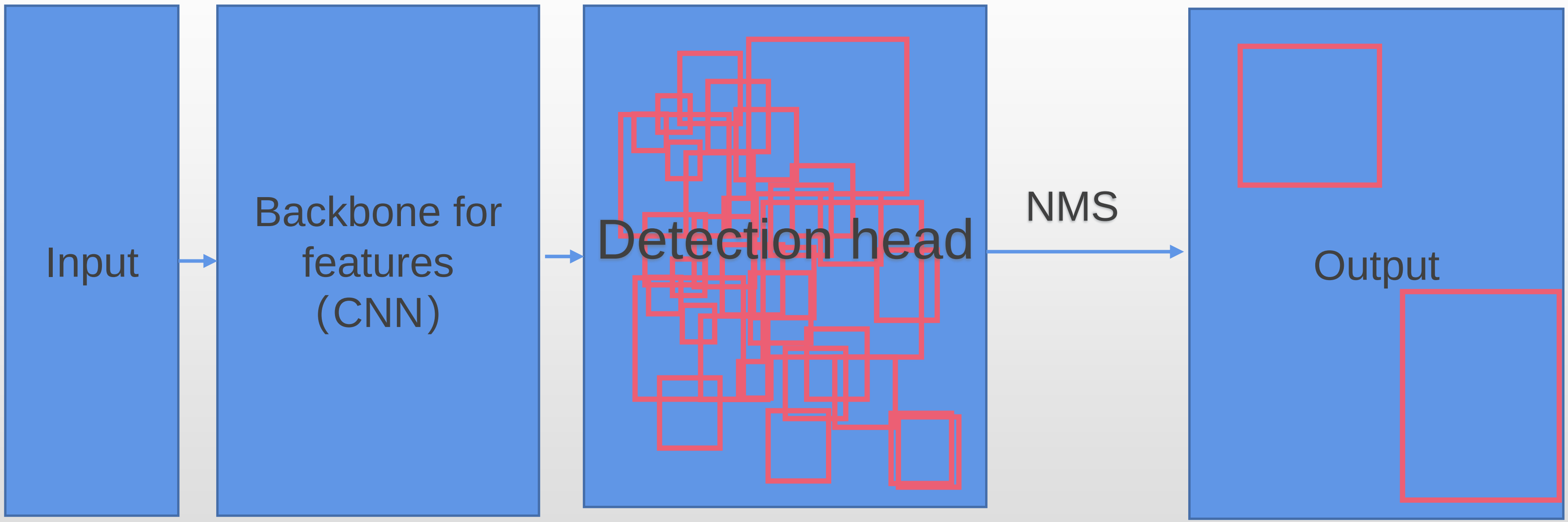
Survey of small object detection

汇报人：王浩宇

传统目标检测pipline

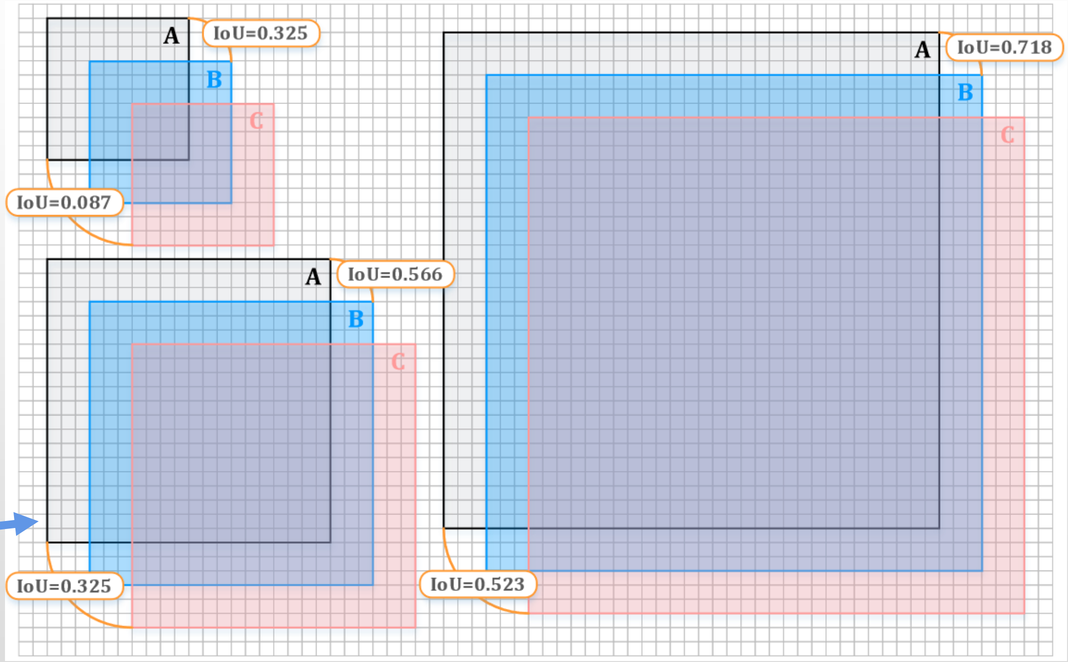


Main Challenges

模型问题:
CNN经过多次下采样丢失很多细粒度信息

数据问题:
实例尺寸过小，特征少，容易被噪音影响，容易被遮挡；样本数量不均衡（现有数据集都是偏少）

定位信息要求高:



数据增强方法

数据数量：

复制增强：单纯复制小目标实例；自适应采样（AdaResampling：基于实例分割预训练模型，比单纯复制强在结合语义和上下文信息，减少新实例放在错误位置的概率)

尺度变化： Scale match 参考预训练模型中小目标的目标大小的概率分布，调整待检测数据集中目标大小，使两者目标大小概率分布尽可能一致；Mosica数据增强，四张图片缩放拼接为一张图片，可以一定程度上增加小目标的数量。

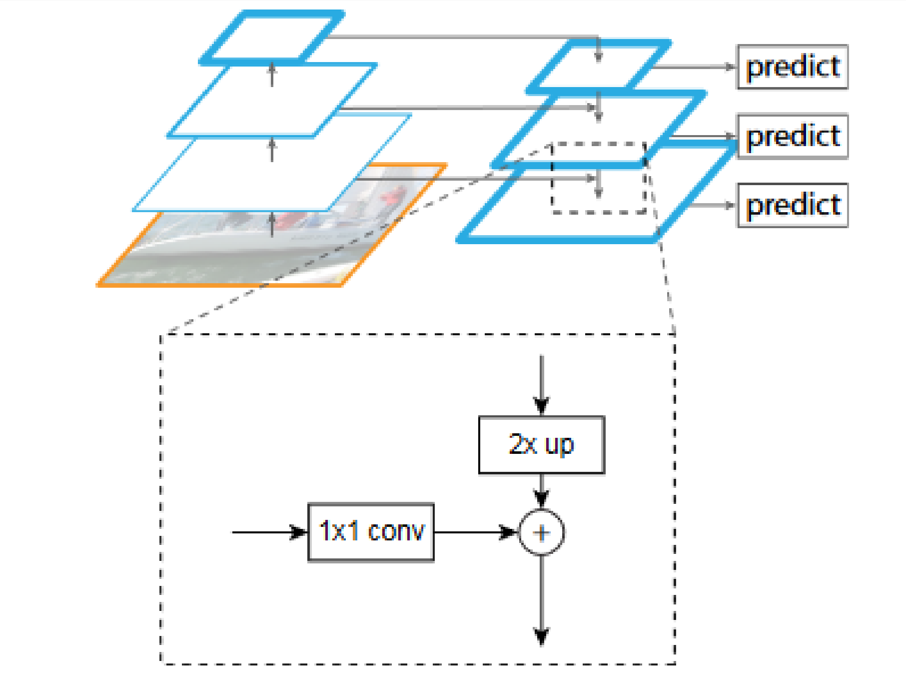
自学习数据增强：通过强化学习选择最佳数据增强策略

数据质量：

提升图像分辨率（大体类似超分辨率的思路）：插值算法；转置卷积进行上采样；基于GAN的超分辨率算法等

多尺度信息-特征融合方法

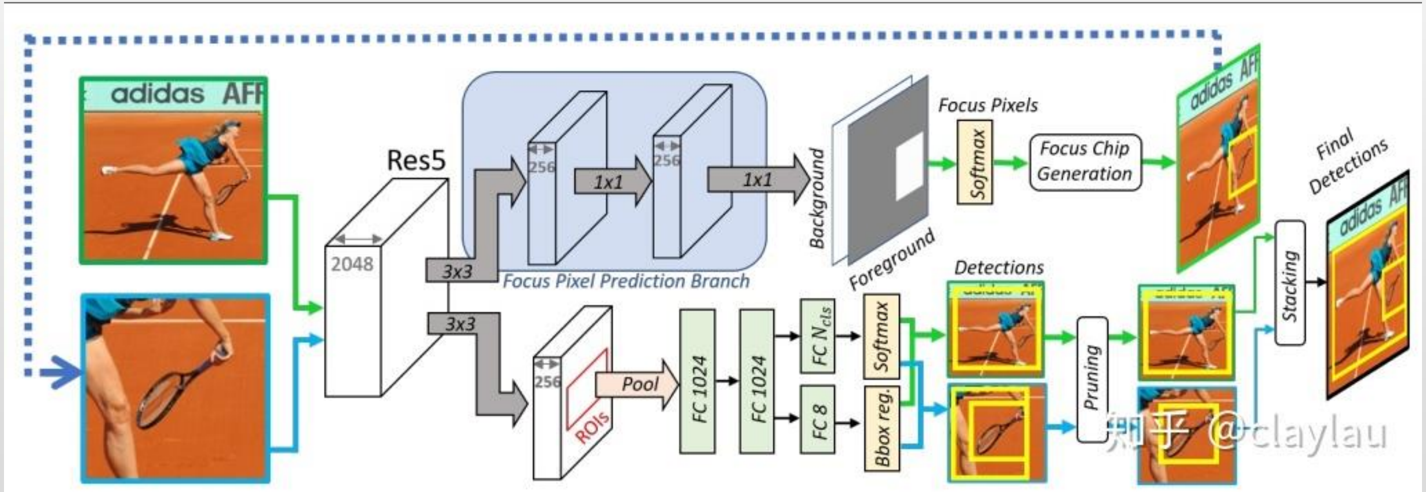
经典思想：FPN（特征金字塔）
后续工作主要基于FPN进行改进，有一些精细化的融合方法



多尺度信息-尺度匹配方法

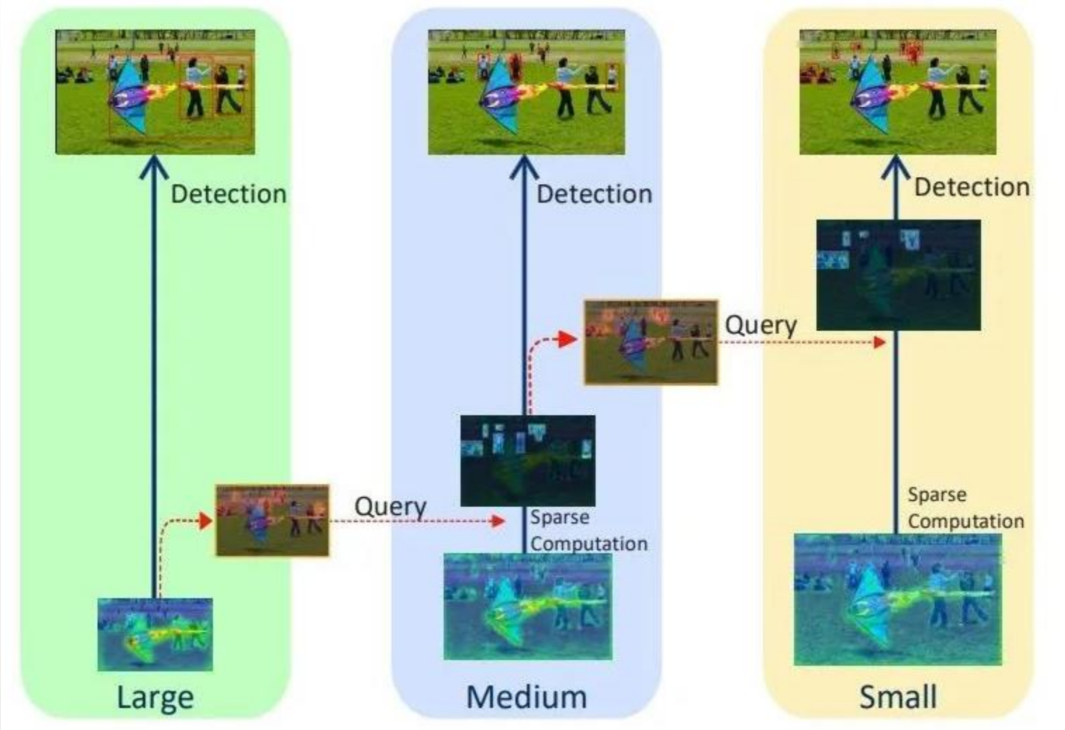
TridentNet：对于不同尺寸的物体，通过与其适配的感受野进行特征提取，得到尽可能一致的特征。（不同stride的空洞卷积，参数共享使模型特征更为统一

AutoFcous：



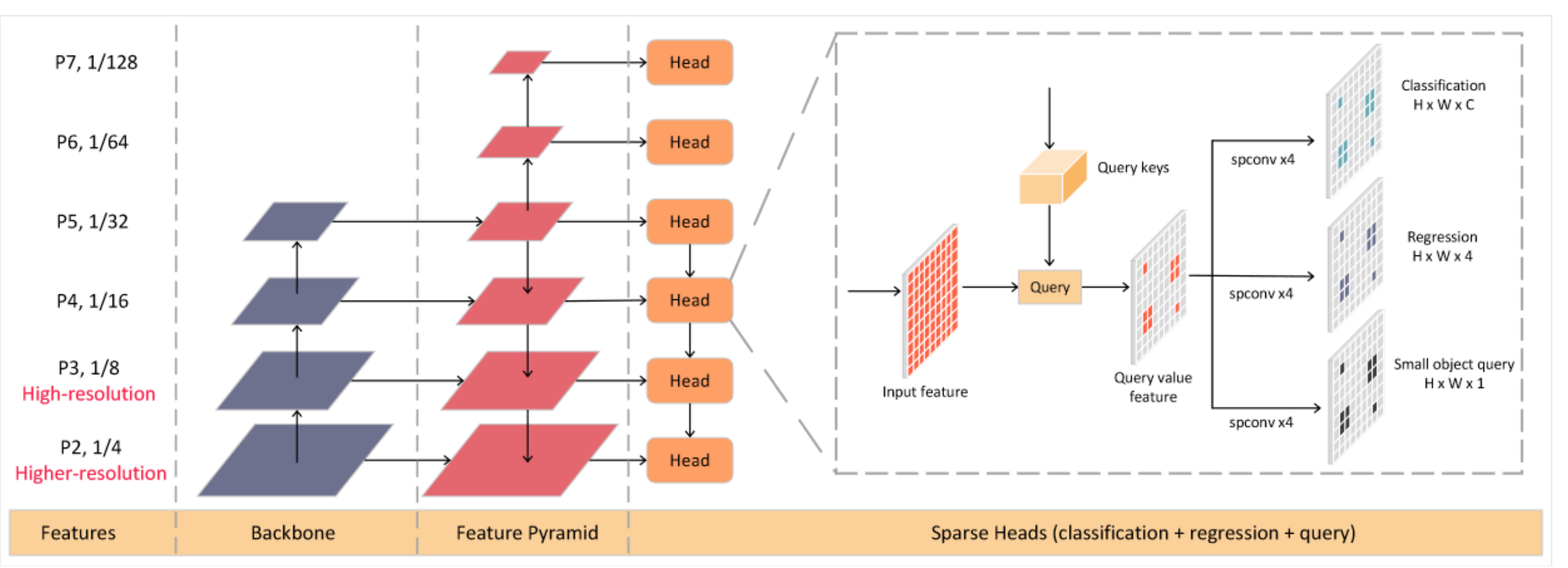
多尺度信息-尺度匹配方法

QueryDet：使用级联稀疏query加速高分辨率下的小目标检测（CVPR2022）



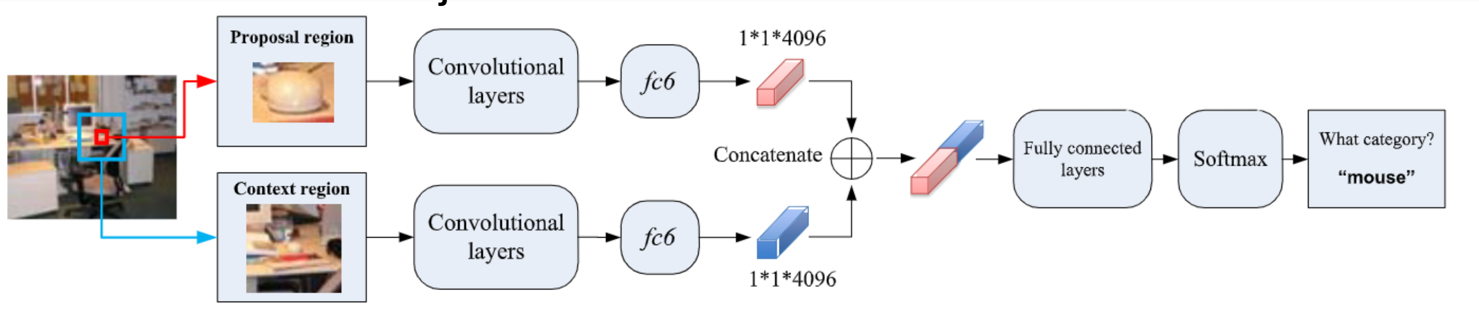
多尺度信息-尺度匹配方法

QueryDet：使用级联稀疏query加速高分辨率下的小目标检测（CVPR2022）

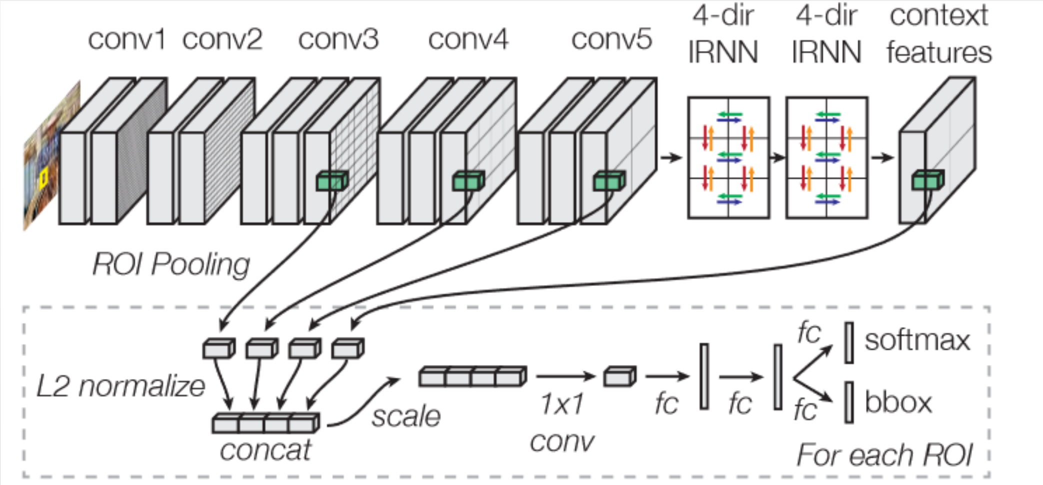


多尺度信息-上下文方法

R-CNN for Small Object Detection：



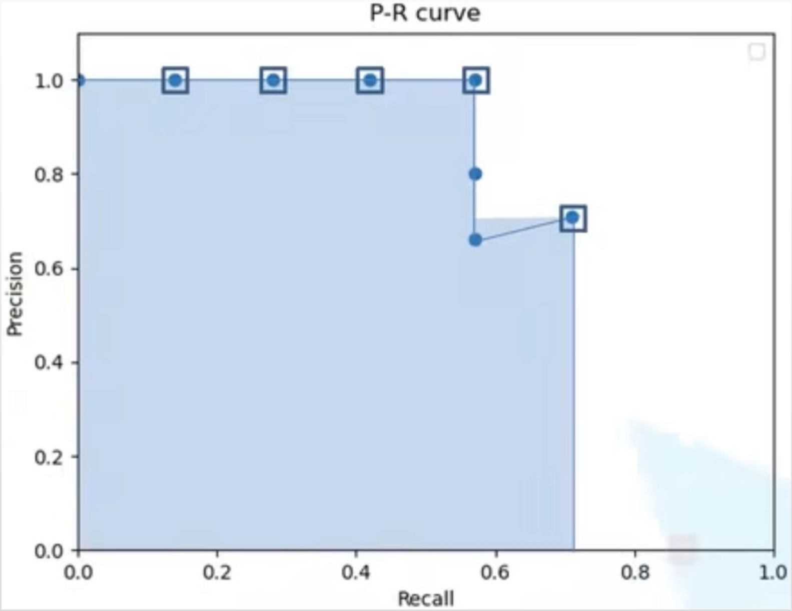
Inside-Outside Net (ION)



评估指标

正常的输出：种类，置信度，预测框（预测框与标签计算出交并比（IOU））

True Postive（TP）
预测种类正确，IOU大于设定好的阈值
False Postive（FP）
预测种类正确，IOU小于设定好的阈值
Falese Negative（FN）
未在实例上产生预测框
Precision：
TP/（TP+FP）所有预测结果中，预测正确的比例
Recall：
所有真实目标中，预测正确的比例
AP：
P-R曲线面积：recall逐渐升高的同时，precision越高越好



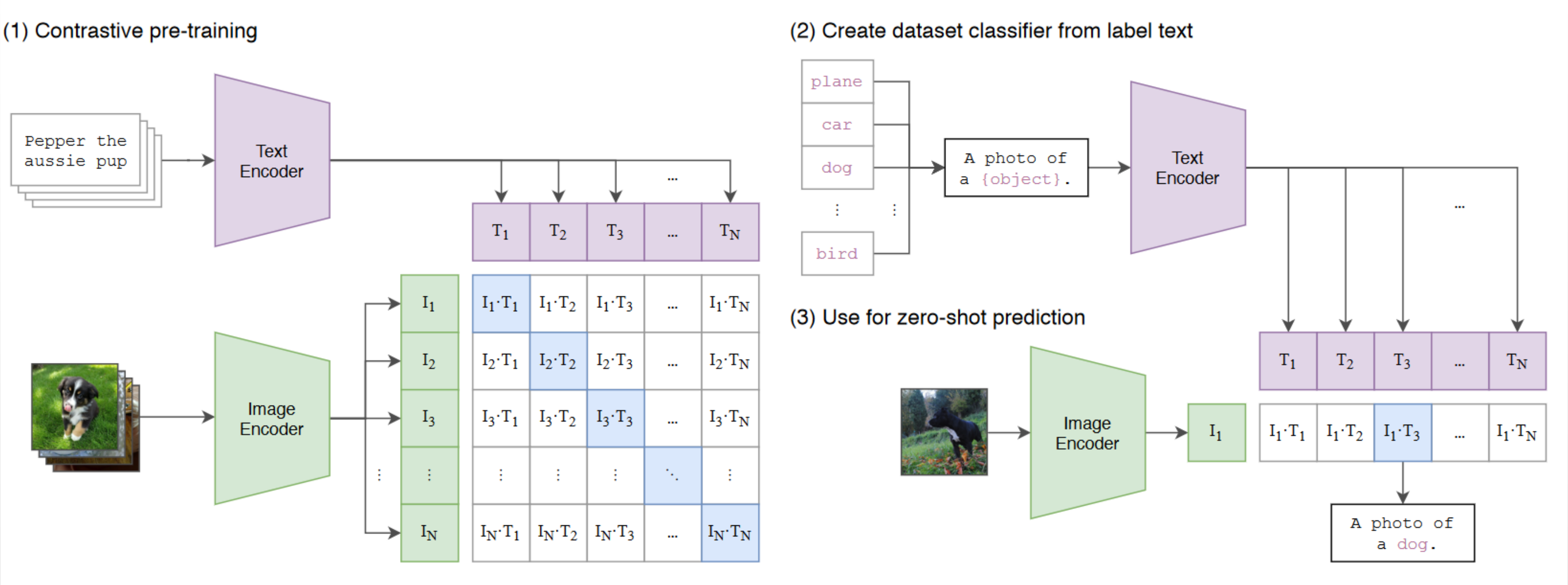
数据集

Dataset name	Task field	Publication	#Images	#Instances	Descriptions and Characteristics
COCO [6]	ODNI	ECCV 2014	123K	886K	One of the most popular datasets for generic object detection
SOD [28]	ODNI	ACCV 2016	4925	8393	A small-scale dataset for small object detection
WiderFace [8]	Face detection	CVPR 2016	32K	393K	A large-scale benchmark with rich annotations for face detection
EuroCity Persons [113]	Pedestrian detection	TPAMI 2019	47K	219K	The largest dataset for pedestrian detection captured from dozens of Europe cities
WiderPerson [114]	Pedestrian detection	TMM 2020	13K	39K	Pedestrian detection benchmark in traffic scenarios
TinyPerson [7]	Pedestrian detection	WACV 2020	1610	72K	The first dataset dedicated to tiny-scale pedestrian detection
TT100K [115]	Traffic sign detection	CVPR 2016	100K	30K	A realistic and large-scale benchmark for traffic sign detection
DIOR [20]	ODAI	JPRS 2020	23K	192K	One of the most frequently used benchmarks for object detection in aerial images
DOTA [30]	ODAI	TPAMI 2021	11K	1.79M	The largest remote sensing detection dataset including considerable small objects
AI-TOD [116]	ODAI	ICPR 2021	28K	700K	A tiny object detection dataset based on previous available datasets
NWPU-Crowd [117]	Crowd counting	TPAMI 2021	5109	2.13M	The largest dataset for crowd counting and localization to date

SOTA

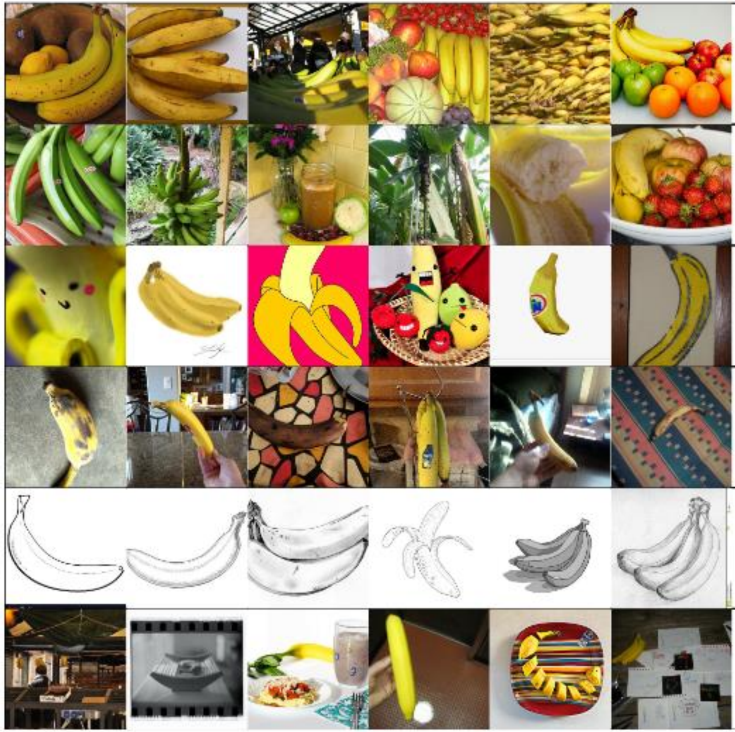
Rank	Model	box AP	AP50	AP75	APS	APM	APL	Params (M)	Extra Training Data	Paper	Code	Result	Year	Tags
1	**EVA** CLIP based	64.7	81.9	71.7	48.5	67.7	77.9		✓	EVA: Exploring the Limits of Masked Visual Representation Learning at Scale			2022	
2	**Group DETR v2** DETR based	64.5	81.8	71.1	48.4	67.2	77.1		×	Group DETR v2: Strong Object Detector with Encoder-Decoder Pretraining			2022	Group DETR, DINO, ViT-Huge
3	**GLIP** (Swin-L, multi-scale) CLIP based	61.5	79.5	67.7	45.3	64.9	75.0		×	Grounded Language-Image Pre-training			2021	multiscale, Vision Language, Dynamic Head, BERT-Base
4	**PyCenterNet** (Swin-L, multi-scale)	57.1	73.7	62.4	38.7	59.2	71.3		×	CenterNet++ for Object Detection			2022	End-to-End, Swin-Transformer, multiscale

CLIP



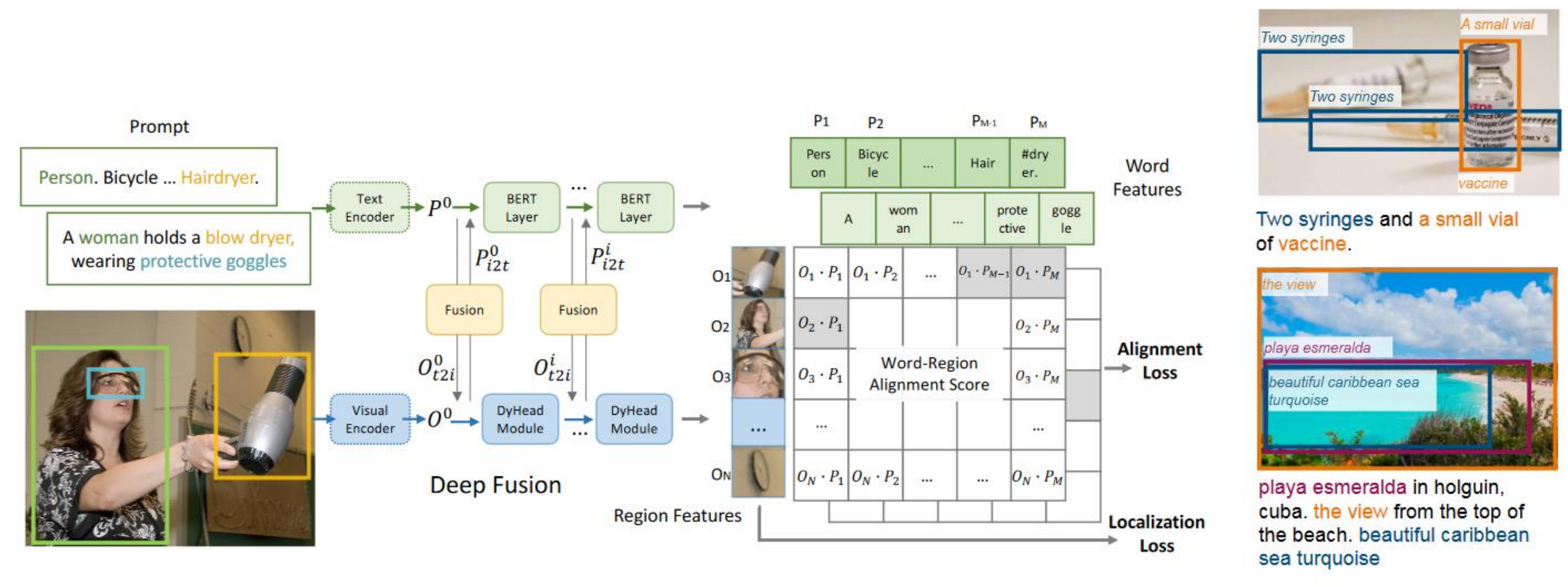
(1) Contrastive pre-training
Pepper the aussie pup
Text Encoder
Image Encoder
T1 T2 T3 ... TN
I1 I2 I3 ... IN
I1·T1 I1·T2 I1·T3 ... I1·TN
I2·T1 I2·T2 I2·T3 ... I2·TN
I3·T1 I3·T2 I3·T3 ... I3·TN
IN·T1 IN·T2 IN·T3 ... IN·TN
(2) Create dataset classifier from label text
plane
car
dog
bird
A photo of a {object}.
Text Encoder
(3) Use for zero-shot prediction
Image Encoder
I1
I1·T1 I1·T2 I1·T3 ... I1·TN
A photo of a dog.

CLIP



Dataset Examples	ImageNet ResNet101	Zero-Shot CLIP	Δ Score
ImageNet	**76.2**	**76.2**	0%
ImageNetV2	64.3	**70.1**	+5.8%
ImageNet-R	37.7	**88.9**	+51.2%
ObjectNet	32.6	**72.3**	+39.7%
ImageNet Sketch	25.2	**60.2**	+35.0%
ImageNet-A	2.7	**77.1**	+74.4%

GLIP



Prompt
Person. Bicycle ... Hairdryer.
A woman holds a blow dryer, wearing protective goggles
Text Encoder
Visual Encoder
BERT Layer
BERT Layer
Fusion
Fusion
DyHead Module
DyHead Module
Deep Fusion
Region Features
Word Features
Word-Region Alignment Score
Alignment Loss
Localization Loss
Two syringes
A small vial
vaccine
Two syringes and a small vial of vaccine.
the view
playa esmeralda
beautiful caribbean sea turquoise
playa esmeralda in holguin, cuba. the view from the top of the beach. beautiful caribbean sea turquoise

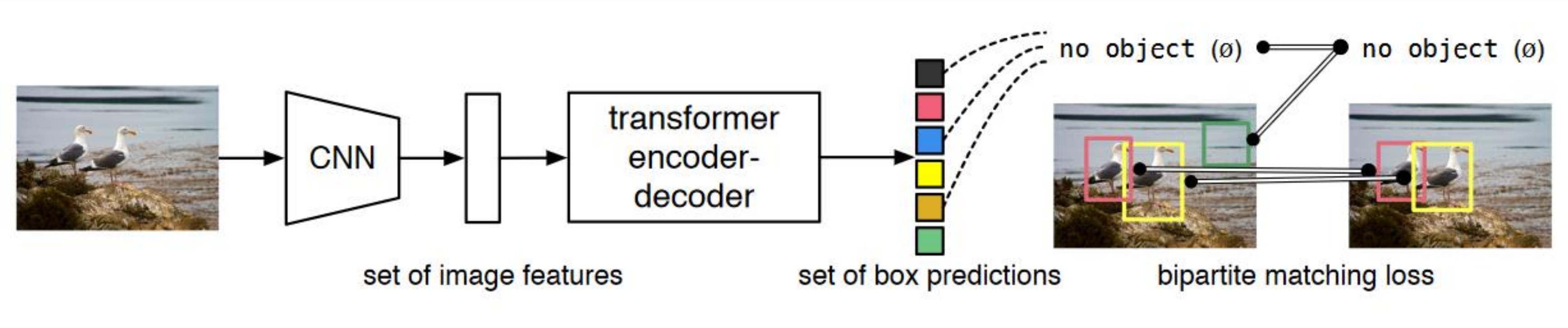
GLIP

Model	Backbone	Deep Fusion	Pre-Train Data		
			Detection	Grounding	Caption
GLIP-T (A)	Swin-T	✗	Objects365	-	-
GLIP-T (B)	Swin-T	✓	Objects365	-	-
GLIP-T (C)	Swin-T	✓	Objects365	GoldG	-
GLIP-T	Swin-T	✓	Objects365	GoldG	Cap4M
GLIP-L	Swin-L	✓	FourODs	GoldG	Cap24M

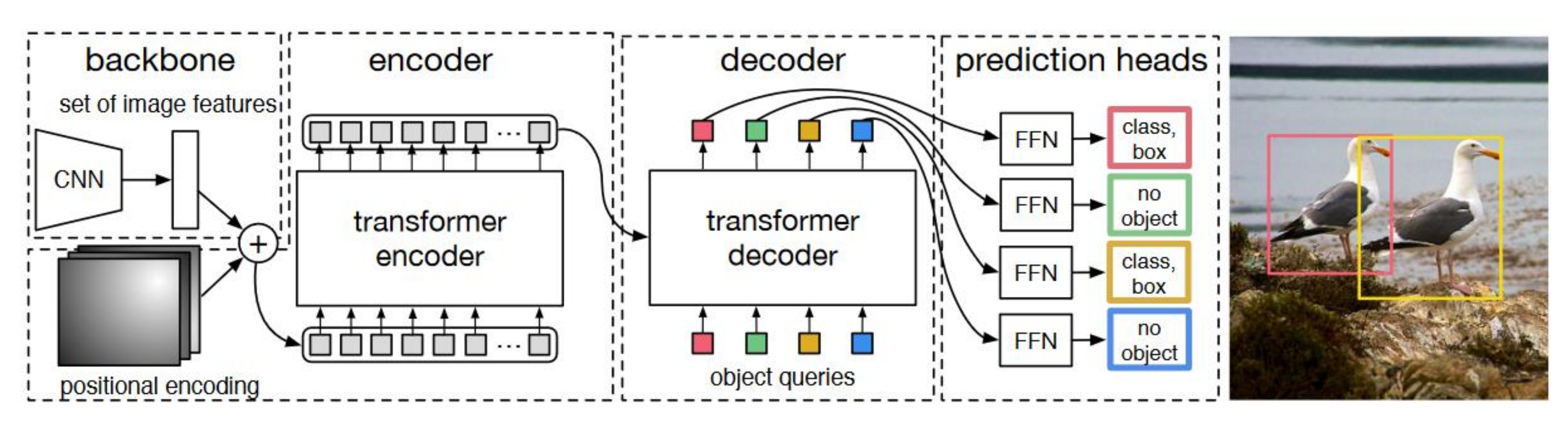
Table 1. A detailed list of GLIP model variants.

Model	Backbone	Pre-Train Data	Zero-Shot 2017val	Fine-Tune 2017val / test-dev
Faster RCNN	RN50-FPN	-	-	40.2 / -
Faster RCNN	RN101-FPN	-	-	42.0 / -
DyHead-T [9]	Swin-T	-	-	49.7 / -
DyHead-L [9]	Swin-L	-	-	58.4 / 58.7
DyHead-L [9]	Swin-L	O365,ImageNet21K	-	60.3 / 60.6
SoftTeacher [58]	Swin-L	O365,SS-COCO	-	60.7 / 61.3
DyHead-T	Swin-T	O365	43.6	53.3 / -
GLIP-T (A)	Swin-T	O365	42.9	52.9 / -
GLIP-T (B)	Swin-T	O365	44.9	53.8 / -
GLIP-T (C)	Swin-T	O365,GoldG	**46.7**	55.1 / -
GLIP-T	Swin-T	O365,GoldG,Cap4M	46.3	54.9 / -
GLIP-T	Swin-T	O365,GoldG,CC3M,SBU	46.6	**55.2** / -
GLIP-L	Swin-L	FourODs,GoldG,Cap24M	**49.8**	**60.8** / 61.0
GLIP-L	Swin-L	FourODs,GoldG+,COCO	-	- / **61.5**

DETR

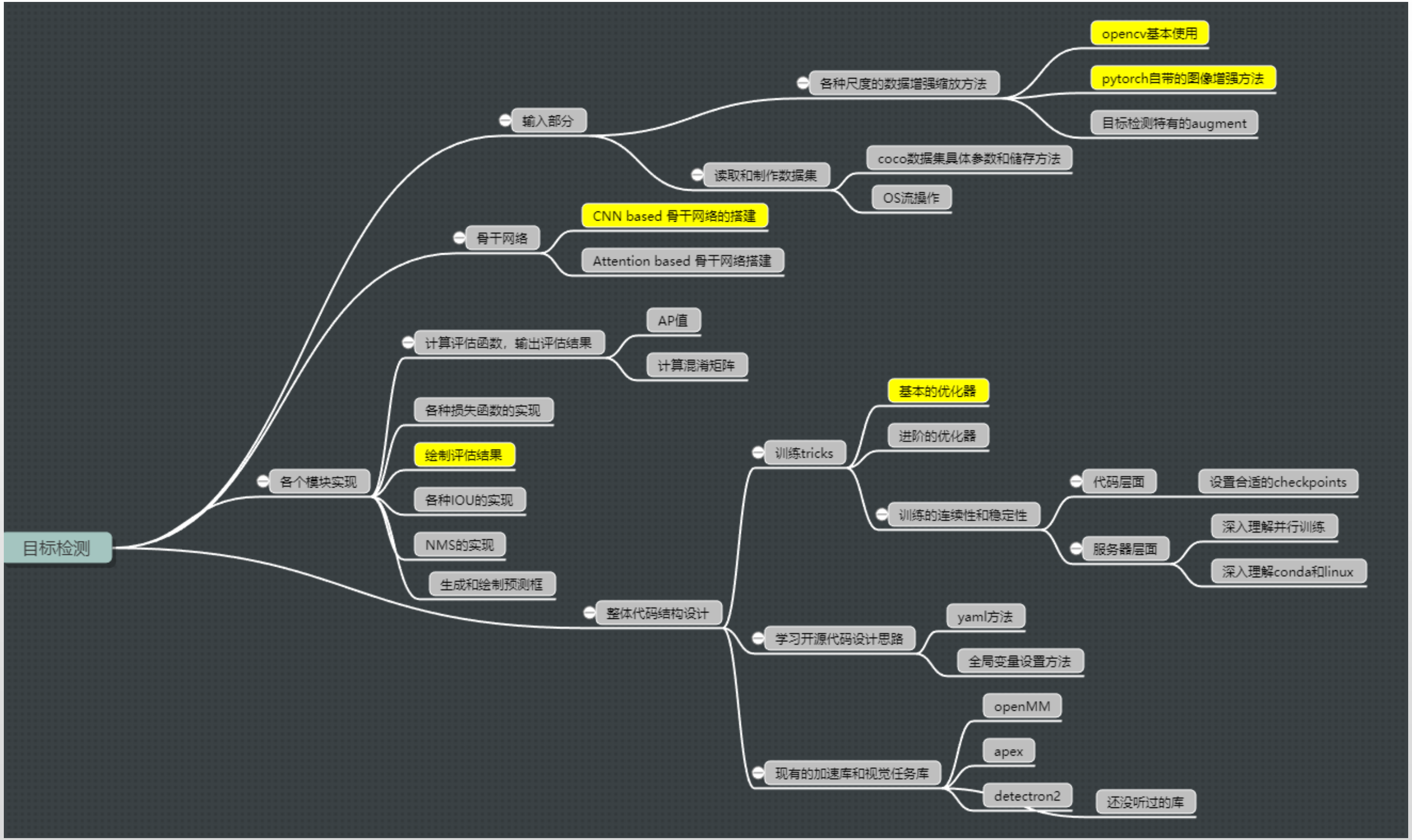


CNN
transformer encoder-decoder
no object (ø)
no object (ø)
set of image features
set of box predictions
bipartite matching loss



backbone
set of image features
CNN
positional encoding
encoder
transformer encoder
decoder
transformer decoder
object queries
prediction heads
FFN
class, box
no object
class, box
no object

需要点亮的技能树



目标检测
输入部分
各种尺度的数据增强缩放方法
opencv基本使用
pytorch自带的图像增强方法
目标检测特有的augment
读取和制作数据集
coco数据集具体参数和储存方法
OS流操作
骨干网络
CNN based 骨干网络的搭建
Attention based 骨干网络搭建
各个模块实现
计算评估函数，输出评估结果
AP值
计算混淆矩阵
各种损失函数的实现
绘制评估结果
各种IOU的实现
NMS的实现
生成和绘制预测框
整体代码结构设计
训练tricks
基本的优化器
进阶的优化器
训练的连续性和稳定性
代码层面
设置合适的checkpoints
服务器层面
深入理解并行训练
深入理解conda和linux
学习开源代码设计思路
yaml方法
全局变量设置方法
现有的加速库和视觉任务库
openMM
apex
detectron2
还没听过的库

Thank you！