

# mmdetection 解析

Sisyphes,yehao

2019 年 8 月 1 日

## 目录

<b>第一节 结构设计</b>	<b>3</b>
1.1 总体逻辑	4
1.2 Configs	4
1.3 Backbone	5
1.4 Necks	5
1.4.1 多尺度结构	5
1.5 Heads	5
1.5.1 AnchorHead	6
1.5.2 SSDHead	9
1.6 Losses	10
1.6.1 基本认识	10
1.6.2 实现解析	11
1.7 Detectors	11
1.7.1 maskrcnn	11
1.7.2 RepPoints	14
1.7.3 RetinaFace	14
<b>第二节 数据处理</b>	<b>14</b>
<b>第三节 FP16</b>	<b>14</b>
<b>第四节 训练 pipeline</b>	<b>14</b>
<b>第五节 更改模型</b>	<b>16</b>

目录	2
第六节 新增模型	16
第七节 numpy,torch 某些基础函数	16
第八节 计划	16

## 第一节 结构设计

- Backbone: 特征提取骨架网络, ResNet, ResNeXt 等.
- Neck: 连接骨架和头部. 多层次特征融合, FPN, BFP 等.
- DenseHead: 处理特征图上的密集框部分, 主要分 AnchorHead, AnchorFreeHead 两大类, 分别有 RPNHead, SSDHead, RetinaHead 和 FCOSHead 等.
- RoIExtractor: 汇集不同层级的特征框, 将其大小统一, 为二步定位, 类别优化服务.
- RoIHead (BBoxHead/MaskHead): 类别分类或位置回归等.
- OneStage: Backbone + Neck + DenseHead
- TwoStage: Backbone + Neck + (DenseHead) + RoIExtractor + RoIHead

代码结构:

configs 网络组件结构等配置信息

tools: 训练和测试的最终包装

mmdet:

apis: 分布式环境设定, 推断和训练基类代码

core: anchor, bbox, mask 等在训练前和训练中的各种变换函数

datasets: coco 和 voc 格式的数据类以及一些增强代码

models: 模型组件, 采用注册和组合构建的形式完成模型搭建

ops: 优化加速代码, 包括 nms, roialign, dcn, gcb, mask, focal\_loss 等

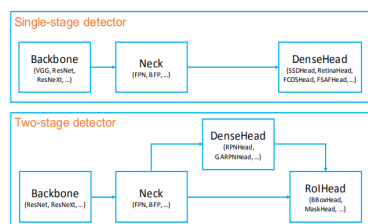


图 1: Framework

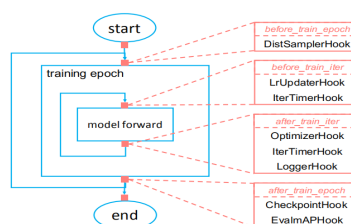


图 2: Training pipeline

## 1.1 总体逻辑

在最外层的 train.py 中能看到:

1. `mmcv.Config.fromfile` 从配置文件解析配置信息, 并做适当更新, 包括预加载模型文件, 分布式相关等
2. `mmdet.models.builder` 中的 `build_detector` 根据配置信息构造模型
  - 2.5 `build` 函数调用 `_build_module`(新版为 `build_from_cfg`) 函数, 按 `type` 关键字从注册表中获取相应的模型对象, 并根据配置参数实例化对象 (配置文件的模型参数只占了各模型构造参数的一小部分, 模型结构并非可以随意更改).
  - 2.6 `registr.py` 实现了模型的注册装饰器, 其主要功能就是将各模型组件类对象保存到 `registry.module_dict` 中, 从而可以实现 2.5 所示功能.
  - 2.7 目前包含 `BACKBONES,NECKS,ROI_EXTRACTORS,SHARED_HEADS,HEADS,LOSSES,DETECTORS` 七个 (容器). 注册器可按 `@NAME.register_module` 方式装饰, 新增. 所有被注册的对象都是一个完整的 `pytorch` 构图
  - 2.9 `@DETECTORS.register_module` 装饰了完整的检测算法 (`OneStage`, `TwoStage`), 各个部件在其 `init()` 函数中实例化, 实现 2.5 的依次调用.
3. 最后是数据迭代器和训练 pipeline四.

## 1.2 Configs

配置方式支持 `python/json/yaml`, 从 `mmcv` 的 `Config` 解析, 其功能同 `maskrcnn-benchmark` 的 `yacs` 类似, 将字典的取值方式属性化.

配置文件模型部分包含模型组件及其可改动模型结构的参数, 比如 `backbone` 的层数, 冻结的 `stage`; `bbox_head` 的 `in_channel`, 类别, 损失函数等; 训练部分主要包括 `anchor` 采样相关系数; 测试包括非极大抑制等相关参数; 剩下数据, 优化器, 模型管理, 日志等相关信息, 一看即明.

### 1.3 Backbone

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112x112					
7x7, 64, stride 2						
3x3 max pool, stride 2						
conv2.1	56x56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv2.5	28x28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 2$
conv3.4	14x14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 26$
conv5.5	7x7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1x1			average pool, 000-kfc, softmax		
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

图 3: resnet

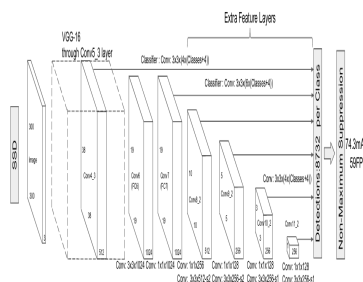


图 4: ssd

## 1.4 Necks

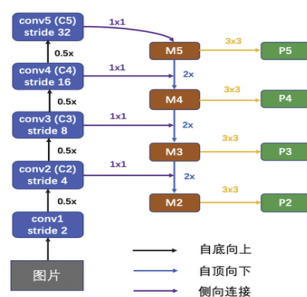


图 5: fpn

### 1.4.1 多尺度结构

主要分析 fpn 结构的变体, 比如 ssd 相关的 FaceBoxes, HRNet 等.

## 1.5 Heads

Heads 主要包含了三部分,anchor\_heads, bbox\_heads, mask\_heads,anchor是 two-stage 的 rpn 部分或者 one-stage 的头部. 最终的损失函数均在这里实现.

### 1.5.1 AnchorHead

anchor head 主要包括 AnchorGenerator, anchor\_target, 后者包含了 assign,sample. 因为不同尺度操作雷同, 于是作者设计了一个公共函数 multi\_apply. 这一小节主要说明以上函数, 损失函数见1.6.

AnchorGenerator 类为不同特征层生成 anchor, 输入参数 base\_size, scales, ratios 分别表示:anchor 在特征层上的基础大小 (特征层相对于原图的 stride),anchor 在特征层上的尺度大小 (可以多个, 增加感受野),anchor 在保持基础大小不变的情况下的长宽比.  $\text{len}(\text{scales}) * \text{len}(\text{ratios})$  即为一格的 anchor 个数.

比如输入图像大小 (640\*640), 选择 (p2, p3) 作为其特征层, 则 p2 大小为 (160\*160),base\_size=4, 若设定 ratios=[0.5,1.0,2.0], scales=[8, 16], 则在 p2 上一格对应的 base\_anchor 的 (w,h) 为 [(45.25,22.63), (90.51, 45.25), (32.00, 32.00), (64.00, 64.00), (22.63, 45.25),(45.25, 90.51)]. 其中  $64 = 4 * 16 * 1$ ,  $90.51 = 4 * 16 * \sqrt{2}$ ,  $22.63 = 4 * 8 / \sqrt{2}$ . 那么每一格所对应的 6 个 base\_anchor 相对于中心点的偏移量即为:

```
1
2      [[-21.,  -9.,  24.,  12.],
3       [-43., -21.,  46.,  24.],
4       [-14., -14.,  17.,  17.],
5       [-30., -30.,  33.,  33.],
6       [ -9., -21.,  12.,  24.],
7       [-21., -43.,  24.,  46.]]
```

因为这里都是相对于自身一格的偏移量, 所有当要算在原图上的中心偏移量时, 直接根据 torch 或者 numpy 的广播机制, 首先得到特征图在原图的网格坐标, 然后和此相加, 即得此特征层的所有 all\_anchors. 见 grid\_anchors 函数. 这里面涉及到一些操作技巧, 比如:

```
1 ws = (w * w_ratios[:, None] * self.scales[None, :]).view(-1)
2 hs = (h * h_ratios[:, None] * self.scales[None, :]).view(-1)
```

包括 nms,assign 等部分都有很多细节操作, 这些实现检测算法的基础函数, 需要多加练, 才能复现新的算法. 可见七. 如上不难看到, 输入图像的所有 anchor 数量巨大.

模块 anchor 中还有一个 anchor\_target 文件, 其主要是为了给设计的 anchor 标定 label 以及筛选 anchor(便于优化等).

这里利用的 anchor\_target\_single 函数来解释 multi\_apply 函数:

```

1 from functools import partial
2 from six.moves import map, zip
3 def multi_apply(func, *args, **kwargs):
4     pfunc = partial(func, **kwargs) if kwargs else func
5     map_results = map(pfunc, *args)
6     return tuple(map(list, zip(*map_results)))

```

partial 函数的功能就是：把一个函数的某些参数给固定住，返回一个新的函数（这里就是将 \*\*kwargs 中的参数固定住）。当函数参数太多，需要固定某些参数时，可以使用 functools.partial 创建一个新的函数 map(function, sequence)，对 sequence 中的 item 依次执行 function(item)，并将结果组成一个迭代器返回，最后 zip 并行循环。

在 anchor\_target 中

```

1 result = multi_apply(
2     anchor_target_single,
3     anchor_list,
4     valid_flag_list,
5     gt_bboxes_list,
6     gt_bboxes_ignore_list,
7     gt_labels_list,
8     img_metas,
9     target_means=target_means,
10    target_stddevs=target_stddevs,
11    cfg=cfg,
12    label_channels=label_channels,
13    sampling=sampling,
14    unmap_outputs=unmap_outputs)

```

是先将配置文件，采样与否的 flag，anchor 于 gt\_box 在映射空间（平移归一，放缩 log）中各偏差量的滑动平均处理（见 core/bbox/transforms/bbox2delta 函数最后一行）固定住。然后 anchor\_target\_single 对 \*args 参数：anchor\_list, gt\_bboxes\_list, img\_metas 等并行处理。得到最终的 label, bbox, 正负样本 index。这里 label\_weights 用来记录筛选后的信息位置。

\*args 参数的说明：

- anchor\_list: 见 anhor\_head 中的 get\_anchor 函数，是一个 list[list[Tensors]] 结构，最外层是图片个数，再内一层是尺度个数，里面的 Tensors 的 shape 是 [H\*W\*4, 4]，其中 H 和 W 代表对应尺度特征图的高和宽。
- img\_metas 有五个字段: ori/orig\_img\_shape, scale\_factor, flip

在调用 `anchor_target_single` 函数之前:

```

1 num_imgs = len(img metas)
2 assert len(anchor_list) == len(valid_flag_list) == num_imgs
3 # anchor number of multi levels
4 # 选出第一张图的各个特征层的anchor数量, 比如上面给的p2即为160*160*6, 一格点对
   应的anchor shape为(6,4)
5 # 因为有多层, 比如p2,p3, 则[(160*160*4,4), (80*80*6,4)]
6 # 为images_to_levels函数, 切片用: 将所有以图片为第一维度的结果, 转换成以特征图
   尺度个数为第一维度的结果(算loss)
7 num_level_anchors = [anchors.size(0) for anchors in anchor_list[0]]
8 # concat all level anchors and flags to a single tensor
9 #在调用multi\_apply之前需要将anchor\_list的不同层级的anchor cat在一起, 最终变
   为list [Tensor] 结构.
10 #最外层为图像个数. 这样就能利用multi\_apply中的map在每张图上做做同样的操作了.
11 #将所有图的anchor合在一起
12 for i in range(num_imgs):
13     assert len(anchor_list[i]) == len(valid_flag_list[i])
14     anchor_list[i] = torch.cat(anchor_list[i])
15     valid_flag_list[i] = torch.cat(valid_flag_list[i])
16 # valid_flag_list结合meta信息, 进一步筛选无效anchor(其实筛掉的很少)

```

`anchor_target_single` 函数: 主要涉及 `assign_and_sample`, `bbox2delta`, `unmap` 四个函数.

`assign_and_sample` 根据 `cfg` 从配置信息拿到对应的 `assign`, `sample` 类对象, `one-stage` 模型没有 `sample` (不代表损失函数利用所有的 `anchor`), `sample` 主要为 `two-stage` 服务.

采样主要包含 `RandomSampler`, `OHEMSampler`, `InstanceBalancedPosSampler`, `IoUBalancedNegSampler`, 均继承自 `BaseSampler` 为一抽象类, 继承自它的类必须完成 `_sample_pos`, `_sample_neg` 两函数. 按照 python 的语法, 子类可以使用 `raise NotImplementedError` 来避免不能实例化的问题. `PseudoSampler` 即是如此, 它重写了 `sample` 函数, 该函数并没有做任何筛选. `RandomSampler` 的意思自明, 其他采样后补. 标定 `MaxIoUAssigner` 见 1.7.1.

`bbox2delta` 主要注意一点的是相对位置是相对于 `anchor` 的, 所以除的是 `anchor` 的  $w, h$

$$d_x = (g_x - p_x)/p_w, d_y = (g_y - p_y)/p_h, d_w = \log(g_w/p_w), d_h = \log(g_h/p_h).$$

其中  $x, y$  是中心坐标. 这里涉及到一些优化目标函数的变量的”变换”问题, 比如这里若  $g_w - p_w = 1 + \epsilon$ , 则  $d_w = \epsilon$ . 如此似乎可以将其替换为



$d_w = g_w/p_w - 1$ , 实际上这样的效果可能还是不及  $\log$ , 因为一开始它就将变化幅度压缩了. 另外, 相对位置的归一化差异是关键, 比如人脸识别中, landmarks 的优化, 通常损失函数为欧氏距离, 如何对变量进行处理? 若是先预测 bbox, 再预测 landmarks, 可以相对于 bbox 的中心坐标做变换, 若直接预测呢?

### 1.5.2 SSDHead

ssd 结构的检测网络, 目前已有 ssd300, ssd512, 结构细节参考 1.3. 从配置文件中可有看到, 它没有 neck, 因层级结构在 backbone 实现.

ssdhead 继承自 anchorhead, 主要功能为处理多层级特征上的 anchor 构造和 target 标定与筛选, 基本的 feauturemap 上的 acnhor 生成由 mmdet.core.anchor 中的 AnchorGenerator 完成, 优化目标 anchor 由 anhor\_target 完成. ssdhead 中 forward 前向返回各层级对应的类别分数和坐标信息, loss 函数则得到对应的损失函数, 以字典的形式返回, 最终求导时, 汇总成一个值, 同时也能计算各个部分损失函数的均值, 方差, 方便优化, debug.

此处的难点在于 anchor 的设定和 target 的标定, 筛选. 现就 anchor 这一块细说如下:

anchor 基本介绍: anchor 设计和 caffe ssd anchor 设计一致, 假设 min\_size 为  $a$ , max\_size 为  $b$ , 则先生成 ratio 为 1, 宽度和高度为  $(a, a), (\sqrt{ab}, \sqrt{ab})$  的两个 anchor, ratio 为 2, 1/2, 3, 1/3 则分别生成宽度和高度为  $(a * \sqrt{ratio}, a/\sqrt{ratio})$  的 anchor, mmdetection 中必须设定每一层的 min\_size, max\_size, 因此 ratios 为 [2] 则对应 4 个 anchor, ratios 为 [2, 3] 则对应 6 个 anchor.

在 init() 函数中, 先生成 min\_size, max\_size, 注意它这里是必须要指定 max\_size (和 caffe SSD 不同, 无法生成奇数个 anchor), 确保  $\text{len}(\text{min\_size}) = \text{len}(\text{max\_size})$ , 调用 AnchorGenerator() 类生成了 base\_anchors, 数量是 6 或者 10, 使用 indices 操作从 6 个 anchor 里选择 (0, 3, 1, 2) 或者从 10 个 anchor 里选择 (0, 5, 1, 2, 3, 4) → 最终生成 4 个或者 6 个 anchor. 于在多个 feature map 上生成 anchor, 因此使用了一个 for 循环操作, 将 anchor\_generator 放入到 anchor\_generatos[] 中.

AnchorGenerator 类, init() 函数需要如下参数:

- base\_size: 即设置的 min\_size

- scales: 是  $(1, \sqrt{\max\_size/\min\_size})$ , 用来生成 ratio 为 1 的两个 anchor
- ratios: 是  $(1, 2, 1/2)$  或者  $(1, 2, 1/2, 3, 1/3)$
- ctr: ctr 由 stride 生成, 是 anchor 的中心坐标,  $(\frac{stride-1}{2}, \frac{stride-1}{2})$  在 `gen_base_anchor()` 函数里, 使用上面的参数来计算 `base_anchor`, 计算流程如下:
  - 根据 ratios 来计算 `h_ratios` 和 `w_ratios`, 即上面所述的  $(1/\sqrt{ratios}, \sqrt{ratios})$ .
  - 根据 scales 来计算 `base_size`, 一共有 2 个分别是
 
$$(\min\_size, \sqrt{\min\_size * \max\_size}) = \min\_size * scales$$
  - 计算 anchors 的宽度和高度, 只以宽度举例:  $w = base\_size * w\_ratios$ , 以 ratios 是  $(1, 2, 1/2)$  举例, `base_size` shape 为  $(2, 1)$ , `w_ratios` shape 为  $(1, 3)$ , 计算出的 `w` 是  $(2, 3)$  一共生成了 6 个 anchor, 如果 ratios 是  $(1, 2, 1/2, 3, 1/3)$ , 则生成 10 个 anchor (此处 anchor 数量和标准 `ssd` anchor 数量不一致 → 再筛选 (即 `ssd_head.py` 中使用 `indices` 操作进行筛选))

## 1.6 Losses

### 1.6.1 基本认识

**Focalloss**. 想像一下特征层上的锚框, 远离 `gt` 的必然占绝大多数, 围绕在 `gt` 周围的 `bbbox` 有模棱两可的 (正负), 故有正负难易之分. `FocalLoss` 是为解决难易样本的不平衡问题. 公式如下:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \text{ 其中 } p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{else.} \end{cases} \quad (1)$$

$\alpha_t$  缓解样本不平衡现象,  $(1 - p_t)^\gamma$  为降低易分样本的损失值 (考虑到易分样本比例高). 公式可以这样理解: 对于正样本分对了, 则  $p_t \rightarrow 1$ ,  $(1 - p_t)^\gamma$  有减缓效果, 当分错了, 即  $p_t \ll 0.5$ , 则  $1 - p_t \gg 0.5$ , 效果和 `CrossEntropy` 没差, 对于负样本也是如此. 因此起到压缩图 6 中的左图右端, 右图左端的效果.

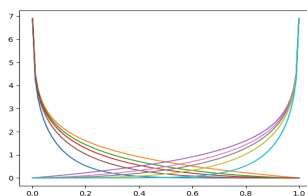


图 6: FocalLoss

**GHM**. GHM 算是对 Focal 改进, 作者统计了样本的梯度信息, 提出梯度均衡机制, 让各种难度类型的样本有均衡的累计贡献. 具体细节看论文即可.

**IoULoss**. **BoundedIoULoss**.

### 1.6.2 实现解析

CrossEntropyLoss, SmoothL1Loss, Focalloss, GHMLoss, BalancedL1Loss, IoULoss, BoundedIoULoss, ArcLoss

## 1.7 Detectors

这节主要分析 maskrcnn 和 reppoints, retinaFace 三个算法.

### 1.7.1 maskrcnn

以配置文件 mask\_rcnn\_r50\_fpn\_1x.py 为例说说 twao\_stage 的实现过程. 配合 two\_stage 的 forward\_train() 函数和配置文件, 即可.

首先 backbone 为 resnet50, (resnet 系列结构参见3), 其以 tuple 形式返回 4 个 stage 的特征图, 片段代码如下:

```

1
2 outs = []
3 for i, layer_name in enumerate(self.res_layers):
4     res_layer = getattr(self, layer_name)
5     x = res_layer(x)
6     if i in self.out_indices:
7         outs.append(x)

```

然后 neck 为 fpn, 结构参见5, fpn 根据 config 中的 out\_indices 取出以 resnet50 输出的对应 stage, 分别构造输出 channel 维度统一的卷积算子, 然后按照5所示融合方式进行不同尺度的特征融合, 以元组形式输出结果. 在

配置信息里有一条 num\_outs=5, 是为 mask-rcnn 在最顶层特征增加的最大池化特征输出. 以上两块为提取特征, 被 extract\_feat 整合在一块,

紧接着 forward\_train 中包含了剩下的所有流程.

*rpn\_head* → *rpn\_head.loss* → *rpn\_head.get\_bboxes* → *assign* → *sample*  
 → *bbox\_roi\_extractor* → *bbox\_head* → *bbox\_head.get\_target*. → *bbox\_head.loss* → *mask\_roi\_extractor* → *mask\_head* → *mask\_head.get\_target*

这里梳理一下部分函数.

候选框层 RPN, RPNHead 继承 AnchorHead, 它的几个核心操作都在 anchor\_head.py 中实现, 主要包括 get\_anchors, anchor\_target 见 1.5.1, 函数 get\_bboxes 结合配置参数从 rpn 前向得到的 2 分类和位置预测结果中筛选出最终的 proposals.

get\_bboxes 中先通过 self.anchor\_generators[i].grid\_anchors() 这个函数取到所有的 anchor\_boxes, 再通过 self.get\_bboxes\_single() 根据 rpn 前向的结果选出候选框, 在 self.get\_bboxes\_single() 中, 先在每个尺度上取 2000(配置) 个 anchor 出来, concat 到一起作为该图像的 anchor, 对这些 anchor boxes 作 nms(thr=0.7) 就得到了所需的候选框. 需注意预测的 bbox 是对数化了的, 在做 iou 计算之前需用 delta2bbox() 函数进行逆变换. bbox\_head 中的 bbox2roi 类似.

得到的候选框最终由配置中 train\_cfg 的 rcnn.assigner, rcnn.sampler 进行标定和筛选, 保持正负样本平衡和框的质量, 方便优化.

MaxIoUAssigner:

1. 所有候选框置-1
2. 将与所有 gtbbbox 的 iou 小于 neg\_iou\_thr 置 0
3. iou 大于 pos\_iou\_thr 的将其匹配
4. 为了避免标定框无训练目标, 将 gtbbbox 匹配于与它 iou 最近的 bbox(会导致部分正样本的匹配 iou 值很小).

```
1  # 交并比矩阵(n,m), gt=n, bboxes=m
2  overlaps = bbox_overlaps(gt_bboxes, bboxes)
3  # 每个bbox和所有gt的最大交并比,(m,)
4  max_overlaps, argmax_overlaps = overlaps.max(dim=0)
5  # 每个gt和所有bbox的最大交并比
6  gt_max_overlaps, gt_argmax_overlaps = overlaps.max(dim=1)
```

```

7  # 1将所有bbox赋值为-1,注意new_full操作
8  assigned_gt_inds = overlaps.new_full(
9      (num_bboxes, ), -1, dtype=torch.long)
10 # 2交并比大于0同时小于负阈值的赋值为0
11 assigned_gt_inds[(max_overlaps >= 0)
12                  & (max_overlaps < self.neg_iou_thr)] = 0
13 # 将与gt交并比大于正阈值的赋值为1(可能没有)
14 pos_inds = max_overlaps >= self.pos_iou_thr
15 assigned_gt_inds[pos_inds] = argmax_overlaps[pos_inds] + 1
16 # 保证每个gt至少对应一个bbox
17 # 遍历gt,将与gt最近(max(iou))的bbox,将gt的label赋值给此bbox
18
19 for i in range(num_gts):
20     if gt_max_overlaps[i] >= self.min_pos_iou:
21         # 此判断较迷
22         max_iou_inds = overlaps[i, :] == gt_max_overlaps[i]
23         assigned_gt_inds[max_iou_inds] = i + 1
24         # 与gt最大iou的bbox 赋值为i+1

```

RandomSampler, 保持设定的平衡比例, 随机采样.

然后通过 SingleRoIExtractor(roi\_extractors/single\_level.py) 统一 RoI Align 四个尺度且大小不同的 proposals, 使其大小为  $7*7(\text{bbox})$  或  $14*14(\text{mask})$ . 配置信息 rpn\_head 中的 anchor\_strides 为 5 个尺度, 包含了 fpn 额外加入的最大池化层, 而 bbox\_roi\_extractor 的 featmap\_strides 却只包含四个尺度, 表明只需对前四层进行 align. 最终送入 bbox head 和 mask head 做第二次优化 (two stage).

RoIAlign 在 ops 中, 经 cuda 加速, 详解待后. 其中 roi\_extractors 中的特征层级映射函数如下:

```

1  def map_roi_levels(self, rois, num_levels):
2      """Map rois to corresponding feature levels by scales.
3          self.finest_scale = 56, 映射到0级的阈值
4          $(0, 56, 56*2, 56*4, \infty) \rightarrow (0, 1, 2, 3)$
5          bbox2roi变换后的 rois
6
7      Returns:
8          Tensor: Level index (0-based) of each RoI, shape (k, )
9          因不同层级对应不同的ROIAlign
10     """
11     scale = torch.sqrt(
12         (rois[:, 3] - rois[:, 1] + 1) * (rois[:, 4] - rois[:, 2] + 1))
13     target_lvls = torch.floor(torch.log2(scale / self.finest_scale + 1e-6))
14     target_lvls = target_lvls.clamp(min=0, max=num_levels - 1).long()
15     # 这个变换在原始论文中有.

```

```
16 return target_lvls
```

### 1.7.2 RepPoints

利用 DCN 特性, 实现了检测的结构点表示方式的优化.

### 1.7.3 RetinaFace

## 第二节 数据处理

coco 数据等格式, 多格式的转换, 增强检测算法的预处理.

## 第三节 FP16

模型预测加速等.

## 第四节 训练 pipeline

图见2注意它的四个层级. 主要查看 api/train.py, mmdcv 中的 runner 相关文件. 主要两个类: Runner 和 Hook Runner 将模型, 批处理函数 batch\_processor, 优化器作为基本属性, 是为训练过程中记录相关节点信息, 这些信息均被记录在 mode, \_hooks, \_epoch, \_iter, \_inner\_iter, \_max\_epochs, \_max\_iters 中. 从而实现训练过程中插入不同的操作, 也即各种 hook. 理清训练流程只需看 Runner 的成员函数 run. 在 run 里会根据 mode 按配置 (workflow) epoch 循环调用 train 和 val 函数, 跑完所有的 epoch. 其中 train 代码如下:

```
1 def train(self, data_loader, **kwargs):
2     self.model.train()
3     self.mode = 'train'    # 改变模式
4     self.data_loader = data_loader
5     self._max_iters = self._max_epochs * len(data_loader)    # 最大batch循环次数
6     self.call_hook('before_train_epoch')    # 根据名字获取hook对象函数
7     for i, data_batch in enumerate(data_loader):
8         self._inner_iter = i    # 记录训练迭代轮数
9         self.call_hook('before_train_iter')    # 一个batch前向开始
10        outputs = self.batch_processor(
```

```

11     self.model, data_batch, train_mode=True, **kwargs)
12     self.outputs = outputs
13     self.call_hook('after_train_iter')    # 一个batch前向结束
14     self._iter += 1    # 方便resume时,知道从哪一轮开始优化
15
16     self.call_hook('after_train_epoch')    # 一个epoch结束
17     self._epoch += 1    # 记录训练epoch状态,方便resume
    
```

上面让人困惑的是 hook 函数, hook 函数继承自 mmcv 的 Hook 类, 其默认了 6+8+4 个函数, 也即2所示的 6 个层级节点, 外加 2\*4 个区分 train 和 val 的节点记录函数, 以及 4 个边界检查函数. 从 train.py 中容易看出, 在训练之前, 已经将需要的 hook 函数注册到 Runner 的 self.\_hook 中了, 包括从配置文件解析的优化器, 学习率调整函数, 模型保存, 一个 batch 的时间记录等 (注册 hook 算子在 self.\_hook 中按优先级升序排列). 于是只需理解 call\_hook 函数即可.

```

1 def call_hook(self, fn_name):
2     for hook in self._hooks:
3         getattr(hook, fn_name)(self)
    
```

如上看出, 在训练的不同节点, 将从注册列表中调用实现了该节点函数的类成员函数. 比如

```

1 class OptimizerHook(Hook):
2
3     def __init__(self, grad_clip=None):
4         self.grad_clip = grad_clip
5
6     def clip_grads(self, params):
7         clip_grad.clip_grad_norm_(
8             filter(lambda p: p.requires_grad, params), **self.grad_clip)
9
10    def after_train_iter(self, runner):
11        runner.optimizer.zero_grad()
12        runner.outputs['loss'].backward()
13        if self.grad_clip is not None:
14            self.clip_grads(runner.model.parameters())
15        runner.optimizer.step()
    
```

将在每个 train\_iter 后实现反向传播和参数更新.

学习率优化相对复杂一点, 其基类 LrUpdaterHook, 实现了 before\_run, before\_train\_epoch, before\_train\_iter 三个 hook 函数, 意义自明. 这里选一个余弦式变化, 稍作说明:

```

1 class CosineLrUpdaterHook(LrUpdaterHook):
2
3     def __init__(self, target_lr=0, **kwargs):
4         self.target_lr = target_lr
5         super(CosineLrUpdaterHook, self).__init__(**kwargs)
6
7     def get_lr(self, runner, base_lr):
8         if self.by_epoch:
9             progress = runner.epoch
10            max_progress = runner.max_epochs
11        else:
12            progress = runner.iter
13            max_progress = runner.max_iters
14        return self.target_lr + 0.5 * (base_lr - self.target_lr) * \
15            (1 + cos(pi * (progress / max_progress)))

```

从 `get_lr` 可以看到, 学习率变换周期有两种, `epoch`->`max_epoch`, 或者更大的 `iter`->`max_iter`, 后者表明一个 `epoch` 内不同 `batch` 的学习率可以不同, 因为没有什么理论, 所有这两种方式都行. 其中 `base_lr` 为初始学习率, `target_lr` 为学习率衰减的上界, 而当前学习率正如函数的返回表达式.

## 第五节 更改模型

案例

## 第六节 新增模型

RetinaFace, FaceBoxes, RepPoints.(放一个完整代码即可)

## 第七节 numpy,torch 某些基础函数

## 第八节 计划

1. x
2. x
3. RetinaFace, FaceBoxes, RepPoints
4. 损失函数



5. x

6. ops 中的 ROIAlign,DCN