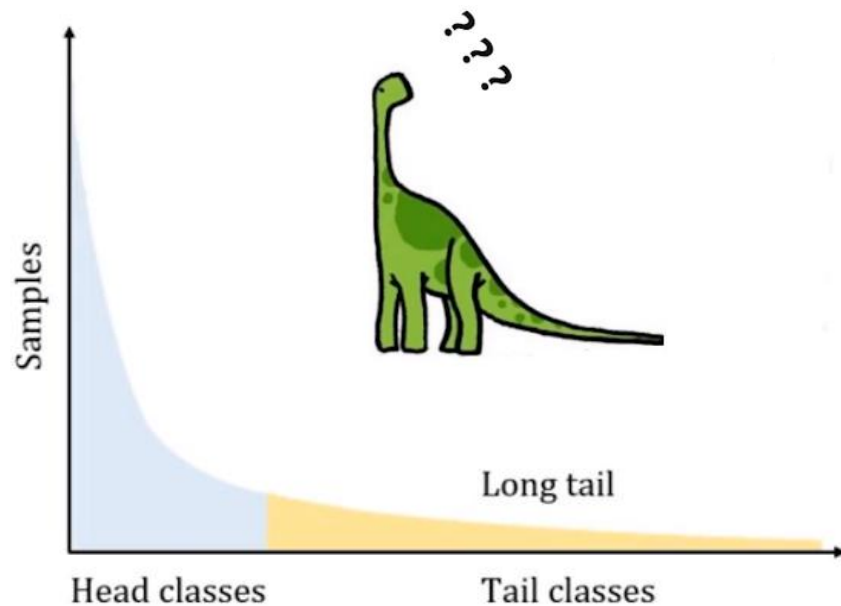


Long-Tailed and scale imbalance/class imbalance

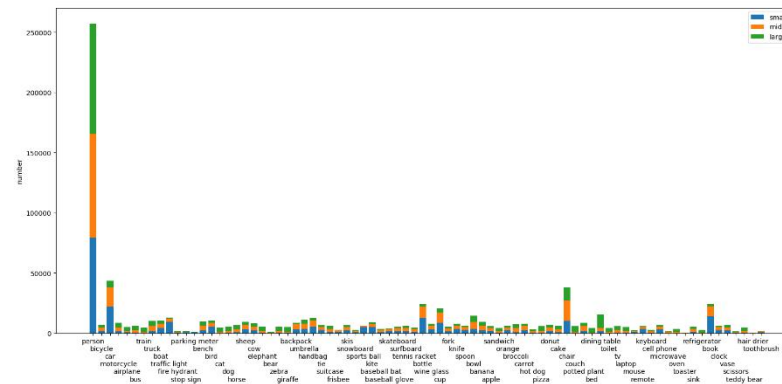
Long-Tailed、scale imbalance与class imbalance总览

- Long-Tailed: 即有很长的“尾巴”的分布。有大量的ID所拥有的样本数量十分稀少(tail class), 有限的样本数量不足以代表这个类别真实的分布情况; 而少量的ID, 其样本量充分, 类内多样性丰富(head class)。 (CIFAR-100-IR50和CIFAR-10-IR50)
- Scale imbalance: 尺度不平衡, 主要由目标的尺度引起, 例如COCO中的小物体过多即是一种尺度不平衡; 又如将物体分配至feature pyramid时的不平衡。
- Class imbalance: 类别不平衡, 主要由样本数量上的差别引起。最典型的是foreground-background imbalance, 即训练过程中的正例数量远远小于负例数量引起的不平衡。

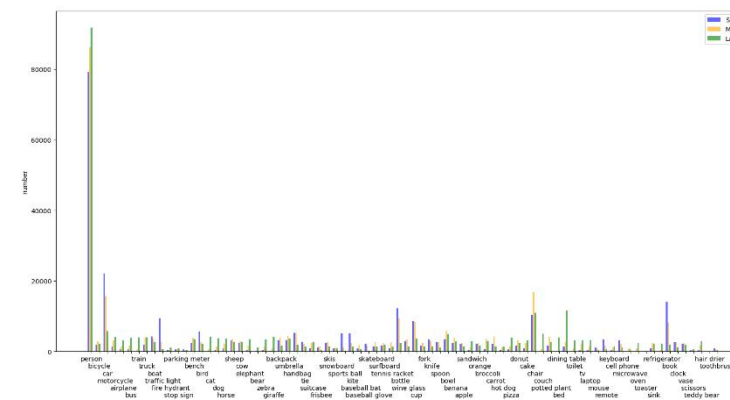
长尾分布、coco类别、尺度不平衡



- 长尾分布是广泛存在于实际的训练数据中



- Coco的类别统计



- 每个类别scale分布

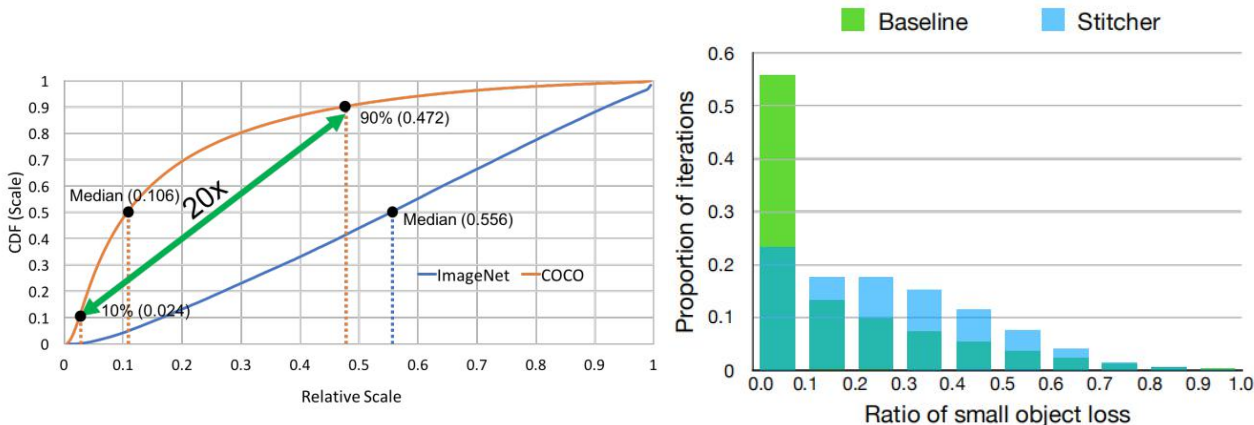
Tridentnet、stitcher解决coco尺度、类别不平衡的方法

- Tridentnet: 通过三支不同大小的receptive field来适应不同尺度大小的物体。
- stitcher: 设置一个阈值, 将大中型物体转化为中小型, 弥补了batch内小物体的数量, 使小物体的分布更加均匀, 达到数据增广的作用。

AP	AP small	AP mid	AP large
36.7 %	21.1 %	39.9 %	48.1 %

	Small	Mid	Large
Ratio of total boxes (%)	41.4	34.3	24.3
Ratio of images included (%)	52.3	70.7	83.0

Backbone	Dilation	AP	AP _s	AP _m	AP _l
ResNet-50	1	0.332	0.174	0.384	0.464
	2	0.342	0.168	0.386	0.486
	3	0.341	0.162	0.383	0.492
ResNet-101	1	0.379	0.200	0.430	0.528
	2	0.380	0.191	0.427	0.538
	3	0.371	0.181	0.410	0.538



- **BBN: Bilateral-Branch Network with Cumulative Learning for Long-Tailed Visual Recognition 2019.CVPR**

BBN: Bilateral-Branch Network

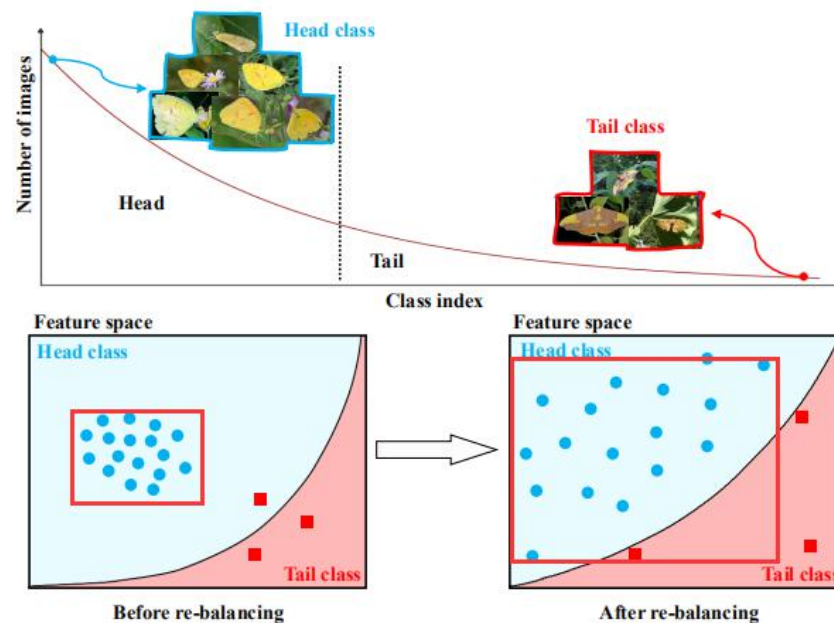
- 通过解耦分类网络的特征学习层和分类层，解决传统的重平衡方法由于刻意改变样本数目（重采样法）或刻意扭曲数据分布（重权重法），在一定程度上会损害深度网络学习到的深度特征的代表能力的问题。
- 提出了一个统一的双边分支网络 Bilateral-Branch Network (BBN)，来同时处理表达学习和分类学习。

传统方法的不足

- **重采样法**：是一种对样本进行有放回的抽样；采用经验性方法而非分析方法生成，它从原始数据样本中提取重复样本，即不使用通用分布来逼近计算概率 P 的值。
- **重赋权法**：是赋权法的一种，是指在训练过程的每一轮中，可根据样本分布为每个训练样本重新赋予权重。

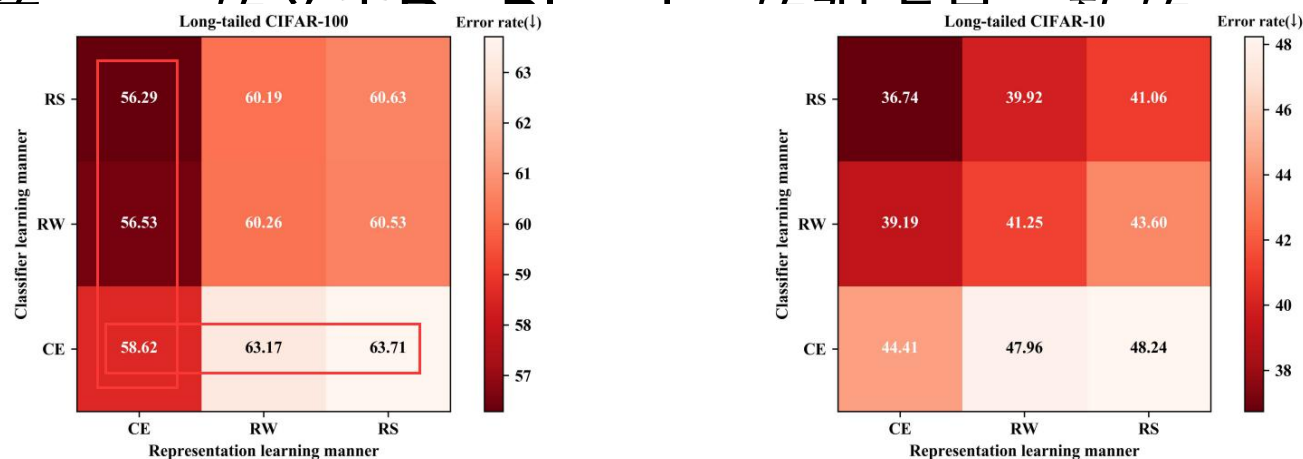
传统方法的不足

- 现在解决Long-Tail问题常用的re-balancing方法（Re-Weighting & Re-Sampling）虽然可以通过提升网络结构中classifier部分的性能，但是却损害了Representation部分（feature extractor backbone）学习的效果。
- Re-blancing之后，模型可以正确分类部分tail类的数据，但是每个类的类内分布变得更加separable。



idea验证

- 作者把训练过程划分为两个阶段：特征表示学习和分类器学习。分别利用CE, RW和RS作为三类。
- 横向比较时**，即Classifier部分的学习策略一致，Representation Learning阶段使用RS和RW会导致error rate变差；
- 纵向比较时**，即Representation Learning阶段使用的训练策略一致，classifier部分使用RS或者RW会提升模型的性能。这与作者在前文提出的观点一致。



Motivation

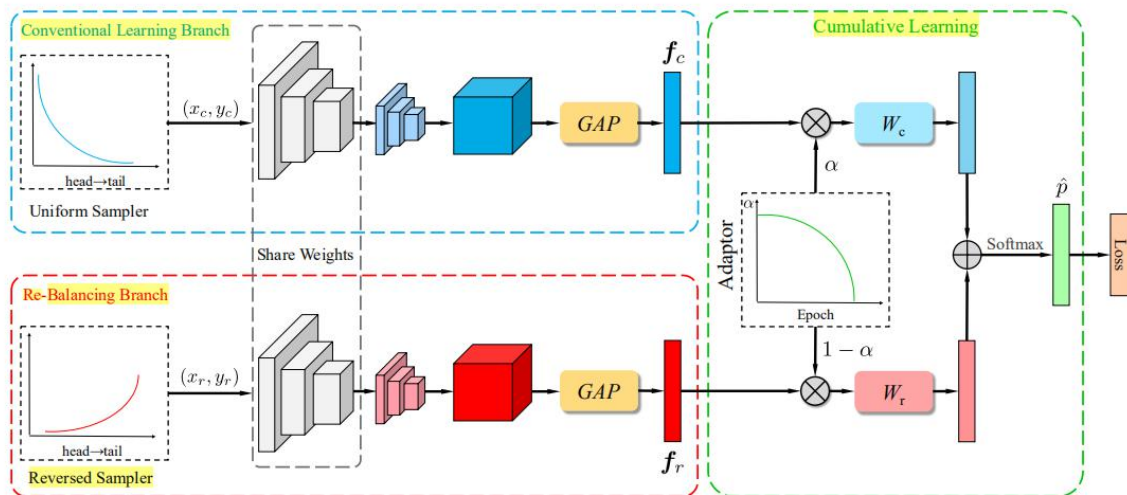
- 即类别重平衡可直接影响分类器权重对深层网络的更新，即促进分类器学习。
- 但重新平衡会在一定程度上意外地损害所学习的深度特征（即表示学习）的代表能力。
- 提出同时兼顾分类器学习和表征学习的双边分支网络（BBN），来同时处理表达学习和分类学习。

主要框架

- 提出了一个双分支神经网络结构用来同时兼顾特征学习和分类器学习，将深度模型的这两个重要模块进行**解耦**。
- 保证两个模块互不影响，共同达到优异的收敛状态，协同促进深度网络在长尾数据分布上的泛化性能。

主要框架

- BBN 模型包含 3 个主要组件：1) 常规学习分支，2) 再平衡分支，3) 累积学习策略。
- BBN使用CE（上）和RS（下）作为训练样本的采样方式，分别进行特征提取。然后使用全局平均池化（GAP）获得每个分支的特征向量。
- 其中CE对原始的长尾数据进行采样（为了保证表示学习的能力），RS则是设计一个反向采样器（构建稳健的分类器）进行采样。



反向采样器

- re-blancing分支使用reversed sampler，即与每个类别样本数量成反比例关系的采样策略。



(\mathbf{x}_c, y_c)





(\mathbf{x}_r, y_r)

$$w_i = \frac{N_{max}}{N_i}$$




$$P_i = \frac{w_i}{\sum_{j=1}^C w_j},$$

$$\alpha = 1 - \left(\frac{T}{T_{max}}\right)^2$$

两个分支的输入:  代表训练样本,  代表某个类别







设第  类的样本数为  , 所有类中数量最多的的样本数为   ; 反向采样器。

某个类的采样概率。

 代表现在的  ,  代表总epoch。

$$\hat{p}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}$$

$$\mathbf{z} = \alpha \mathbf{W}_c^\top \mathbf{f}_c + (1 - \alpha) \mathbf{W}_r^\top \mathbf{f}_r$$

 是  的输出; 上下分支得到  与  , 渐进式学习的策略; 动态调整两个   在的权重。

$$\mathbf{f}_c \in \mathbb{R}^D \quad \mathbf{W}_c \in \mathbb{R}^{D \times C}$$

$$\mathbf{f}_r \in \mathbb{R}^D \quad \mathbf{W}_r \in \mathbb{R}^{D \times C}$$

$$\hat{\mathbf{p}} = [\hat{p}_1, \hat{p}_2, \dots, \hat{p}_C]^\top$$

$$\mathcal{L} = \alpha E(\hat{\mathbf{p}}, y_c) + (1 - \alpha) E(\hat{\mathbf{p}}, y_r)$$

$$L = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$$

$E()$ 为交叉熵损失函数, 累积学习策略 (先学习通用模式, 然后逐渐关注尾部数据的输出) 两个分支 代表两个分类器

BBN算法

- 1. 通过提出的适配器为 α 分配一个值。
- 2. 分别通过均匀采样器和反向采样器对训练样本进行采样。将样本输入=网络后，获得两个独立的特征向量 f_c 和 f_r 。
- 3. 根据等式计算输出对数 z 和预测可能性 \hat{p} 。
- 4. 基于等式计算分类损失函数。通过优化此损失函数来更新模型参数。

Algorithm 1 Learning algorithm of our proposed BBN

Require : Training Dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$; $\text{UniformSampler}(\cdot)$ denotes obtaining a sample from \mathcal{D} selected by a uniform sampler; $\text{ReversedSampler}(\cdot)$ denotes obtaining a sample by a reversed sampler; $\mathcal{F}_{cnn}(\cdot; \cdot)$ denotes extracting the feature representation from a CNN; θ_c and θ_r denote the model parameters of the conventional learning and re-balancing branch; \mathbf{W}_c and \mathbf{W}_r present the classifiers' weights (*i.e.*, last fully connected layers) of the conventional learning and re-balancing branch.

```

1: for  $T = 1$  to  $T_{max}$  do
2:    $\alpha \leftarrow 1 - \left(\frac{T}{T_{max}}\right)^2$ 
3:    $(\mathbf{x}_c, y_c) \leftarrow \text{UniformSampler}(\mathcal{D})$ 
4:    $(\mathbf{x}_r, y_r) \leftarrow \text{ReversedSampler}(\mathcal{D})$ 
5:    $\mathbf{f}_c \leftarrow \mathcal{F}_{cnn}(\mathbf{x}_c; \theta_c)$ 
6:    $\mathbf{f}_r \leftarrow \mathcal{F}_{cnn}(\mathbf{x}_r; \theta_r)$ 
7:    $\mathbf{z} \leftarrow \alpha \mathbf{W}_c^\top \mathbf{f}_c + (1 - \alpha) \mathbf{W}_r^\top \mathbf{f}_r$ 
8:    $\hat{\mathbf{p}} \leftarrow \text{Softmax}(\mathbf{z})$ 
9:    $\mathcal{L} \leftarrow \alpha E(\hat{\mathbf{p}}, y_c) + (1 - \alpha) E(\hat{\mathbf{p}}, y_r)$ 
10:  Update model parameters by minimizing  $\mathcal{L}$ 
11: end for

```

实验

- 在不平衡比分别为10、50、100的长尾CIFAR数据集上进行广泛的实验。在CIFAR-10 和 CIFAR-100 数据集上实现 ResNet 的 top-1 错误率。

Dataset	Long-tailed CIFAR-10			Long-tailed CIFAR-100		
Imbalance ratio	100	50	10	100	50	10
CE	29.64	25.19	13.61	61.68	56.15	44.29
Focal [17]	29.62	23.28	13.34	61.59	55.68	44.22
Mixup [35]	26.94	22.18	12.90	60.46	55.01	41.98
Manifold Mixup [29]	27.04	22.05	12.97	61.75	56.91	43.45
Manifold Mixup (two samplers)	26.90	20.79	13.17	63.19	57.95	43.54
CE-DRW [3]	23.66	20.03	12.44	58.49	54.71	41.88
CE-DRS [3]	24.39	20.19	12.62	58.39	54.52	41.89
CB-Focal [5]	25.43	20.73	12.90	60.40	54.83	42.01
LDAM-DRW [3]	22.97	18.97	11.84	57.96	53.38	41.29
Our BBN	20.18	17.82	11.68	57.44	52.98	40.88

实验

- 在 iNaturalist 2018 和 iNaturalist 2017 这两个大规模长尾数据集上实验。以ResNet-50为backbone，在 iNaturalist 2018 和 iNaturalist 2017 上的 top-1 错误率BBN 仍然更优。

Dataset	iNaturalist 2018	iNaturalist 2017
CE	42.84	45.38
CE-DRW [3]	36.27	40.48
CE-DRS [3]	36.44	40.12
CB-Focal [5]	38.88	41.92
LDAM-DRW* [3]	32.00	–
LDAM-DRW [3]	35.42	39.49
LDAM-DRW [3] (2×)	33.88	38.19
Our BBN	33.71	36.61
Our BBN (2×)	30.38	34.25

实验

- 由于 iNaturalist 数据集规模很大，作者还使用 $2\times$ 调度器进行了实验。
- 为了公平，作者也使用 $2\times$ 调度器训练了之前最佳的 LDAM-DRW。可以明显看到，使用 $2\times$ 调度器的 BBN 的表现显著优于未使用 $2\times$ 调度器的 BBN 的表现。此外，BBN($2\times$) 的表现也明显优于 LDAM-DRW ($2\times$)。

Dataset	iNaturalist 2018	iNaturalist 2017
CE	42.84	45.38
CE-DRW [3]	36.27	40.48
CE-DRS [3]	36.44	40.12
CB-Focal [5]	38.88	41.92
LDAM-DRW* [3]	32.00	—
LDAM-DRW [3]	35.42	39.49
LDAM-DRW [3] ($2\times$)	33.88	38.19
Our BBN	33.71	36.61
Our BBN ($2\times$)	30.38	34.25

启发性或普适性

- 1.作者探索了长尾数据问题中类别再平衡机制，进一步发现这些方法可以显著地促进分类器学习，同时也会影响原始数据分布的表示学习。
- 2.作者提出了一个BBN模型，兼顾表示学习（特征提取模块）和分类器学习，以最大限度地提高长尾数据识别。此外，作者还提出了一种新的累积学习策略来调整双边学习。

总结

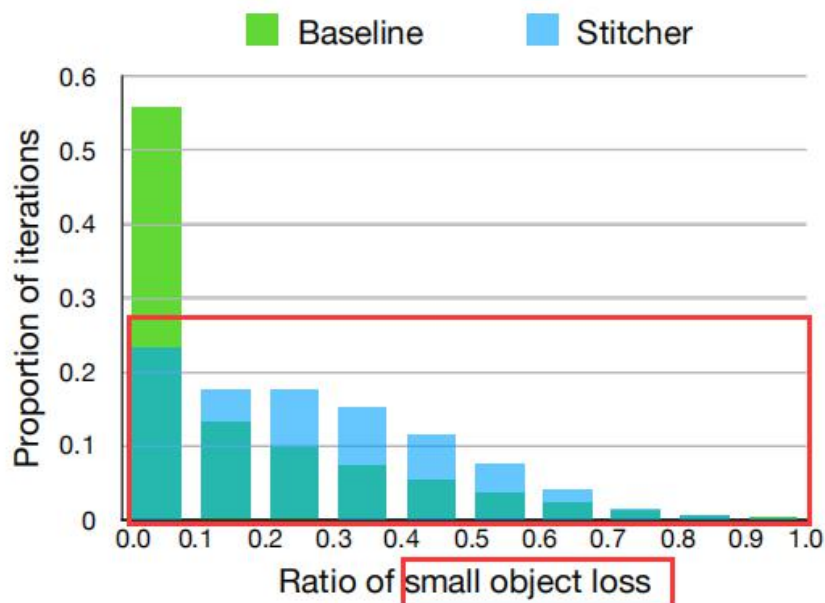
- BNN = 常规学习分支 + 再平衡分支(倒置采样器和特征加权聚合) + 累积学习策略(先学习通用模式，然后逐渐关注尾部数据。)

我的观点

- 重采样和重权重法这类类别重平衡的方法提升了分类器学习能力。但由于刻意改变样本数目（重采样法）或刻意扭曲数据分布（重权重法），在一定程度上损害表示特征。
- BBN的再平衡分支(倒置采样器和特征加权聚合) + 累积学习策略(先学习通用模式，然后逐渐关注尾部数据)是一种分类器学习和表征学习恰当适中的学习方法。

Loss不平衡

- 对于Faster R-CNN baseline，在超过50%的迭代中，小物体所产生的loss贡献非常低（不到总loss的0.1）。采用Stitcher时，损失分配变得平衡。
- 这说明在模型训练过程中，网络中对小物体的监督是不足的。



- 设计一个实验，关于不同尺度的物体在不同尺度的物体在epoch中的loss分布
- 复现：
- 中型物体在的loss分布；
- 大型物体在的loss分布；