[2019] Bag of tricks and a strong baseline for deep person re-identification

2021年5月12日 11:06

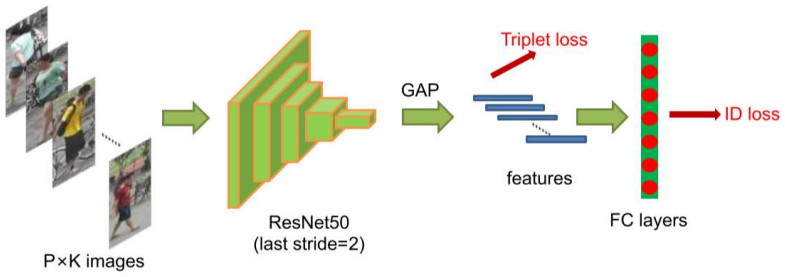
基本信息

```
@inproceedings{luo2019bag,
  title={Bag of tricks and a strong baseline for deep person re-identification},
  author={Luo, Hao and Gu, Youzhi and Liao, Xingyu and Lai, Shenqi and Jiang, Wei},
  affiliation={Zhejiang University, Chinese Academy of Sciences, Xi' an Jiaotong University},
  booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops},
  pages={0--0},
  year={2019}
}
```

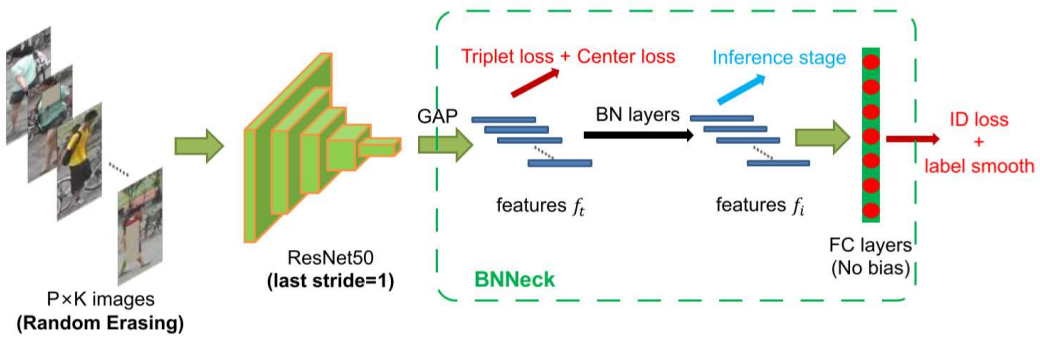
主要贡献

1. 汇聚了众多重识别模型的 tricks，搭建了一个简单有效的行人重识别 baseline。
2. 提出了一个 neck structure（BNNeck）

Framework



(a) The pipeline of the standard baseline.



(b) The pipeline of our modified baseline.

基线 Baseline

1. 初始化 ResNet50（基于 ImageNet 的预训练模型） -> 通过一个全连接层将其输出的维度变为N，N为训练集中的行人实体（identities）数。
2. 随机采样 P 个行人实体，对于每个人选取 K 张照片，构成一个训练的 batch。因此，batch size = P × K. 在本文中，设置$P = 16, K = 4$。
3. 将每幅图像 resize 至 256 × 128，此后用 0 元素 padding 10个像素值，最后再随机将其裁剪为 256 × 128 大小的图片。
4. 每张图片有 0.5 的概率被水平翻转。
5. 每张图片被编码为 32 bit 的 float 类型的点，并且将像素值缩放至 [0,1] 大小。然后通过对 RGB 三个通道各减去 $0.485, 0456, 0.406$，并各除以 $0.229, 0.224, 0.225$ 进行规范化。
6. 该模型的输出为 ReID 的特征 f，以及 ID prediction logits p。

7. ReID 的特征 f 被用来计算 triplet loss，ID prediction logits p 被用来计算 cross entropy loss。Triplet loss 中的 margin 被设置为 0.3.
8. 采用了 Adam 的优化方法，初始学习率为 0.00035，40个 epoch 和 70个 epoch 时各衰减 0.1，总共训练 120 个 epoch。

训练技巧

（一）Warmup Learning Rate

$$\mathrm{lr(t)} = \begin{cases} 3.5 \times 10^{-5} \times \frac{t}{10}, & if\ t < 10 \\ 3.5 \times 10^{-4}, & if\ 10 < t < 40 \\ 3.5 \times 10^{-5}, & if\ 40 < t < 70 \\ 3.5 \times 10^{-6}, & if\ 70 < t < 120 \end{cases}$$

（二）Random Erasing Augmentation （REA）

Figure 4. Sampled examples of random erasing augmentation. The first row shows five original training images. The processed images are presented in the second low.

1. 目的：解决遮挡问题以及提高泛化能力
2. 符号：
   (1) 对于 mini-batch 中的图片 I，以概率 $\mathrm{p_c}$ 对其进行随机擦除操作
   (2) REA 在图片 I 中随机选择一个矩形区域 $\mathrm{I_e}$，其大小为$(\mathrm{W_e}, \mathrm{H_e})$，并且用随机值进行替换擦除
   (3) 定义擦除区域的面积率为 $\mathrm{r_c} = \frac{\mathrm{S_e}}{\mathrm{S}} = \frac{\mathrm{W_e} \times \mathrm{H_e}}{\mathrm{W} \times \mathrm{H}}$
   (4) 擦除区域 $\mathrm{I_e}$ 的长宽比为 $(\mathrm{r_1}, \mathrm{r_2})$ 之间随机初始化的值
3. 方法：
   (1) 随机选择一个点 $\mathrm{P} = (\mathrm{x_e}, \mathrm{y_e})$，如果 $\mathrm{x_e} + \mathrm{W_e} \leq \mathrm{W}$ 并且 $\mathrm{y_e} + \mathrm{H_e} \leq \mathrm{H}$，则将区域 $\mathrm{I_e} = (\mathrm{x_e}, \mathrm{y_e}, \mathrm{x_e} + \mathrm{W_e}, \mathrm{y_e} + \mathrm{H_e})$ 作为选择的区域；否则，重复上述操作直至选择了一个合适的区域。
   (2) 选择了该区域后，使用整张图片的像素均值来填充该区域的像素值
4. 经验值：$\mathrm{p} = 0.5, 0.02 < \mathrm{S_e} < 0.4, \mathrm{r_1} = 0.3, \mathrm{r_2} = 3.33.$

（三）Label Smoothing

1. 目的：提高模型的泛化性能，防止过拟合。
2. 符号：y 为真实标签，$\mathrm{p_i}$ 为预测为 i 类的概率
3. 标准交叉熵损失：

$$\mathrm{L(ID)} = \sum_{i=1}^{N} -q_i \log(p_i) \begin{cases} q_i = 0, y \neq i \\ q_i = 1, y = i \end{cases}$$

4. 标签平滑：

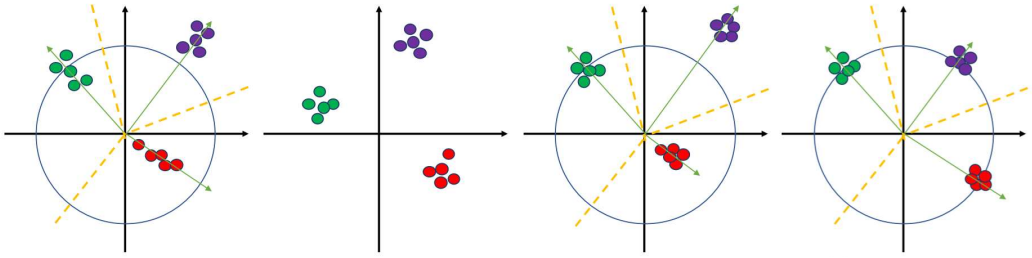
$$\mathrm{q_i} = \begin{cases} 1 - \frac{N-1}{N}\varepsilon & if\ i = y \\ \frac{\varepsilon}{N} & otherwise \end{cases}$$

$\varepsilon = 0.1$，促使模型在训练集上的置信度降低，当训练集不大时，该方法能有效提高性能 。

（四）Last Stride

1. 目的：提高模型的表征能力
2. 方法：将 ResNet50 的最后一个 stride 由 2 设置为 1，对于256×128 的图片使输出的特征 由 (8×4) 变为 (16×8)
3. 该方法仅会增加少量的计算量，且不会引入额外的训练参数，但可以带来显著的性能提升。

（五）BNNeck



1. 目的：解决不同损失优化目标不一致的问题。
2. 原因：ID loss 为交叉熵损失，将特征空间划分为不同的区域，因此在推断时使用余弦距离更为合理；但 Triplet loss 是令类内距离小，类间距离大，因此使用欧式距离更为合理。两者合二为一后，往往导致一个损失下降，另一个损失震荡或上升。
3. 方法：在 ResNet50 输出的特征 f_t 后增加了一个 batch normalization（BN）层，对应的特征为 f_i。在训练时，f_t 与 f_i 分别被用来计算 triplet loss 和 ID loss。在推断时，使用 f_i 以及余弦相似度。
4. 结果：BN 使特征 f_i 高斯分布于超平面，该分布使得 ID loss 更容易收敛，以及减少了对其的约束。这也使得 triplet loss 更为容易收敛。归一化也保证了同一个人的特征的紧密分布。
5. 技巧：由于超平面是对称的，所以一个技巧是移除 FC 层的偏置

（六）Center Loss

1. Triplet Loss：$L_{Tri} = [d_p - d_n + \alpha]_+$
   d_p 和 d_n 分别是 positive pair 和 negative pair 的距离，$\alpha = 0.3$ 是 triplet loss 的 margin
   $[z]_+$ 等价于 $\max(0, z)$
   存在的问题：仅考虑了一个 batch 之中的 d_p 和 d_n 距离，忽略了两者的绝对值，难以保证在全局中类内距离小于类间距离。
2. Center Loss：$L_c = \frac{1}{2}\sum_{j=1}^{B}\left\|f_{t_j} - c_{y_j}\right\|_2^2$
   c_{y_j} 为类 j 的中心，B 为 batch size大小
3. 本文中使用了三者，构建损失
   $L = L_{ID} + L_{Triplet} + \beta L_C$
   β 设置为了 0.0005

实验结果

1. 每个 trick 的影响
   （1）Same Domain（逐一叠加加入各种 trick）

Model	Market1501 r = 1	Market1501 mAP	DukeMTMC r = 1	DukeMTMC mAP
Baseline-S	87.7	74.0	79.7	63.7
+warmup	88.7	75.2	80.6	65.1
+REA	91.3	79.3	81.5	68.3
+LS	91.4	80.3	82.4	69.3
+stride=1	92.0	81.7	82.6	70.6
+BNNeck	94.1	85.7	86.2	75.9
+center loss	94.5	85.9	86.4	76.4

(2) Cross Domain

Model	M→D r = 1	M→D mAP	D→M r = 1	D→M mAP
Baseline	24.4	12.9	34.2	14.5
+warmup	26.3	14.1	39.7	17.4
+REA	21.5	10.2	32.5	13.5
+LS	23.2	11.3	36.5	14.9
+stride=1	23.1	11.8	37.1	15.4
+BNNeck	26.7	15.2	47.7	21.6
+center loss	27.5	15.0	47.4	21.4
-REA	41.4	25.7	54.3	25.5

REA mask 了训练集中的图片，使得模型学习了更多训练集领域的知识，造成泛化性降低

2. BNNeck 网络的分析

Feature	Metric	Market1501 r = 1	Market1501 mAP	DukeMTMC r = 1	DukeMTMC mAP
f (w/o BNNeck)	Euclidean	92.0	81.7	82.6	70.6
f_t	Euclidean	94.2	85.5	85.7	74.4
f_t	Cosine	94.2	85.7	85.5	74.6
f_i	Euclidean	93.8	83.7	86.6	73.0
f_i	Cosine	**94.1**	**85.7**	**86.2**	**75.9**

3. 与 SOTA 的比较

Type	Method	N_f	Market1501 r = 1	Market1501 mAP	DukeMTMC r = 1	DukeMTMC mAP
Pose-guided	GLAD[19]	4	89.9	73.9	-	-
	PIE [23]	3	87.7	69.0	79.8	62.0
	PSE [13]	3	78.7	56.0	-	-
Mask-guided	SPReID [7]	5	92.5	81.3	84.4	71.0
	MaskReID [9]	3	90.0	75.3	78.8	61.9
Stripe-based	AlignedReID [21]	1	90.6	77.7	81.2	67.4
	SCPNet [3]	1	91.2	75.2	80.3	62.6
	PCB [16]	6	93.8	81.6	83.3	69.2
	Pyramid[22]	1	92.8	82.1	-	-
	Pyramid[22]	21	95.7	88.2	89.0	79.0
	BFE[1]	2	94.5	85.0	88.7	75.8
Attention-based	Mancs [18]	1	93.1	82.3	84.9	71.8
	DuATM [14]	1	91.4	76.6	81.2	62.3
	HA-CNN [8]	4	91.2	75.7	80.5	63.8
GAN-based	Camstyle [28]	1	88.1	68.7	75.3	53.5
	PN-GAN [10]	9	89.4	72.6	73.6	53.2
Global feature	IDE [25]	1	79.5	59.9	-	-
	SVDNet [15]	1	82.3	62.1	76.7	56.8
	TriNet[6]	1	84.9	69.1	-	-
	AWTL[12]	1	89.5	75.7	79.8	63.4
	Ours	1	**94.5**	**85.9**	**86.4**	**76.4**
	Ours(RK)	1	**95.4**	**94.2**	**90.3**	**89.1**

4. Batch Size 的影响

Batch Size	Market1501		DukeMTMC	
$P \times K$	r = 1	mAP	r = 1	mAP
8×3	92.6	79.2	84.4	68.1
8×4	92.9	80.0	84.7	69.4
8×6	93.5	81.6	85.1	70.7
8×8	93.9	82.0	85.8	71.5
16×3	93.8	83.1	86.8	72.1
16×4	93.8	83.7	86.6	73.0
16×6	94.0	82.8	85.1	69.9
16×8	93.1	81.6	86.7	72.1
32×3	94.5	84.1	86.0	71.4
32×4	93.2	82.8	86.5	73.1

5. 图像大小的影响

	Market1501		DukeMTMC	
Image Size	r = 1	mAP	r = 1	mAP
256×128	93.8	83.7	86.6	73.0
224×224	94.2	83.3	86.1	72.2
384×128	94.0	82.7	86.4	73.2
384×192	93.8	83.1	87.1	72.9