

Pedestrian Alignment Network for Large-scale Person Re-identification

摘要

Re-ID被划分为图像检索问题，主要存在两个挑战：1、复杂背景；2、局部缺失。这两个误差都会破坏行人对准的质量，并可能由于位置和比例差异而降低行人匹配准确率。对准也是可以学习的，本文提出一种行人对准网络（PAN），网络可进行有区别的嵌入学习，而无需额外的标注。



Fig. 1 Sample images influenced by detector errors (the first row) which are aligned by the proposed method (the second row). Two types of errors are shown: excessive background and part missing. We show that the pedestrian alignment network (PAN) corrects the misalignment problem by 1) removing extra background or 2) padding zeros to the image borders. PAN reduces the scale and position variance, and the aligned output thus benefit the subsequent matching step.

正文

简介

PAN中，给定检测到行人的图像，该网络同时学习重新定位该人并将该人归类为预定义的身份。因此，PAN利用了人员对齐和重新识别的互补性。

该算法有以下步骤：

(1) 一个预测输入图片id的网络

(2) 一个重新框取人物的方法

(3) 另一个根据输入图片预测id的网络

(1) 和 (3) 共享低层次的特这个，在测试的时候联结起来用全连接层来构建行人的特征描述符。

(2) 中的精炼参数由高层次的特征图来估计。这一步构建了一个不一样的局部网络：空间转换网络 (spatial transformer network, STN)

本文的方法解决了由检测错误引起的未对准问题，而常用的色块匹配策略旨在在良好对准的图像中发现匹配的局部结构

相关研究

手工特征Re-ID系统

之前的比较鲁棒的手工特征有：LBP、Gabor、LOMO。

还有一系列的工作围绕这32-D LAB颜色值房图和128-D SIFT算子。

除了手工寻找特征，矩阵学习也很多，如KISSME是基于马氏距离的且将对比较表述为对数似然比检验的方法。然后扩展了贝叶斯和KISSME以学习具有度量的判别子空间。

还有的就是一组弱RankSVMs来组成一个强排序器。有用AdaBoost算法来将不同的特征归结到单一相似度的方法。

基于深度学习的方法

将行人图片水平划分为三块，然后训练三个part-CNN来获取特征。

类似的，将卷集图分为四块将局部特征与全局特征结合。

一种基于注意力的软模型，以关注零件并将CNN与LSTM零件选择性地结合在一起；它的局限性还包括计算效率低下。

对象对准

还有STN网络来精炼鸟类识别和房屋树木识别。

Pedestrian Alignment Network (行人提取网络, PAN)

网络使用ResNet-50作为基准模型，每个block包括conv+batchnorm+ReLU+maxpooling，每个block后特征图降维到边长为原来一半。

模型中由两个主要的conv分支，每个都是用于预测id值的分类网络。给一个原始的检测图像，base分支不仅学习如何辨别id，同时对图像的外观进行编码。对准分支使用一个类似的网络，但是处理由仿射估计分支生成的对齐特征图。

使用 W_1, W_2 表示两个conv分支的参数。输入一张图片 x ，使用 $p(k|x)$ 来表示 x 是第 $k \in \{1 \dots K\}$ 类。特别的 $p(k|x) = \frac{\exp(z_k)}{\sum_{k=1}^K \exp(z_i)}$ 。其中 z_i 是从CNN模型中输出的配额能行

对这两个分支，cross-entropy loss如下：

$$l_{base}(W_1, x, y) = - \sum_{k=1}^K (\log(p(k|x))q(k|x))$$

$$l_{align}(W_2, x, y) = - \sum_{k=1}^K (\log(p(k|x_a))q(k|x_a))$$

其中 x_a 是aligned的输入，给一个label y ，对于所有 $k \neq y$ 的真实分布 $q(y|x) = 1$ 且 $q(k|x) = 0$ 。代入上面的式子得到：

$$l_{base}(W_1, x, y) = -\log(p(y|x))$$

$$l_{align}(W_2, x_a, y) = -\log(p(y|x_a))$$

仿射估算分支

从ResNet截出Res2和Res4，affine branch包括一个双线性插值采样和网格网络，Grid Network包括一个ResBlock和一个avgpooling。Res4通过Grid Network来回归到一个6维的转换参数。这个学习到的转换参数 θ 用于表示图片网格。

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

其中 (x_i^t, y_i^t) 是输出要素图上的目标坐标， (x_i^s, y_i^s) 是输出要素图上源的坐标。 $\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}$ 处理规模和旋转变换。 θ_{13}, θ_{23} 处理平移。

$$U_{(m,n)}^c = \sum_{x^s} \sum_{y^s}^W V_{(x^s, y^s)}^C \max(0, 1 - |x^t - m|) \max(0, 1 - |y^t - n|)$$

$U_{(m,n)}^c$ 是输出特征图(m,n)处在c通道的值， $V_{(x^s, y^s)}^C$ 是输入特征图在 (x_s, y_s) 处c通道的值。

行人描述符

$$f_i = [\alpha |f_i^1|^T, (1 - \alpha) |f_i^2|^T]^T$$

Re-ranking 重排列

首先通过将画廊图像的欧几里德距离排序到查询中来获得等级列表 $N(q, n) = [x_1, x_2, \dots, s x_n]$ 。

文章还考虑了Jaccard similarity。k个倒数最近的邻居 $R(p, k)$ 由出现在查询p的前k个检索排名中的元素组成，而查询也位于元素的前k个排名中。

$$R(p, k) = \{x | x \in N(p, k), p \in N(x, k)\}$$

将 R 拓展到 R^* 来给出更多的正样本，Jaccard similarity即为：

$$D_{similarity} = 1 - \frac{|R^*(q, k) \cap R^*(x_i, k)|}{|R^*(q, k) \cup R^*(x_i, k)|}$$

Method	Single Query		Multi. Query	
	rank-1	mAP	rank-1	mAP
DADM [Su et al., 2016]	39.4	19.6	49.0	25.8
BoW+kissme [Zheng et al., 2015]	44.42	20.76	-	-
MR-CNN [Ustinova et al., 2015]*	45.58	26.11	56.59	32.26
MST-CNN [Liu et al., 2016b]	45.1	-	55.4	-
FisherNet [Wu et al., 2016a]*	48.15	29.94	-	-
CAN [Liu et al., 2016a]*	48.24	24.43	-	-
SL [Chen et al., 2016]	51.90	26.35	-	-
S-LSTM [Varior et al., 2016b]	-	-	61.6	35.3
DNS [Zhang et al., 2016]	55.43	29.87	71.56	46.03
Gate Reid [Varior et al., 2016a]	65.88	39.55	76.04	48.45
SOMAnet [Barbosa et al., 2017]*	73.87	47.89	81.29	56.98
PIE [Zheng et al., 2017a]*	78.65	53.87	-	-
Verif.-Identif. [Zheng et al., 2016d]*	79.51	59.87	85.84	70.33
SVDNet [Sun et al., 2017]*	82.3	62.1	-	-
DeepTransfer [Geng et al., 2016]*	83.7	65.5	89.6	73.8
GAN [Zheng et al., 2017b]*	83.97	66.07	88.42	76.10
APR [Lin et al., 2017]*	84.29	64.67	-	-
Triplet [Hermans et al., 2017]*	84.92	69.14	90.53	76.42
Triplet+re-rank [Hermans et al., 2017]*	86.67	81.07	91.75	87.18
Basel.	80.17	59.14	87.41	72.05
Ours	82.81	63.35	88.18	71.72
Ours+re-rank	85.78	76.56	89.79	83.79
Ours (GAN)	86.67	69.33	90.88	76.32
Ours (GAN)+re-rank	88.57	81.53	91.45	87.44



Market-1501

DukeMTMC-reID

CUHK03