

# Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline)

## 摘要

使用行人图像描述的部分功能可提供细粒度的信息，本文的目标是学习具有区别性的部分特征以供检索，并做出两个贡献（1）一个被称为基于部分的卷积基线(Part-based Convolutional Baseline, PCB)，输入一张图片，它将输出一系列part级别特征的卷积描述。（2）一个精炼部分卷积(refined part pooling, RPP)方法，统一的分区不可避免地会在每个部分中产生离群值，而离群实际上更类似于其他部分。RPP将这些离群值重新分配给它们最接近的部分，从而使精制部分的内部一致性得到增强。

## 简介

## 相关研究

### 手工设计的用于行人检索的部分特征

1、捕捉颜色与纹理特征的水平条带，2、将人体分为许多三角形来取得部分特征描述；3、采用图片结构将行人解析为语义部分；4、在头部，躯干和腿部应用HSV直方图以捕获空间信息

### 基于深度学习的部分特征

首先，学习出的特征有更强的描述能力；其次，深度学习为解析行人提供了更好的工具，这进一步使部分特征发挥作用。特别是人体姿势估计和关键点检测取得了令人瞩目的进展。

将最大激活的坐标聚集在特征图上，以找到几个感兴趣的区域。

将注意力机制嵌入网络中，允许模型自行决定将注意力集中在哪里。

### 具有注意力机制的深度学习部分

对比本文的RPP与PAR

**动机：** PAR旨在直接学习对齐的part，而RPP则旨在完善预先分区的part

**工作机制：**PAR使用注意力方法，以无监督的方式训练part分类器，而RPP的训练可以看作是半监督的过程。

**训练过程：**RPP首先训练具有均匀划分的身份分类模型，然后利用所学知识来诱导part分类器的训练。

## PCB：一个健壮的卷积基线

### PCB的结构

#### 骨架网络

PCB可以使用任何去掉隐藏全连接层的图像分类网络作为骨架网络，本文使用ResNet50

#### 从骨架网络到PCB

如下图所示，在全局平均池化(global average pooling, GAP)之前的结构与骨架网络一致。不同点在于GAP层及其后的部分被移除了。图片经过这些层后成为一个3D的张量 $T$ ，文中设定为列向量。然后通过一个卷积平均池化，PCB将张量 $T$ 转换为 $p$ 个水平条带且平均每一列的向量值到一个条带能作为一个单独的part级别的列向量 $g_i (i = 1, 2, \dots, p)$ 。接下来，PCB使用一个卷积层来为 $g$ 降维为 $h$ 。最后每个 $h_i$ 被送入由全连接层接上softmax函数的分类器来预测输入的ID值。

在训练期间，通过最小化 $p$ 个ID预测的交叉熵损失之和来优化PCB

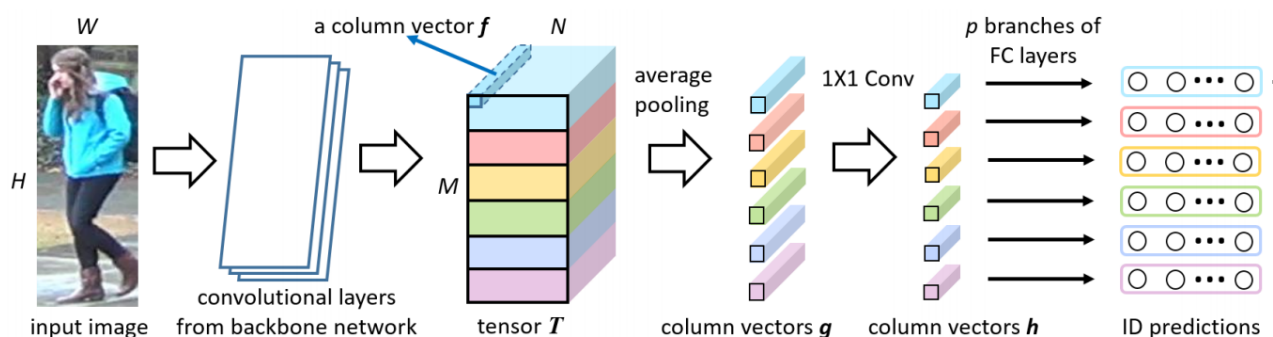


Figure 2. Structure of PCB. The input image goes forward through the stacked convolutional layers from the backbone network to form a 3D tensor  $T$ . PCB replaces the original global pooling layer with a conventional pooling layer, to spatially down-sample  $T$  into  $p$  pieces of column vectors  $g$ . A following  $1 \times 1$  kernel-sized convolutional layer reduces the dimension of  $g$ . Finally, each dimension-reduced column vector  $h$  is input into a classifier, respectively. Each classifier is implemented with a fully-connected (FC) layer and a sequential Softmax layer. During training, each classifier predicts the identity of the input image and is supervised by Cross-Entropy loss. During testing, either  $p$  pieces of  $g$  or  $h$  are concatenated to form the final descriptor of the input image.

#### 一些可用于替代的结构

平均所有的 $h_i$ 为单一向量 $\bar{h}$ ，然后对 $\bar{h}$ 用全连接分类预测ID值

### 精炼部分池化(Refined Part Pooling)

PCB的均匀分区简单，有效，但有待改进。本节首先解释均匀分隔伴随的不一致现象，然后提出改进的part合并作为加强part内部一致性的补救措施。

## part内不一致

着眼于要在空间上划分的张量 $T$ ，我们对部分内不一致的直觉是： $T$ 的相同部分中的列向量 $f$ 应该彼此相似，并且与其他部分中的列向量不相似；否则会出现part内部不一致的现象，这意味着part分配不当。所做的就是对每个条带中的各部分进行聚类。

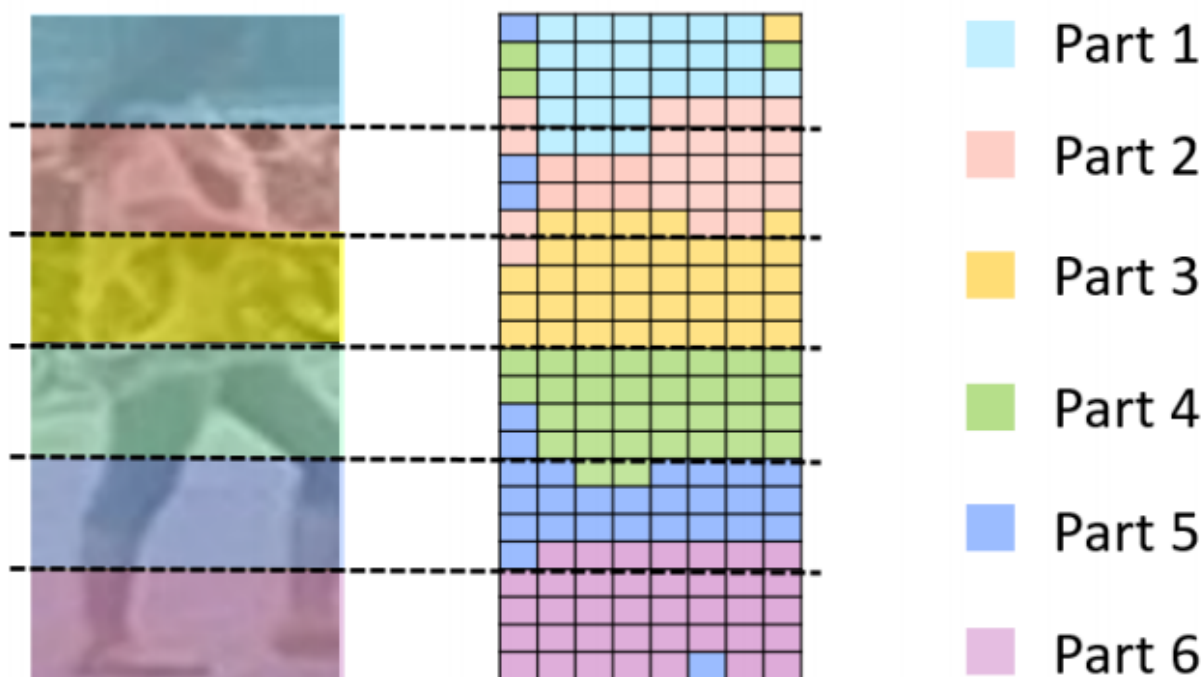


Figure 3. Visualization of within-part inconsistency.  $T$ . Left:  $T$  is equally partitioned to  $p = 6$  horizontal stripes (parts) during training. Right: Every column vector in  $T$  is denoted with a small rectangle and painted in the color of its closest part.

## 重新定位离群值

提出改进的part池化以纠正part内部的不一致。目标是根据所有列向量的相似度将它们分配给每个部分，以便对异常值进行重新定位。

基于学习到的张量 $T$ ，使用softmax后接一个线性层的结构作为part分类器：

$P(P_i|f) = \text{softmax}(W_i^T f) = \frac{\exp(W_i^T f)}{\sum_{j=1}^p \exp(W_j^T f)}$ ，其中 $P(P_i|f)$ 是 $f$ 属于 $P_i$ 部分的预测概率， $p$ 是与定义parts的个数， $W$ 是分类器的可训练的权重矩阵。

给定 $T$ 中的列向量 $f$ 和属于部分 $P_i$ 的 $f$ 的预测概率，我们将 $P(P_i|f)$ 作为置信度将 $f$ 分配给部分 $P_i$ 。每一个部分 $P_i (i = 1, 2, \dots, p)$ 是从所有的列向量 $f$ 中以 $P(P_i|j)$ 作为采样权重采样出来的，如： $P_i = \{P(P_i|f) * f, \forall f \in F\}$ ，其中 $F$ 是 $T$ 中的列向量的一个复杂集合， $\{.\}$ 代表采样操作。通过执行此操作，建议的精炼part池化将进行“软”和自适应分区，以精炼原始的“硬”和统一分区，并且源自统一分区的异常值将被重新定位。

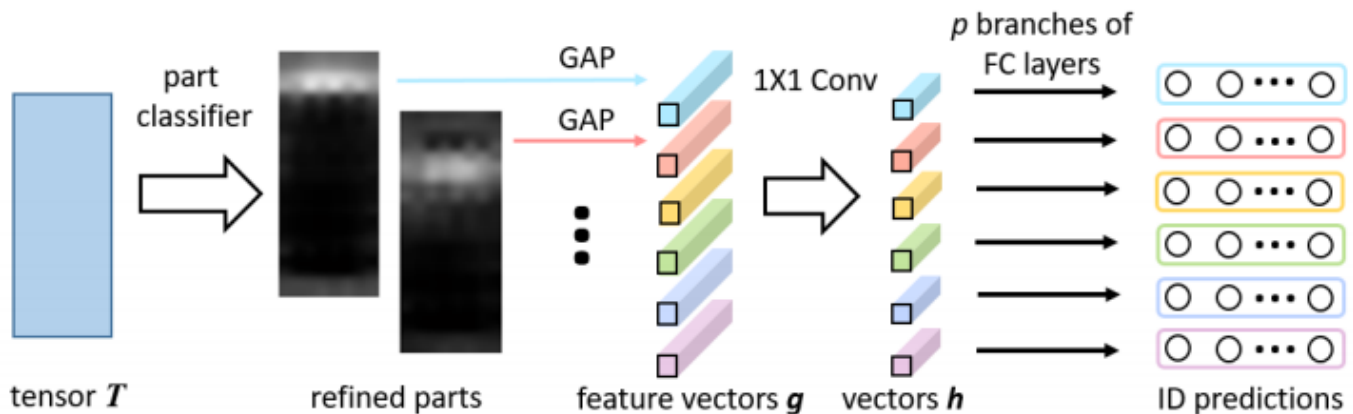


Figure 4. PCB in combination with refined part pooling. The 3D tensor  $T$  is denoted simply by a rectangle instead of a cube as we focus on the spatial partition. Layers before  $T$  are omitted as they remain unchanged compared with Fig. 2. A part classifier predicts the probability of each column vector belonging to  $p$  parts. Then each part is sampled from all the column vectors with the corresponding probability as the sampling weight. GAP denotes global average pooling.

## part分类器的归纳训练

$W$ 的训练如下：

- 1.训练一个标准的PCB模型来，使其与等分的 $T$ 收敛
- 2.在 $T$ 之后移除原始的平均池化层，增加一个 $p$ 类的part分类器
- 3.固定已训练好的PCB层，只训练part分类器。然后将此模型重新在训练集上训练。
- 4.最后，所有的层都允许更新，整个个PCB被微调

---

## Algorithm 1: Induced training for part classifier

---

**Step 1.** A standard PCB is trained to convergence with uniform partition.

**Step 2.** A  $p$ -category part classifier is appended on the tensor  $T$ .

**Step 3.** All the pre-trained layers of PCB are fixed. Only the part classifier is trainable. The model is trained until convergence again.

**Step 4.** The whole net is fine-tuned to convergence for overall optimization.

---

## 实验

### 数据集和设置

实验在Market-1501, DukeMTMC-reID, CUHK03数据集分别进行

### 实现细节

#### 实现IDE进行比较

Models	Feature	dim	Market-1501				DukeMTMC-reID				CUHK03			
			R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
IDE	pool5	2048	85.3	94.0	96.3	68.5	73.2	84.0	87.6	52.8	43.8	62.7	71.2	38.9
IDE	FC	256	83.8	93.1	95.8	67.7	72.4	83.0	87.1	51.6	43.3	62.5	71.0	38.3
Variant 1	$\mathcal{G}$	12288	86.7	95.2	96.5	69.4	73.9	84.6	88.1	53.2	43.6	62.9	71.3	38.8
Variant 1	$\mathcal{H}$	1536	85.6	94.3	96.3	68.3	72.8	83.3	87.2	52.5	44.1	63.0	71.5	39.1
Variant 2	$\mathcal{G}$	12288	91.2	96.6	97.7	75.0	80.2	88.8	91.3	62.8	52.6	72.4	80.9	45.8
Variant 2	$\mathcal{H}$	1536	91.0	96.6	97.6	75.3	80.0	88.1	90.4	62.6	54.0	73.7	81.4	47.2
PCB	$\mathcal{G}$	12288	92.3	97.2	98.2	77.4	81.7	89.7	91.9	66.1	59.7	77.7	85.2	53.2
PCB	$\mathcal{H}$	1536	92.4	97.0	97.9	77.3	81.9	89.4	91.6	65.3	61.3	78.6	85.6	54.2
PCB+RPP	$\mathcal{G}$	12288	<b>93.8</b>	<b>97.5</b>	<b>98.5</b>	<b>81.6</b>	<b>83.3</b>	<b>90.5</b>	<b>92.5</b>	<b>69.2</b>	62.8	79.8	86.8	56.7
PCB+RPP	$\mathcal{H}$	1536	93.1	97.4	98.3	81.0	82.9	90.1	92.3	68.5	<b>63.7</b>	<b>80.6</b>	<b>86.9</b>	<b>57.5</b>

Table 1. Comparison of the proposed method with IDE and 2 variants. Both variants are described in Section 3.3. pool5: output of Pool5 layer in ResNet50. FC: output of the appended FC layer for dimension reduction.  $\mathcal{G}$  ( $\mathcal{H}$ ): feature representation assembled with column vectors  $g$  ( $h$ ). Both  $g$  and  $h$  are illustrated in Fig. 2.



PCB是一个强健的baseline

RPP提高了PCB的mAP

使用 $p$ 损失的好处

对每个part采用各自的损耗对于学习区分part的特征至关重要

在分类时不共享参数的好处

与前沿研究的对比

Methods	R-1	R-5	R-10	mAP
BoW+kissme [39]	44.4	63.9	72.2	20.8
WARCA[16]	45.2	68.1	76.0	-
KLFDA[17]	46.5	71.1	79.9	-
SOMAnet[1]	73.9	-	-	47.9
SVDNet[28]	82.3	92.3	95.2	62.1
PAN[42]	82.8	-	-	63.4
Transfer [10]	83.7	-	-	65.5
Triplet Loss [14]	84.9	94.2	-	69.1
DML [36]	87.7	-	-	68.8
MultiRegion [30]	66.4	85.0	90.2	41.2
HydraPlus [22]	76.9	91.3	94.5	-
PAR [37]	81.0	92.0	94.7	63.4
MultiLoss [19]	83.9	-	-	64.4
PDC* [27]	84.4	92.7	94.9	63.4
PartLoss [35]	88.2	-	-	69.3
MultiScale [3]	88.9	-	-	73.1
GLAD* [31]	89.9	-	-	73.9
PCB	92.3	97.2	98.2	77.4
PCB+RPP	<b>93.8</b>	<b>97.5</b>	<b>98.5</b>	<b>81.6</b>

Table 2. Comparison of the proposed method with the art on Market-1501. The compared methods are categorized into 3 groups. Group 1: hand-crafted methods. Group 2: deep learning methods employing global feature. Group 3: deep learning methods employing part features. \* denotes those requiring auxiliary part labels. Our method is denoted by “PCB” and “PCB+RPP”.

Methods	DukeMTMC-reID		CUHK03	
	rank-1	mAP	rank-1	mAP
BoW+kissme [39]	25.1	12.2	6.4	6.4
LOMO+XQDA [20]	30.8	17.0	12.8	11.5
GAN [43]	67.7	47.1	-	-
PAN [42]	71.6	51.5	36.3	34.0
SVDNet [28]	76.7	56.8	41.5	37.3
MultiScale [3]	79.2	60.6	40.7	37.0
TriNet+Era [45]	73.0	56.6	55.5	50.7
SVDNet+Era [45]	79.3	62.4	48.7	43.5
PCB (UP)	81.8	66.1	61.3	54.2
PCB (RPP)	<b>83.3</b>	<b>69.2</b>	<b>63.7</b>	<b>57.5</b>

Table 3. Comparison with prior art on DukeMTMC-reID and CUHK03. Rank-1 accuracy (%) and mAP (%) are shown.

文章算法达到的结果为：mAP = 81.6%, 69.2%, 57.5% and Rank-1 = 93.8%, 83.3% and 63.7% for Market-1501, Duke and CUHK03,

## 参数分析

首先，较大的图像尺寸有利于学习的part特征。mAP和rank-1精度都随图像大小而增加，直到达到稳定的性能。

其次，较小的下采样率，即张量T的较大空间大小会增强性能，尤其是在使用相对较小的图像作为输入时。

## 归纳和注意机制

通过归纳学习的精炼部分优于通过注意力机制学习的部分。具有归纳和注意力机制的分区结果如下所示：

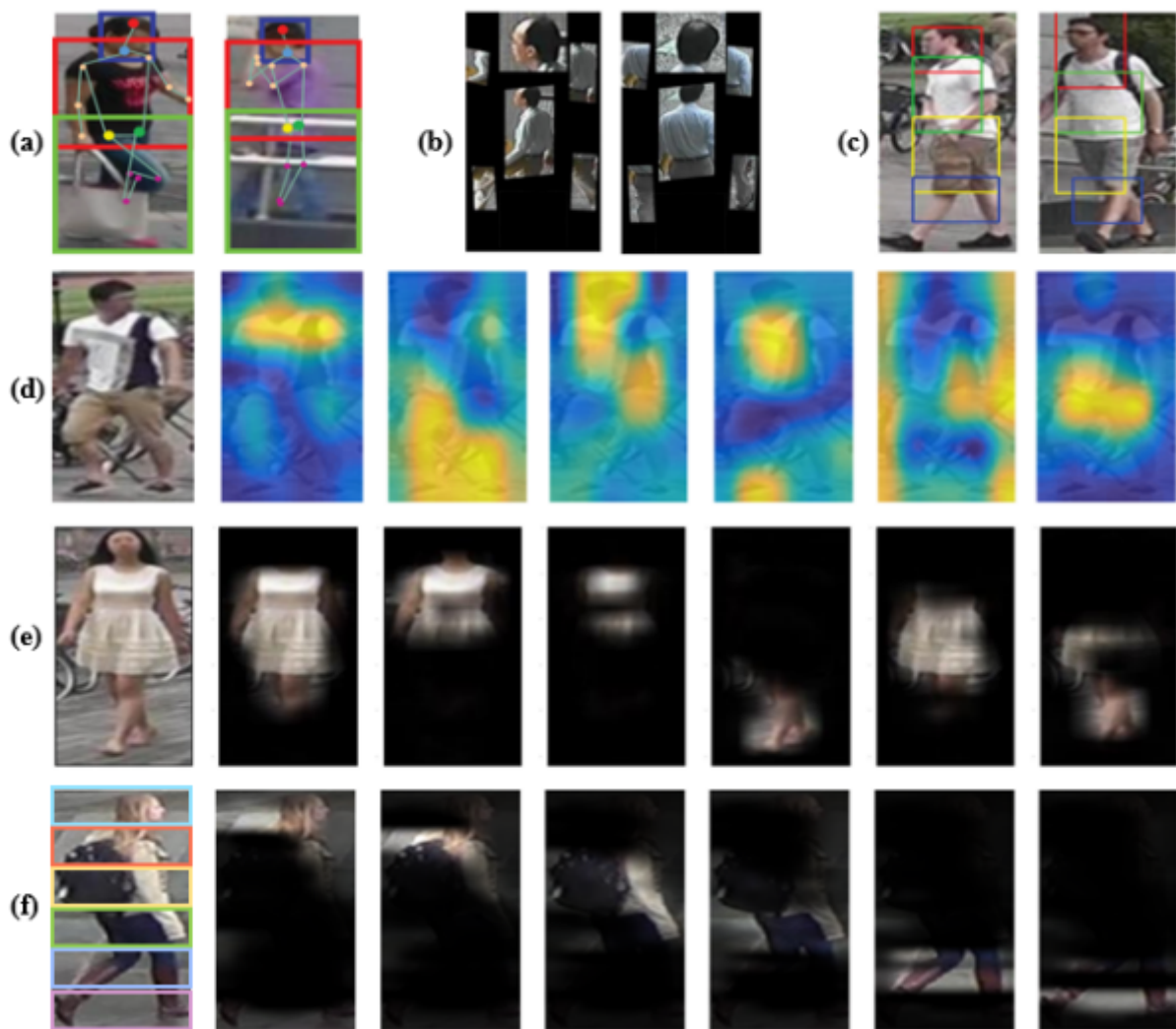


Figure 1. Partition strategies of several deep part models in person retrieval. (a) to (e): Partitioned parts by GLAD [31], PDC [27], DPL [35], Hydra-plus [22] and PAR [37], respectively, which are cropped from the corresponding papers. (f): Our method employs a uniform partition and then refines each stripe. Both PAR [37] and our method conduct “soft” partition, but our method differs significantly from [37], as detailed in Section 2.

## 总结

本文首先提出用于学习part特征的PCB，使用了简单的统一分配策略和结合part特征到卷积描述符中。文章使用精炼部分池化的方法，以增强每个part的part内部一致性。经过精炼之后，相似的列向量被统计到一个part中，使每个part更加内联。精制的part池化不需要part标签信息，并大大改善了PCB。