# 阅读纲要

## 1 自己的总结、评价以及应用

本文主要介绍了基于deep learning的Person Re-Identification发展与分类。

主要分为发展/研究趋势：the closed-world and open-world settings，其中the closed-world得到了广泛研究，它可以根据deep feature representation learning, deep metric learning and ranking optimization这三个角度进行进一步分类。另外还有research-oriented scenarios 和practical applications的区别（gap）。

## 2 文章的主要问题（abstract、疑问句中）

Person Re-Identification的目的：

Given a query person-ofinterest, the goal of Re-ID is to determine whether this person has appeared in another place at a distinct time captured by a different camera（给定查询感兴趣的人，Re-ID的目标是确定此人是否在不同的相机捕获的不同时间出现在另一个地方）。

Re-ID现实需要：

Due to the urgent demand of public safety and increasing number of surveillance cameras in university campuses, theme parks, streets, etc,

person Re-ID is imperative in intelligent video surveillance system designs（智能视频监控系统设计）.

Re-ID存在的挑战：

Person Re-ID is a challenging task due to the presence of different viewpoints [10], [11], varying low-image resolutions [12], [13], illumination changes [14], unconstrained poses [15], [16], [17], occlusions [18], [19], heterogeneous

modalities [9], [20], etc.

构建Re-ID系统的五个步骤：

building a person Re-ID system for a specific scenario requires five main steps.

①收集原始数据

Step 1: Raw Data Collection:Obtaining raw video data from surveillance cameras is the primary requirement of practical video investigation. These cameras are usually located in different places under varying environments [39]. Most likely, this raw data contains a large amount of complex and noisy background clutter.

②生成边界框

Step 2: Bounding Box Generation: Extracting the bounding boxes which contain the person images from the raw video data.

③训练数据标签

Step 3: Training Data Annotation: Annotating the crosscamera labels. Training data annotation is usually indispensable for discriminative Re-ID model learning due to the large cross-camera variations. In the existence of large domain shift （区域差别很大）[44], we often need to annotate the training data in every new scenario.

④模型训练

Step 4: Model Training: Training a discriminative and robust Re-ID model with the previous annotated person images/videos. This step is the core for developing a Re-ID system and it is also the most widely studied paradigm in the literature. Extensive models have been developed to handle the various challenges, concentrating on feature representation learning [45], [46], distance metric learning [47], [48] or their combinations.

⑤行文检索

Step 5: Pedestrian Retrieval: The testing phase conducts the pedestrian retrieval.

## 3 结论（abstract以及conclusion中）

## 4 思路脉络（小标题中的关键句）

2 CLOSED-WORLD PERSON RE-IDENTIFICATION

a standard closed-world Re-ID system contains three main components（三个主要组成部分）：

①Feature Representation Learning

②Deep Metric Learning

③Ranking Optimization

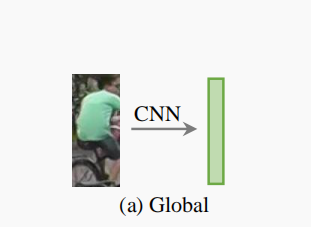
2.1 Feature Representation Learning

There are four main categories/feature learning strategies.（主要有三类）：

1. Global Feature
2. Local Feature
3. Auxiliary Feature
4. Video Feature

2.1.1 Global Feature Representation Learning

Global feature representation learning extracts a global feature vector for each person image.



在发展过程中，人们提出了很多模型、算法，其中这个模型值得注意一下：

通过将每个person视为一个class，将Person Re-ID问题视为一个多类别分类问题。

An ID-discriminative Embedding (IDE) model is presented in [46], which views the training process of person Re-ID as a multi-class classification problem by treating each identity as a distinct class.

Attention Information（这是一个小标题）－－一种强化（增强）表示representation学习的策略，又可以分为以下两类：

1. For attention within the person image

Harmonious Attention CNN (HA-CNN) model [79] jointly learns the soft pixel attention and hard regional attention to enhance the robustness of feature representation against misalignment.

1. For attention across multiple person images

A context-aware attentive feature learning method.

Architecture Modification（这是另一个小标题）

主要有这几种调整方法：  
①An improved bottleneck layer utilizes the orthogonality constraint to reinforce global feature representation learning.

②Using a simple architecture with standard networks, embeddings from multiple layers are aggregated into a single embedding.

③A Class Activation Maps (CAM) augmentation model [90] expands the activation scope to explore rich visual cues in a multi-branch network.

2.1.2 Local Feature Representation Learning

Local feature representation usually learns part/region aggregated features（局部聚合特征）， The body parts are either generated by human pose estimation or roughly horizontal division.

①The main trend is to combine the full body representation and local part features.

②Some works have also studied the robustness of part level feature learning against background clutter.

③For specific part attention design

④For horizontal region features without pose estimation

**2.1.3 Auxiliary Feature Representation Learning**

辅助特征表示，本文提出了以下几种辅助特征：

Semantic Attributes、Viewpoint Information、Domain Information、Generation/Augmentation

**2.1.4 Video Feature Representation Learning**

**2.1.5 Architecture Design**

算法结构多采用神经网络的结构作为骨架，即Framing person Re-ID as a specific pedestrian retrieval problem, most existing works adopt the network architectures [32], [70], [71], [72] designed for image classification as

the backbone.

## 5 难理解点

## 6 本文所用方法有什么不同

主要有三点不同：

Our survey makes three major differences:

①进行了深入、综合分析

1. We provide an in-depth and comprehensive analysis of existing deep learning methods by discussing their advantages and limitations, rather than a simpleoverview. This provides insights for future algorithm designand new topic exploration.

②提出了新的AGW准则和新的评估标准

1. We design a new powerful AGW baseline and a new evaluation metric (mINP) for future developments. AGW achieves state-of-the-art performance on both single- and cross-modality Re-ID tasks. mINP provides a supplement metric to existing CMC/mAP, indicating the cost to find all the correct matches.

③探讨未来的研究方向

3) We make an attempt to discuss several important research directions with under-investigated open issues to narrow the gap between the closed-world and open-world applications, taking a step towards real-world Re-ID system design.

专业术语：

bilinear-pooling：双线性池

undesirable background features：不良背景特征

Semantically：语义地

Aligned：对齐的 well-align：非常对齐的

specific part：特定部位

high-order polynmial predictor：高阶多项式预测器

scale maps ：比例图

discriminative features：判别特征

Long Short-Term Memory (LSTM) architecture：长短期记忆架构

Discriminability：判别能力，可分辨性

contextual information：上下文信息

partition strategy：分配策略 partition ：分区

Baseline：基准

heavy occlusions：重度咬合

semantic attributes：语义属性

Consistency：一致性

view-generic and view-specific learning：通用视图和特定视图的学习

pose-normalized image generation approach：姿势归一化图像生成方法

Deep Metric Learning：深度度量学习