Pedestrian Re-identification

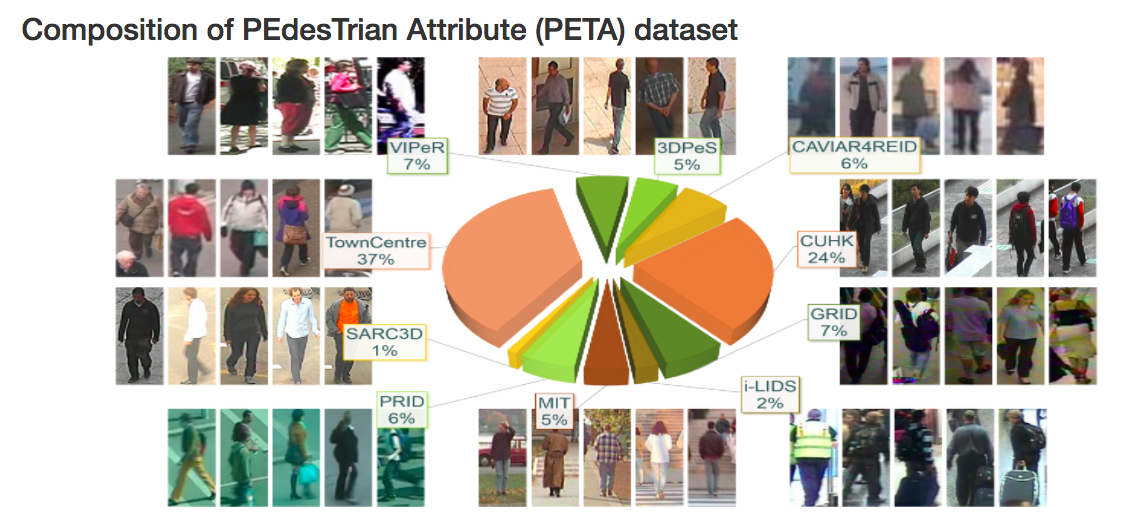
Previous

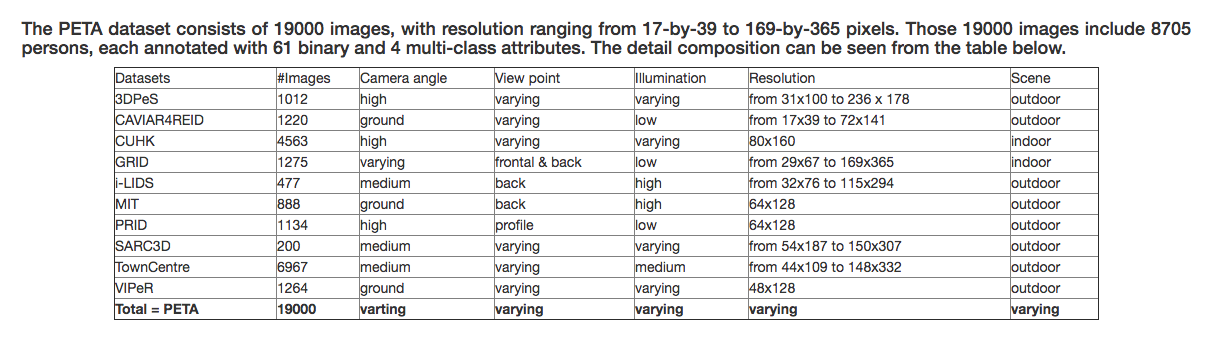
* ECCV2016
* Benchmark
  1. MARS: A Video Benchmark for Large-Scale Person Re-identification (video)
* Non-typical approach
  1. Human-In-The-Loop Person Re-Identification
  2. Person Re-identification by Unsupervised L1 Graph Learning
  3. Human Re-identification in Crowd Videos using Personal, Social and Environmental Constraints (video)
* Typical approach
  1. Embedding Deep Metric for Person Re-identification A Study Against Large Variations
  2. .Person Re-Identification via Recurrent Feature Aggregation (video)
  3. Temporal Model Adaptation for Person Re-Identification
  4. Deep Attributes Driven Person Re-identification
* Siamese Network
  1. Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification
  2. A Siamese Long Short-Term Memory Architecture for Human Re-Identification

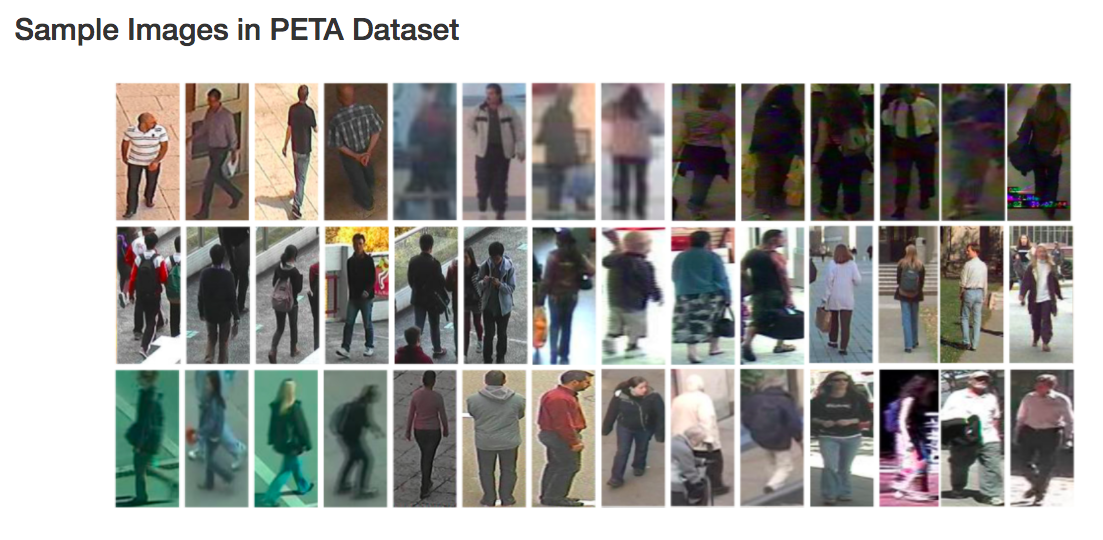
Supplement

* Benchmark
  1. PATE：Pedestrian Attribute Recognition At Far Distance ( <http://mmlab.ie.cuhk.edu.hk/projects/PETA.html> )
* Non-typical approach
  1. Hierarchical Gaussian Descriptor for Person Re-Identification (not CNN) (CVPR2016)
  2. Video-Based Pedestrian Re-Identification by Adaptive Spatio-Temporal Appearance Model. (IEEE Transactions on Image Processing, 2017.)
  3. Enhancing Person Re-identification in a Self-trained Subspace (ACM TOMM2017)
  4. Relevance Subject Machine: A Novel Person Re-identification Framework (IEEE.PRML)
* Typical approach（Presentation, 2017/4/28）
  1. Person Re-Identification Using CNN Features Learned from Combination of Attributes (ICLR2016)
  2. Improving Person Re-identification by Attribute and Identity Learning (L Zheng)
  3. Person Re-Identification by Multi-Channel Parts-Based CNN with ImprovedTriplet Loss Function (CVPR2016)
  4. Beyond triplet loss: a deep quadruplet network for person re-identification (CVPR2017)
  5. Person Search with Natural Language Description (Sensetime)
  6. Video-based Person Re-identification with Accumulative Motion Context (颜水成)
  7. Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro
  8. Pose Invariant Embedding for Deep Person Re-identification
  9. SVDNet for Pedestrian Retrieval
  10. Re-ranking Person Re-identification with k-reciprocal Encoding (CVPR2017)

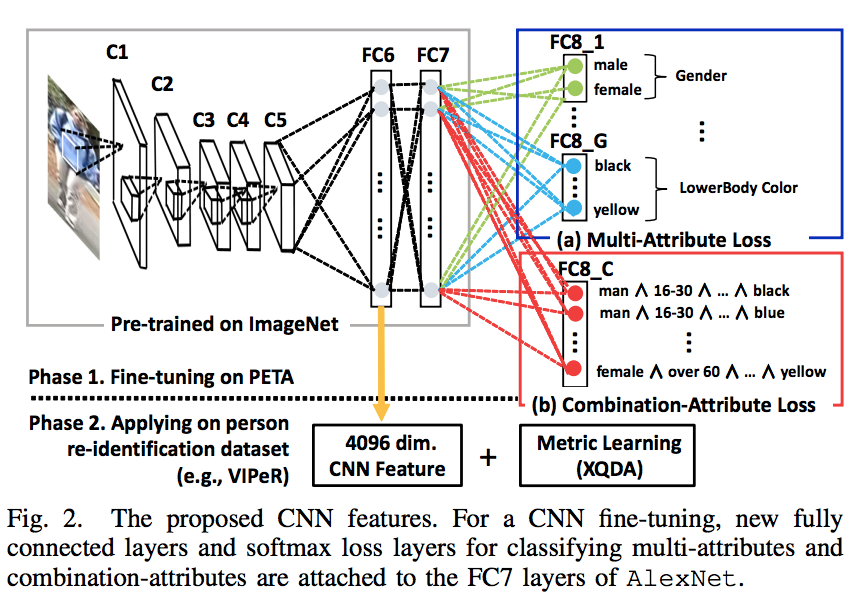
PETA数据集：<http://mmlab.ie.cuhk.edu.hk/projects/PETA.html>





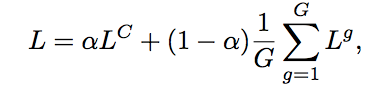


* Person Re-Identification Using CNN Features Learned from Combination of Attributes(ICLR2016)

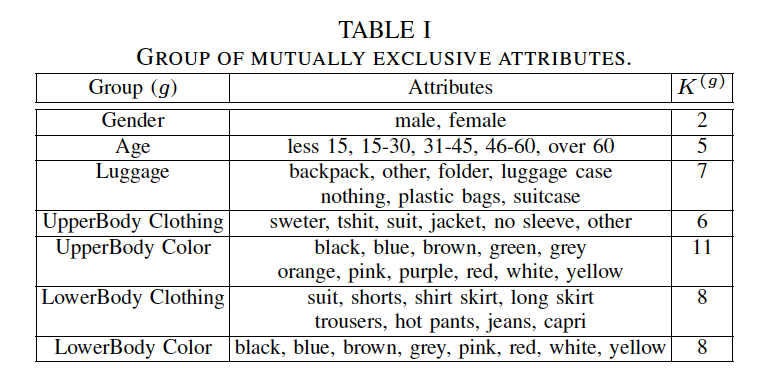


论文核心思想是对Alexnet进行fine-tuning，Alexnet模型是在ImageNet数据集上预训练的。Fine-tuning分为两步，第一步是在相关数据集上对整个网络进行训练，所有的attributes都会计算loss用于训练所有层，第二步是把训练的网络的FC6层作为输出，得到的特征向量再次进行一次metric learning来实现Re-id的任务。

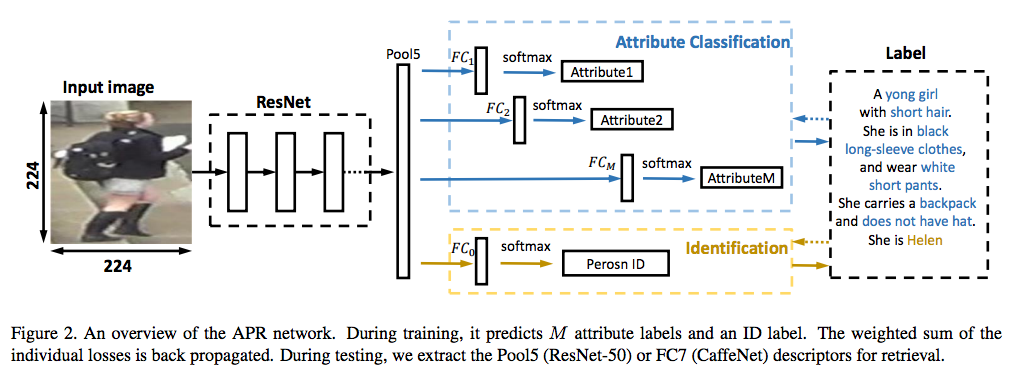
* + 标签损失
    - 全标签LC，集成所有属性：属性1类别×属性2类别数×…×属性N类别数
    - 第G个属性的标签损失LG
  + 损失函数
    - 全标签的交叉熵损失和每个属性的交叉熵损失平均的加权



* + Label

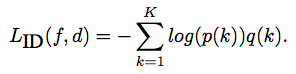


* Improving Person Re-identification by Attribute and Identity Learning (L Zheng)

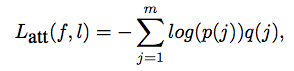


这篇论文和上一篇论文几乎完全一样，最主要的不同点在于计算损失的时候Loss增加了ID的分类loss，另外一个小的不同点是CNN使用的网络也不一样。

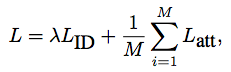
* + ID loss



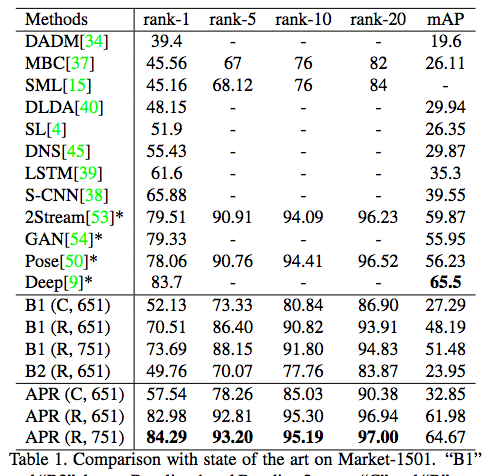
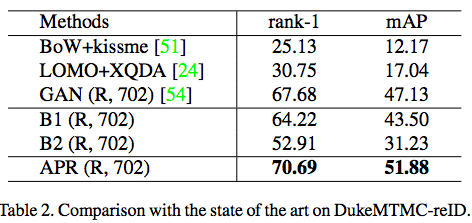
* + Attributes loss



* + Final loss

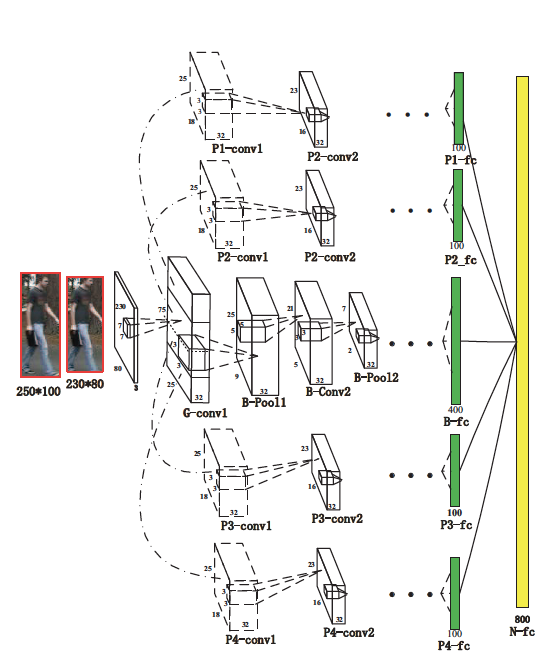


* + Result

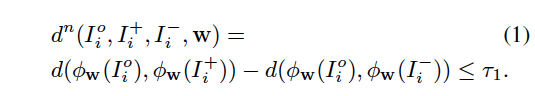
 

* Person Re-Identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function (CVPR2016)

思路使用典型的Triplet loss，一个原始样本，一个正样本和一个负样本，共享CNN网络层。改进点在于CNN层融合了多个CNN网络层进行Ensemble learning。另外通过crop的方式进行了data argumentation。最大改进点在于设计了loss function，同时考虑了类内损失和类间损失。结果显示基本都提高了5~10%



* + Loss function
    - Inter-class-constraint



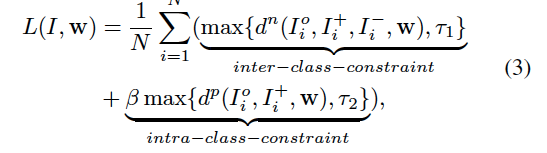
* + - Intra-class-constraint



* + - L2-norm distance



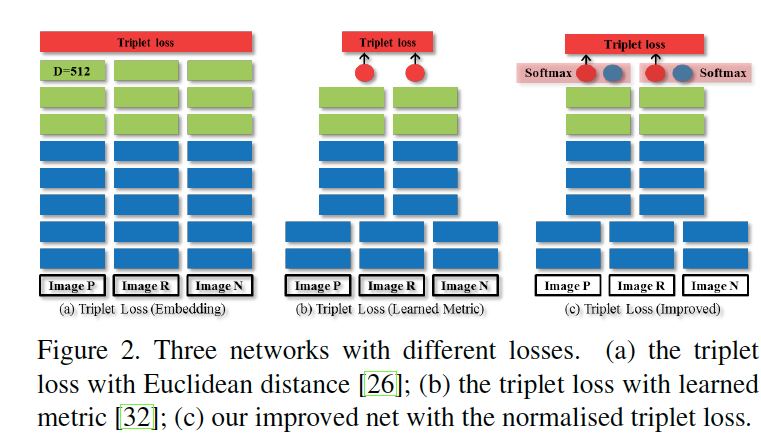
* + - Improved triplet loss function



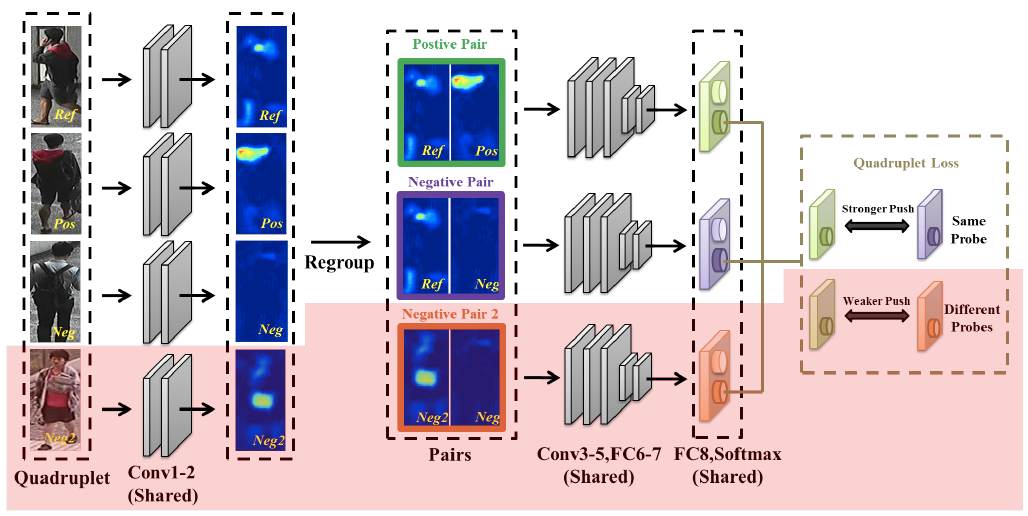
* Beyond triplet loss: a deep quadruplet network for person re-identification (CVPR2017)

本文主要贡献在于设计了一个新的loss function，改变了传统的triplet loss，设计了一个新的quadruple loss，并且声称击败了绝大部分的state-of-the-art。

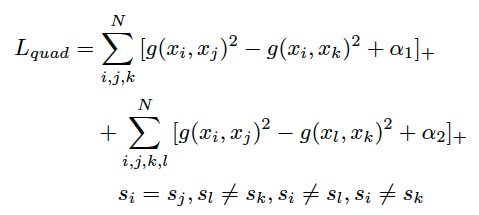
* + Loss流程，Triplet Loss (Improved)为论文提出



* + Network

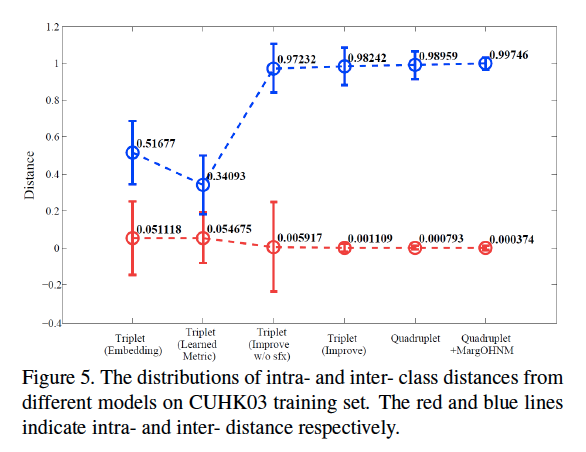


* + Quadruplet loss

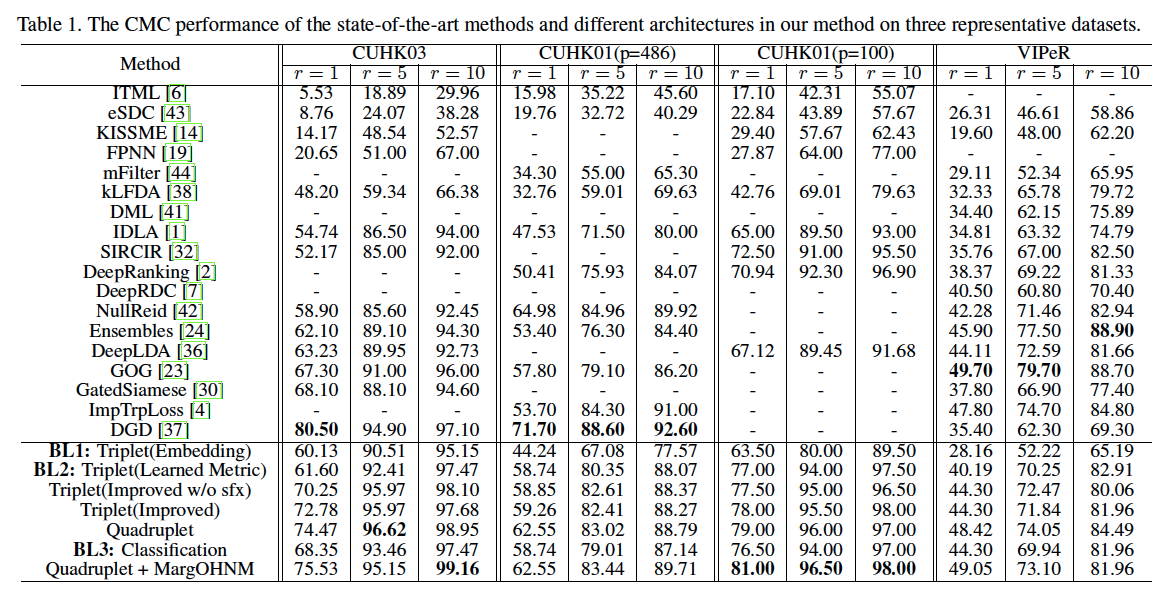


其中i,j是正样本，k,l分别是两个不一样的负样本

* + Distance visualization



* + Result



* Person Search with Natural Language Description (Sensetime)

这篇文章和别的文章最重要的不同在于不再使用离散的属性标签，而是使用信息量更加丰富同时也更加杂乱的NLP标签，这样的标签更加符合人类的描述，但是也同时处理难度更大，更加适合用于互联网搜索、智能客服机器人等场景。本文给出了商汤自己标注的语义Person Re-ID数据集CHUK-PEDES，总共有80000+语句，40000+图片，13000+人。本文的算法使用的是带Attention model的RNN网络GNA-RNN。论文里提高还特意屏蔽了语句中的某些单词，来观测哪些单词的重要度更高，从而达到提升标准性能的目的。

* + 总体性能

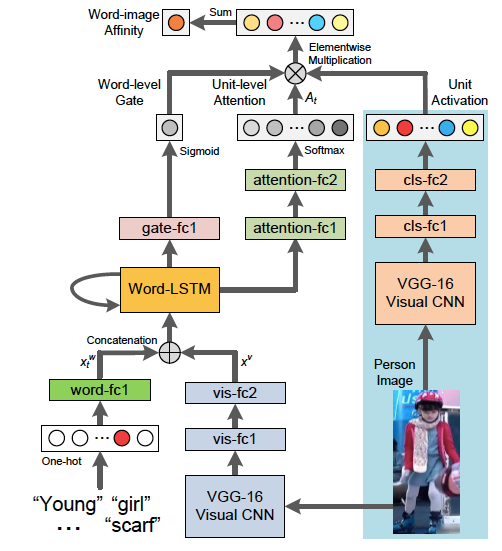
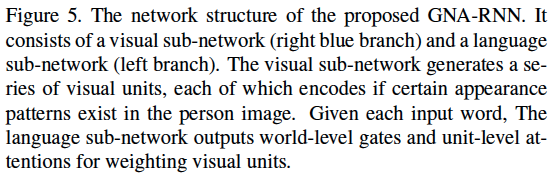
Attributes: Top1:33.3%; Top5:74.7%; time:81.84s

Description: Top1:58.7%; Top5:92.0%; time:62.18s

* + 数据集示例



* + GNA-RNN model

* + - Visual sub-network

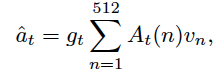
视觉网络是上图中蓝色的部分，用的是VGG-16和512维输出的FC7，之后添加了两个FC1,FC2进行fine-tuning，输出一个visual feature。

* + - Language sub-network

NLP网络的输入时word vector和image feature，网络是LSTM结合attention unit，每次输入一个单词都会在image上更新attention，网络输出为两个通道，一个为attention vector，和视觉网络输出的visual feature的维度一样，另一个输出是word gate level，表示当前单词和图像的联系度，比如a,the这类单词可能值就很小。

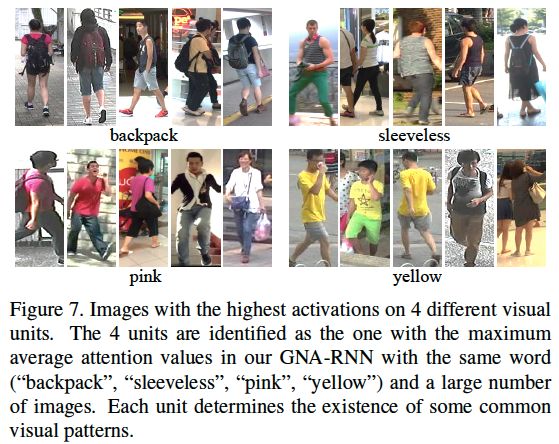
* + - 融合

最后把word gate level和attention vector和visual feature相乘，最后得到的feature放到最后相加然后进行end to end训练，最后就可以得到每个单词的affinity value，然后把所有单词的affinity value加起来就可以得到最后sentence的affinity value，这个值在0~1之间，最后和label（0表示不匹配，1表示匹配）计算交叉熵损失，训练网络。

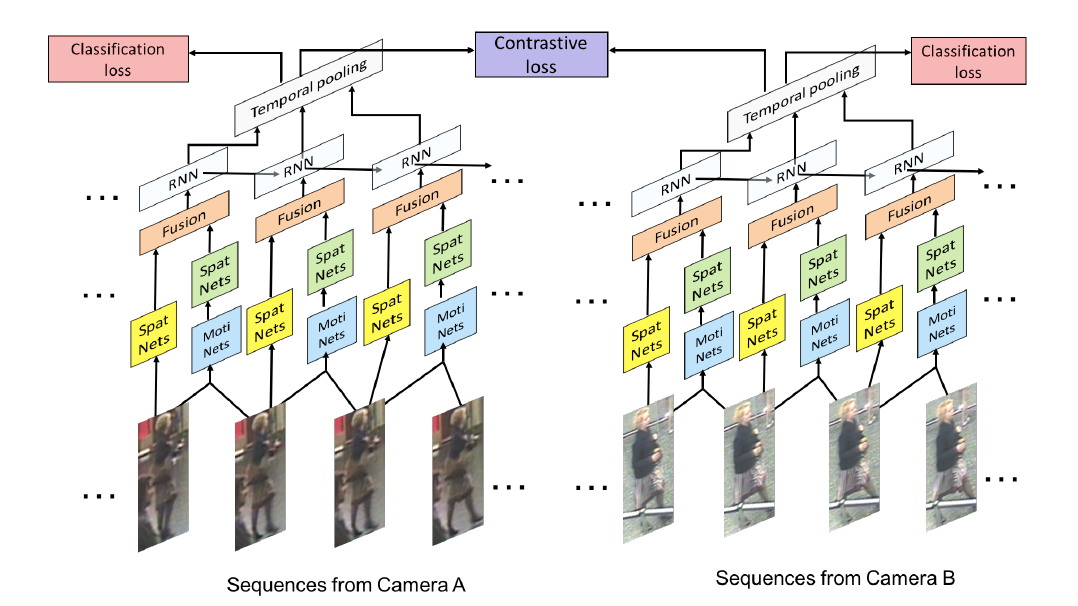
* + GNA-RNN model



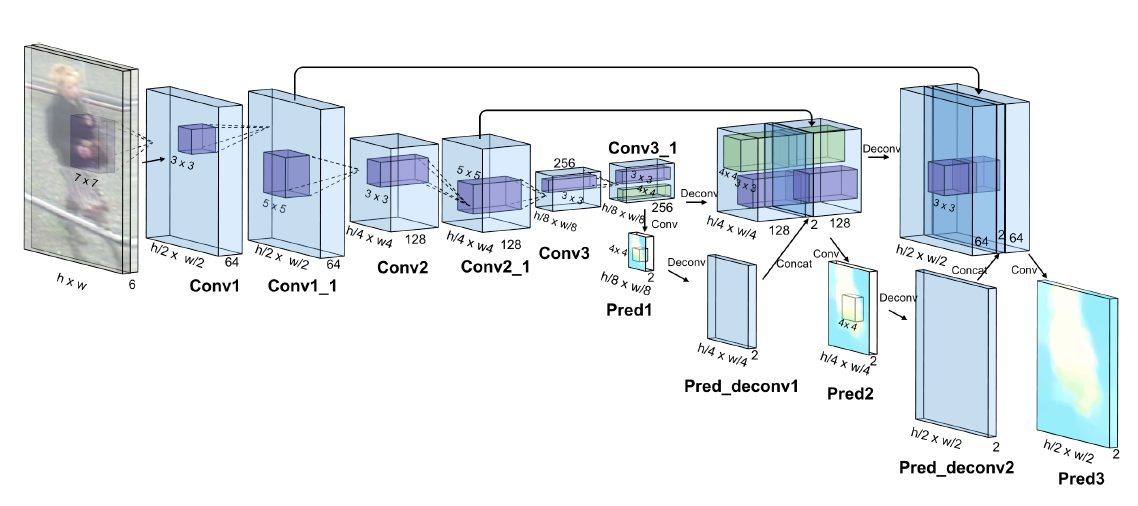


* Video-based Person Re-identification with Accumulative Motion Context (颜水成)

这篇文章最主要的思想是融合的视频的上下文信息，用了一个空间网络提取单帧图像特征，用了一个运动网络提取相邻帧运动信息，运动网络用的是FLownet (当然运动网络也可以用光流图+CNN)，运动网络的监督信号来自于传统光流法，最后把融合这些特征放到LSTM上去实现上下文特征提取（水滴科技一直致力于步态辅助Re-Id的研究），当然增加了LSTM层之后会带来计算时间的增加。

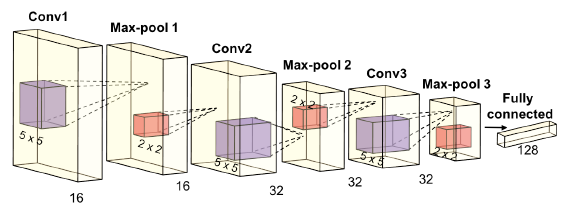


* + Moti Nets (Flownet)

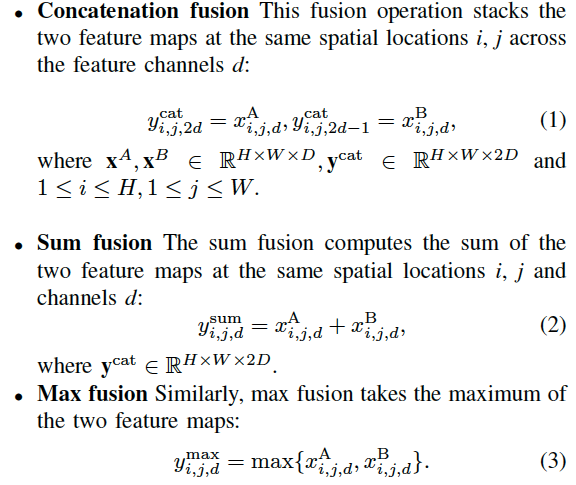
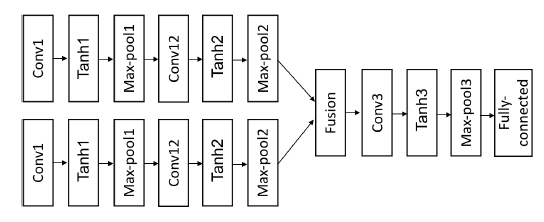


FLownet有两种模式，一种是将图像叠层输入到网络，另外一种是作为两幅图像输入到双通道的网络。本文使用的是第一种模式，并且用了三个尺度的光流图做最终的训练。

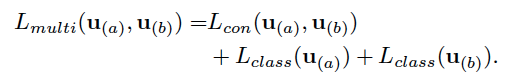
* + Spat Nets：一个简单的三层CNN网络



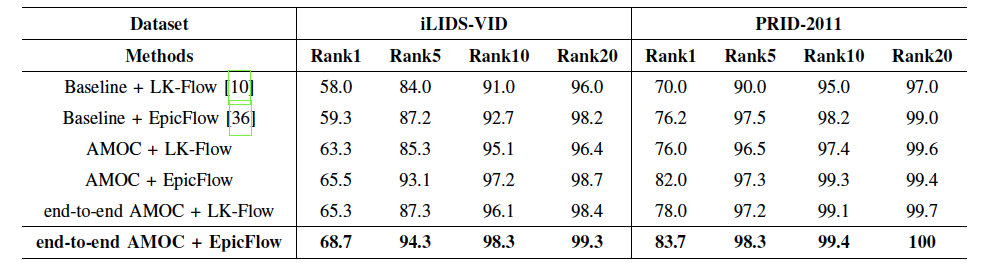
* + Fusion：三种fusion模式，实验显示concatenation fusion在max-pool2层的准确度最高

* + Loss：融合了contrastive loss和classification loss



* + Result



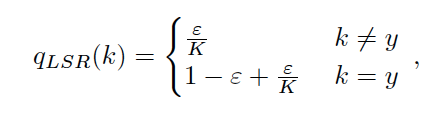
* + 训练细节

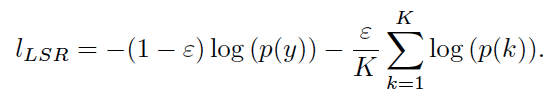
用了crop和mirror等操作进行数据增广，网络初始化细节可见原论文。

* Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro

这篇文章主要思路是用GAN网络进行数据生成，另外因为用的是无条件的随机GAN生成，所以论文提出给生成的数据用一个标签平滑正则化的技术（LSRO），所谓LSRO技术根据公式推导，最后的结果就是每个标签最后的值是1/K。CNN网络用的在ImageNet上预训练的ResNet50，然后进行fine-tuing。

* + LSRO (Label & Loss)







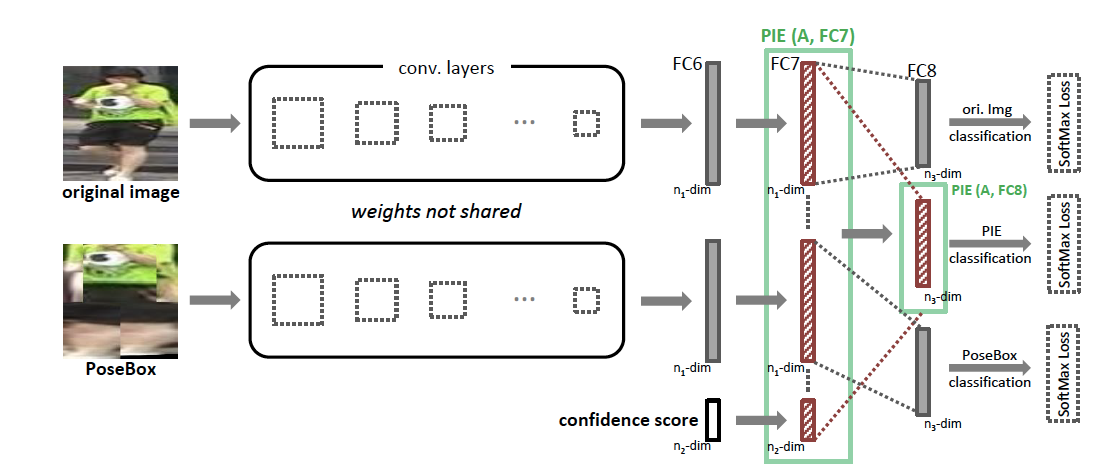
等于0的话，那么就是真实图像的one-hot标签，等于1的话就是随机的平滑标签，即1/K。

* + Result



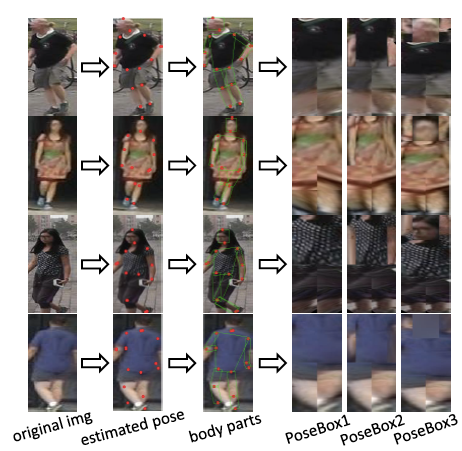
* Pose Invariant Embedding for Deep Person Re-identification

这篇文章主要思想是作者认为影响re-ID很重要的一个因素是行人的姿态不对齐，如果能够对姿态进行一个标准的对齐，就能够提高re-ID的准确度。姿态估计采用convolutional pose machines (CPM)，然后对图片进行了仿射变换，CNN网络是用ResNet-50和Alexnet的fine-tuning，原始图片是Alexnet，pose图片是ResNet-50，另外pose置信度得分的14维向量和FC7的输出向量合并。



训练网络时候用了三个loss，原始图片CNN的loss，pose图片CNN的loss，融合三个输入的FC8层loss，测试的时候只用了PIE分类时的feature，用了欧拉距离计算相似度。

* + Result



PoseBox1: torso + legs;

PoseBox2:PoseBox1 + arms;

PoseBox3: PoseBox2 + head

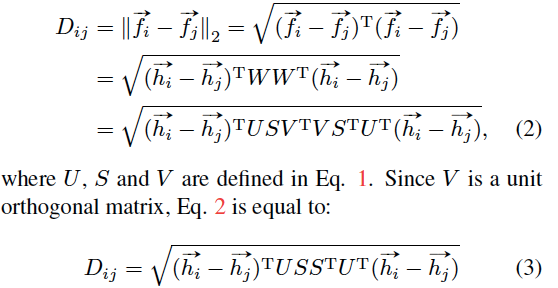
14点pose，pose图片进行了分块的放射变换

实验显示通常PoseBox2的准确度最高

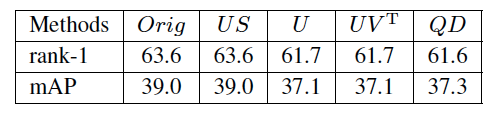
* + 结果显示基本在各大数据集都得到了最高准确度，验证了此方法可以一定程度上解决因相机变化而带来的姿态变化的问题，但是该方法错误多半出现在人体侧面无法定位到较好pose的情况下。
* SVDNet for Pedestrian Retrieval

这篇文章最主要的就是对网络进行了一个SVD分解，先用US代替W，这个替换是不掉准确度的，但是把网络权重在空间中的映射方向给改变了，然后固定US对网络进行fine-tuning，之后解放US，对整个网络进行fine-tuning，以提高准确度。

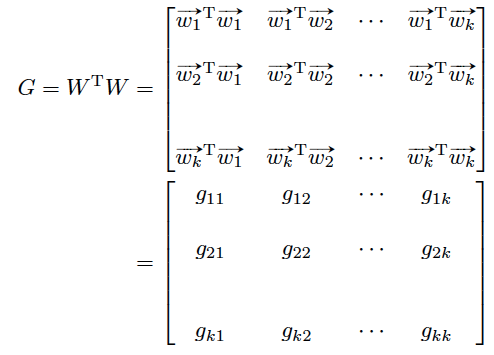
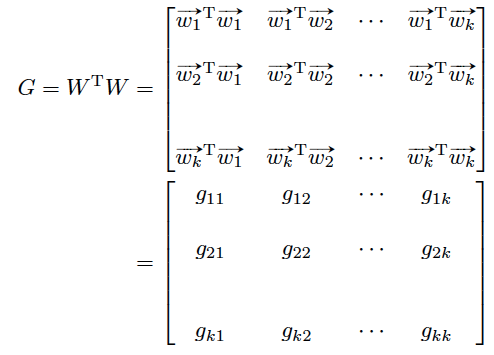
* + 公式：对权重W矩阵进行SVD分解

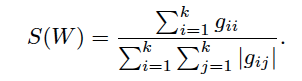


* + 初始实验：对W矩阵进行了五次替换（证明直接替换会丢失准确度，于是要进行下一步“主成分提取”）

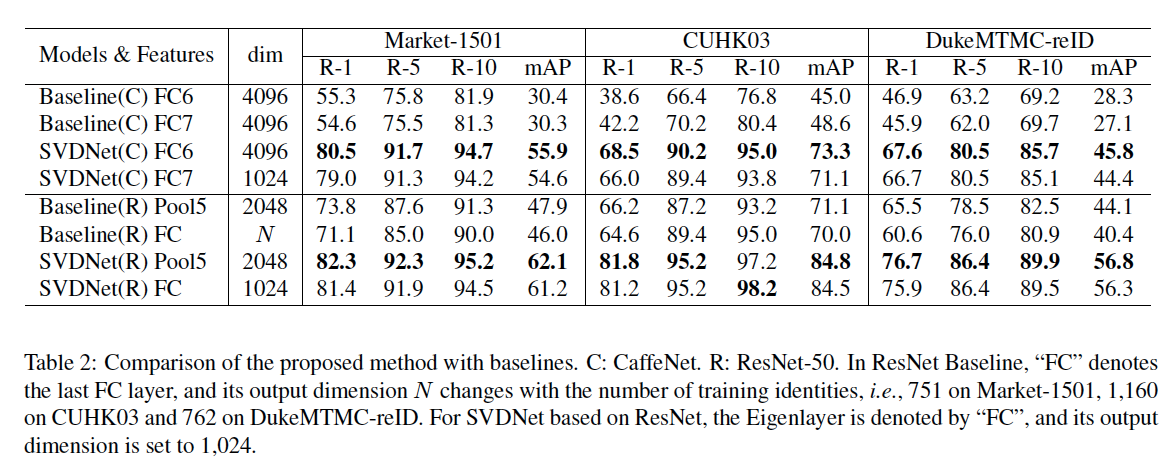


* + 成分贡献度：论通过计算成分贡献度可以了解某层网络的贡献度



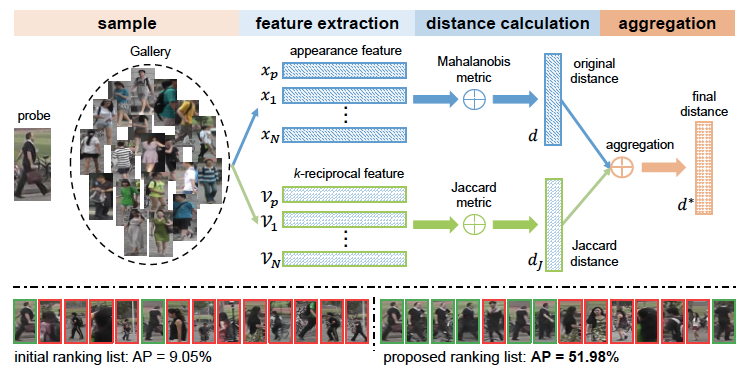


* + Result：以CaffeNet和ResNet-50为基准

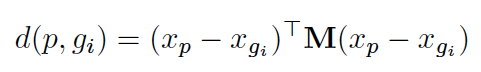


* Re-ranking Person Re-identification with k-reciprocal Encoding (CVPR2017)

这篇文章最主要的工作就是在计算相似度的距离里面引入了jaccard distance，另外一个工作是在搜索环节引入了K近邻算法，降低了搜索的计算复杂度。



* + Origin distance (Mahalanobis distance): M是个半正定矩阵，通常就是单位阵



* + K-reciprocal Nearest Neighbors

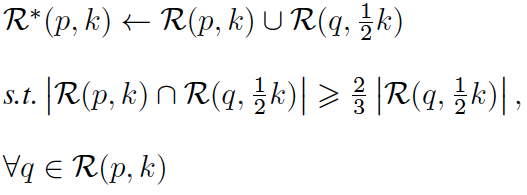
p最近的K个样本：

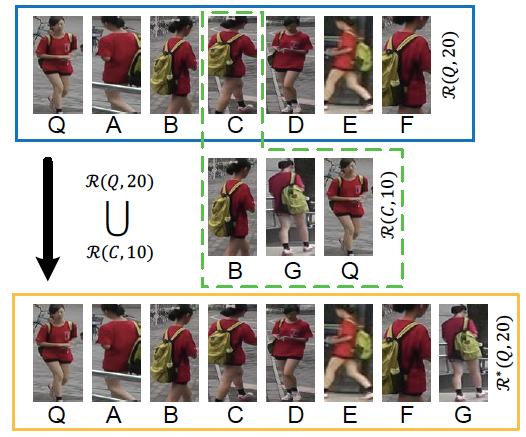


定义关系集：



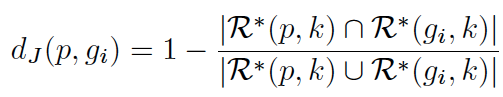
定义更加鲁棒的关系集，因为背景、角度、姿态等的变化：





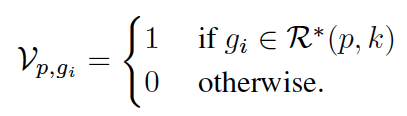
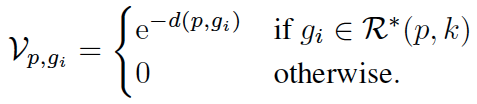
* + Jaccard distance

原始的公式

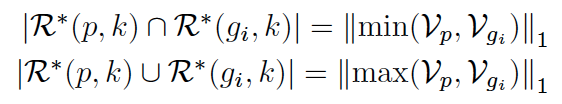


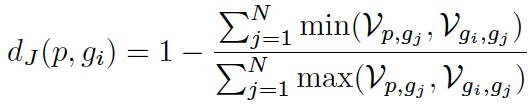
但是这个公式是很耗时并且重复计算的，所以要改进，先定义一个向量，有hard模式和soft模式，hard模式不考虑内容相似度，soft模式考虑内容相似度，所以论文最终使用soft模式

k-reciprocal feature: 

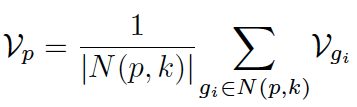
 

之后Jaccard distance近似为：





进一步化简，可以用子集中的feature平均来近似作为probe p的feature



* + Final distance



* + 在KISSME和XQDA度量方法，在LOMO和IDE特征在CaffeNet和ResNet-50上都做了对比实验，在Market-1501，CUHK03，Mars，PRW数据集上准确度均有提高

