

Multi-Level Factorisation Net for Person Re-Identification

from CVPR 2018

abstract

本文利用多层结构提取不同在不同卷积深度下的表现最优的特征，然后对特征进行整合用来实现对行人的Re-ID。在Re-ID和物体检测上都表现出了很好的效果，尤其在Re-ID上达到了sota

introduction

对于Re-ID问题，难点在于同一人的姿态、外观等细小特征表现不同，同时对于输入的数据中还包含了摄像头的角度不同等因素，导致Re-ID成为难题。对于每一个人，可以通过图像进行一些信息的提取，例如下图：

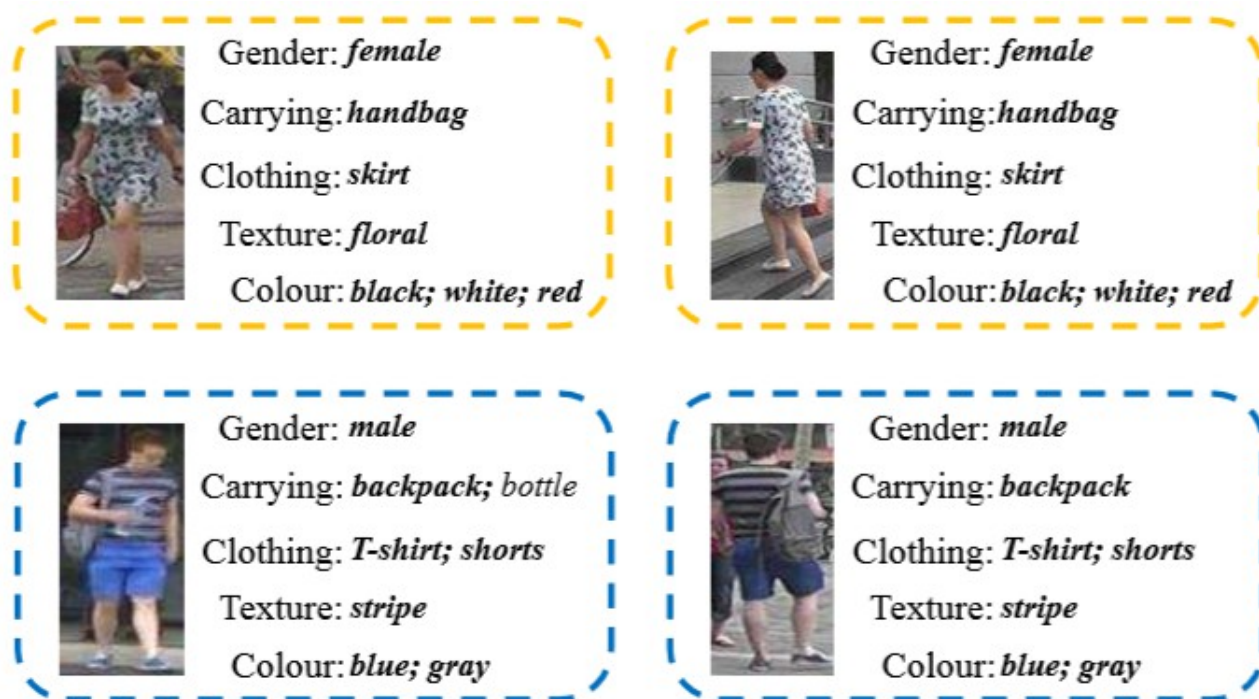
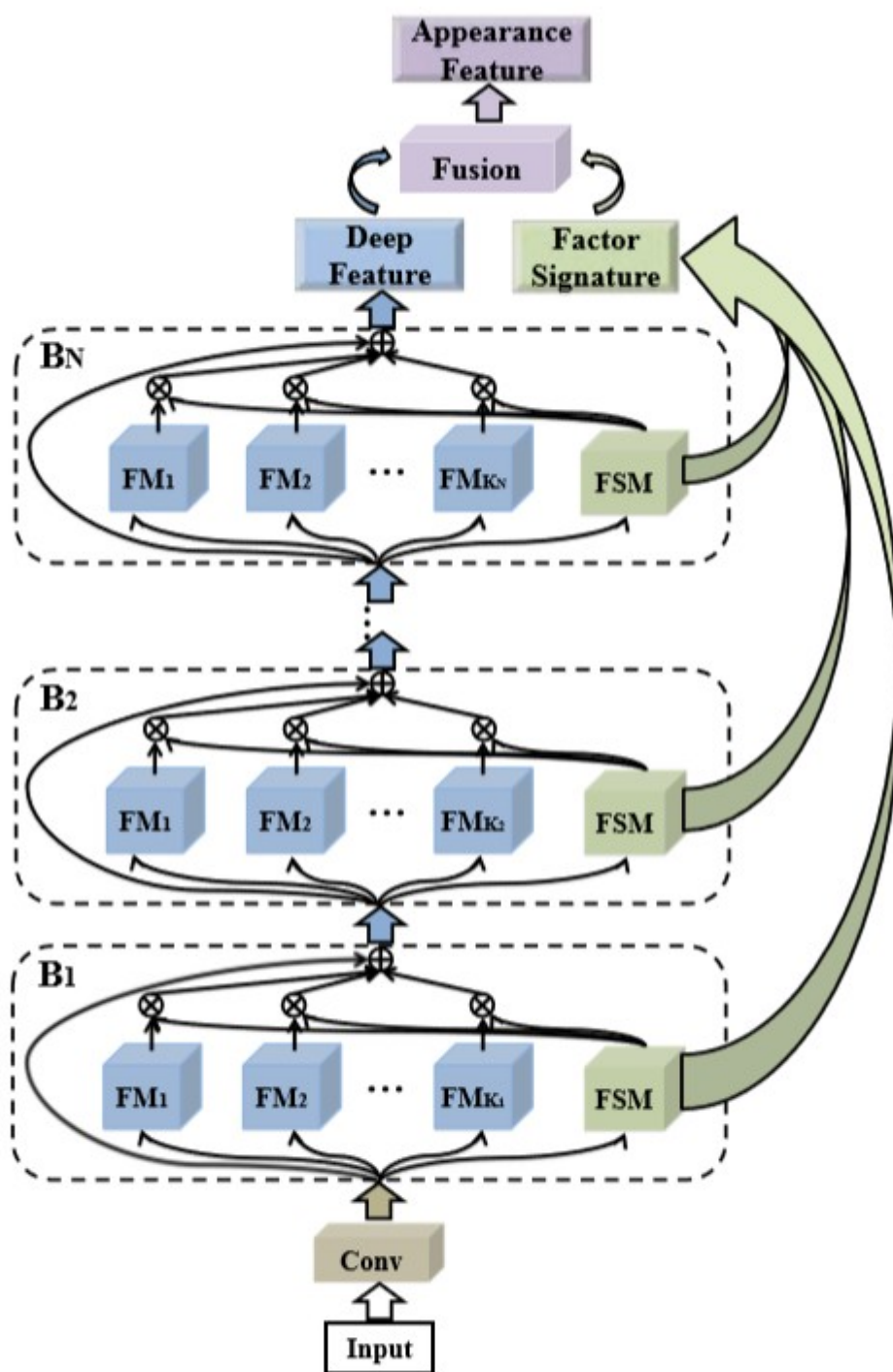


Figure 1. A person's appearance can be described by appearance factors of multiple semantic levels. Modelling the view-invariant discriminative factors is important for matching people across views. Each row shows a person captured by two camera views.

其中很多方法是通过对于这可见的语义进行深层的特征提取，然后通过这些通过高层次的语义进行识别，但这样存在高成本和高错误的问题。另外的方法是通过注意力机制对特征进行融合，但这样的做法很盲目，无法获取每层能够获取的最有效的因素特征。(例如:浅层的特征一般用于提取衣服颜色，高层的特征用于提取性别)本文中采用了多级别因素提取网络，网络结构如下：



利用该

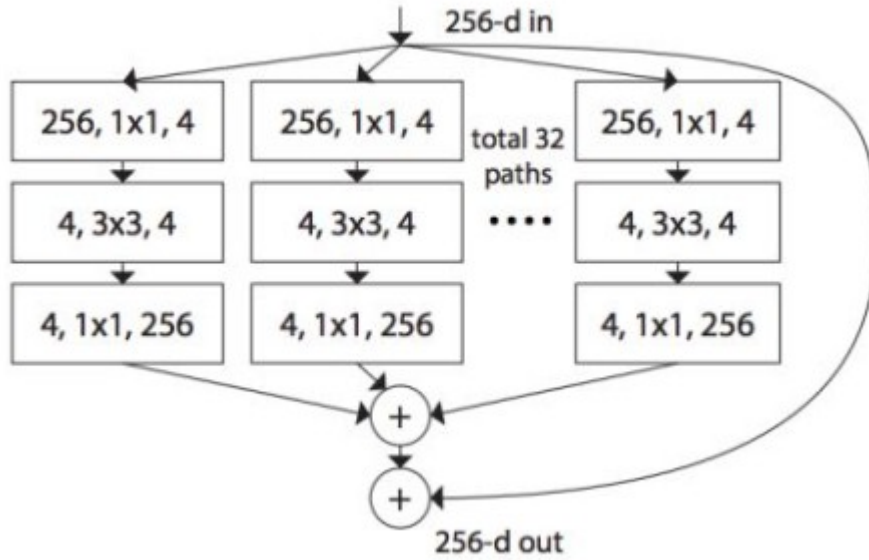
Figure 2. Illustration of Multi-Level Factorisation Net (MLFN) Architecture. Best viewed in colour.

网络将得到不同层次中可以用来分类的因素，和与分类无关的因素。

network architecture and optimisation

network architecture

MLFN(Multi-Level Factorisation Net)网络由多个block构成，每个block(B)表示一个语义级别；每个block中包含了多个Factor Module(FM)用来提取每个因素的特征，每个FM中包含了独立多个卷积层与pooling层。Factor Selection Module(FSM)因素选择模型，用来选取该block中生效的因素，对于不生效的因素，网络进行冻结处理。网络结构与ResNeXt相似，但不同的是FM之间没有共享权重。



1.FM

对于每个block中，包含了 K_n 个FM模型，每个block中的FM的个数不同，对于每个FM的输出如下

2.FSM

每个block中的FSM用来输出一个FM选择向量

3.Block

因此每一层block的输出为 $\{M_n, S_n\}$ ，对于输出的详细格式如下：

每层的block的输出特征都会被下一层所接收，这样所有的因素特征都会集中在最顶层的输出中。

4.Factor Signature

在经过blocks的特征提取后，通过Facotr Signature层将每层的FSM进行融合，融合方式如下：

$$\hat{S} = [S_1, \dots, S_N], \quad (5)$$

where $\hat{S} \in \mathbb{R}^{1 \times K}$, $K = \sum_{n=1}^N K_n$ represents the feature dimension of \hat{S} . The value of K depends on the architecture of MLFN, i.e., both the total number of blocks N and the number of FMs K_n in each block. However, it is inde-

这里的输出向量大小与输出的特征图大小无关，只与block的个数与block中FM的个数相关

5.Fusion

该结构的特点之一就是不需要修改维度就可以进行融合，因为所有因素的特征都在最顶层block的输出 Y_N 中，所以我们只需要将 Y_N 与FS的输出结可，这里采用了一种类似全连接的方式，使用函数 T 将两个维度不同的张量映射到维度为 $1 \times d$ 的张量上

$$\begin{aligned} R &= \frac{1}{2}(\phi_Y + \phi_{\hat{S}}), \\ \begin{cases} \phi_Y = T(Y_N, d) \\ \phi_{\hat{S}} = T(\hat{S}, d) \end{cases} \end{aligned} \quad (6)$$

Optimisation

对于模型的优化，采用BP为基础，但由于loss通过BP传播到底层时，梯度值的影响降低，所以采用了short connection的方式减缓BP从顶到底的衰弱。由于FS与FSM的输出相连接，所以每层的FM与FSM相连接来使BP能够通过FSM传到到FM中。BP求导公式如下：

其中对于无效得因素，其FM得参数会通过 $S_{n,j} \rightarrow 0$ 来判断是否冻结参数。

对于Re-ID问题，每一个人就是一个分类，所以对于我们得到得最终输出结果 R 还需要映射到 $1 * n_class$ 的空间内，然后通过L2距离来进行分类

train

在训练的时候，包含了16个blocks，每层的FM个数为32个，最后输出的 R 的维度 d 采用1024 对于mini-batch=64时，采用Adam优化器，mini-batch=256时，采用SGD优化器

improvement

本文中在进行对比的时候提到了DPFL模型，其效果与本文结果相近，其中采用了multi-branch来获取不同感受野下的模型，作者认为如果将该部分融入模型模型会有更好的效果。

对于MLFN的输出结果 R ，可以通过无监督聚类的方式，增强输出向量的关系。

对于本文的损失函数，仅有输出 R 的L2距离，而对于每个因素的特征训练，可以通过控制其他因素相同，单独训练某一个因素，进而提高每个FM对于因素的分辨能力。