

引用格式:张伟信,刘斌. 基于残差网络的特征加权行人重识别研究[J]. 微电子学与计算机, 2020, 37(4): 26-31. [ZHANG W X, LIU B. Person re-identification based on residual network with weighted features[J]. Microelectronics & Computer, 2020, 37(4): 26-31.]

# 基于残差网络的特征加权行人重识别研究

张伟信, 刘 斌

(南京工业大学 计算机科学与技术学院, 江苏 南京 210000)

**摘 要:** 文章讨论了基于视频到视频的行人重识别的深度学习方法, 提出的网络模型主要由特征表示子网络和相似性度量子网络两部分组成. 首先利用残差网络提取视频的每帧图像的特征, 再将该特征输入到长短期记忆网络中获取时空特征, 在长短期记忆网络层后添加权重模块, 在该模块中使用帧质量注意力机制为视频的每一帧分配适当的权重. 进一步将加权后的特征向量传入相似性度量子网络进行距离度量学习, 在该框架中, 将特征表示与相似性度量使用全连接层进行连接, 同时学习和优化特征表示和相似性度量学习. 最后在两个公共数据集上进行实验, 通过一系列对比实验验证了该网络模型能提高行人重识别准确率和性能.

**关键词:** 残差网络; 行人重识别; 特征加权; 注意力机制; 相似性度量

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1000—7180(2020)04—0026—06

## Person Re-identification based on residual network with weighted features

ZHANG Wei-xin, LIU Bin

(Nanjing Tech University, School of Computer Science and Technology, Nanjing 210000, China)

**Abstract:** In this paper, we address the video-to-video person re-id. The proposed network is mainly composed of the feature representation subnetwork and the similarity measure network. First, the residual network is used to extract features from each frame of video, and then the features are inputted into the Long Short-Term Memory network to obtain features which contain information of time and space. A weight module is applied on the upper layer of Long Short-Term Memory network. In this model, applying an attentive quality mechanism to assign appropriate weight for each frame. Then, the weighted feature of each video sequence will be inputted into the similarity measure sub-network to measure similarity. In this framework, use fully connected layers to connect feature representation subnetwork and the similarity measure network, so the feature representation and similarity metric learning can be learned and optimized at the same time. Finally, we do experiments on two public datasets to prove that our network model can improve pedestrian recognition accuracy and performance.

**Key words:** residual network; person re-identification; weighted features; attentive mechanism; similarity measure

### 1 引言

行人重识别是用来判断一个摄像头下出现的行人是否在另一个摄像头下再次出现的方法. 现阶段,

行人重识别技术面对的难点有<sup>[1,2]</sup>: 首先光照变化, 视角位置, 背景复杂, 行人被遮挡导致获得的行人图像分辨率较差, 获取不到清晰行人信息. 其次, 不同摄像机的拍摄角度不同, 行人在不同的摄像机下姿

态会发生变化。

目前,大多数的研究是基于图像-图像的行人重识别问题<sup>[3-5]</sup>。现在基于视频<sup>[6]</sup>的行人重识别也得到越来越广泛地研究。与图像相比视频序列中包含了大量的人的外貌样本信息,这些样本信息包含了同一行人在不同的姿势、视角和背景下的外貌信息。此外,还可以利用视频序列的时间信息,建立更好的行人外貌分类模型。本文主要研究基于视频-视频(Video-to-Video)的行人重识别问题。基于深度学习的行人重识别方法的主要思想是首先使用卷积神经网络(CNN)或循环神经网络(RNN)<sup>[4]</sup>从视频帧中提取特征,然后对获得的深度学习特征进行相似性度量<sup>[7]</sup>。网络越深,提取的特征更具鲁棒性,但是随着网络层数不断加深,会出现梯度消失和退化两个问题,导致不能获得鲁棒性的特征。

在视频序列中存在被遮挡的不完整的行人图片,这些图片会降低重识别模型的准确度和性能。本文中提出了如图1所示的方法,该方法包括了特征表示子网络和相似性度量子网络。在特征表示子网络中采用深度残差网络来提取视频的特征,克服了随着网络深度增加和训练样本有限导致的分类精度下降的问题。进一步采用长期短期记忆(LSTM)网络来嵌入视频的上下文时空特征,增强了模型区分具有相似外观但是并非同一人的能力。在视频中存在大量的冗余信息和噪声,因此重识别模型需要能够关注那些具有判别性,高质量的特征。因此为了降低视频中被严重遮挡的行人图片对行人重识别模型性能的影响,在LSTM层后添加一个特征加权模块,该模块中使用了一种注意力机制,自适应为视频的帧分配合适的权重,最后,使用一种广义相似性度量学习方法来度量两段视频序列之间的相似度。

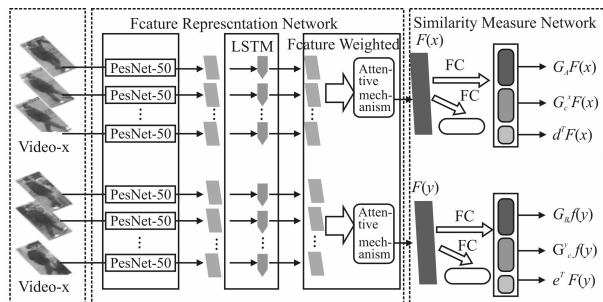


图1 全局网络结构

## 2 特征表示子网络

### 2.1 深度残差网络

随着深度学习的发展,神经网络模型的深度不断增加。增加网络深度会带来梯度消失或退化导致识别准确率下降等问题,本文使用残差神经网络模块(ResNet)<sup>[8]</sup>来解决网络深度的问题。将添加的网络层使用残差的方式进行映射,使得深层网络层的损失降低。

假设输入是  $x$ , 输出的实际映射是  $y$ 。通过堆叠的多层非线性网络层去拟合映射关系  $F(x) := y - identity(x)$ , 从而将原始映射替换为  $F(x) + identity(x)$ 。堆叠残差块构建一个具有非常深的结构的 ResNet。而且拟合残差往往比拟合原始映射更便捷。整个 ResNet 只需要学习输入、输出的差别部分,简化了学习难度。残差块的定义为:

$$y = F(x, W_i) + identity(x) \quad (1)$$

式中,  $x$  和  $y$  分别为网络层之间的输入和输出向量。函数  $F(x, \{W_i\})$  表示待学习的残差映射,  $W_i$  表示第  $i$  层网络的权重参数。

深度残差网络可以由多个依次连接的残差块构成。本文使用去除了最后一层全连接层的 ResNet-50 来提取视频的特征。本文使用的 ResNet 块如图2所示。残差块由批量归一化(BN)层<sup>[9]</sup>、激活层<sup>[10]</sup>、卷积层(Conv),随机失活(Dropout)<sup>[11]</sup>层组成。卷积核的大小为  $3 \times 3$ , 此时网络的效率更高,精度更高。在每个卷积层之后,使用 BN 层对卷积层的输出数据进行归一化处理,然后再传递到激活层,非线性激活函数所采用的是修正线性单元(ReLU)激活函数。在两个卷积层之间,激活层之后添加了一个 Dropout 层。增加 BN 层和 Dropout 层的目的是抑制和均衡视频序列中行人某些突出的特征表示,减少过拟合,从而获得稳定的外观特征信息。设  $y = \{y^{(t)} \mid t \in [1, \dots, N]\}$  为视频序列,其中  $N$  为视频序列的长度;  $y^{(t)}$  表示  $t$  时刻的行人的帧图片。视频的的每一帧经过 ResNet 产生特征向量:  $f_y^{(t)}, t \in [1, \dots, N]$ 。然后每一帧的特征向量  $f_y^{(t)}$  被传递到 LSTM 层,提取时序特征。

### 2.2 特征加权模块

假设  $t$  时刻视频帧的特征向量为  $f_y^{(t)}$ , 将该特征向量输入到长短期记忆网络(LSTM)中 LSTM 中并由此获得了该在  $t$  时刻视频帧的状态信息。

LSMT 的隐藏单元  $h_y^{(t+1)}$  能够使用视频序列的时序信息,但是 LSTM 的输出可能倾向于较晚的时

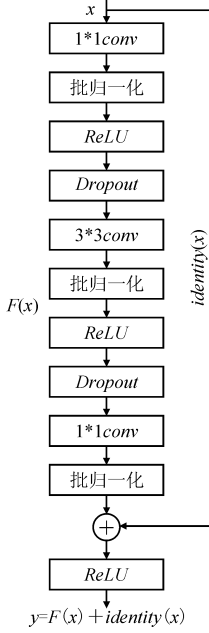


图2 本文中使用的残差块

间步长,这会降低 LSTM 处理整个视频序列信息时的性能.在文献[12]中,作者使用时间维度上的平均池策略来生成平均特征向量.这种平均策略可以看作是对视频中的每一帧的特征都赋予相同的权重.平均权重表示方法如下

$$\bar{h} = \frac{1}{N} \sum_{t=1}^N h_y^t \quad (2)$$

式中, $N$  表示视频序列的长度.

这种平均策略可以看作是为 LSTM 在  $t$  时刻的输出向量  $h_y^t$  赋予相等的权值.但是在实际情况下,严重遮挡的视频帧会对行人重识别产生很大的负面影响.与此同时,某些视频帧的特征又可以提供对身份识别提供鉴别性的信息.式(2)的平均池策略并不能充分利用这些特征的高判别性.所以在特征加权模块,提出一种质量注意力机制<sup>[13]</sup>,在获取时域特征时,将特征选择的注重点放到具有高判别性,高质量的的特征上.特征的重要性完全取决于特征自身质量.从特征的自身出发,学习出此特征在聚合成 LSTM 输出特征时所占的权重.

输入视频序列  $x$  和视频序列  $y$ , 获取的 LSTM 特征分别为:  $h_x^{(N \times d)} = [h_y^1, h_y^2, h_y^3, \dots, h_y^N]$  和

$$h_y^{(N \times d)} = [h_y^1, h_y^2, h_y^3, \dots, h_y^N]. \text{ 维度为 } N \times d.$$

视频帧质量注意力机制矩阵  $A$  计算如下:

$$A = \tanh((h_x^{(N \times d)})P(h^{(N \times d)})^T), \quad A \in \mathbb{R}^{N \times N} \quad (3)$$

$P \in \mathbb{R}^{d \times d}$  是通过特征  $h_x^{(N \times d)}$  和  $h_y^{(N \times d)}$  获得的信息共享矩阵<sup>[14]</sup>,使用注意力机制的主要目的是找出

视频中高质量的被外在因素影响较小的帧,这些帧在相似性度量时将起到关键性的作用.对矩阵  $A$  分别进行列最大池化操作和行最大池化操作得到输入序列  $x$  和  $y$  的权重向量  $\omega_x \in \mathbb{R}^N, \omega_y \in \mathbb{R}^N$ .

对权重向量  $\omega_x \in \mathbb{R}^N$  和  $\omega_y \in \mathbb{R}^N$  应用 softmax 函数进行归一化操作得到第  $i$  帧图片注意力权重:

$$[\alpha_x]_i = \frac{e^{[\omega_x]_i}}{\sum_{i=1}^N e^{[\omega_x]_i}} \quad (4)$$

$$[\alpha_y]_i = \frac{e^{[\omega_y]_i}}{\sum_{i=1}^N e^{[\omega_y]_i}} \quad (5)$$

注意力权重代表着每个区域向量在最后的视频特征聚合中的重要程度.将注意力权重向量与 LSTM 特征  $h(N \times d)_x$  和  $h(N \times d)_y$  进行点乘操作,获得加权后的视频特征:

$$F(x) = [h_x^{(N \times d)}]^T \odot \alpha_x \quad (6)$$

$$F(y) = [h_y^{(N \times d)}]^T \odot \alpha_y \quad (7)$$

特征  $F(x)$  和特征  $F(y)$  的维度为  $d$ .对于获得加权后的视频序列后,分别经过两个全连接层,第一个全连接层用来计算加权后的视频序列的分类损失函数.输入视频序列  $x$  使用 softmax 函数预测输入的样本属于各类的概率,即:

$$\hat{p} = \text{softmax}(W_{f1}^d F(x) + b_{f1}) \quad (8)$$

然后使用交叉熵损失函数计算损失:

$$\text{Loss}_{I(F(x))} = - \sum_{i=1}^K p_i \log(\hat{p}_i) \quad (9)$$

式中,  $\hat{p}_i$  表示预测出的视频身份为  $i$  的概率;  $id$  表示真实的身份类别;  $p_i$  表示视频所属人的身份为  $i$  的概率;  $K$  表示样本个数.

第二个全连接层连接相似性度量模块.

使用相同的方法,获得  $F(y)$  的分类损失函数为:

$$\text{Loss}_{I(F(y))} = - \sum_{i=1}^K p_i \log(\hat{p}_i) \quad (10)$$

### 3 相似性度量量子网络

本文中使用了一种融合了马氏距离和余弦相似度的广义相似性度量方法<sup>[15]</sup>来计算两个视频序列之间的相似度,该方法将特征表示和相似性度量学习统一在一个框架中,共同对这两个部分进行优化.假设  $f(x) \in \mathbb{R}^d, f(y) \in \mathbb{R}^d$  分别是探针视频和图库视频的特征,文献[15]中提出的广义相似性度量方法定义如下所示:

$$S(f(x), f(y)) = \|G_A f(x)\|^2 + \|G_B f(y)\|^2 + 2d^T f(x) - 2(G_C^x f(x))^T (G_C^y f(y)) + 2e^T f(y) + \rho \quad (11)$$

式中,  $G_A f(x)$ ,  $G_C^x f(x)$ ,  $d^T f(x)$  表示  $x$  的相似性分量;  $G_B f(y)$ ,  $G_C^y f(y)$ ,  $e^T f(y)$  表示  $y$  的相似性分量。

相似性度量量子网络包含两个由全连接层连接分支,每个分支都通过全连接层连接.将  $F(x)$ ,  $F(y)$  输入到全连接层,输出各自相似度分量:  $G_A F(x)$ ,  $G_C^x F(x)$ ,  $d^T F(x)$ ,  $G_B F(y)$ ,  $G_C^y F(y)$ ,  $e^T F(y)$ .

假设  $\{(x_i, y_i), k_i\}_{i=1}^K$  是样本的训练集对,  $k_i$  表示  $\{x_i, y_i\}$  对应的标签,如果是同一人则  $k_i = -1$ ,否则为 1. 式(11)表示的相似度函数是一个表示差异度的函数。

使用 Hinge 损失函数作为相似性度量的优化函数:

$$Loss_{SV(F(x_i), F(y_i))} = \sum_{i=1}^K (1 - k_i S(F(x_i), F(y_i)))_+ \quad (12)$$

引入正则化项  $\Psi(W, \Phi) = \lambda \|W\|^2 + \mu \|\Phi\|^2$ , 此时相似性度量量子网络中全连接层的权重作为模型损失函数的一部分. 这样就使得相似性度量量子网络就与特征表示子网络相互紧密结合,在模型训练过程中共同优化。

将训练集合  $\{(x_i, y_i), k_i\}_{i=1}^K$  输入到网络中,得到最终目标优化函数为:

$$LOSS = Loss_{SV(F(x_i), F(y_i))} + \Psi(W, \Phi) + Loss_{I(F(x))} + Loss_{I(F(y))} \quad (12)$$

使用标准反向传播算法对提出的网络进行训练. 其中,  $W$  表示第二个全连接层的参数矩阵.  $\Phi = (G_A, G_B, G_C^x, G_C^y, d, e, f)$  表示特征  $F(x)$  和  $F(y)$  的相似度参数矩阵。

## 4 方法验证与实验对比

### 4.1 行人重识别数据集和训练策略

本文使用的数据集如下:

iLIDS-VID 数据集<sup>[16]</sup>: iLIDS-VID 数据集包含了随机 300 个人的 600 个视频序列,由两个非重叠相机在不同的场景下获得. 每个人的视频序列长度从 23 帧到 192 帧不等,平均为 73 帧. 由于获取行人视频时存在交叉相机视角变化,行人外观相似,以及混乱的背景造成的光照的变化等问题,所以该数据集是非常具有挑战性的。

PRID-2011 数据集<sup>[17]</sup>: PRID-2011 数据集包

含 200 人的 400 个图像序列,这些图像序列是从相邻的两个相机中获取的. 每个人的视频序列的长度在 5 到 675 帧之间,平均为 100 帧. PRID-2011 拍摄的主要是户外场景,背景纯净很少有遮挡. 但是,有些视图中会有明显的颜色变化和阴影存在。

在数据集 iLIDS-VID 和 PRID-2011 上进行实验,数据集被随机分为两部分:训练集和测试集. 其中 50% 作为训练集,50% 作为测试集. 使用不同的训练集和测试集重复进行 10 次实验,并对最后的结果取平均值,保证结果的稳定性. 由于本文研究的内容是针对 Video-to-Video 的行人重识别问题. 训练阶段,第一个摄像头下的视频最为查询集(probe),第二个人摄像头下的视频作为候选集(gallery),正样本对包含的视频对是属于同一人的,负样本对包含的视频对对应两个不同身份的行人. 训练中,正负样本对交替输入. 同时,通过对同一视频裁切同一区域来做数据增强处理. 如果视频序列的长度  $N$  小于 10,则弃用该视频序列. 实验过程中使用小批量随机梯度下降法<sup>[18]</sup> (mini-batch Stochastic Gradient Descent, SGD) 来训练本文提出的网络模型. 在训练阶段的每次迭代中,这个方法可以得到更快的反向传播速度也能获得更加平滑的收敛效果. 训练网络时参数设定如下:初始学习率设为  $\beta = 1e-4$ ,网络训练的样本数(batch-size)大小为 100 (50 对),最大迭代为 100 000 次,每经过 2 000 次迭代,学习率就以  $\alpha_{new} = 0.02 \times \beta$  的方式降低. 最终,得到可以基于深度残差网络的行人重识别模型. 为了真实地反映匹配度的统计数据,每个实验重复 10 次,并计算平均结果值. 在实验中,行人重识别任务就是要将查询集(probe)中每一个视频序列,在候选集(gallery)中找到最相似的视频. 采取排序- $k$  识别率(rank- $k$ ,  $k$  的值一般为 1、5、10、20)来进行对比. rank- $k$  表示在候选集中相似度最高的前  $k$  个视频中正确找到查询行人的概率。

### 4.2 实验结果和分析

实验结果使用 CMC 进行评估,且都使用相同的数据集和相同的训练和测试策略. 将本文提出的方法与以下几种行人重识别方法进行对比实验: LOMO+XQDA<sup>[19]</sup>, CNN+RNN<sup>[12]</sup>, DNN+Reranking<sup>[14]</sup>, JAFN<sup>[13]</sup>. 实验结果验证了本文提出的重识别模型较现在的方法在准确度和性能上是有一定的提高的。

表 1 iLIDS-VID 数据集上实验结果对比

| iLIDS-VID                     |       |       |       |       |
|-------------------------------|-------|-------|-------|-------|
| rank-k                        | 1     | 5     | 10    | 20    |
| LOMO+XQDA <sup>[19]</sup>     | 38.5% | 60.2% | 73.7% | 85.8% |
| CNN+RNN <sup>[12]</sup>       | 58.4% | 70.1% | 75.7% | 89%   |
| DNN+Reranking <sup>[14]</sup> | 60.2% | 77.2% | 85. % | 92.7% |
| JAFN <sup>[13]</sup>          | 63.7% | 80.6% | 88.8% | 92.9% |
| DRN+LSTM                      | 66.5% | 83.4% | 90.7% | 93.8% |

表 2 PRID-2011 数据集上实验结果对比

| PRID-2011                     |       |       |       |       |
|-------------------------------|-------|-------|-------|-------|
| rank-k                        | 1     | 5     | 10    | 20    |
| LOMO+XQDA <sup>[19]</sup>     | 67.8% | 74.7% | 83.7% | 90.3% |
| CNN+RNN <sup>[12]</sup>       | 74.6% | 80.3% | 90.1% | 93.1% |
| DNN+Reranking <sup>[14]</sup> | 76.8% | 90.1% | 92.2% | 94.5% |
| JAFN <sup>[13]</sup>          | 80.1% | 90.6% | 92.8% | 95.5% |
| DRN+LSTM                      | 82.8% | 92.3% | 93.1% | 96.1% |

从上述表格可以看出相对 LOMO+XQDA 方法,方法 CNN+RNN,DNN+Reranking, JAFN, DRN+LSTM,识别的准确率都有一定的提高,因为本文讨论的行人重识别问题是基于视频到视频的,CNN+RNN,DNN+Reranking, JAFN, DRN+LSTM 四种方法都使用了循环神经网络,获取视频序列的时序特征,构建了更好的行人外貌分类模型.

对比方法 DNN+Reranking,JAFN 与本文提出的方法 DRN+LSTM 的结果可以看出,本文提出的基于残差网络的模型的准确率更高,在特征表示子网络,本文使用残差网络的残差网络,解决了随着网络深度的不断增加,神经网络梯度消失,退化的问题,构建了一个更深的神经网络,可以获取更具鲁棒性的行人特征.

在数据集 iLIDS-VID 上进行试验获得的识别准确率比在数据集 PRID-2011 上进行实验准确率提高的更明显,是因为本文使用的特征加权模块更适用于应对遮挡的情况,注意力机制注重视频中的高质量特征和干扰特征,并对其进行分辨和分配相应权重.降低了低质量的视频帧对识别准备率的影响,提高了高质量帧在识别过程中的作用.在 iLIDS-VID 数据集中存在多数遮挡情况,行人的部分帧质量较低,而 PRID-2011 数据集中几乎没有遮挡情况,由此可以看出本文所提出的方法针对于遮挡情况有较好的改进效果.

在 iLIDS-VID 数据集上的对比实验的 CMC 曲

线绘制结果如图 4 所示,在 PRID-2011 数据集上的对比实验 CMC 曲线绘制结果如图 5 所示.

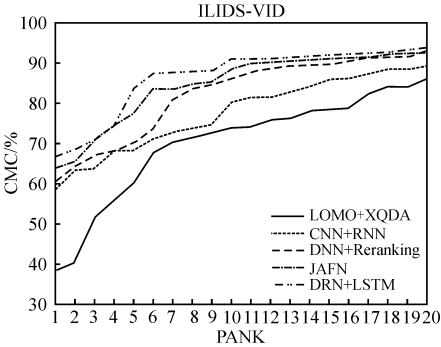


图 4 iLIDS-VID CMC 曲线图

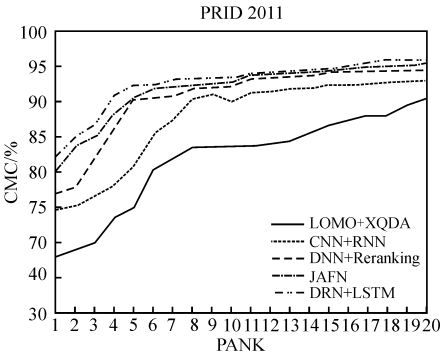


图 5 PRID-2011 CMC 曲线图

5 结束语

在行人重识别问题的研究中,主要目的是要学习到具有判别性和鲁棒性的行人特征,本文为了突出行人特征的每一维度对行人身份匹配结果的影响,使用了深度残差网络解决网络深度增加导致的梯度消失和网络性能退化问题对行人重识别性能的影响,并使用了注意力机制让模型更加关注高质量的视频帧,并且使用了一种广义的相似性度量方法实现特征表示与相似性度量共同优化.在数据集 iLIDS-VID 和 PRID2011 再识别领域数据集的实验结果也证明了本文模型算法相比其他算法更具优势.目前关于行人重识别的研究主要基于采集好的数据集,不考虑行人检测,未来将进一步研究行人检测技术与行人再识别的结合使用.

参考文献:

[1] ZHENG L, SHEN L, TIAN L, et al. Scalable Person Re-identification: A Benchmark[C]// 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015.

- [2] ZHANG W, MA B, LIU K, et al. Video-Based Pedestrian Re-Identification by Adaptive Spatio-Temporal Appearance Model[J]. IEEE Transactions on Image Processing, 2017, 26(4):2042-2054.
- [3] ZUO W, WANG F, ZHANG D, et al. Distance Metric Learning via Iterated Support Vector Machines[J]. IEEE Transactions on Image Processing, 2017:1-1.
- [4] ZHANG D, LIN L, CHEN T, et al. Content-Adaptive Sketch Portrait Generation by Compositional Representation Learning[J]. IEEE Transactions on Image Processing, 2016:1-1.
- [5] ZUO W, WU X, LIN L, et al. Learning Support Correlation Filters for Visual Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018:1-1.
- [6] ZHANG D, WU W, CHENG H, et al. Image-to-Video Person Re-Identification with Temporally Memorized Similarity Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017:1-1.
- [7] DING S, LIN L, WANG G, et al. Deep feature learning with relative distance comparison for person re-identification[J]. Pattern Recognition, 2015, 48(10): 2993-3003.
- [8] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.
- [9] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2015.
- [10] HINTON G E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010.
- [11] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [12] MCLAUGHLIN N, RINCON J M D, MILLER P. Recurrent Convolutional Network for Video-Based Person Re-identification[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
- [13] CHEN L, YANG H, GAO Z. Joint Attentive Spatial-Temporal Feature Aggregation for Video-Based Person Re-Identification[J]. IEEE Access, 2019:1-1.
- [14] SAHA B, RAM K S, MUKHOPADHYAY J, et al. Video Based Person Re-Identification by Re-Ranking Attentive Temporal Information in Deep Recurrent Convolutional Networks[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1663-1667.
- [15] LIN L, WANG G, ZUO W, et al. Cross-Domain Visual Matching via Generalized Similarity Measure and Feature Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016:1-1.
- [16] WANG T, GONG S, ZHU X, et al. Person Re-Identification by Discriminative Selection in Video Ranking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016:1-1.
- [17] WANG T, GONG S, ZHU X, et al. Person Re-Identification by Video Ranking[C]// European Conference on Computer Vision. Springer, Cham, 2014.
- [18] NICOL N. Schraudolph. Centering Neural Network Gradient Factors[M]. 1997.
- [19] LIAO S, HU Y, ZHU X, et al. Person re-identification by Local Maximal Occurrence representation and metric learning [C]. Computer Vision and Pattern Recognition. IEEE, 2015: 2197-2206.

### 作者简介:

**张伟信** 男, (1994-), 硕士研究生. 研究方向为深度学习、行人重识别.

**刘 斌(通讯作者)** 男, (1966-), 副教授. 研究方向为模糊技术、人工智能、数据挖掘. E-mail: nj\_liubin@163.com