

一种用于多源人脸图像检索的损失函数改进策略

任国印, 吕晓琪, 李宇豪

引用本文:

任国印, 吕晓琪, 李宇豪. 一种用于多源人脸图像检索的损失函数改进策略[J]. 液晶与显示, 2021, 36(11): 1583–1596.

REN Guo-yin, LYU Xiao-qi, LI Yu-hao. Improved loss function strategy for multi source face image retrieval[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1583–1596.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0003>

您可能感兴趣的其他文章

Articles you may be interested in

融合多尺度特征及注意力机制的医学图像检索

Medical image retrieval with multiscale features and attention mechanisms

液晶与显示. 2021, 36(8): 1174–1185 <https://doi.org/10.37188/CJLCD.2020-0248>

改进非负矩阵分解的神经网络人脸识别

Face recognition based on improved NMF and neural network

液晶与显示. 2017, 32(3): 213–218 <https://doi.org/10.3788/YJYXS20173203.0213>

基于残差整流增强卷积神经网络的表情识别

Expression recognition based on residual rectifier enhanced convolution neural network

液晶与显示. 2020, 35(12): 1299–1308 <https://doi.org/10.37188/YJYXS20203512.1299>

基于SURF特征的栈式自编码网络人脸对齐算法

Stacked auto-encoder networks face alignment algorithm based on SURF features

液晶与显示. 2018, 33(3): 254–260 <https://doi.org/10.3788/YJYXS20183303.0254>

一种基于LVQ神经网络模型的人脸检测方法

Face detection algorithm based on LVQ neural network model

液晶与显示. 2021, 36(7): 1027–1034 <https://doi.org/10.37188/CJLCD.2020-0295>

文章编号:1007-2780(2021)11-1583-14

一种用于多源人脸图像检索的损失函数改进策略

任国印^{1,2}, 吕晓琪^{1,2,3*}, 李宇豪²

(1. 内蒙古科技大学 机械工程学院, 内蒙古 包头 014010;

2. 内蒙古科技大学 信息工程学院, 内蒙古 包头 014010;

3. 内蒙古工业大学, 内蒙古 呼和浩特 010051)

摘要:目前,公安机关刑侦部门大量的离线视频存储在监控网络的服务器中,为了在这些海量视频帧中提取嫌疑人目标人脸,设计了人脸检索系统。通过改变 CNN 网络的 RELU 结构,训练新的四分支孪生网络来获取深度特征,构造了一个新的四分支孪生网络。结合网上发布的逃犯人脸图片发起通缉,借助深度特征对比展开基于内容的图像检索(CBIR)。新的四分支孪生网络比熟悉的网络,如 Alexnet、GoogLeNet、VGGNet 和 ResNet 等收敛得更快,系统鲁棒性好。网络的共享权重设计使得检索具有较高的模型训练精度和检索精度。图像深度特征可以在摄像机之间快速在线共享。实验结果表明,该方法的平均检索精度(ARP)为 98.74%,模型训练精度为 99.51%,帧率为 28 FPS。

关键词:四孪生网络;人脸检索;跨摄像机;基于内容的图片检索

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0003

Improved loss function strategy for multi source face image retrieval

REN Guo-yin^{1,2}, LYU Xiao-qi^{1,2,3*}, LI Yu-hao²

(1. School of Mechanical Engineering, Inner Mongolia University of Science & Technology, Baotou 014010, China;

2. School of Information Engineering, Inner Mongolia University of Science & Technology, Baotou 014010, China;

3. Inner Mongolia University of Technology, Hohhot 010051, China)

Abstract: At present, a large number of offline videos are stored in the servers of the surveillance network in criminal investigation departments of public security organs, and the face retrieval system is designed. The new Quadruplet Network converges faster than the familiar networks such as Alexnet,

收稿日期:2021-01-05; **修订日期:**2021-04-09.

基金项目:国家自然科学基金(No.61179019, No.81571753);赛尔网络下一代互联网技术创新项目(No. NGII20170705);包头市青年创新人才项目(No.0701011904)

Supported by National Natural Science Foundation of China(No.61179019, No.81571753);CERNET Innovation Project (No. NGII20170705); Baotou Youth Innovative Talent Project(No.0701011904)

* 通信联系人, E-mail:635302395@qq.com

GoogLeNet, VGGNet and ResNet. Because of the shared weight design of the network, the retrieval has a high precision, Average Retrieval Precision(ARP) and model training accuracy, and the system has good robustness. The image depth features can be shared quickly online between the cameras. The proposed method is effective, with ARP of 98.74% and a model training accuracy of 99.51%, and a frame rate of 28 FPS.

Key words: quadruplet network; face retrieval; cross camera; content-based image retrieval

1 引言

目前,从监控摄像机上获取目标嫌疑人人脸图像是公安民警的重要技术手段之一。几乎所有街道都安装了高清摄像头,并且摄像机的覆盖范围很广。高清监控摄像机每天不停地采集视频数据,每天都会产生大量的离线视频。以高清摄像机为例,每秒 50~60 帧,24 h 能够产生 500 万帧左右的视频录制图片,这样一台摄像机录一个月的视频数据非常庞大。如果将一个区域的所有摄像机数据加起来,就会更加庞大。对于刑侦部门来说,一帧一帧地对可疑视频进行人工过滤,显然是低效和痛苦的。

人脸识别一直是图像识别中的一个重要问题。首先,人脸识别的任务比图像分类要复杂得多。在人脸识别技术中,最重要的工作是人脸图片的特征提取^[1-3]。直到 2013 年,人脸识别还处于手工特征提取阶段,主要依靠图像处理专家对人脸进行仔细分析,最终得到人脸的特征,也就是构建人脸局部描述符。目前,基于深度学习的人脸识别已逐渐取代传统的人工人脸特征识别方法,传统借助人眼来识别人脸的方法已经不能满足刑侦部门的需要。

研究和发 展 图 像 检 索 技 术 已 经 进 行 了 近 20 年^[4-6]。但是目前基于内容的人脸图像检索系统还很少,主要的瓶颈是检索精度低和存在语义鸿沟。人脸检索是人脸识别和基于内容的图像检索的结合,它通过匹配人脸特征来搜索现有的人脸照片数据库^[7-10]。基于深度学习的图像识别和检索技术逐渐取代现有方法。为了找到一组与被侦查人面部最为相似的照片,进而缩小刑事侦查范围,协助刑事侦查人员确定犯罪嫌疑人的身份。目前基于深度学习的跨多摄像机检索技术的研究还很少,所以基于深度学习的跨多摄像机的人脸检索技术也越来越迫切^[11-14]。

M. Zhang、Y. Zhao 和 M. Sardoğan 等人提出了基于 CNN 的图像分类方法,它和传统图像分类方法相比明显提高了分类精度^[15-17]。特别地,W.Ge 等人使用孪生网络输入同一人脸图像对(正样本对)或不同人脸对(负样本对),每对训练图像都有一个标签。以上两个样本对完成训练,通过最小化相似度损失函数,使同一类人脸之间的距离最小,不同类人脸之间的距离最大。尽管这些数据集预处理方法比以往常用的 PaSC、LFW、PubFig、FERET、AR 和 YaleB 等基准数据集有了显著的改进,但更多的研究者试图通过改进损失函数和孪生网络的网络结构来探索改进的可能性。例如 Triplet loss 是一种广泛使用的学习损失度量函数,它通过添加多个网络分支来提高孪生网络的性能^[18-19]。三支孪生网络 Triplet 训练使用 Triplet loss,这需要 3 个图像作为输入。与对比度损失不同的是,Triplet loss 输入三元组由一对正采样对和一对负采样对组成。这 3 幅图像分别被命名为锚样本(anchor)、正样本(positive)和负样本(negative)。人脸 a 和人脸 p 是一对正样本对,人脸 a 和人脸 n 是一对负样本对。三元组可以缩短正样本对之间的距离,扩大负样本对之间的距离。最后,在特征空间中对相同的人脸图像进行聚类。通过类比,Triplet loss 只考虑正负样本对之间的相对距离,而不考虑正样本对之间的绝对距离。W. Chen 等人使用四分支孪生网络 Quadruplet Network 和一个新的损失函数来训练局部图像描述符学习模型,该模型可应用于孪生网络 Siamese 和三分支孪生网络 Triplet^[20]。四分支孪生网络 Quadruplet 损失函数产生一个特征向量,该特征向量使属于同一类的正图像之间的距离方差最小化,使属于同一类的样本图像之间的平均距离最小化,并使不同类的图像之间的平均距离最大化。四分支孪生网络损失函数增加负样本的绝对距离,以确保四分支孪生网络不仅在特征空间中扩大正样本和负样本,而且使正样本或负样本彼此

靠近。

Trihard loss 是 Triplet loss 的一个改进版本。传统的 Triplet loss 随机抽取训练数据中的 3 幅图像作为训练输入。该方法简单,但大多数样本是易于识别的简单样本对。如果大量的训练样本对都是简单的样本对,则不利于网络学习更好的表征。大量的研究发现,使用更难的样本对网络进行训练可以提高网络的泛化能力,对更难的样本对进行训练的方法有很多种。用一种基于训练批 Batch 的在线困难样本抽样方法,称为难样本损失法,即 Trihard loss^[20]。

Trihard loss 通常比传统的 Triplet loss 更有效。边界样本挖掘损失 (Boundary sample mining loss, MSML) 是一种度量学习方法,它引入了难样本采样的思想^[21]。Triplet loss 只考虑正负样本对之间的相对距离。总之, Trihard loss 为 Batch 中的每个图像选择一个正负样本,而 MSML 损失函数仅选择最困难的正样本对和最困难的负样本对来计算损失。因此, MSML 比 Trihard 更难采样,可以看作是正样本对之间距离的上界和负样本对之间距离的下界。MSML 是将正负样本对的边界分开,因此称为边界样本挖掘损失。一般来说, MSML 是一种同时考虑相对距离和绝对距离的度量学习方法, MSML 引入了困难样本抽样的思想。

2 本文所用方法

2.1 MSML

MSML 是一种度量学习方法,它引入了困难样本抽样的思想。

Triplet loss 只考虑正负样本对之间的相对距离。为了引入正负样本对之间的绝对距离, MSML 加上了负样本对比较,形成三元数模型。MSML 损失函数可定义为公式(1):

$$L_q = (d_{a,p} - d_{a,n_1} + \alpha)_+ + (d_{a,p} - d_{n_1,n_2} + \beta)_+, \quad (1)$$

如果忽略参数的影响,可以用更通用的形式公式(2)来表示 MSML 损失:

$$L_q = (d_{a,p} - d_{m,n} + \alpha)_+, \quad (2)$$

其中 m 和 n 是一对负样本对, m 和 a 是一对正负样本对。然后介绍了三孪生损失的难样本挖掘思想,公式(3)中引入了 MSML 损失函数:

$$L_{\text{msml}} = (\max_{a,p} d_{a,p} - \min_{m,n} d_{m,n} + \alpha)_+, \quad (3)$$

其中: a, p, m, n 是输入的所有图片, a, p 是该 Batch 中最不相似的正样本对, m, n 是该 Batch 中与 a 最相似的负样本对。总之, Trihard loss^[22] 为网络选择一个三元难样本输入,而 MSML loss^[23] 仅选择一个难四元组,即最困难的正样本对和困难的负样本对来计算损失。所以 MSML loss 比 Trihard loss 更难取样。

2.2 改进型四孪生神经网络

在介绍网络模型设计之前,说明模型的输入是一项必要的工作。首先模型有 4 个样本作为输入,分别是 anchor、positive、negative1、negative2,具体说明如下:

anchor: anchor 是锚定样本也就是基准图像,一个随机样本是正样本还是负样本取决于随机样本的分类是否和该锚定样本同属一类,如果属于一类就是正样本,如果不同类就是负样本。

positive: 该样本和锚定样本同属一类的样本,因此称之为正样本。

negative: negative1 和 negative2 样本和锚定样本不属于一类样本,由于本模型需要输入两个负样本,因此称之为 negative1 和 negative2。

而样本对泛指两个样本,正负样本是指样本对中一个为正样本和一个负样本,正样本对是指样本对由两个正样本构成,而负样本对是指样本对由两个负样本构成。

按照样本对的难度可以进一步细划分为难正样本对、难负样本对、难正负样本对。这里的难度是相对 Siamese、Triplet 和 Trihard 输入样本的复杂度而言的。Siamese 是随机抽取两个样本, Triplet 是随机抽取两个正样本和一个负样本, Trihard 是将两个最不像的正样本和一个与正样本最像的负样本构成一个三元组。本模型需要输入 4 个样本,这 4 个样本可构成 3 组样本对,即难正样本对、难负样本对、难正负样本对,具体说明如下:

难正样本对是同属锚定样本类的人脸样本中差别较大的两张人脸图片,虽然来自一个人但看上去却像不同人的脸。

难负样本对是不属锚定样本类的人脸样本中差别较大的两张人脸图片,虽然来自一个人但看上去却像不同人的脸。

难正负样本对是锚定样本类的人脸样本和非

锚定样本类的人脸样本构成的样本对,这两个样本的差别较小,虽然样本不是一个人,但这两张图片看上去却像同一个人的脸。

本文所述方法基于 MSML。为了改进

MSML 损失函数输入的复杂度,从样本输入难度上本文模型输入的难度比以上提到的所有损失函数的输入都要大,因此本文模型可以更加充分拉近同类模型间距离而拉远不同类别间的

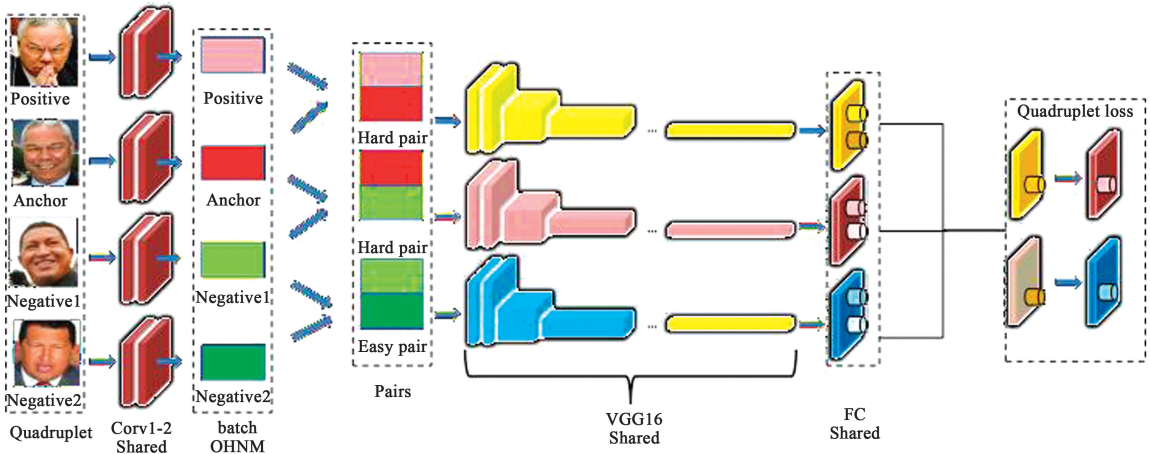


图 1 改进的 MSML 模型结构图
Fig.1 Improved MSML model structure diagram

距离。

(1) 网络结构:本文使用的网络如图 1 所示。图像输入仍然是 anchor, positive, negative1, negative2 四种样本,其中 positive 和 anchor 是正样本,negative1 和 negative2 是负样本。通过两个卷积层提取图像特征,形成 3 对难样本。采用

3 层 VGG16 网络对困难样本进行训练,完成样本分类。因此,前 3 层(1~3 层)构成两个正样本和一个负样本面对三分支孪生网络 T1,后 3 层(2~4 层)构成两个负样本和一个正样本面对三分支孪生网络 T2,即两个平行互相传递参数的三分支孪生网络,如图 2 所示。

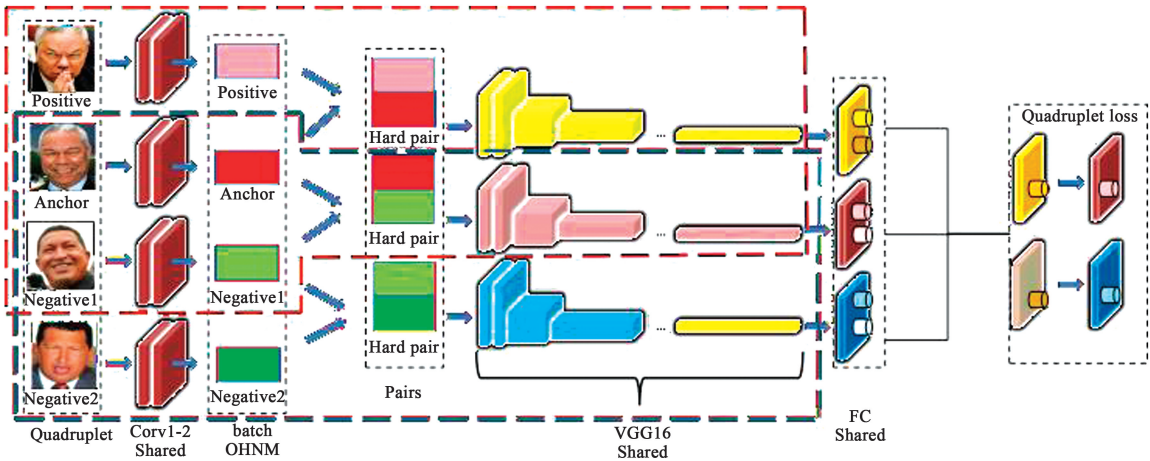


图 2 将改进的 MSML 模型看作双三重网络(DTN)
Fig.2 Improved MSML model regarded as the double triple networks (DTN)

(2) MSML 改进:从图 2 可以看出,四分支孪生网络可以看作是两个并行的三分支孪生网络,红色虚线部分是三分支孪生网络,蓝色虚线部分

也是三分支孪生网络。本文的创新之处之一是从这一角度对 MSML 的损失函数进行改进,同时增加输入四元组样本对难度,输入最不相似的正样

本对和最相似的正负样本对,同时负样本对之间挑选同一类别中距离最大的负样本对作为样本输入。

MSML 在第一层和第二层输入难正样本对(两个最不相似的正样本对),在第 3 层和第 4 层输入难负样本对(同一个人的两个最不相似的负样本对)。正负样本对使用了不同人群最相似的两张脸。该方法考虑了正样本距离的上界和负样本距离的下界,同时考虑了相对距离和绝对距离。

在双三重网络(DTN)中,每个三重实现不同的目标,而 T1 和 T2 通过 Trihard loss 计算正采样距离和负采样距离。这样可以防止同一张人脸被误认为其他人的脸。T1 和 T2 的功能是计算具有最难 Trihard loss 距离,即最不相似的正样本之间的距离,并扩大具有最近 Trihard loss 距离,即最可能的负样本之间的距离。这将使负样本与正样本完全分离。这相当于同时使用 Triphard 两次。这样做使得正负样本的模糊边缘变得逐渐清晰,如图 3 所示。

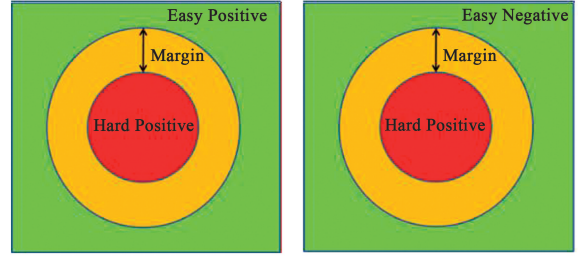


图 3 从边缘模糊样本向难样本和简单样本转换
Fig.3 Conversion from margin blurred sample to hard sample and simple sample

双三分支孪生网络(DTN)实现了不同的目标,即红圈区域变大,绿色区域也变大,黄色区域逐渐缩小直到黄圈区域消失,如图 3 所示。

3 训练策略

3.1 基于子空间聚类的 MSML

子空间聚类的计算过程定义为公式(4)和公式(5):

$$\sum_{i=1}^{|B|} [\|f(x_i^a) - f(x_i^b)\|^2 - \|f(x_i^a) - f(x_i^n)\|^2 + \alpha]_+, s.t. x_i^n = \arg \min_x \|f(x_i^a) - f(x)\|^2, \\ I(x_i^a) \neq I(x) \forall x \in B, x_i^a, x_i^b \in T_m, m = 1, \dots, M, \quad (4)$$

$$\sum_{i=1}^{|B|} [\|f(x_i^{n1}) - f(x_i^{n2})\|^2 - \|f(x_i^{n1}) - f(x_i^a)\|^2 + \beta]_+, s.t. x_i^n = \arg \min_x \|f(x_i^{n1}) - f(x)\|^2, \\ I(x_i^{n1}) \neq I(x) \forall x \in B, x_i^{n1}, x_i^{n2} \in T_m, m = 1, \dots, M. \quad (5)$$

Y. Sun 等人^[23]认为只选择两对样本来形成四元组 Batch 是不合适的,并提出用 OHNM 通过聚类来优化 Batch。因为这个正负 Batch 也是从整个样本中选择的,所以当有太多的人脸样本时,很难匹配两个相似的人脸。这也大幅降低了获取难样本的成功率。一个简单的做法是把同一个人的人脸聚在一起。最后,在每个聚类空间中发现难以区分或非常相似的样本。虽然特征提取和聚类比较耗时,但在聚类空间中的搜索速度得到了很大的提高,因此提升了整体系统的实时性。

聚类子空间的个数决定了困难样本提取的成功率。如果子空间的数目太少,那么子空间的划分将与批处理方法非常相似,大量不相似的人脸被聚类成一类;如果子空间数目太多,会将同一个人的人脸图片划分到多个子空间中,这样挖掘的难度就会加大。

采用子空间聚类方法对所有人脸样本进行聚类,将相似的人脸分为一组。在这些聚类空间中形成一个四元数输入比较容易,挖掘效率较高,如图 4 所示。

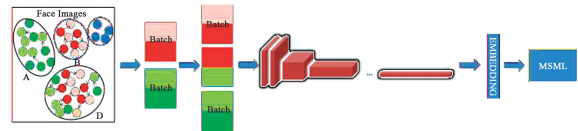


图 4 聚类子空间中负样本和正样本的选择
Fig.4 Selection of negative and positive samples from cluster subspace

3.2 聚焦损失

聚焦损失(Focal loss)用来解决正样本和负样本比例之间的严重不平衡问题。Focal loss 降低了训练中大量简单负样本的权重,这也可以理

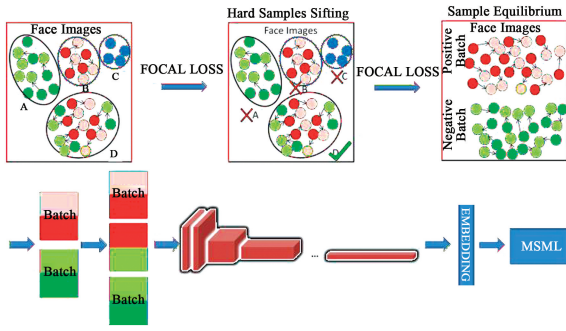


图 5 通过焦损平衡难阴性正样品和难负样品

Fig.5 Balance of hard negative samples and hard positive samples by focal loss

解为挖掘难样本,如图 5 所示。

Focal loss^[24]是交叉熵损失函数的改进,从二分类交叉损失来看,其表达如公式(6)所示:

$$L = \begin{cases} -\log y', & y = 1 \\ -\log(1 - y'), & y = 0 \end{cases}, \quad (6)$$

y' 是激活函数的输出,所以它介于 0 和 1 之间。对于正样本,输出概率越大,损失越小;对于负样本,输出概率越小,损失越小。当 y' 减小时,交叉熵损失函数越小,正样本的收敛速度越慢,不适合选择难样本,损失函数没有参数控制正负样本的比例。

$$L_{fl} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & y = 1 \\ - (1 - \alpha)y'^\gamma \log(1 - y'), & y = 0 \end{cases}. \quad (7)$$

Focal loss 通过增加困难样本控制参数和正负样本平衡控制参数,改进了交叉熵损失函数,并用公式(7)计算了 L_{fl} 。

如果 α 是常数,其中 $\gamma > 0$,则对于正样本,如果 γ 接近 1,则所选样本是简单样本,因此 $-\alpha(1 - y')^\gamma$ 的值将很小,然后损失函数的值将较小。如果 γ 接近 0,则采样困难,损失较大,收敛速度较快。对于负样本来说,它将选择最难的样本,即如果 y' 接近 0。

如果 γ 为常数,其中 $\alpha > 0$,对于正负样本,平衡控制参数 α 是用来平衡正负样本比例,让正负样本本身更加均匀,正样本比例 α 越大,负样本越少。相反, α 越小,正样本比越小,负样本越多。

4 系统设计

4.1 多摄像机监控的特征共享

在现实中,人脸目标可能会出现在几个监视区域。如果要从这些区域的监控视频中检索出所

有的相似人脸帧,就需要建立一个具有数据交换功能的监控网络。

因此,有必要实现多摄像机信息互联和数据共享。多摄像机协作不仅可以扩大监控范围,而且可以从不同角度捕捉人脸信息。如果在摄像机与服务器之间连接多个数据通道来同步人脸特征信息,智能视频监控(IVS)就有可能完成特征交换。在摄像机之间进行人脸特征交换,利用 DTN 神经网络对摄像机之间的人脸特征相似度进行匹配,就可以完成局域网中的多摄像机人脸识别和相似人脸检索。

为了更清楚地描述上述多摄像头特征同步方法的原理,这里设置了一个 2.5D 虚拟智慧城市场景,如图 6 所示。虚拟场景中有 10 个主要建筑物,假设 4 个建筑物监控范围内有行人经过,所以一个特征中心和 4 个建筑物组成一个特征交换监控网络。在理想的情况下,我们假设特征中心没有安装摄像头,其他 4 栋普通建筑都有监控摄像头。从图 6 可以看出,每台监控摄像机都给出了自己的编号和定位标志,分别是 A、B、C、D。可以看到,在每栋楼的摄像机视野范围内,只有一名行人经过,被拍摄的 4 名行人也给出了自己的编号和定位标志。其中, A 摄像机拍摄行人 1, B 摄像机拍摄行人 2, C 摄像机拍摄行人 3, D 摄像机拍摄行人 4。当摄像头 A 捕捉到行人 1 的面部特征时,首先通过 PUSH 将其同步到特征中心。同样,在第一时间, B、C 和 D 摄像机分别将行人 2、3 和 4 的面部特征通过 PUSH 推送到特征中心。特征中心从不同的摄像头接收到这些面部特征后,首先要做的是打包这些面部特征,并在第一时间将它们发送到每个摄像头,以完成 SHARE 共享同步。在从特征中心接收到 SHARE 完成的命令后,每个相机将开始 PULL 同步,即将其他摄像机的人脸特征拉到本地,并将其存储在自己的本地特征库中。完成 PUSH、SHARE 和 PULL 的同步操作后,每个摄像头的特征库将保持一致。局部特征库中的特征用于匹配捕获的人脸特征。因此,当行人 1 通过 B、C 和 D 摄像机的视野时,每个区域的监控摄像机都可以识别和检索行人 1。类似地,每个摄像机也可以识别和检索行人 2、3 和 4。

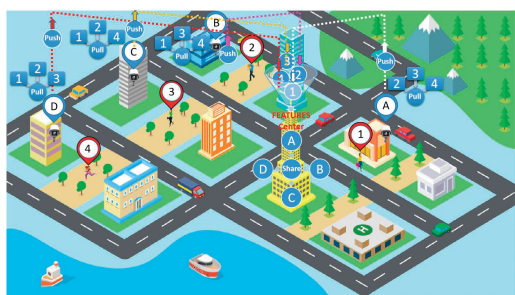


图 6 智慧城市数据通道特征共享立体图

Fig.6 Stereogram of data channel feature sharing in smart city

4.2 联合人脸检索

4.2.1 训练 DTN

(1) 人脸训练数据集 LFW

户外标记人脸(Wild LFW)是人脸识别中常用的一种测试集,它提供了生活中自然场景下的人脸图像,因此识别起来会比较困难,特别是由于多姿态、光线、表情、年龄、遮挡等因素的影响,甚至同一个人的照片也有很大的不同。而且有些照片可能有多张人脸,对于这些多人脸图像只能选择图片中心坐标的人脸作为目标,其他区域作为背景干扰。目前,LFW 数据库的性能评价已成为人脸识别算法的一个重要指标。

(2) 图片预处理

采用子空间聚类方法对 LFW 数据集的所有样本进行聚类,将相似的人脸都聚类到一个子空间中。在局部聚类空间中选择训练的 Batch 样本对将更容易、更有效。

在每个相似的人脸子空间中包含了简单样本和难样本。所有包含简单样本的子空间都应该被剔除。本文采用 Focal loss 难样本筛选的方法,剔除简单样本,留下难样本子空间,从中选取本文网络输入的四元组输入。这个四元数输入包含 3 个难样本对,一个是难正样本,另一个是难负样本,还有一个难正负样本对。系数参见公式(4)所示。

样本方程可以在同一子空间中选取正样本对 anchor 和 positive,负样本对 negative1 和 negative2,Focal loss 可以平衡正样本和负样本,解决了正样本和负样本的不平衡问题。

4.2.2 使用改进的 MSML 训练 DTN

子空间聚类 and Focal loss 可以解决以下问

题。子空间聚类方法能更有效地将相似人脸的样本聚类到一个子空间中。对于一个子空间中的所有 Batch,预测错误的 Batch 交叉熵损失不会减少。对于得分好的样本,交叉熵损失将大为减小,而 Focal loss 使得分好的难样本的损失增大。同时,该模型将更加注重优化样本数较少的损失,训练权重将更加准确。测试分类结果 DTN 加载训练好的权值,通过样本预处理将每个视频的视频帧输入 DTN,对所有人脸进行分类。

4.2.3 联合人脸图像检索

每个视频监控区域将要检索的图像转换为图像的深层特征,用于图像特征匹配,将特征交换和存储层的人脸特征共享给其他视频监控区域后,完成联合人脸检索。

5 实验和结果分析

5.1 改进损失函数有效性的比较分析

LFW 数据集主要测试人脸识别的准确性。数据库随机抽取 6 000 对人脸组成人脸识别对,其中 3 000 对属于同一个人,3 000 对属于不同的人。LFW 给出一对照片,询问测试中的 2 张照片是否是同一个人,然后 DTN 神经网络给出是或否的答案。将 6 000 对人脸测试结果的系统答案与真实答案进行比较,可以得到人脸识别的准确率。

DTN 在人脸识别方面有很大的优势。在 LFW 人脸数据库中,人脸识别的准确率为 99.51%。虽然在精确度上没有超越功能强大的 Facenet 网络,但 DTN 也有很好的表现。DTN 除了网络结构的设计外,最重要的是损失函数的选择。

本文采用 20 种损失函数对 Siamese、Triplet 和 Quadruplet、DTN 网络进行训练,比较了哪种网络模型的损失函数性能最好。本文的方法 1~12 从文献[25]训练精度数据中参考,在此基础上加入各种 DTN 损失函数,然后从多个角度分析训练的最优结果。从表 1 中的 20 种方法的准确度可以看出, $\text{acc}(\text{Siamese}) < \text{acc}(\text{Triplet}) < \text{acc}(\text{Quadruplet}) < \text{acc}(\text{Triphard}) < \text{acc}(\text{MSML}) < \text{acc}(\text{DTN})$ 。从网络结构上,通过对网络结构和损失函数的综合分析,我们可以从表中的数据中得到一些规律:

(1)通过方法 1 和方法 2 的横向对比可以发现,随着网络分支的增加,人脸识别的准确率有了很大的提高。

(2)从方法 3 到 7,可以看到使用难样本作为输入提高了 Triplet 训练精度。然而,当用子空间聚类代替 Batch OHNM 时,其精度不如 softmax。这主要是因为 softmax 降低了训练难度,使得多分类问题更加收敛。Softmax 鼓励增多在实际目标类别中的特征种类。Softmax 鼓励在类别之间分离特征,但不鼓励太多的特征分离。然而,通过难 Trihard 在 OHNM 上预先使用子空间聚类,可以大幅提高收敛速度和精度。如果从方法 8~11 中添加 softmax,子空间可以更有效地分离。

(3)在使用 OHNM 和子空间聚类的训练方

法 5、6、7 的过程中,过小子空间会增加负样本干扰,过小子空间会分离出相似的人脸图像,搜索效率会下降。因此,选择合适的子空间数目更为重要。

(4)方法 15~17 与方法 5~7 比较,方法 8 与方法 13 比较,发现 DTN 比 Triphard 更准确。

(5)方法 9~11 和方法 8~20 表明,在排序方面,Focal loss 优于 softmax。通过分析,得出 DTN、Batch OHNM、子空间聚类和 Focal loss 相结合的分类方法是最佳的分类方法,人脸识别效果最为准确。

5.2 损失函数受试对象工作特征曲线(ROC)综合比较

本文从不同角度对表 1 中的 20 种损失函数进行了比较。将 20 种方法中最具代表性的 3 组

表 1 LFW 数据集上多种损失函数训练精度对比
Tab.1 Comparison of training accuracy of various loss functions on LFW data sets

Method -N (LFW DATASET)	Acc/%
1-Siamese+contrastive loss ^[25]	96.21
2- Quadruplet loss ^[21]	98.85
3-Triplet+Softmax (Baseline) ^[25]	99.36
4-Triplet+Batch OHNM ^[21]	98.98
5-Triplet+Batch OHNM+Subspace-5 ^[25]	99.25
6-Triplet+Batch OHNM+Subspace-10 ^[25]	99.38
7-Triplet+Batch OHNM+Subspace-20 ^[25]	99.33
8-Triplet+Batch OHNM+Softmax ^[21]	99.33
9-Triplet+Batch OHNM+Subspace-5 + Softmax ^[25]	99.38
10-Triplet+Batch OHNM+Subspace-10 + Softmax ^[25]	99.48
11-Triplet+Batch OHNM+Subspace-20 + Softmax ^[25]	99.41
12-Face Net (Triplet+OHNM) ^[25]	99.63
13-Double Triplet Network(MSML)+Batch OHNM + Softmax	99.37
14-Double Triplet Network(MSML)+Batch OHNM + Focal LOSS	99.38
Double Triplet Network(MSML)+Batch OHNM+Subspace-5	99.28
Double Triplet Network(MSML)+Batch OHNM+Subspace-10	99.39
Double Triplet Network(MSML)+Batch OHNM+Subspace-20	99.35
Double Triplet Network(MSML)+Batch OHNM+Subspace-5 +Focal LOSS	99.39
Double Triplet Network(MSML)+Batch OHNM+Subspace-10 +Focal LOSS	99.51
Double Triplet Network(MSML)+Batch OHNM+Subspace-20 +Focal LOSS	99.43

作为判断依据,每组中有 6 种方法。各组分别建立真阳性率(TPR)和假阳性率(FPR)的 ROC 曲线。每个模型都有相似阈值的调整。曲线越靠近左上方,分类器就越好。如图 7~9 所示,每组 ROC 曲线由 6 条 ROC 曲线组成。在 FPR 都取 0.1 的情况下,每个 ROC 曲线中的 red 分类器产生更高的 TPR。这表明 ROC 越高,分类器越好。

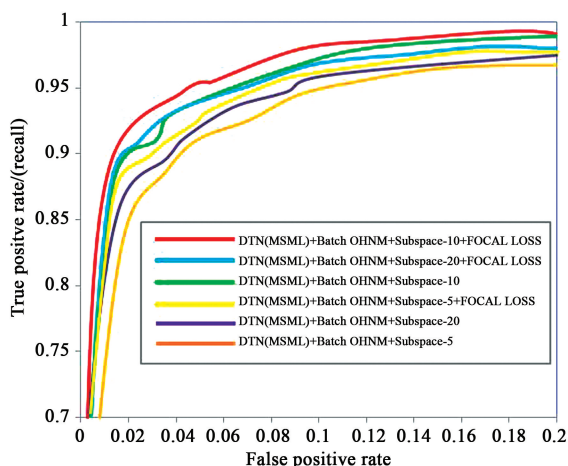


图 7 不同子空间数 DTN 的 ROC 曲线图

Fig.7 ROC for DTN of different subspace numbers

图 7 中使用了两种方法:DTN + MSML + Batch OHNM + Subspace 和 DTN + MSML + Batch OHNM + Subspace + Focal loss。为 3 个子空间绘制 ROC 曲线,子空间数分别为 5, 10, 20。如图所示,为训练添加焦点损失函数可以提高分类精度。同时,子空间个数的选择是关键,子空间不能太多或太少。太多的子空间会使相似的人脸分开;子空间数目太少,许多不相似的人脸被划分到相似的人脸子空间中,降低了遍历效率。

图 8 使用两种类型的方法:Triple + Batch OHNM + Subspace 和 Triple + Batch OHNM + Subspace + Softmax。为 3 种子空间编号中的每一种绘制 ROC 曲线,分别设置子空间数为 5, 10, 20。从图中可以看出,利用 Softmax 函数进行训练可以提高 Triple 的分类精度。此外,子空间数目的选择也很关键,子空间不能太多或太少。太多的子空间会使相似的人脸分开;子空间数目太少,许多不相似的人脸被划分到相似的人脸子空间中,降低了遍历效率。

图 9 将不同网络与 Facenet 在性能上进行了比较,结果表明,虽然 DTN 的精度比 Facenet 低,

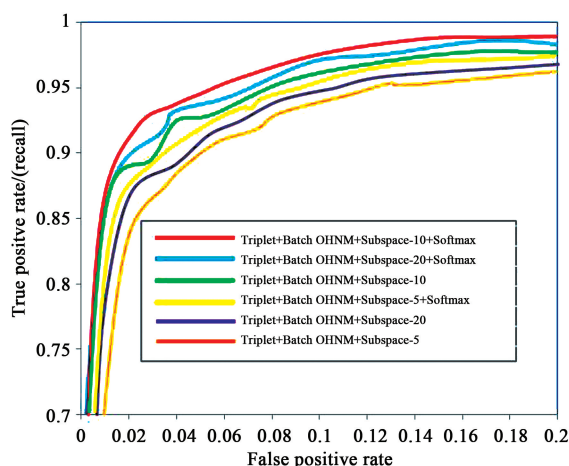


图 8 3 种不同子空间数的 ROC 曲线图

Fig.8 ROC for Triple of different subspace numbers

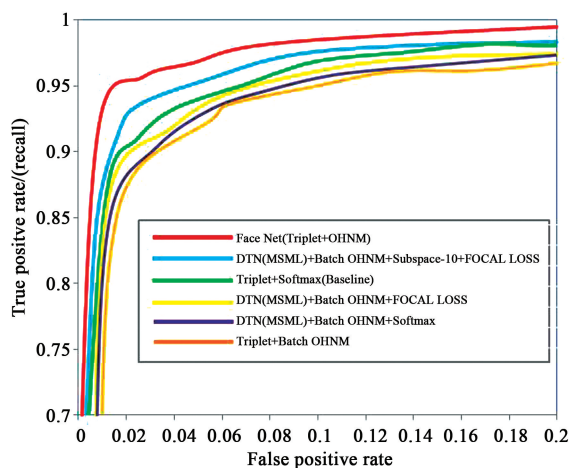


图 9 不同网络与 FaceNet 的 ROC 曲线比较

Fig.9 ROC comparison between different networks and FaceNet

但与其他模型相比,DTN 的训练效果有了明显提高。

5.3 联合人脸检索的有效性分析

本文选取表 1 中 5、6、7、9、10、11、15、16、17、18、19、20 的方法在 PaSC、LFW、PubFig、FERET、AR and YaleB 数据集上完成训练和测试。

在数据集 PASC、LFW、PubFig、FERET、AR 和 YaleB 中,本文利用这些方法计算了平均搜索精度 ARP 的性能评价。

采用平均检索查准率对人脸检索结果进行评价。每个数据库中的每幅图像都作为一个查询图像来检索所有剩余的图像,以记录相应网络的检

索性能。对数据库中所有数据的检索精度和检索率进行平均,得到平均检索精度(ARP)。

采用 Triplet + Batch OHNM + Subspace、Triplet + Batch OHNM + Subspace + Softmax 和 DTN MSML + OHNM + Batch + Subspace + Focal loss 三种训练模式对 PaSC、LFW、PubFig、FERET、AR 和 YaleB 数据集的 ARP 进行测试。

表 2 使用上述 3 种训练模式计算了上述 6 个数据库中平均检索精度最高的最佳匹配人脸。从表 2 中可以看出,这 3 类子空间的 ARP 在 LFW 上最高,在 YaleB 上最低。这是因为扩展的 YaleB 增加了很多干扰因素,如光照、姿态等,在这 3 种方法中,DTN-MSML+Batch-OHNM+Subspace+Focal 方法的性能最好。同样地,选择子空间数为 10 可以提高训练精度。

表 2 最顶层匹配的平均检索精度(ARP)

		Tab.2 Average retrieval precision (ARP) for topmost match										(%)
Dataset	M.5	M.6	M.7	M.9	M.10	M.11	M.15	M.16	M.17	M.18	M.19	M.20
PaSC	94.16	94.25	94.21	94.18	94.36	94.25	94.31	94.43	94.36	95.05	95.32	95.28
LFW	99.22	99.36	99.30	99.30	99.45	99.39	99.26	99.36	99.33	99.35	99.49	99.40
PubFig	98.18	98.30	98.24	98.26	98.40	98.34	98.31	98.34	98.32	98.33	98.44	98.35
FERET	96.12	96.16	96.15	96.23	96.27	96.25	96.27	96.32	96.29	96.41	96.50	96.45
AR	99.76	99.78	99.77	99.82	99.87	99.83	99.82	99.84	99.79	99.85	99.87	99.86
YaleB	87.01	87.09	87.07	87.09	87.12	87.10	87.13	87.17	87.14	87.24	87.27	87.25

5.4 消融实验

通过以上对比实验,发现模型损失函数的最佳组合为:DTN(MSML)+Batch OHNM+subspace-10+Focal loss。在本节中,我们将建立该模型的消融实验^[26]。消融实验的目的是从一个单独角度探讨每个变量的作用。这里以 DTN(MSML)为基本模型,控制变量主要包括 Batch OHNM、Subspace 和 Focal loss。我们一次只减少一个变量来观察检索精度、实时性、鲁棒性的变化。

5.4.1 控制变量 Batch OHNM

从 DTN + Batch OHNM + Subspace-10 + Focal loss 的组合中去除 Batch OHNM,选择随机样本配对。从表 3 可以看出,正样本对和负样本对没有难样本配对机制,使得 DTN 模型失去了拉开正样本和负样本之间相对距离的意义,导致检索率大幅降低。通过 Batch OHNM 机制,可以在 DTN 模型中加入两个不同模式的样本,使得模型具有异构性,大幅提高了检索精度。然而,由于样本选择的困难,预处理时间比随机样本时间长,因此严重影响了实时性。

表 3 在 PaSC、LFW、PubFig、FERET、AR 和 YaleB 上使用 DTN+ Subspace-10 + Focal Loss 的 10 个检索图像的 ARP。Batch OHNM 对检索精度和样品预处理时间的影响

Tab.3 ARP for 10 numbers of retrieved images using DTN + Subspace-10 + Focal Loss over the PaSC, LFW, PubFig, FERET, AR and YaleB. Influence of Batch OHNM on retrieval accuracy and sample Pre-processing time

Dataset	Random sample		Batch OHNM	
	Acc/ %	Pre-processing time/s	Acc/ %	Prep-rocessing time/s
PaSC	81.36	12.6	84.98	16.9
LFW	95.78	13.5	97.77	18.3
PubFig	92.26	67.5	95.41	91.2
FERET	79.16	13.2	81.48	17.8
AR	80.55	2.9	82.39	3.9
YaleB	68.74	6.7	71.39	9.1

5.4.2 控制变量 Subspace

从 DTN+Batch OHNM+Subspace-10+Focal loss 组合中去除子空间(Subspace),不生成聚类子空间。从表 4 可以看出,没有子空间聚类的 DTN 模型检索精度大幅降低,焦点丢失无法找到模糊样本集,进而无法平衡困难样本对的比例。同时,由于没有子空间聚类,使得困难样本的配对效率低下,导致预处理速度慢。加入子空间聚类机制后,可以快速发现相似样本,包括相似正样本和相似负

样本。但是从表 4 来看,太少的聚类子空间划分会减慢搜索速度,如子空间为 5,而太多的聚类子空间划分会丢失出许多相似的样本对,如子空间为 20,子空间为 40,这也破坏了样本的内部关系。因此,选择合适的聚类子空间数目是非常重要的。从实验中可以看出,子空间为 10 是最好的子空间划分,而且每个数据集的检索精度最高,而随着子空间数目的增加从更小的范围内搜寻目标样本的时间明显降低,因此实时性会有所提高。

表 4 10 个使用 DTN+批次 OHNM+在 PaSC、LFW、PUBFIG、FERET、AR 和 YaleB 上使用 DTN+批次 OHNM+焦耗检索的图像的 ARP。聚类子空间的数目对检索精度和实时性的影响

Tab.4 ARP for 10 numbers of retrieved images using DTN+Batch OHNM+ Focal Loss over the PaSC, LFW, PubFig, FERET, AR and YaleB. Influence of the number of clustering subspaces on retrieval accuracy and real-time

Dataset	Subspace=0	FPS	Subspace=5	FPS	Subspace=10	FPS	Subspace=20	FPS	Subspace=40	FPS
PaSC	81.35	18	82.15	24	84.98	27	81.98	29	81.49	34
LFW	95.25	15	97.49	26	97.77	28	96.45	31	95.96	32
PubFig	92.26	14	94.67	24	95.41	28	94.45	34	93.49	36
FERET	79.49	16	80.95	25	81.48	26	80.12	30	79.89	34
AR	80.19	17	81.69	27	82.39	27	81.47	31	81.02	32
YaleB	68.68	15	70.66	26	71.39	26	70.34	29	69.48	31

5.4.3 控制变量 Focal loss

从 DTN+Batch OHNM+Subspace-10+Focal loss 组合中消除 Focal loss,这样做的后果是正、负样本既不均衡,也不利于难样本筛选。从表 5 可以看出,DTN 模型输入端缺乏难样本筛选,在很大程度上是从简单样本聚类子空间中选择样本对,不能充分发挥模型的相对优势。同时,正样本与负样本之间不平衡,导致正负样本混淆,进而导致模型训练中的过拟合问题,大幅降低了检索精度。Focal loss 的增加使得难样本进行过滤成为可能,也可以分离模糊样本带来的过拟合问题。通过调整 Focal loss 函数参数,发现当 $\alpha=0.5$ 时,即正负样本均衡,训练精度随样本难度的增加而提高。然而,当样本数相等时,训练精度提高,即样本难度最大时,训练精度提高 $\gamma=0.9$ 。样本不平衡引起的过拟合问题使得检索精度降低

(例如,当 $\alpha=0.1$ 和 $\alpha=0.9$ 时),从实时性角度未做样本均衡的输入会增加正负样本的搜索时间,影响实时性表现,而样本均衡设置的不好的情况(例如,当 $\alpha=0.1$ 和 $\alpha=0.9$ 时)也会使帧频降低,实时性表现不好,只有正负样本均衡的情况实时性才表现最佳。

以上消融实验结果表明,各变量的作用是不可或缺的。同时,每个变量都有参数调整状态,以达到最优解。从表 3、表 4 和表 5 中可以得到最佳参数,最佳参数如下:①选择批量 OHNM 预处理;②子空间为 10 是子空间聚类的最佳聚类数;③ $\alpha=0.5, \gamma=0.9$ 是焦损的最佳参数。因此在最佳参数设置下,本文选择 DTN(MSML)+Batch OHNM+Subspace-10+Focal loss 的最佳组合,共同配合提高了检索精度,且实时性可达 28 FPS 左右。

表 5 10 个使用 DTN+批次 OHNM+在 PaSC、LFW、PUBPIG、FERET、AR 和 YaleB 上使用 DTN+批次 OHNM+焦耗检索的图像的 ARP。Focal loss 对检索精度和实时性的影响

Tab.5 ARP for 10 numbers of retrieved images using DTN+Batch OHNM+ Subspace=10 over the PaSC, LFW, PubFig, FERET, AR and YaleB. Influence of Focal loss and its corresponding parameters on retrieval accuracy and realtime

Dataset	Without Focal loss	Acc%/FPS									
		$\alpha = 0.5$	$\gamma = 0.1$	$\alpha = 0.5$	$\gamma = 0.5$	$\alpha = 0.5$	$\gamma = 0.9$	$\alpha = 0.1$	$\gamma = 0.9$	$\alpha = 0.9$	$\gamma = 0.9$
PaSC	81.28/15	81.96/35		82.89/31		84.98/26		83.16/25		83.24/21	
LFW	95.17/19	95.63/36		96.59/29		97.77/28		96.92/23		96.86/25	
PubFig	92.05/14	95.39/35		93.92/30		95.41/25		94.34/24		94.44/23	
FERET	79.18/16	81.32/39		80.49/29		81.48/27		80.79/26		80.82/24	
AR	79.98/14	82.25/34		81.13/31		82.39/25		81.88/24		81.76/24	
YaleB	68.56/13	69.10/35		70.26/32		71.39/28		70.72/22		70.83/25	

5.5 系统鲁棒性测试

为了验证系统的鲁棒性,应用 Silktest 测试工具建立测试,将系统监控数目和鲁棒性之间建立内部关联项,在最优化组合、消除变量 Batch OHNM、消除变量 Subspace、消除变量 Focal loss

属性上评估系统鲁棒性。从图 10 测试结果中可以看出,随着监控压力逐渐加大,最优解组合的鲁棒性表现最优,而其他每个变量都将对鲁棒性造成严重影响,其中 Focal loss 是对系统鲁棒性影响最大的变量。

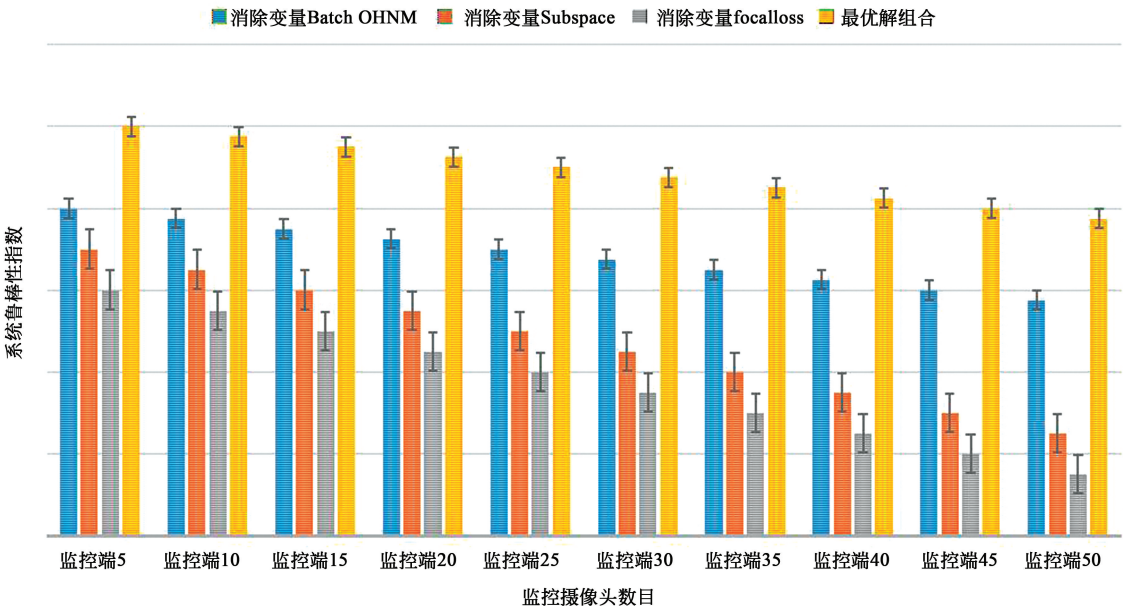


图 10 系统鲁棒性分析图

Fig.10 Robust analysis of the system

6 结 论

本文提出了一种新型网络结构 DTN。通过

对 Siamese, Triplet 和 Quadruplet 的比较,我们发现网络结构与网络分支的通道数有关。在分支数相同的情况下,最重要的是如何设计有效损失函数。本文对 DTN 网络进行了改进。网络

结构是一种双三支孪生结构,以两个样本对为输入,可以拓宽正负样本之间的距离,使正负样本完全分离。本文设计的损失函数是几种损失函数的结合。为了提高搜索速度,选择了子空间聚类方法。在 DTN 网络的样本预处理阶段,为了解决增加训练难度和均衡样本的问题,在训练阶段采用了 Focal loss 函数。基于 MSML,比较了 20 种方法的网络损失。实验结果表明,DTN + OHNM + Batch + Subspace +

Focal loss 具有较高的 ARP、ARR 和 F-score 值。因此,该方法的平均检索精度(ARP)为 98.74%,模型训练精度为 99.54%,帧率为 28 FPS,能够满足人脸检索的实际应用,并且系统考虑了远程视频检索在刑侦部门远程通缉嫌疑人目标人脸应用中的实际困难。设计了一个多摄像机人脸联合检索系统,解决了多摄像机特征共享和特征联合的问题,实现了多局域网多摄像机人脸联合检索。

参 考 文 献:

- [1] WU H, YOKOYAMA T, PRAMADIHANTO D, *et al.* Face and facial feature extraction from color image [C]// *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*. Killington, VT, USA: IEEE, 1996: 345-350.
- [2] SHI Y X, REN X X, YANG S Z, *et al.* A generalized Kernel Fisher Discriminant framework used for feature extraction and face recognition [C]// *Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. Changsha, China: IEEE, 2016: 1487-1491.
- [3] LU Z, JIANG X D, KOT A. Deep coupled ResNet for low-resolution face recognition [J]. *IEEE Signal Processing Letters*, 2018, 25(4): 526-530.
- [4] SCHMID C, MOHR R. Local grayvalue invariants for image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(5): 530-535.
- [5] RAZA A, DAWOOD H, DAWOOD H, *et al.* Correlated primary visual Texton histogram features for content base image retrieval [J]. *IEEE Access*, 2018, 6: 46595-46616.
- [6] XIANG J, ZHANG N, PAN R R, *et al.* Fabric image retrieval system using hierarchical search based on deep convolutional neural network [J]. *IEEE Access*, 2019, 7: 35405-35417.
- [7] PHAM T T D, KIM S, LU Y C, *et al.* Facial action units-based image retrieval for facial expression recognition [J]. *IEEE Access*, 2019, 7: 5200-5207.
- [8] DONG Z, JING C C, PEI M T, *et al.* Deep CNN based binary hash video representations for face retrieval [J]. *Pattern Recognition*, 2018, 81: 357-369.
- [9] NAGPAL S, SINGH M, SINGH R, *et al.* Regularized deep learning for face recognition with weight variations [J]. *IEEE Access*, 2015, 3: 3010-3018.
- [10] ABUZNEID M A, MAHMOOD A. Enhanced human face recognition using LBPH descriptor, Multi-KNN, and back-propagation neural network [J]. *IEEE Access*, 2018, 6: 20641-20651.
- [11] 肖康,田立君,王中阳. 基于深度学习的低信噪比下的快速超分辨率荧光显微成像[J]. *中国激光*, 2020, 47(10): 222-232.
XIAO K, TIAN L J, WANG Z Y. Fast super-resolution fluorescence microscopy imaging with low signal-to-noiseratio based on deep learning [J]. *Chinese Journal of Lasers*, 2020, 47(10): 222-232. (in Chinese)
- [12] 胡麟苗,张湧,楼晨凤. 基于内容特征提取的短波红外-可见光人脸识别[J]. *光学精密工程*, 2021, 29(1): 160-171.
HU L M, ZHANG Y, LOU C F. Shortwave infrared visible-light face recognition based on content feature extraction [J]. *Optics and Precision Engineering*, 2021, 29(1): 160-171. (in Chinese)
- [13] 明悦,王绍颖,范春晓,等. 对齐特征表示的跨模态人脸识别[J]. *光学精密工程*, 2020, 28(10): 2311-2322.
MING Y, WANG S Y, FAN C X, *et al.* Exploring aligned latent representations for cross-domain face recognition [J]. *Optics and Precision Engineering*, 2020, 28(10): 2311-2322. (in Chinese)
- [14] 秦智康,张衡阳. 基于变换域中的自适应纹理图像检索[J]. *液晶与显示*, 2020, 35(3): 281-286.
QIN Z K, ZHANG H Y. Adaptive texture image retrieval based on transform domain [J]. *Chinese Journal of*

- Liquid Crystals and Displays*, 2020, 35(3): 281-286. (in Chinese)
- [15] SUN Y, CHEN Y H, WANG X G, *et al.* Deep learning face representation by joint identification-verification [C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada: MIT Press, 2014: 1988-1996.
 - [16] ZHANG M M, LI W, DU Q. Diverse region-based CNN for hyperspectral image classification [J]. *IEEE Transactions on Image Processing*, 2018, 27(6): 2623-2634.
 - [17] ZHAO Y D, HAO K R, HE H B, *et al.* A visual long-short-term memory based integrated CNN model for fabric defect image classification [J]. *Neurocomputing*, 2020, 380: 259-270.
 - [18] SARDOGAN M, TUNCER A, OZEN Y. Plant leaf disease detection and classification based on CNN with LVQ algorithm [C]//*2018 3rd International Conference on Computer Science and Engineering (UBMK)*. Sarajevo, Bosnia-Herzegovina: IEEE, 2018: 382-385.
 - [19] GE W F, HUANG W L, DONG D K, *et al.* Deep metric learning with hierarchical triplet loss [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 272-288.
 - [20] CHEN W H, CHEN X T, ZHANG J G, *et al.* Beyond triplet loss: A deep quadruplet network for person re-identification [C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 1320-1329.
 - [21] LV X B, ZHAO C R, CHEN W. A novel hard mining center-triplet loss for person re-identification [C]//*Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Xi'an, China: Springer, 2019: 199-210.
 - [22] YAO H P, FU D Y, ZHANG P Y, *et al.* MSML: A novel multilevel semi-supervised machine learning framework for intrusion detection system [J]. *IEEE Internet of Things Journal*, 2019, 6(2): 1949-1959.
 - [23] SUN Y, LIANG D, WANG X G, *et al.* DeepID3: Face recognition with very deep neural networks [J]. *arXiv preprint arXiv*: 1502.00873, 2015.
 - [24] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE, 2017: 2999-3007.
 - [25] WANG C, LAN X P, ZHANG X. How to train triplet networks with 100 K identities? [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice, Italy: IEEE, 2017: 1907-1915.
 - [26] ZHANG H J, JI Y Z, HUANG W, *et al.* Sitcom-star-based clothing retrieval for video advertising: a deep learning framework [J]. *Neural Computing and Applications*, 2019, 31: 7361-7380.

作者简介:



任国印(1985—),男,内蒙古包头人,博士研究生,讲师,2013 年于内蒙古科技大学获得硕士学位,主要从事模式识别和智能图像处理方面的研究, E-mail: 1712152231@qq.com



吕晓琪(1963—),男,内蒙古包头人,博士,教授,2003 年于北京科技大学获得博士学位,主要从事模式识别和智能图像处理的研究。 E-mail: 635302395@qq.com