

DOI: 10.13382/j.jemi.B2003459

# 利用判别采样的视频人脸亲属关系验证\*

王仕伟 闫海滨

(北京邮电大学 自动化学院 北京 100876)

**摘要:**针对人脸视频数据中包含大量低质量数据和冗余信息的问题,提出了一种利用深度强化学习实现人脸视频的关键帧和困难负样本的判别采样,实现了视频人脸亲属关系的验证。设计了3个深度网络来实现特征的学习与关键帧的采样,其中基于残差网络(Resnet)训练的亲属关系验证网络(KVN)用于验证亲属分类;通过深度强化学习方法分别设计了两个采样网络:关键帧采样网络(KSN)和负样本采样网络(NESN),用于实现视频关键帧的选取以及负样本的筛选。实验结果表明,相比于现有的主流人脸亲属关系验证算法,方法有效提高了亲属关系验证识别率。

**关键词:**视频亲属关系验证;深度学习;强化学习;判别采样

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4050

## Discriminative sampling for video-based facial kinship verification

Wang Shiwei Yan Haibin

(School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** In this paper, we propose a discriminative sampling method to select most effective samples via deep reinforcement learning for video-based kinship verification. Unlike most existing facial kinship verification methods which focus on extracting effective features with the random sampling strategy, we develop two deep reinforcement learning methods to select samples which are more suitable for learning discriminative features, so that the overall performance can be improved. Compared with the conventional kinship verification problem, video-based kinship verification has received less attention. However, this work has greater value in practical applications. When we try to use kinship verification to solve, for example, the problem of missing population, we often get video data. Compared with images, videos contain more information, and through reasonable use we will get better performance than image-based kinship verification. Specifically, our method uses three subnetworks to achieve the kinship verification task: one DQN-based sampling network to filter the key frame, one DQN-based sampling network to filter the negative sample, and one multi-layer convolutional network to verify the kin relationship. Experimental results on the KFVW datasets show the superiority of our proposed approach over the state-of-the-arts.

**Keywords:** video-based kinship verification; deep learning; reinforcement learning; discriminative sampling

## 0 引言

随着人工智能的飞速发展,计算机视觉在视频监控<sup>[1]</sup>、人机交互<sup>[2]</sup>、社交媒体<sup>[3]</sup>和人脸识别<sup>[4]</sup>等领域得到了广泛的应用。此外,由于计算机视觉和社会科学的紧密结合,使得年龄估计<sup>[5]</sup>、人类情感识别<sup>[6]</sup>和社会关系分析<sup>[7]</sup>等应用也得到了学者们的广泛关注。受失踪人员

搜索、社交媒体分析和模仿识别<sup>[6,8-9]</sup>的实际应用需求所启发,人脸亲属关系验证已经成为计算机视觉领域又一个研究热点。与传统的人脸识别<sup>[10-11]</sup>相比,亲属关系验证面临更多挑战,亲属之间存在的年龄和性别差异、样本数据的采集条件等因素都会对亲属关系验证的性能产生巨大的影响。另外,实验数据集的缺乏也影响了亲属关系验证性能的提升。如何从有限的数据中提取关键信息是亲属关系验证需要解决的一个课题。

收稿日期: 2020-09-03 Received Date: 2020-09-03

\* 基金项目: 国家自然科学基金(61976023)、北京市自然科学基金(4174101)项目资助

近几年来,亲属关系验证方面的研究大多基于静态图像。在实际应用中,随着我国天网系统的日趋完善,视频已逐渐成为获取数据的一种常用方式,例如可以通过监控获取的视频来对失踪人员进行搜索。另外,视频相比图像包含更多的信息,如同一对象不同角度的人脸信息等。合理利用视频中的信息将会更好地提升人脸亲属关系的验证性能。视频数据虽然可以提取出大量不同的帧,但这些帧并不全都适用于亲属关系的验证。因此,本文将针对这一问题对基于视频的人脸血缘关系进行研究。

深度卷积网络已广泛应用于图像识别领域<sup>[12-13]</sup>,在人脸亲属关系识别中也得到了较好的验证。通过将亲属图片对输入深度卷积网络的方式,可以得到二者之间相似的特征,进而可以得到亲属关系的识别结果。通常来说,为保证深度卷积网络的训练效果,需要提供大量的正负样本。正样本数据为具有正确亲属关系的样本对,获取难度大,数量有限;除正样本外,其余样本均可认为是负样本。由于负样本数量巨大,为保证正负样本的均衡性,传统构建负样本数据集的做法采取从大量负样本中随机挑选与正样本同等数量负样本的原则。但这些随机选取的负样本并不全都利于网络的梯度下降。如何选取有利于提升网络训练性能的负样本为本文将要研究的第二个问题。

本文提出采用深度强化学习方法解决选取视频关键帧及有效负样本的问题,以便获得更好的网络性能。主要创新点包括应用深度卷积网络研究了基于视频的亲属关系验证问题;设计了亲属关系验证网络(KVN),基于KFVW数据集,直接通过亲属视频验证人物间的亲属关系,具有更强的实际应用价值。基于深度强化学习设计了关键帧采样网络(KSN),解决了亲属视频关键帧提取问题,提高了训练数据的质量。基于深度强化学习设计了负样本采样网络(NESN),解决了适用于亲属关系验证的负样本筛选问题;聚焦数据的筛选问题,为改善亲属关系验证方法的性能提出了一种新的途径。

## 1 相关工作

### 1.1 亲属关系验证

亲属关系验证方法可以分为基于特征的方法和基于学习的方法两个大类。基于特征的方法主要包括梯度直方图<sup>[14]</sup>、Gabor 梯度方向金字塔<sup>[15]</sup>、肤色<sup>[14]</sup>、自相似性<sup>[16]</sup>等。基于学习的方法主要包括子空间学习、度量学习、迁移学习、多核学习等。这些常用于亲属关系验证中的方法主要专注于如何构建更复杂的系统以获得更详细的特征信息,但却忽略了样本本身的问题。对于亲属关系验证中的大量负样本,现有方法仅执行简单的随机选

择,忽略了负样本的质量。而且,这些方法主要应用于基于图像的亲属关系验证,很少被用于解决基于视频的亲属关系验证。随着亲属关系视频数据库的建立和发布,基于视频的亲属关系验证逐渐成为研究热点。解决视频数据中关键帧的选择问题可以作为提高亲属验证性能的一个研究方向。

### 1.2 困难样本挖掘

困难样本挖掘是机器学习中处理样本的常用操作,同样适用于深度卷积网络。常用的困难样本挖掘算法主要包含两种,第1种方法用于优化支持向量机(SVM),第2种方法用于非SVM,已应用于各种模型,如浅层神经网络和增强决策树。困难样本挖掘在许多有效模型中得到广泛应用,如在深度嵌入式学习中,研究人员使用困难样本挖掘方法在嵌入空间中选择困难样本。对训练方法和损失函数的研究也证明了困难样本挖掘的有效性。在深度卷积网络中使用困难样本挖掘的现有方法中,大多数方法都基于损失。本文将采取通过深度强化学习来获得自适应采样的策略。

### 1.3 深度强化学习

深度强化学习将具有感知能力的深度学习和具有决策能力的强化学习结合在了一起。由于其在多个领域展现出来的良好性能,已经成为了人工智能领域新的研究热点。在早期工作中,研究人员探索了与分布式动态规划相关的问题,并使用对等通信研究了异步优化设置下的Q-learning收敛特性。在后续工作中,由于深度网络的发展,研究人员最近做了很多扩展,并在深度Q-learning(DQN)的基础上进行了进一步的扩展。本文将使用DQN作为采样网络的设计框架。

## 2 提出的方法

本文提出应用深度强化学习的方法进行亲属关系验证的研究,所设计网络主要包括3个部分,亲属关系验证网络用于验证输入视频对的亲属关系,关键帧采样网络用于从大量的视频帧中提取包含判别信息的关键帧,负样本采样网络用于从大量负样本中选取有效负样本提升网络的训练性能。本文所提出方法的整体流程如图1所示。在步骤1)中,对关键帧采样网络、负样本采样网络这两个结构相似的采样网络进行训练。为了得到每次采样的评分,需要预训练一个亲属关系验证网络,将这个网络的验证结果作为评分标准。采样网络根据当前状态值得到动作值,然后环境中预训练好的亲属关系验证网络根据动作值得到奖励值以及下一个状态值。接下来先训练关键帧采样网络,再训练负样本采样网络。由于两个网络的训练步骤基本相同,在整体流程上可以将它们视

为并行。在采样网络训练完成后,进入步骤 2), 训练最终的亲属关系验证网络。通过训练好的采样网络得到适合亲属关系验证的样本, 并使用这些样本对亲属关系验证网络进行训练。

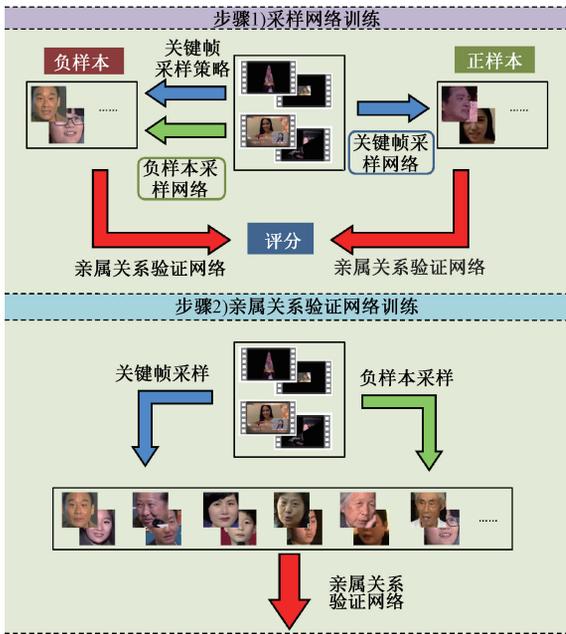


图 1 本文方法流程示意图

Fig. 1 Process diagram

### 2.1 亲属关系验证网络

亲属关系验证网络采取残差网络(Resnet)的结构。残差网络通过堆叠多个残差模块, 保证了网络深层信息一定大于浅层, 有效避免了网络退化的问题, 同时让网络层数大量增加, 便于提取判别性强的特征。

用于亲属关系验证网络中残差模块的结构如图 2 所示。模块包含 3 个卷积层, 在每个卷积层之后, 进行一次批标准化与 Relu 激活函数操作。设卷积层的传递函数为  $G(x)$ , 残差模块的输出为  $C(x)$ , 可得:

$$C(x) = x(1 + G(x)) \quad (1)$$

可以看出, 残差模块不会丢失输入特征原本的信息, 利用这种特质堆叠多个残差模块可以让网络变得非常深。本文使用的网络为 Resnet50, 将残差模块堆叠了 16 次。网络末端, 通过一个全连接层输出二分类的亲属关系验证结果。

进行亲属关系验证前, 需要从待验证的两个视频中分别提取一张关键帧, 将人脸区域检测并裁剪出来, 形成两张大小为  $64 \times 64$  的三通道人脸图片。将两张人脸图片进行堆叠, 得到大小为  $6 \times 64 \times 64$  的数据, 并将其输入亲属关系验证网络进行识别。

### 2.2 关键帧采样网络

关键帧采样网络用于从亲属关系视频数据中挑选关

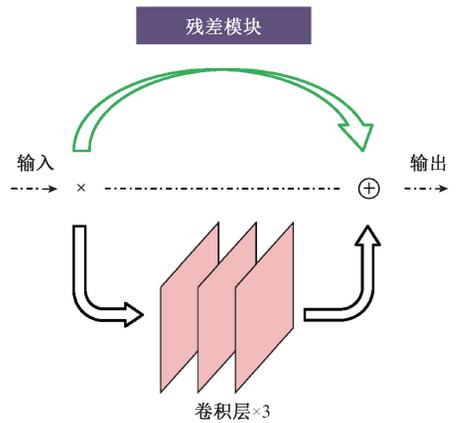


图 2 应用于亲属关系验证网络中的残差模块结构

Fig. 2 Residual module structure diagram

键帧。由于视频拍摄状况各不相同, 不同的帧之间往往存在着较大的差异, 如对象缺失、人脸区域局部遮挡、非正面人脸等。KFVW 数据集的部分样本示例如图 3 所示, 每个样本示例都是分别从两个亲属关系视频中截取出的帧。从图 3 可以看出, 存在遮挡或面部角度过大的人脸图像属于低质量样本。如果仅仅采取简单的随机选取往往无法获得令人满意的样本, 一些效果不佳的示例可能会影响网络性能。为解决这个问题, 本文设计了一个关键帧采样网络, 对每个亲属关系视频的关键帧进行挑选。

	优秀的样本	低质量样本	
正样本			存在遮挡
			面部角度过大
			年龄差过小
			年龄差过大
负样本			年龄差过小
			年龄差过大

图 3 KFVW 数据集的样本示例

Fig. 3 Image example of KFVW dataset

关键帧采样网络采取了基于深度强化学习的结构。每一次筛选流程中, 分别从父母和孩子的视频中挑选出 3 帧。为了评估关键帧的质量, 需要提前训练一个亲属关系验证网络, 通过它的验证结果来判定关键帧的评分。设评分函数为  $S_1(x)$ , 目的是令整个筛选过程获得较高的总分。在实验中, 令  $S_1(x) = \lambda$ , 其中  $\lambda$  是关键帧通过

亲属关系验证网络后得到的损失函数的值。

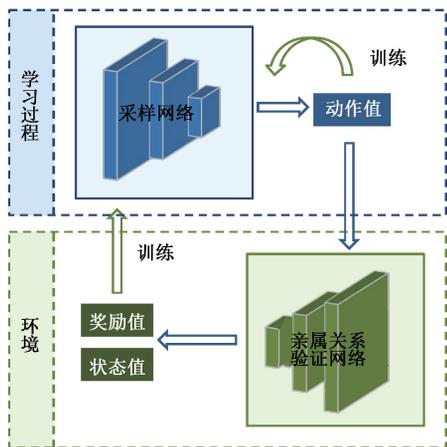


图 4 采样网络的训练流程

Fig. 4 Training process of sampling network

得到评分函数后,采用一个深度强化网络来学习采样策略,训练流程如图 4 所示。关键帧采样网络通过输入的状态值得到下一次的动作值。状态值为当前动作挑选的关键帧的图像数据,动作值为下一个关键帧的编号。使用结构相同但参数不同的两个网络来学习当前状态值和将来状态值,并使用获得的两个  $Q$  值来计算损失函数。采样网络的网络部分学习过程如图 5 所示。这两个网络结构相同,并以一定的频率共享参数。两个网络得到的  $Q$  值加上环境反馈的奖励值可以计算得到损失函数的值,并反馈给行为网络。经过 100 次这样的学习过程后,行为网络向目标网络共享参数值。网络中的  $Q$  值更新如下:

$$Q(s, \alpha) \leftarrow Q(s, \alpha) + \alpha[r + \gamma \max_{\alpha'} Q(s', \alpha')] \quad (2)$$

式中:  $\alpha$  表示学习率,学习率决定了权重迭代的步长,因此是一个非常敏感的参数,低学习率会使模型收敛困难,高学习率容易造成过拟合。 $r$  表示环境反馈激励,也就是评分函数  $S_1(x)$ 。 $\gamma$  是递减的奖励参数,含义是网络将以  $\gamma$  的概率挑选最优动作,( $1 - \gamma$ ) 的概率挑选随机动作。这个值的设定可以防止网络在某种策略下自锁,同时如果取值较低将影响网络收敛效率,实验中将其设置为 0.9。 $Q(s, \alpha)$  是当前状态的  $Q$  值, $Q(s', \alpha')$  是下一个状态的  $Q$  值。

算法中使用的损失函数如下:

$$\text{Target } Q = r + \gamma \max_{\alpha'} Q(s', \alpha'; \theta) \quad (3)$$

$$L(\theta) = E[(\text{Target } Q - Q(s, \alpha, \theta))^2] \quad (4)$$

式中:  $\theta$  表示网络参数是均方差损失。通过筛选,可以获得尽可能有利于深度网络梯度下降的数据。

### 2.3 负样本采样网络

负样本采样网络用于从大量的负样本数据库中选择

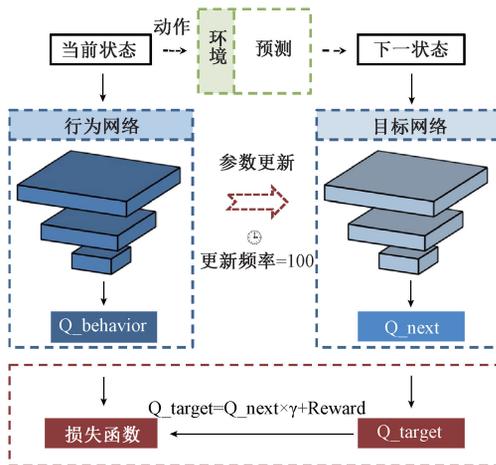


图 5 采样网络的网络部分学习过程

Fig. 5 The learning process of the network part of the sampling network

有利于提升网络训练性能的负样本。进行亲属关系验证时,需要向网络中同时输入均等数量的正样本和负样本。正样本来自具有亲属关系的视频库,父母和子女各一个视频。首先需要分别提取每个视频对的关键帧,然后将父母与孩子的关键帧堆叠得到  $6 \times 64 \times 64$  的数据。负样本来自两个不具有亲属关系的视频库,同样父母和其非子女各一个视频,采取相同的操作进行关键帧的提取和堆叠。

可以看出,在亲属关系验证问题中,正样本的数量非常有限,但是负样本的数量却非常庞大。以 100 个具有亲属关系的视频对举例,可以组成的正样本有 100 组,但是可以组成的负样本却有  $100 \times (100 - 1) = 9900$  组。为保证正负样本的均衡性,用于实验中的正负样本的数量应该相同或差距不大。这就需要从 9900 个样本中筛选出 100 个样本。

大多数方法采用随机的方式进行采样,但这样往往会获得一些效果不佳甚至具有误导性的样本。例如,在正样本中,父母与孩子之间的年龄差通常在 20~40 岁。如果随机挑选的负样本对之间存在较大或较小的年龄差异,如老人与幼儿、两个中年人进行配对,如图 3 的第 2 行样本示例所示,将会导致与正样本的年龄差严重不符,而本文期望这种年龄差与正样本不符的负样本尽量少。其次,负样本的随机挑选还会出现一些具有误导性的样本。如 A 和 B 为一组亲属,A 和 C 为另一组亲属,在随机挑选的时候,可能会将 A 和 C 或者 A 和 B 的配对错误匹配为负样本,实际上它们是正样本。可以看到,随机采样下的负样本无法取得令人满意的结果,因此如何从大量的负样本中采样出合适的样例将是一项值得展开的工作。

在负样本采样网络的设计上,采取了与关键帧采

样网络相似的结构,如图 6 所示,网络包含 3 个卷积层,两个池化层与全连接层。在每个卷积层之后,进行一次批标准化与 Relu 激活函数操作。卷积层的卷积核大小为  $5 \times 5$ ,每一层过后数据的尺寸如图所示。训练过程也与关键帧采样网络的训练过程相同,不同之处在于两个采样网络的动作值。关键帧采样网络中的动作值是当前视频里关键帧的编号,负样本采样网络中的动作值是下一个匹配亲属的编号。每一次筛选过程

中,根据父母视频去寻找与之匹配的孩子视频,再以该孩子的父母作为锚点匹配下一个负样本。评估负样本质量的评分函数与关键帧采样网络相似,首先用一个预先训练好的关键帧采样网络对亲属视频的关键帧进行采样,将关键帧匹配后输入亲属关系验证网络,得到损失函数的平均值就是负样本采样网络的评分函数  $S_2(x)$ 。 $Q$  值更新方式和损失函数与关键帧采样网络相同。

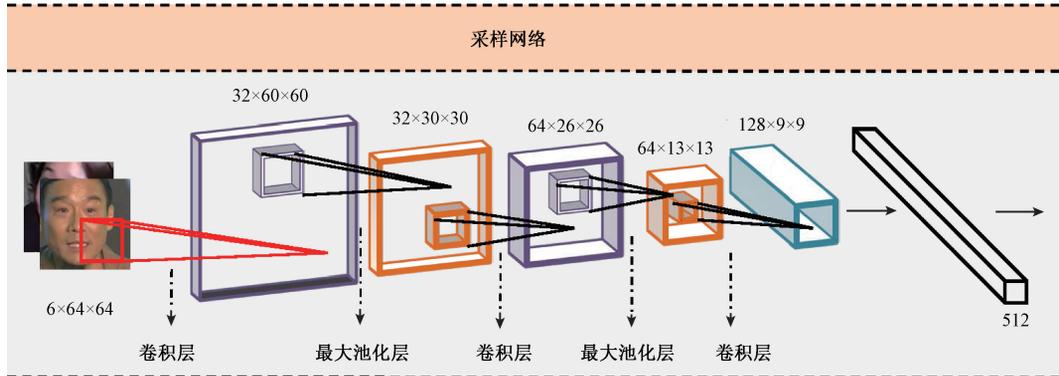


图 6 采样网络的网络结构

Fig. 6 Network structure of sampling network

### 3 实验结果与分析

#### 3.1 数据集

为了验证本文提出算法的有效性,实验所用的数据集为亲属视频数据集 KFWV<sup>[11]</sup>。该数据集中的数据主要来自于网络视频及电视节目,其建立目的为便于对基于视频的亲属关系进行验证研究。与静态图像相比,视频可以提供更多信息描述人脸,可以较为轻松地根据不同的人脸姿势、表情从不同光照条件和复杂背景中捕捉目标人脸。KFWV 数据集共收集了 418 对亲属视频,每个视频包含大约 100~500 帧,这些帧的姿势、光照、背景、遮挡、表情、妆容、年龄等变化很大。一个视频帧的平均大小约为  $900 \times 500$  像素。KFWV 数据集中包含 4 种亲属关系类型,分别是父子(FS)、父女(FD)、母子(MS)和母女(MD),分别有 107、101、100,以及 110 对亲属视频。

#### 3.2 实验设置

本文采取交叉验证的方式,每一次验证中,随机从每一类数据集中选取  $4/5$  的数据作为训练集,剩下  $1/5$  的数据作为测试集。这样的验证过程,重复 10 次,取平均值作为最终的验证结果。

亲属关系验证网络采用 Resnet50 的网络结构,学习率设置为 0.000 1,网络优化器选择 Adam。通过多种不同学习率的尝试和对比,可以发现,当学习率较低时,系

统将获得更好的性能,这与选择的优化器 Adam 有关。在训练过程的迭代中,每学习批量大小为 32 的数据后更新所有权重。由于亲属之间的特征不易提取,需要迭代多次以获得良好的性能,实验中完整的训练过程将遍历数据集 100 次。该参数通过多次实验进行设定,观察可得,当网络在训练 70~90 次的时候将趋于稳定,并得到最佳验证结果。为提升网络的训练性能,在数据增强方面,将图像大小从  $64 \times 64$  扩充到  $73 \times 73$ ,然后使用随机截取的方法裁剪出大小为  $64 \times 64$  的数据。除此之外,还添加了随机的水平翻转。

关键帧采样网络采取了深度强化学习的结构,网络部分包含 3 个卷积层,学习率设置为 0.000 1,优化器选择 Adam,同样每次向网络中输入批量大小为 32 的数据。在选择采样动作时,有 90% 的几率从网络输出中选择最佳项,10% 几率选择随机项。网络的向后递减系数  $\gamma$  为 0.9。目标网络和行为网络之间的参数更新频率为 100,即每进行 100 次训练行为更新一次网络参数。为了获得较好的实验效果,完整的训练过程将遍历数据集 200 次。

负样本采样网络的设置基本上与关键帧采样网络相同,仅在网络输出部分存在差异。负样本采样网络不会遍历整个数据集,而是在采样的负样本数量等于数据集大小的  $4/5$  时结束本轮采样,开始新一轮。这也是为了还原亲属关系验证网络训练时的采样过程。

### 3.3 结果与分析

由于目前在 KFWW 数据集上进行的工作并不多,所以本文主要与文献[11]进行对比,采取 EER 与 AUC 值作为评估标准。对比文献分别提取了 LBP 特征和 HOG 特征,其中 LBP 特征取得了更好的效果,作为本文的实验比对。表 1 为本文提出的方法与几种常用人脸亲属关系验证方法的对比结果。在本文提出的方法中,KVN 表示只使用亲属关系验证网络的验证结果,Final 表示提前用关键帧采样网络与负样本采样网络进行采样后得到的验证结果。两种方法的区别主要在采样方面,KVN 方法的采样方式为传统的均匀随机采样,而 Final 方法基于强化学习使用了判别性采样。可以看出,本文提出的方法在任何一种亲属关系下验证结果均优于以上所有方法。对比 KVN 与 Final 的实验结果可以看出,采样网络的加入实质性地提高了网络整体性能。

表 1 与其他方法间 EER 与 AUC 值的对比

Table 1 Comparison of EER and AUC

values with other methods (%)

Method	Measure	FD	FS	MD	MS	mean
Euclidean	EER	43.8	48.1	43.5	44.1	44.9
	AUC	60.5	56.0	57.8	58.9	58.3
ITML	EER	42.9	44.3	40.5	<b>42.7</b>	42.6
	AUC	59.1	56.8	61.5	<b>63.1</b>	60.1
SILD	EER	42.9	<b>42.9</b>	43.0	44.1	43.2
	AUC	62.6	<b>60.7</b>	58.5	59.0	60.2
KISSME	EER	40.0	44.8	43.5	42.7	42.8
	AUC	63.7	60.1	57.1	58.6	59.9
CSML	EER	<b>38.6</b>	47.1	<b>38.5</b>	43.2	<b>41.9</b>
	AUC	<b>66.2</b>	57.1	<b>64.4</b>	59.6	<b>61.8</b>
KVN	EER	37.1	35.9	35.0	36.8	36.2
	AUC	66.7	69.1	70.7	66.8	68.3
Final	EER	<b>37.0</b>	<b>34.9</b>	<b>33.2</b>	<b>36.1</b>	<b>35.3</b>
	AUC	<b>66.9</b>	<b>71.3</b>	<b>72.1</b>	<b>67.3</b>	<b>69.4</b>

KVN 方法和 Final 方法在 4 种亲属关系上取平均值的 ROC 曲线如图 7、8 所示。可以直观地看出,Final 方法相比起 KVN 方法 ROC 曲线下方的面积更大,也就是说它的整体性能更强。

通过几种亲属关系结果的横向对比可以看到,FS 和 MD 两种亲属关系的验证结果优于另外两种。分析可知,这两种亲属关系由于不存在性别差异,更容易获取亲属面部之间的相似性。

此外,文献[11]还进行了另一项测试,随机选择 20 对正样本视频和 20 对负样本视频,通过 10 个志愿者对数据集进行人为亲属关系验证。测试分为 A 和 B 两部分,测试 A 的亲属视频被裁剪为 64×64 的人脸区域,而测试 B 的亲属视频不进行任何处理,完整地呈现给测试者,测试结果如表 2 所示。将 KVN 方法与 Final 方法对

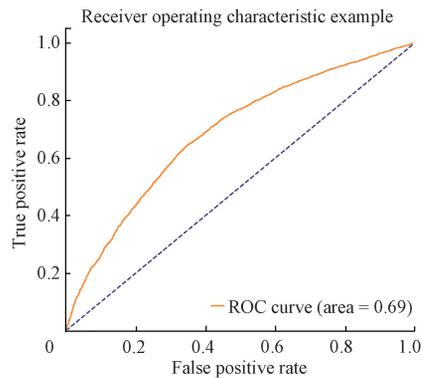


图 7 KVN 方法的 ROC 曲线

Fig. 7 ROC curve of KVN method

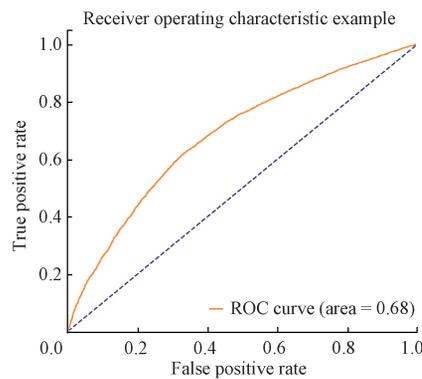


图 8 Final 方法的 ROC 曲线

Fig. 8 ROC curve of final method

比结果如表 2 所示。可以看到,TestB 的验证结果全面优于 TestA,而 KVN 方法全面优于 TestB,Final 方法则全面优于 KVN。

表 2 与人为验证方法之间识别率的对比

Table 2 Comparison of recognition rate with

human verification methods (%)

Method	FD	FS	MD	MS	mean
Test A	70.5	66.5	67.5	70.0	68.6
Test B	75.0	70.5	73.0	73.5	73.0
KVN	75.6	72.3	72.2	76.3	74.1
Final	<b>76.1</b>	<b>74.5</b>	<b>74.5</b>	<b>77.0</b>	<b>75.5</b>

首先,TestB 优于 TestA 说明完整的亲属视频更容易让人判断出亲属关系,因为背景、头发等因素都可以提供一定量的信息。目前本文采取的验证方法与 TestA 类似,输入数据为检测的人脸区域。如果可以有效地利用完整的视频,验证结果也许会进一步提升。本文提出的方法得到的验证结果全面优于 TestB,证明了本文设计的系统在基于视频的亲属关系验证问题上的有效性和实用性。Final 实验中的结果全面优于 KVN,更加说明了本文

设计的系统在基于视频的亲属关系验证问题上的有效性和实用性。

## 4 结 论

本文提出了一种使用判别采样的人脸亲属关系验证的方法,通过深度强化学习实现人脸视频中判别信息的提取与特征表示,提升了验证性能。首先,设计了具备残差结构的深度卷积网络来提取亲属之间的人脸特征,实现亲属关系分类。为了进一步提高验证性能,基于深度强化学习设计了关键帧采样网络和负样本采样网络,用于实现视频数据关键帧和负样本的挑选。实验结果证明该采样网络确实可以提高亲属关系验证的准确性。与其他基于视频的亲属关系验证方法进行的对比也证明了本文所提出方法的实用性和有效性。实验中仍有许多值得进一步探索和研究的,未来的工作将围绕如何更好处理关键帧和困难负样本的采样开展,进一步提升人脸亲属关系的验证率。

## 参考文献

- [ 1 ] SONG Z, NI B, GUO D, et al. Learning universal multi-view age estimator using video context [ C ]. 2011 International Conference on Computer Vision, 2011: 241-248.
- [ 2 ] RAO Y, LU J, LIN J, et al. Runtime network routing for efficient image classification [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(10): 2291-2304.
- [ 3 ] ROTHE R, TIMOFTE R, VAN GOOL L. Deep expectation of real and apparent age from a single image without facial landmarks [ J ]. International Journal of Computer Vision, 2018, 126(2-4): 144-157.
- [ 4 ] LU J, HU J, TAN Y P. Discriminative deep metric learning for face and kinship verification [ J ]. IEEE Transactions on Image Processing, 2017, 26 ( 9 ): 4269-4282.
- [ 5 ] LI W, LU J, FENG J, et al. Bridgenet: A continuity-aware probabilistic network for age estimation [ C ]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1145-1154.
- [ 6 ] YAN H. Collaborative discriminative multi-metric learning for facial expression recognition in video [ J ]. Pattern Recognition, 2018, 75: 33-40.
- [ 7 ] ZHANG Z, LUO P, LOY C C, et al. Learning social relation traits from face images [ C ]. IEEE International Conference on Computer Vision, 2015: 3631-3639.
- [ 8 ] YAN H, HU J. Video-based kinship verification using distance metric learning [ J ]. Pattern Recognition, 2018, 75: 15-24.
- [ 9 ] YAN H, SONG C. Multi-scale deep relational reasoning for facial kinship verification [ J ]. Pattern Recognition, 2021, 110: 107541.
- [ 10 ] LU J, HU J, TAN Y P. Discriminative deep metric learning for face and kinship verification [ J ]. IEEE Transactions on Image Processing, 2017, 26 ( 9 ): 4269-4282.
- [ 11 ] LU J, LIONG V E, ZHOU X, et al. Learning compact binary face descriptor for face recognition [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2041-2056.
- [ 12 ] 韩超,方露,章盛. 一种优化的图像配准算法 [ J ]. 电子测量与仪器学报, 2017, 31(2): 178-184.
- HAN C, FANG L, ZHANG SH. An optimized image registration algorithm [ J ]. Journal of Electronic Measurement and Instrumentation, 2017, 31 ( 2 ): 178-184.
- [ 13 ] 刘敏,陈伟迅,龚蓉,等. 三维图像栈中神经末梢点的自动检测 [ J ]. 电子测量与仪器学报, 2017, 31(8): 1304-1311.
- LIU M, CHEN W, GONG R, et al. Automatic neuron terminal point detection in 3D image stack [ J ]. Journal of Electronic Measurement and Instrumentation, 2017, 31(8): 1304-1311.
- [ 14 ] FANG R, TANG K D, SNAVELY N, et al. Towards computational models of kinship verification [ C ]. 2010 IEEE International Conference on Image Processing, 2010: 1577-1580.
- [ 15 ] ZHOU X, LU J, HU J, et al. Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments [ C ]. 20th ACM International Conference on Multimedia, 2012: 725-728.
- [ 16 ] KOHLI N, SINGH R, VATSA M. Self-similarity representation of weber faces for kinship classification [ C ]. IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems ( BTAS ), 2012: 245-250.

## 作者简介



王仕伟, 2018 年于北京邮电大学获得学士学位, 现为北京邮电大学硕士研究生, 主要研究方向为服务机器人和机器视觉。  
E-mail: 393211032@qq.com

**Wang Shiwei** received his B. Sc. degree from Beijing University of Posts and Telecommunications in 2018. Now he is a M. Sc. candidate at Beijing University of Posts and Telecommunications. His main research interests include service robot and machine vision.



闫海滨(通信作者),分别在 2004 年和 2007 年于西安理工大学获得学士学位和硕士学位,2013 年于新加坡国立大学获得博士学位,现为北京邮电大学副教授,博士生导师,主要研究方向为服务机器人与机器视觉。

E-mail: eyanhaibin@bupt.edu.cn

**Yan Haibin** (Corresponding author) received the B. Sc. and M. Sc. from Xi'an University of Technology in 2004 and 2007, and Ph. D. degree from National University of Singapore in 2013, respectively. Now she is an associate professor and Ph. D. supervisor at Beijing University of Posts and Telecommunications. Her main research interests include service robot and machine vision.