



Face re-identification challenge: Are face recognition models good enough?[☆]

Zhiyi Cheng^{a,*}, Xiatian Zhu^b, Shaogang Gong^a

^aQueen Mary University of London, London, UK

^bVision Semantics Limited, London, UK

ARTICLE INFO

Article history:

Received 19 October 2019

Revised 7 April 2020

Accepted 4 May 2020

Available online 16 May 2020

Keywords:

Face re-identification

Surveillance facial imagery

Low-resolution

Super-resolution

Open-set matching

Deep learning

Face recognition

ABSTRACT

Face re-identification (Re-ID) aims to track the same individuals over space and time with subtle identity class information in automatically detected face images captured by unconstrained surveillance camera views. Despite significant advances of face recognition systems for *constrained* social media facial images, face Re-ID is more challenging due to poor-quality surveillance face imagery data and remains under-studied. However, solving this problem enables a wide range of practical applications, ranging from law enforcement and information security to business, entertainment and e-commerce. To facilitate more studies on face Re-ID towards practical and robust solutions, a *true* large scale Surveillance Face Re-ID benchmark (*SurvFace*) is introduced, characterised by natively low-resolution, motion blur, uncontrolled poses, varying occlusion, poor illumination, and background clutters. This new benchmark is the largest and more importantly the only true surveillance face Re-ID dataset to our best knowledge, where facial images are captured and detected under realistic surveillance scenarios. We show that the current state-of-the-art FR methods are *surprisingly poor* for face Re-ID. Besides, face Re-ID is generally more difficult in an open-set setting as naturally required in surveillance scenarios, owing to a large number of non-target people (distractors) appearing in open ended scenes. Moreover, the low-resolution problem inherent to surveillance facial imagery is investigated. Finally, we discuss open research problems that need to be solved in order to overcome the under-studied face Re-ID problem.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

With the rapid expansion of surveillance multi-camera systems around the world, associating people over space and time becomes an increasingly significant capability for a wide range of applications such as public safety, law enforcement and forensic search [1]. Among the existing visual biometrics for person identity recognition, such as whole-body [2], iris [3], gait [4], and fingerprint [5], facial appearance is considered as one of the most convenient and most reliable non-intrusive visual cues. This is due to one fact that faces, provided they are visible in captured images, are more stable cues for long-term tracking and tracing, whereas other visual appearances, e.g. clothes for whole-body Re-ID [2,6], are easier to change over space and time. In this study, we focus on the

task of tracking people across distributed non-overlapped camera views without any domain prior knowledge, by facial images alone captured under unconstrained surveillance conditions, i.e. *face re-identification* (face Re-ID) [7].

Face recognition (FR) has been extensively studied with significant advance in the literature, and FR based commercial products are increasingly appearing in our daily life, e.g. web photo-album and online e-payment. However, this survey shows that current FR methods generalise poorly to face Re-ID task, given realistic noisy and low-quality facial images captured by *unconstrained wide-field surveillance cameras*, far away from being satisfactory. This is due to low-resolution facial imagery with unconstrained noise, pose, expression, occlusion, lighting and background clutter (Fig. 1).

While being critical for public safety and law enforcement applications, the face Re-ID problem is significantly under-studied in comparison to face recognition. A major reason is lacking a large scale surveillance face Re-ID benchmark, as opposite to the rich availability of high-resolution web photostock face recognition benchmarks (Table 1). For example, there are 4,753,320 web

[☆] This work was supported by Vision Semantics Limited, the Innovate UK Industrial Challenge Project on Developing and Commercialising Intelligent Video Analytics Solutions for Public Safety (98111–571149), the Alan Turing Institute Turing Fellowship, and the Royal Society Newton Advanced Fellowship Programme (NA150459).

* Corresponding author.

E-mail address: z.cheng@qmul.ac.uk (Z. Cheng).



Fig. 1. Example comparisons of (Left) web face images from standard face datasets and (Right) native surveillance face images from typical real-world public surveillance scenes.

Table 1

Statistics of representative publicly available face benchmarks. Celeb: Celebrity.

Challenge	Year	IDs	Images	Videos	Subject	Surv?
Yale [25]	1997	15	165	0	Cooperative	No
QMUL-MultiView [23]	1998	25	4450	5	Cooperative	No
XM2VTS [26]	1999	295	0	1180	Cooperative	No
Yale B [27]	2001	10	5760	0	Cooperative	No
CMU PIE [28]	2002	68	41,368	0	Cooperative	No
Multi-PIE [29]	2010	337	750,000	0	Cooperative	No
Morph [36]	2006	13,618	55,134	0	Celeb (Web)	No
LFW [31]	2007	5749	13,233	0	Celeb (Web)	No
YouTube [37]	2011	1595	0	3425	Celeb (Web)	No
WDRRef [38]	2012	2995	99,773	0	Celeb (Web)	No
FaceScrub [39]	2014	530	100,000	0	Celeb (Web)	No
CASIA [32]	2014	10,575	494,414	0	Celeb (Web)	No
CelebFaces [33]	2014	10,177	202,599	0	Celeb (Web)	No
IJB-A [40]	2015	500	5712	2085	Celeb (Web)	No
VGGFace [11]	2015	2622	2.6M	0	Celeb (Web)	No
UMDFaces [41]	2016	8277	367,888	0	Celeb (Web)	No
MS-Celeb-1M [34]	2016	99,892	8,456,240	0	Celeb (Web)	No
UMDFaces-Videos [42]	2017	3107	0	22,075	Celeb (Web)	No
IJB-B [43]	2017	1845	11,754	7011	Celeb (Web)	No
VGGFace2 [44]	2017	9131	3.31M	0	Celeb (Web)	No
MegaFace2 [8]	2017	672,057	4,753,320	0	Non-Celeb (Web)	No
FERET [7]	1996	1199	14,126	0	Cooperative	No
FRGC [45]	2004	466+	50,000+	0	Cooperative	No
CAS-PEAL [46]	2008	1040	99,594	0	Cooperative	No
PaSC [47]	2013	293	9376	2802	Cooperative	No
SCface [48]	2011	130	4160	0	Cooperative	Yes
COX [49]	2015	1000	1000	3000	Cooperative	Yes
EBOLO [50]	2016	Unknown	6,135	0	Cooperative	Yes
FaceSurv [51]	2019	252	0	460	Cooperative	Yes
UCCS [9]	2017	1732	14,016+	0	Uncooperative	Yes
SurvFace	2019	15,573	463,507	0	Uncooperative	Yes

face images from 672,057 face IDs in MegaFace2¹ [8], which is made possible by easier collection and labelling of large scale facial images in the public domain from the Internet. On the contrary, it is prohibitively expensive and less feasible to construct large scale *native* (i.e. non-simulated) surveillance facial imagery data as a benchmark for wider studies, due both to largely restricted data access and very tedious data labelling at high costs. Currently, the largest surveillance face dataset is the UnConstrained College Students (UCCS) dataset² [9], which contains 100,000 face images from 1732 face IDs, at a significantly smaller scale than the MegaFace celebrities photoshoot dataset. However, the UCCS is limited and only *semi-native* due to being captured in a man-made, simulated surveillance setup with a high-resolution camera at a single location. In this study, we show that: (1) The state-of-the-art FR models trained on large scale high-quality benchmark datasets such as the MegaFace generalise poorly to face Re-ID task on native low-quality surveillance facial images; (2) The performance of face Re-ID on artificially synthesised low-resolution images does not well reflect the true challenges of native surveillance facial images in system deployments; (3) The image super-resolution models suffer from the lack of pixel-aligned low- and high-resolution

surveillance image pairs which are necessary for model training, apart from the domain distribution shift between web and surveillance data. To facilitate solving the aforementioned problems and limitations, we introduce a realistic and large scale *Surveillance Face Re-ID Challenge*, where a model is expected to associate people in the multi-camera systems, by surveillance facial images taken from unconstrained public scenes.

We make three contributions: (I) We construct a large scale face Re-ID benchmark with *native* surveillance facial imagery data for enabling scalable model development and evaluation. Specifically, we introduce the *SurvFace* challenge, containing 463,507 face images of 15,573 unique facial identities³. To our best knowledge, this is the largest and only dataset for *native* face Re-ID challenge. *SurvFace* is constructed by data-mining 17 public domain person re-identification datasets (Table 2) using a deep face detection model, so to assemble a large pool of labelled surveillance face images in a *cross-problem* data re-purposing principle. The unique features of the proposed face Re-ID benchmark, compared with the conventional FR datasets, are the provision of cross-location (cross

¹ <http://megaface.cs.washington.edu/>.

² <http://vast.uccs.edu/OpenSetface/>.

³ The *SurvFace* benchmark has been used for the challenge track of an IEEE ICCV 2019 workshop entitled *Real-World Recognition from Low-Quality Images and Videos*. See the details at: <https://www.foriq.org/> and <https://evalai.cloudcv.org/web/challenges/challenge-page/392/overview>.

Table 2
Person re-identification datasets utilised in constructing the *SurvFace* challenge.

Person Re-Identification Dataset	IDs	Detected IDs	Bodies	Detected Faces	Nation
Shinpuhkan [91]	24	24	22,504	6883	Japan
WARD [92]	30	11	1436	390	Italy
RAiD [93]	43	43	6920	3724	US
CAVIAR4ReID [94]	50	43	1221	141	Portugal
SARC3D [95]	50	49	200	107	Italy
ETHZ [96]	148	110	8580	2681	Switzerland
3DPeS [97]	192	133	1012	366	Italy
QMUL-GRID [98]	250	242	1275	287	UK
iLIDS-VID [85]	300	280	43,800	14,181	UK
SDU-VID [99]	300	300	79,058	67,988	China
PRID 450S [100]	450	34	900	34	Austria
ViPeR [101]	632	456	1264	532	US
CUHK03 [102]	1467	1380	28,192	7911	China
Market-1501 [103]	1501	1429	25,261	9734	China
Duke4ReID [104]	1852	1690	46,261	17,575	US
CUHK-SYSU [105]	8351	6694	22,724	12,526	China
LPW [106]	4584	2655	590,547	318,447	China
Total	20,224	15,573	881,065	463,507	Multiple

camera views) ID label annotations, and the more realistic open-set evaluation protocol in typical surveillance scenarios. **(II)** While showing increasing generalisation to more unconstrained identity matching scenarios, existing FR models have not been tested for large scale face Re-ID in surveillance scenarios. We fill this gap by benchmarking representative deep learning FR models [10–12] on the *SurvFace* challenge. They are particularly evaluated in a more realistic *open-set* scenario, originally missing in the previous studies. In contrast to the more common *closed-set* setting, the open-set test considers the cases of *no* true-matches of a probe in the gallery, respecting the realistic large surveillance search scenarios. **(III)** We investigate extensively the performance of existing models on *SurvFace* by exploiting simultaneously image super-resolution (SR) [13–17] and FR models. We further compare the model performances on MegaFace and UCCS benchmarks to give better understanding of the unique characteristics of *SurvFace*. We finally provide extensive discussions on future research directions for face Re-ID.

2. Related work

We review representative face challenges (Section 2.1) and methods (Section 2.2), and existing face Re-ID systems (Section 2.4) in the literature. More general and extensive reviews can be found in other surveys [18–20] and books [21–24].

2.1. Face recognition challenges

An overview of representative face challenges and benchmarks are summarised in Table 1. Early challenges focus on *small-scale constrained* scenarios [3,25–30], with neither sufficient appearance variation for robust model training, nor practically solid test benchmarks. The seminal LFW [31] started to shift the community towards recognising unconstrained web faces, followed by even larger face benchmarks, such as CASIA [32], CelebFaces [33], VG-GFace [11], MS-Celeb-1M [34], MegaFace [35] and MegaFace2 [8].

With such large benchmarks, FR accuracy in good quality images has reached an unprecedented level, e.g., 99.83% on LFW and 99.80% on MegaFace. However, this does not scale to *native* surveillance faces captured in unconstrained camera views (Section 4.1), due to: (1) Existing datasets have varying degrees of data selection bias (near-frontal pose, less blur, good illumination); and (2) Deep methods are often domain-specific (only generalise well to test data similar to training set). On the other hand, there is a gap of facial images quality between a web photostock view and a surveillance view in-the-wild (Fig. 1).

Research on face Re-ID has slightly advanced since 1996 when the well-known FERET was launched [7]. It is under-studied with a very few benchmarks available. One of the major obstacles is the difficulty of establishing a large scale surveillance face dataset due to the high cost and limited feasibility in collecting surveillance faces and exhaustive ID annotation. Even in the FERET dataset, only simulated (framed) surveillance faces were collected in most cases with carefully controlled imaging settings, therefore it provides a much better facial image quality than those from native surveillance videos.

A notable recent study introduces the UCCS challenge [9], the current largest public surveillance face dataset, where faces were captured from a long-range distance without subjects' cooperation (unconstrained), with various poses, blurriness and occlusion (Fig. 5(b)). This benchmark represents a relatively realistic surveillance scenario compared to FERET. However, the UCCS images were captured at high-resolution from a single camera view⁴, therefore providing significantly more facial details and less viewing angle variations. Moreover, UCCS is small in size, particularly the ID numbers (1,732), statistically limited for face Re-ID evaluation (Section 4.1). This study addresses such limitations by constructing a larger scale native surveillance face Re-ID challenge, the *SurvFace* benchmark. It consists of 463,507 real-world surveillance face images of 15,573 different IDs captured from a diverse source of public spaces (Section 3).

2.2. Face recognition methods

We provide a brief review on the existing FR algorithms, including models specially designed for low-resolution faces. We also discuss super-resolution models for image fidelity and discriminability enhancement.

2.2.1. Face recognition models

Early FR methods adopt hand-crafted features (e.g. Color Histogram, LBP, SIFT, Gabor) and matching model learning (e.g. discriminative margin mining, subspace learning, dictionary based sparse coding, Bayesian modelling) [25,38,52–55]. They suffer from sub-optimal recognition generalisation, particularly with significant facial appearance variations, due to weak representation power (limited and incomplete human domain knowledge for hand-

⁴ A single Canon 7D camera equipped with a Sigma 800mm F5.6 EX APO DG HSM lens.

crafted features) and lack of end-to-end interaction learning between feature extraction and model inference.

Recently, deep learning based FR models [10–12,40,56–60] have achieved remarkable success. This paradigm benefits from superior network architectures [61–64] and optimisation algorithms [10,33,57]. Deep FR methods naturally address the limitations of hand-crafted alternatives by jointly learning face representation and matching model end-to-end. A large set of labelled face images is usually necessary to train the millions of parameters of deep models. This can be commonly satisfied by large scale web face data collected and labelled (filtered) from Internet. Consequently, modern FR models are often trained, evaluated and deployed on web face datasets (Table 1).

Despite advances in web FR, it remains unclear how well the state-of-the-art methods generalise to surveillance faces. Intuitively, face Re-ID is extreme challenging due to three reasons: (1) Surveillance faces contain much less appearance details with poorer quality and lower resolution (Fig. 1). (2) Deep models are highly domain-specific and likely yield big performance degradation in cross-domain deployments, especially with large train-test domain gap, e.g. web and surveillance faces. In such cases, transfer learning is challenging [65]. The scarcity of labelled surveillance data makes the problem even more challenging. (3) Instead of closed-set search considered by most existing methods, face Re-ID is intrinsically open-set where the probe face ID is not necessarily presented in the gallery. It brings about a significant challenge by additionally requiring the system to reject non-target (distractors) whilst not missing target IDs, especially when the distractors are of arbitrary variety.

2.2.2. Recognising low-resolution faces

An inherent challenge of face Re-ID is rooted in low-resolution [18]. Generally, existing low-resolution FR methods fall into two categories: (1) image super-resolution [66–70], and (2) resolution-invariant learning [71–75]. The first category is based on two learning criteria: pixel-level visual fidelity and ID discrimination. Existing models often focus more on appearance enhancement [66,67]. Recent studies [68–70] attempt to unite the two sub-tasks for more discriminative learning. The second category aims to learn resolution-invariant features [71,72] or a cross-resolution structure transformation [73–75]. The data-driven deep models can be conceptually categorised into this strategy whenever suitable training data is available for model optimisation.

However, all the existing methods have a number of limitations: (1) Considering small scale and/or artificial low-resolution face images in the closed-set setting, therefore unable to reflect the genuine face Re-ID challenge at scales. (2) Relying on hand-crafted features and linear/shallow model structures with suboptimal generalisation. (3) Requiring pixel-aligned low- and high-resolution training image pairs, which are unavailable for surveillance faces.

2.3. Image super-resolution

SR methods have significantly advanced thanks to the strong capacity of deep models in regressing the pixel-wise loss between reconstructed and ground-truth images [13,15–17,76,77]. Mostly, FR and SR researches advance independently, both assuming the availability of large high-resolution training data. In surveillance, high-resolution images are typically unavailable, which in turn resorts existing methods to transfer learning. When the distributions of training and test data are very different, SR becomes extremely challenging due to an extra need for domain adaptation.

Specially, face super-resolution (face hallucination) is dedicated for facial appearance restoration [78–82]. A common approach is to transfer high-frequency details and structure information

from exemplar high-resolution images, by mapping low- and high-resolution training pairs. Existing models require noise-free inputs, assuming stringent part detection and dense correspondence alignment, or may introduce overwhelming artifacts. Such assumptions significantly limit their usability to surveillance faces with uncontrolled noise and the absence of paired high-resolution images.

2.4. Person re-identification

Existing person Re-ID methods [83–87] assume that the whole-body visual appearance are stationary [2]. This significantly limits their scalability for long-range identity tracking over space and time, since the clothing and associated objects can change easily. In contrast, the facial appearance is intrinsically much more stable therefore providing reliable representations and evidences for large scale forensic search.

2.5. Surveillance face re-identification

Surveillance face Re-ID remains under-studied in the literature, with very limited dedicated attempts. Dantcheva et al. [88] constructed facial representations from patches of hair, skin and clothes, for frontal-to-profile faces matching in video surveillance systems. Farinella et al. [89] adopted Local Ternary Patterns as representation for Re-ID task. These models, relying on hand-crafted facial representations and shallow recognition models, have limited generalisation power to large scale realistic surveillance data. Li et al. [50] explored facial information in person Re-ID by showing face as a more reliable biometric for long-term tracking with body clothes changes over time. This work was conducted on a very small dataset with ideal front-view person images captured in constrained scenarios. A more recent work [90] adopted deep model for representation learning and clustering for ID recognition, which however is also designed for constrained face images. Generally, the advance of face Re-ID systems is largely limited by lacking of a large, unconstrained, and realistic benchmark. We in this work introduce a large scale surveillance dataset for face Re-ID, and conduct an extensive set of experiments to test the performance of the state-of-the-art face recognition and image super-resolution methods. The empirical results reveal that these methods are surprisingly limited in face Re-ID. This phenomenon is thought provoking, making the researchers to reevaluate existing algorithms and motivating them to develop practically effective solutions.

3. Face re-ID challenge

3.1. A native surveillance face dataset

To our best knowledge, there is no large native surveillance face Re-ID challenge in the public domain. To stimulate the research on this problem, we construct a new large scale benchmark (challenge) by extracting faces of the uncooperative general public appearing in real-world surveillance videos and images. We call this challenge **SurvFace**. Unlike most existing FR challenges using either high-quality web or simulated surveillance images captured in controlled conditions therefore *failing* to evaluate the true surveillance face Re-ID performance, we explore real-world native surveillance imagery from a combination of 17 person re-identification benchmarks which were collected in different surveillance scenarios across diverse sites and multiple countries (Table 2).

3.1.1. Dataset statistics

The **SurvFace** challenge contains 463,507 face images of 15,573 unique person IDs with uncontrolled appearance variations in pose, illumination, motion blur, occlusion and background (Fig. 2).



Fig. 2. Matched (Left) and unmatched (Right) face image pairs from SurvFace.

Among all, there are 10,638 (68.3%) people each associated with ≥ 2 detected face images. This is the largest native surveillance face benchmark to date (Table 1).

3.1.2. Faces collection

We automatically extracted faces with TinyFace detector [107] in re-identification surveillance images. Manually labelling is not scalable due to the huge amount of surveillance video data. Note that not all faces in source images can be successfully detected given imperfect detection, poor image quality and extreme head poses. The average detection recall is 77.0% (15,573 out of 20,224) in ID and 52.6% (463,507 out of 881,065) in image. Table 2 summarises face detection statistics across all person re-identification datasets.

3.1.3. Face image cleaning and annotation

We manually cleaned SurvFace data by filtering out false detections, with two independent annotators and a subsequent mutual cross-check. All non-surveillance images in CUHK-SYSU dataset are thrown away. For face ID annotation, we used the person labels available in sources assuming no ID overlap across datasets. This is rational since they were independently created over different time and surveillance venues, i.e., the possibility that a person appears in multiple source datasets is extremely low.

3.1.4. Face characteristics

In contrast to existing face datasets, SurvFace is uniquely characterised by low resolution typical in surveillance (Fig. 4) – one major source making face Re-ID challenging. The bi-modal distribution in resolution sources from the 17 independent person Re-ID datasets we used collectively, mainly due to that CUHK03 provides relatively higher-resolution images. The face spatial resolution ranges from 6/5 to 124/106 pixels in height/width, with average 24/20. It exhibits a power-law distribution in frequency ranging from 1 to 558 (Fig. 3).

3.2. Evaluation protocols

3.2.1. Data partition

We first split the SurvFace data into training and test sets. We divide the 10,638 IDs each with ≥ 2 face images into two halves: one half (5,319) for training, one half (5,319) plus the remaining 4935 single-shot IDs (in total 10,254) for test (Table 3). We benchmark only one train/test data split since the dataset is sufficiently large to support a statistically stable evaluation. All face images of training IDs are used for models training. Additional imagery from other sources may be used subject to no facial images of test IDs.

3.2.2. Closed-set

We first set up the closed-set face Re-ID evaluation on SurvFace. For each of the 5319 multi-shot test IDs, we randomly sample the corresponding images into probe or gallery. The gallery set

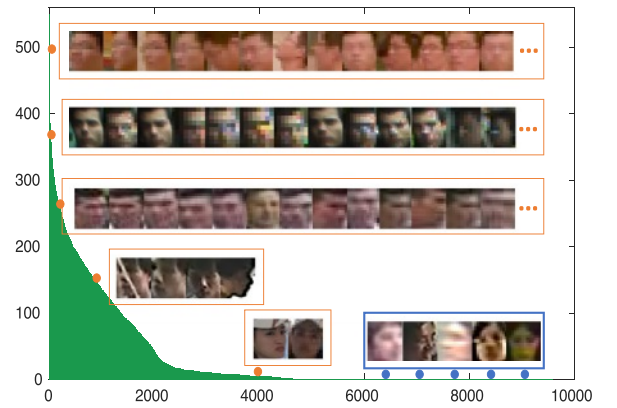


Fig. 3. Image frequency over all SurvFace IDs.

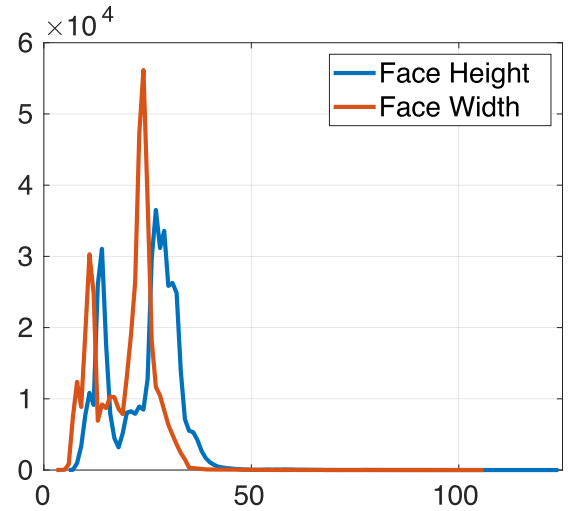


Fig. 4. Scale distributions of SurvFace images.

Table 3

Statistics of SurvFace. Numbers in parentheses: per-identity image number range.

Split	All	Training	Test
IDs	15,573	5319	10,254
Images	463,507 (1 ~ 558)	220,890 (2 ~ 558)	242,617 (1 ~ 482)

represents imagery involved in an operational database, e.g., access control system's repository. For any unique person, we generate a single ID-specific face template from one or multiple gallery images [40]. This makes the ranking list concise and more efficient for post-rank manual validation, e.g., no case that a single ID takes

multiple ranks. The probe set represents imagery used to query a face Re-ID system.

The Cumulative Matching Characteristic (CMC) [40] measure is selected for *closed-set* face Re-ID. CMC reports the fraction of searches returning the mate (true match) at rank r or better, with the rank-1 rate as the most common summary indicator of an algorithm's efficacy. It is a non-threshold rank based metric. Formally, the CMC at rank r is defined as:

$$\text{CMC}(r) = \sum_{i=1}^r \frac{N_{\text{mate}}(i)}{N} \quad (1)$$

where $N_{\text{mate}}(i)$ denotes the number of probe images with the mate ranked at position i , and N the total probe number.

3.2.3. Open-set

In realistic surveillance applications, however, most faces captured by CCTV cameras are not of any gallery ID therefore should be detected as unknown, leading to the *open-set* protocol [108,109]. It is often referred to the *watch-list identification* (forensic search) scenario where only persons of interest are enrolled into the gallery, typically each ID with several different images such as the FBI's most wanted list⁵. To enable the *open-set* face Re-ID test, we construct a watch list identification protocol where only IDs of interest are enrolled in the gallery. Specifically, we create the probe and gallery sets as: (1) Out of the 5319 multi-shot test IDs, we randomly select 3000 and sample half face images for each selected ID into the gallery set, i.e. the watch list. (2) All the remaining images including single-shot ID imagery are used to form the probe set. As such, the majority of probe people are *unknown* (not enrolled gallery IDs), more accurately reflecting the open space forensic search nature.

For the *open-set* face Re-ID evaluation, we must quantify two error types [108]. The first type is *false alarm* – a face image from an unknown person (i.e. nonmate search) is incorrectly associated with one or more enrollees' data. This error is quantified by the *False Positive Identification Rate* (FPIR):

$$\text{FPIR}(t) = \frac{N_{\text{nm}}^m}{N_{\text{nm}}} \quad (2)$$

which measures the proportion of nonmate searches N_{nm}^m (of no mate IDs in the gallery) that produce one or more enrolled candidates at or above a threshold t (i.e. false alarm), among a total of N_{nm} nonmate searches attempted.

The second type of error is *miss* – a search of an enrolled target person's data (i.e. mate search) does not return the correct ID. We quantify this by the *False Negative Identification Rate* (FNIR):

$$\text{FNIR}(t, r) = \frac{N_{\text{m}}^{\text{nm}}}{N_{\text{m}}} \quad (3)$$

which is the proportion of mate searches N_{m}^{nm} (of IDs in the gallery) with enrolled mate found outside top r ranks or matching similarity score below the threshold t , among N_{m} mate searches. By default, we set $r=20$ ($\text{FNIR}(t, 20)$) assuming small workloads by human reviewers employed to review the candidates returned from identification searches [108]. In practice, a more intuitive measure may be the "hit rate" or *True Positive Identification Rate* (TPIR):

$$\text{TPIR}(t, r) = 1 - \text{FNIR}(t, r) \quad (4)$$

which is the complement of FNIR offering a positive statement of how often mated searches are succeeded. $\text{TPIR}(t, 1)$ corresponds to the Detection and Identification Rate (DIR) as defined in [110]. In SurvFace, we adopt the TPIR@FPIR measure as the open-set face Re-ID performance metrics. TPIR -vs- FPIR can similarly generate an ROC curve, the AUC of which stands for an overall measurement.

Table 4
Benchmark data partition of SurvFace.

Scenario	Open-Set	
Partition	Probe	Gallery
IDs	10,254	3,000
Images	182,323	60,294
Metrics	TPIR@FPIR, ROC	

3.2.4. Considerations

Existing FR challenges often adopt the closed-set evaluation protocol [35]. While being able to evaluate FR model in large scale search, it does not fully generalise to face Re-ID. For face Re-ID, human operators are often assigned with a list of target IDs with face images enrolled in gallery. The task is then to search the faces of target IDs across camera views. This is an open-set scenario. Therefore, we adopt the open-set protocol as the main setting of SurvFace (Table 4). Besides, we still consider closed-set experiments to enable like-for-like comparisons with existing benchmarks.

4. Experimental evaluations

We present the experimental evaluations of face Re-ID, with top FR models on both native surveillance faces (Section 4.1) and super-resolved faces (Section 4.2). We choose three representative FR models, CentreFace [10], VggFace [11] and SphereFace [12], for native face Re-ID benchmarking; and three image super-resolution methods, SRCNN [13], VDSR [16] and LapSRN [15], to evaluate the face Re-ID performance on super-resolved surveillance faces. The feature vectors are L_2 normalised before face Re-ID matching. This is equivalent to using cosine similarity.

4.1. Native surveillance face re-ID

We evaluated face Re-ID on the *native* SurvFace images. Besides the low-resolution issue, there are other uncontrolled covariates, e.g. illumination variations, expression, occlusions, background clutter, and compression artifacts. All of these factors cause inference uncertainty to varying degrees (Fig. 2).

4.1.1. Model training and test

We adopted three strategies for model training: (1) Only using SurvFace training set (220,890 images from 5319 IDs). (2) Only using CASIA web data (494,414 images from 10,575 IDs). We will test the effect of using different web source datasets such as MegaFace2 [8] and MS-Celeb-1M [34]. (3) First pre-training a FR model on CASIA, then fine-tuning on SurvFace (default strategy). The trained model is deployed with Euclidean distance. In both training and test, we rescaled facial images by *bicubic* interpolation to the required input model size. Note that such interpolation process does not change the underlying resolution, i.e. the visual information.

4.1.2. Evaluation settings

We considered both closed-set and open-set scenarios. By default, we adopt the more realistic open-set evaluation, unless stated otherwise. For open-set, we used TPIR (Eqn. (3)) at varying FPIR rates (Eqn. (2)). The true match ranked in top- r ($r=20$ in Eqn. (4)) is considered as success.

4.1.3. Implementation details

We used the codes released by the original authors for models implementation [10–12]. Throughout the experiments, we adopted the suggested parameter setting by the authors if available, or

⁵ <http://www.fbi.gov/wanted>.

Table 5

Face Re-ID results on *SurvFace*. Protocol: Open-Set. Metrics: TP1R20@FPIR ($r=20$) and AUC. “-”: No results available due to failure of model convergence.

Train Data	SurvFace (Ours)					CASIA [32]					CASIA + SurvFace				
	Metrics					Metrics					Metrics				
	TP1R20(%)@FPIR					AUC(%)					TP1R20(%)@FPIR				
	30%	20%	10%	1%		30%	20%	10%	1%		30%	20%	10%	1%	
CentreFace	26.2	20.0	12.2	2.8	34.6	5.7	4.4	2.3	0.2	7.6	27.3	21.0	13.8	3.1	37.3
VggFace	-	-	-	-	-	6.5	4.8	2.5	0.2	9.6	5.1	2.6	0.8	0.1	14.0
SphereFace	18.8	13.5	7.0	0.7	26.6	5.9	4.2	2.2	1.7	9.0	21.3	15.7	8.3	1.0	28.1



(a) SurvFace

(b) UCCS

(c) CASIA

Fig. 5. Quality comparison of example faces from (a) SurvFace, (b) UCCS, and (c) CASIA.**Table 6**

Open-Set (TP1R20(%)@FPIR) vs Closed-Set (CMC (%)) on *SurvFace*.

Metrics	TP1R20 FPIR			CMC		
	30%	20%	10%	Rank-1	Rank-10	Rank-20
CentreFace	27.3	21.0	13.8	29.9	53.4	61.1
SphereFace	21.3	15.7	8.3	29.3	50.0	55.4

carefully tuned the hyper-parameters by grid search. Data augmentation was applied to *SurvFace* training data, including flipping, Gaussian kernel blurring, colour shift, brightness and contrast adjustment. We excluded cropping and rotation transformation which bring negative influence due to tight face bounding boxes.

4.1.4. Face re-ID evaluation

(I) Benchmark SurvFace. We benchmarked face Re-ID on *SurvFace* in Table 5. We make four observations: (1) Not all FR models converge when directly training on *SurvFace*, e.g. VggFace fails. As opposite, all the models are well trained using CASIA data. Whilst CASIA is larger, we conjecture that the scale is not a key obstacle as *SurvFace* training data should be arguably sufficient for generic deep learning. Instead this may be more due to extreme challenges posed by poor resolution especially when the model requires high-scale inputs like 224×224 by VggFace. This indicates the dramatic differences between native surveillance and web facial images. (2) The poorest results are yielded by the models trained with only CASIA faces. This is expected due to the big domain gap between CASIA and *SurvFace* (Fig. 5). (3) Most models are notably improved once pre-trained using CASIA faces. This suggests a positive effect of web data based model initialisation. (4) CentreFace is the best performer. This indicates the efficacy of restricting intra-class variation in training for face Re-ID, consistent with web data FR [10].

(II) Open-Set vs Closed-Set. We compared open-set and closed-set face Re-ID on *SurvFace*. All distractors in the gallery are removed for closed-set test. The top-2 models, CentreFace and SphereFace, are evaluated. Table 6 suggests that *closed-set face Re-ID is clearly easier than the open-set counterpart*. For instance, CentreFace achieves 27.3% TP1R20@FPIR30% in open-set vs 61.1% Rank-20 in closed-set. The gap is even larger at lower false alarm rates. This means that with the attack of distractors, face Re-ID becomes much harder.

(III) SurvFace vs WebFace. We compared face Re-ID with web face identification in the closed-set test. For example, CentreFace achieves Rank-1 29.9% (Table 6) on *SurvFace*, much inferior to the

Table 7

Image quality in face Re-ID: UCCS vs *SurvFace*(1090ID).

Dataset	Model	TP1R20(%) FPIR				AUC (%)
		30%	20%	10%	1%	
UCCS	CentreFace	96.1	94.6	90.4	80.7	96.1
	VggFace	71.0	60.3	46.6	15.0	77.0
	SphereFace	74.0	67.5	58.0	26.8	76.5
SurvFace	CentreFace	52.0	46.0	35.0	13.0	60.3
	VggFace	42.0	32.0	21.0	5.0	51.0
	SphereFace	59.9	56.0	49.0	20.0	64.0

rate of 65.2% on MegaFace [10], i.e. a 54% (1–29.9/65.2) performance drop. This indicates that face Re-ID is significantly more challenging, especially so when considering that one million distractors are used to additionally complicate the MegaFace test.

(IV) SurvFace Image Quality. We evaluated the effect of surveillance image quality in open-set face Re-ID. To this end, we qualitatively contrasted *SurvFace* with UCCS [9] that provides surveillance face images with clearly better quality in Fig. 5. For a quantitative comparison, we adopted the Fréchet Inception Distance (FID) [111], a widely used image quality metric in GAN model evaluations. To obtain a face-specific FID measure, instead of a typical ImageNet trained Inception network, we used a CentreFace model [10] trained with CASIA [32] high-quality image data. In this test, we randomly selected 10,000 face images from *SurvFace* and UCCS respectively; Against 10,000 random CASIA images, we then computed and obtained the FID score of 262.33 on *SurvFace* and 94.86 on UCCS. It shows that UCCS images have much higher quality (lower FID, so closer to high-quality CASIA images) than *SurvFace* images.

Setting. For UCCS, face images from the released 1090 IDs is randomly split into 545/545 IDs train/test set, resulting in a 6,948/7,068 image split. For a like-for-like comparison, we constructed a *SurvFace*(1090ID) dataset by randomly picking 545/545 *SurvFace* train/test IDs. For evaluation, we designed an open-set test setting using 100 random IDs for gallery and all 545 IDs for probe.

Results. Table 7 shows that *SurvFace* poses more challenges than UCCS, with varying degrees of performance drops experienced by FR models. This suggests that image quality is an important factor, and UCCS is less accurate in reflecting the face Re-ID challenges due to *artificially* high image quality.

(V) Test Scalability. We examined the test scalability by comparing *SurvFace*(1090ID) (Table 7 bottom) and *SurvFace* (Table 5),

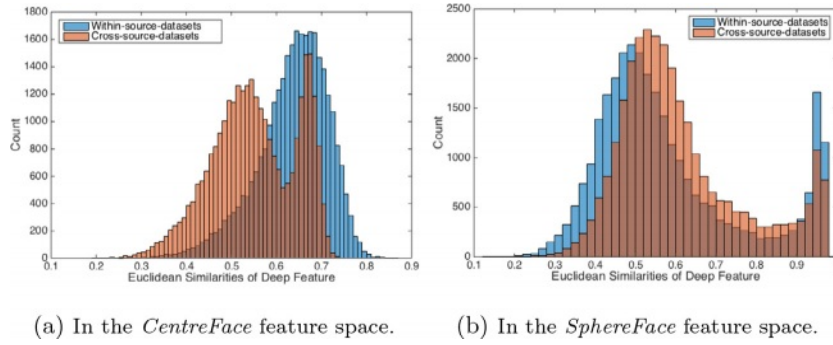


Fig. 6. Euclidean distance distributions of Intra-Domain and Inter-Domain false pairs in the feature space established by (a) CentreFace and (b) SphereFace.



Fig. 7. Face Re-ID examples by CentreFace on SurvFace. True matches are in red box.



Fig. 8. Super-resolved images on SurvFace by independently (left box) and jointly (right box) trained super-resolution models. CentreFace is used in joint training.

Table 8
Selection of web training data source.

Web Dataset	Model	TPIR20(%) FPIR				AUC (%)
		30%	20%	10%	1%	
MS-Celeb-1M	CentreFace	28.0	21.9	14.1	3.1	37.4
	SphereFace	20.1	13.6	5.4	0.8	27.1
MegaFace2	CentreFace	27.7	21.9	15.0	3.5	37.6
	SphereFace	20.0	13.0	5.4	0.7	26.4
CASIA	CentreFace	27.3	21.0	13.8	3.1	37.3
	SphereFace	21.3	15.7	8.3	1.0	28.1

and found significantly higher performances on the smaller 1090ID test set. This suggests that a large benchmark is crucial for true performance evaluation in practical face Re-ID.

(VI) Web Image Source. We tested the effect of web training dataset by comparing three benchmarks, CASIA [32], MS-Celeb-1M [34] and MegaFace2 [8], using CentreFace and SphereFace. Table 8 shows that the selection of web data only leads to neglectable changes in face Re-ID performance. For training, CASIA is tens of times more cost-effective (cheaper) than the other two larger datasets, so we use it in the main experiments. Moreover, it is more challenging to train a FR model given a vast ID class space such as MegaFace2.

Table 9
Effect of SurvFace image resolution.

Width (Pixels)	Model	TPIR20(%) FPIR				AUC (%)
		30%	20%	10%	1%	
≤ 20	CentreFace	32.9	23.3	15.2	4.0	40.0
	SphereFace	13.9	9.4	4.3	0.6	22.4
> 20	CentreFace	25.0	19.7	13.5	3.4	34.6
	SphereFace	26.2	21.6	14.8	2.7	32.2
All	CentreFace	27.3	21.0	13.8	3.1	37.3
	SphereFace	21.3	15.7	8.3	1.0	28.1

(VII) SurvFace Resolution. We examined the effect of test image resolution. Given the bi-modal distribution of SurvFace images, we divided all test *probe* faces into two groups at the threshold of 20 pixels in width. Table 9 shows that whilst the face resolution matters, the performance on all test images summarises the average of each group rather well. The performance variation across groups relies on both the applied models and other imaging factors, suggesting that the resolution alone does not bring a consistent performance bias.

(VIII) Domain Separation. Given that SurvFace is composed of faces captured from multiple data sources (domains), we tested whether face images from one source are overly different from the others, i.e. the domain separation effect. Domain separation may overwhelm subtle facial identity differences, therefore reducing the effective gallery size – larger separation, smaller gallery. To that end, we examined the distance statistics of *intra-domain* and *inter-domain* false pairs. Specifically, we formed 60,000 intra-domain and 60,000 inter-domain probe-gallery false pairs, and profiled their Euclidean distance measured by the CentreFace and SphereFace features, respectively. Fig. 6 shows clear and substantial overlaps between inter-domain and intra-domain although model-dependent. This implies that the domain separation effect is not severe in SurvFace.

(IX) Qualitative Evaluation. We show face Re-ID examples by CentreFace on SurvFace in Fig. 7. The model succeeds in finding

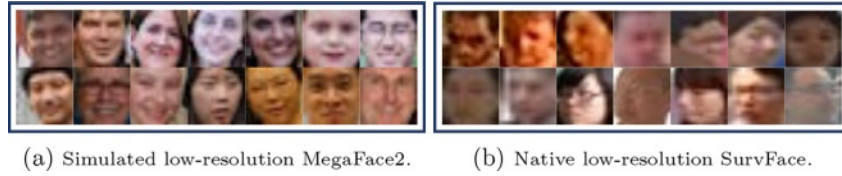


Fig. 9. (a) Simulated vs (b) native low-resolution face images.

Table 10

Effect of image super-resolution (SR) in face Re-ID on *SurvFace*. Protocol: Open-Set. Jnt/Ind: Joint/Independent training. '-': Fail to train.

Metrics		TPIR20(%) FPIR				AUC	TPIR20(%) FPIR				AUC	TPIR20(%) FPIR				AUC
		30%	20%	10%	1%		30%	20%	10%	1%		30%	20%	10%	1%	
SR / FR		CentreFace					VggFace					SphereFace				
No SR		27.3	21.0	13.8	3.1	37.3	5.1	2.6	0.8	0.1	14.0	21.3	15.7	8.3	1.0	28.1
SRCNN	Ind	25.0	20.0	13.1	3.0	35.0	6.2	3.1	1.0	0.1	15.3	20.0	14.9	6.2	0.6	27.0
	Jnt	25.5	20.5	12.0	2.9	35.0	-	-	-	-	-	-	-	-	-	-
VDSR	Ind	25.5	20.1	12.8	3.0	35.1	5.8	2.9	1.0	0.1	15.0	20.1	14.5	6.1	0.8	27.3
	Jnt	26.7	20.4	12.6	3.1	35.3	-	-	-	-	-	-	-	-	-	-
LapSRN	Ind	25.6	20.0	12.7	3.0	35.1	5.7	2.8	0.9	0.1	15.0	20.2	14.7	6.3	0.7	27.4
	Jnt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

the true match among top-20 in the top three tasks, and fails the last.

4.2. Super-resolution in surveillance face re-ID

Following the face Re-ID evaluation in raw low-resolution surveillance data, we tested *super-resolved* face images. The aim is to examine the effect of image super-resolution (SR) in addressing the low-resolution problem in face Re-ID. We only evaluated the native low-resolution SurvFace benchmark, since UCCS images are of *artificial* high resolution therefore excluded (Fig. 5).

Model Training and Test. We consider two training strategies as follows.

- (1) **Independent Training:** We first pre-train FR models on CASIA [32] and then fine-tune on SurvFace, same as in Sec. 4.1. Given no access to high-resolution SurvFace data, we *independently* train image SR models with CASIA data alone, where the low- and high-resolution training image pairs are generated by down-sampling. We deployed the learned SR models to restore SurvFace images before performing face Re-ID by deep features and Euclidean distance.
- (2) **Joint Training:** Training SR and FR models *jointly* in a hybrid pipeline to improve their compatibility. Specifically, we unit SR and FR models by connecting the former's output with the latter's input so allowing end-to-end training with both. In practice, we first performed joint learning with CASIA and then fine-tuned FR part with SurvFace. But joint training is not always doable due to additional challenges such as over-large model size and more difficulties to converge. In our experiments, we achieved joint training of two hybrid pipelines using two SR (SRCNN [13] and VDSR [16]) with one FR (CentreFace [10]) models. At test time, we deployed the hybrid pipeline on SurvFace images to perform face Re-ID using Euclidean distance.

Evaluation Settings. TPIR (Eqn. (3)) and FPIR (Eqn. (2)) are used as performance metrics, same as Section 4.1.

Implementation Details. We performed a $4 \times$ upscaling restoration for super-resolution, with models implemented by the public released codes. We followed the parameter setting as suggested, or carefully tuned them during training.

Table 11

Effect of super-resolution on down-sampled MegaFace2 data. CentreFace is used.

Metrics	TPIR20(%)@FPIR (open-set)				AUC (%)
	30%	20%	10%	1%	
No SR	39.9	28.0	14.0	5.8	46.0
SRCNN	26.6	19.2	10.0	5.0	36.5
VDSR	40.0	28.3	14.1	6.0	47.5

4.2.1. Face re-ID evaluation

(I) **Effect of Super-Resolution.** We tested the effect of image super-resolution on SurvFace. From Table 10, we have two observations: (1) Surprisingly, SR often brings slightly *negative* effect. The plausible reasons are threefold. The *first* is that conventional SR models usually favor visual fidelity instead of perceptual measurements for recognition. The *second* is that SR models are trained on web data, which has a domain gap against SurvFace (Fig. 1). The *third* is the negative effect of artifacts of SR (Fig. 8). An exception case (similar as in Section 4.1) is VggFace given the need for higher-resolution inputs therefore somewhat preference to SR. And VggFace has the weakest performance. (2) Joint training of FR and SR is not necessarily superior than independent training. This suggests that it is non-trivial to effectively propagate the FR discrimination capability into SR learning. Therefore, it is worth further in-depth investigation on how to integrate a super-resolution ability into face Re-ID.

(II) **SurvFace vs WebFace.** We evaluated SR on down-sampled web faces as a comparison to surveillance data.

Setting. We built a low-resolution web face identification test as SurvFace (Table 3) by sampling MegaFace2 [8]. MegaFace2 was selected since it contains non-celebrity people thus ensuring ID non-overlap with the training data CASIA. We down-sampled the selected MegaFace2 images to the mean SurvFace size 24×20 (Fig. 9), and built an open-set test setting with 3000 gallery IDs (51,949 images) and 10,254 probe IDs (176,990 images) (Table 4). We further randomly sampled an ID-disjoint training set with 81,355 images of 5319 IDs. In doing so, we created a like-for-like setting with low-resolution MegaFace2 against SurvFace. We adopted the most effective joint training strategy.

Results. Table 11 shows that SR brings very marginal gain to low-resolution face identification, suggesting that contemporary models are still far from satisfactory to boost facial discriminabil-



Fig. 10. Super-resolved MegaFace2 images. CentreFace is used jointly with the super-resolution models.

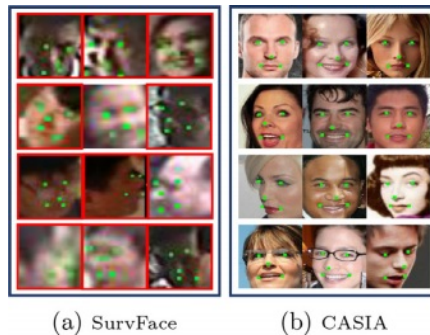


Fig. 11. Facial landmark detection [112] on (a) SurvFace and (b) CASIA web faces. Red box: failure cases.

ity (Fig. 10). In comparison, the model assisted by VDSR on simulated low-resolution web faces is better than on surveillance images, similarly reflected in super-resolved images (Fig. 10 vs Fig. 8). This indicates that super resolving surveillance faces is more challenging due to the lack of low- and high-resolution image pairs for model training.

5. Discussions and conclusion

In this study, we presented a large surveillance face re-identification benchmark *SurvFace*, with extensive benchmarking results, in-depth discussions and analysis. This challenge shows that existing models remain unsatisfactory in handling poor quality image based face Re-ID. In concluding remarks, we discuss research directions worthwhile investigating in the future researches.

5.1. Transfer learning

The benchmarking results (Table 10) show that knowledge transferred from auxiliary web faces boosts the face Re-ID performance, while more effective transfer methods are needed. Domain adaptation [65,113–116] is important given the surveillance-web domain discrepancy. In particular, style transfer [117–120] is a straightforward approach, by transforming source images with target domain style so that the labelled source data can be used for supervised learning. Whilst style transfer is inherently challenging, it promises potential for face Re-ID.

5.2. Resolution restoration

The low-resolution nature hinders the face Re-ID performance. While SR is one natural solution, our evaluations show that current models remain ineffective. Two main reasons are: (1) No access to native low- and high-resolution surveillance image pairs required

for SR training [121]. (2) Difficult to generalise models learned on web data due to domain gap [65]. Although SR models have been adopted for low-resolution faces [18], they rely on hand-crafted representations tested on small simulated data. It remains unclear how effective SR methods are for native face Re-ID.

5.3. Face alignment

Face alignment by landmark detection is an indispensable pre-processing in FR [10,38,112]. Despite the great progress [122–124], aligning face remains a formidable challenge in surveillance images (Fig. 11), suffering from domain shift. To construct a large surveillance face landmark dataset and integrate landmark detection with SR could be interesting topics.

5.4. Contextual constraints

Given incomplete and noisy observation in surveillance face data, it is important to utilise context information as extra constraint. For example, in social events, people often travel in groups. The group structure provides useful social force for model inference [125–128].

5.5. Open-set recognition

Face Re-ID is an open-set recognition problem [129,130]. In reality, most probes are non-target persons. It is hence beneficial that the model learn to construct a decision boundary for the target people [131]. Whilst open-set recognition techniques evolve independently, we expect more future attempts at jointly solving the two problems.

5.6. Imagery data scalability

Compared to existing web FR benchmarks [8,32,35], *SurvFace* is smaller in scale. An important future effort is to expand this challenge for more effective model training and larger scale open-set test.

5.7. Final remarks

This work presents timely a more challenging benchmark *SurvFace* for stimulating further innovative algorithms. This calls for more research efforts for under-studied and crucial face Re-ID.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] S. Gong, M. Cristani, C.C. Loy, T.M. Hospedales, The Re-identification Challenge, in: *Person Re-Identification*, Springer, 2014a, pp. 1–20.
- [2] S. Gong, M. Cristani, S. Yan, C.C. Loy, *Person Re-Identification*, Springer, 2014b.
- [3] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe, Frvt 2006 and ice 2006 large-scale experimental results, *TPAMI* 32 (5) (2010) 831–846.
- [4] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer, The humanid gait challenge problem: data sets, performance, and analysis, *TPAMI*, 27 (2) (2005) 162–177.
- [5] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, Springer Science & Business Media, 2009.
- [6] M. Li, X. Zhu, S. Gong, Unsupervised tracklet person re-identification, *IEEE Trans. Pattern Anal. Mach. Intell.* (2019).
- [7] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The Feret evaluation methodology for face-recognition algorithms, *TPAMI*, 22 (10) (2000) 1090–1104.
- [8] A. Nech, I. Kemelmacher-Shlizerman, Level playing field for million scale face recognition, in: *CVPR*, 2017, pp. 7044–7053.

- [9] M. Günther, P. Hu, C. Herrmann, C.-H. Chan, M. Jiang, S. Yang, A.R. Dhamija, D. Ramanan, J. Beyerer, J. Kittler, et al., Unconstrained face detection and open-set face recognition challenge, in: *IJCB*, IEEE, 2017, pp. 697–706.
- [10] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: *ECCV*, Springer, 2016, pp. 499–515.
- [11] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: *BMVC*, 2015, ????
- [12] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, SpheroFace: deep hypersphere embedding for face recognition, in: *CVPR*, 2017, pp. 212–220.
- [13] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: *ECCV*, Springer, 2014, pp. 184–199.
- [14] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: *ECCV*, Springer, 2016, pp. 391–407.
- [15] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep Laplacian pyramid networks for fast and accurate super-resolution, in: *CVPR*, 2017, pp. 624–632.
- [16] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: *CVPR*, 2016, pp. 1646–1654.
- [17] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: *CVPR*, 2017, pp. 3147–3155.
- [18] Z. Wang, Z. Miao, Q.J. Wu, Y. Wan, Z. Tang, Low-resolution face recognition: a review, *Vis. Comput.* 30 (4) (2014) 359–386.
- [19] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3d face recognition: a survey, *Pattern Recognit. Lett.* 28 (14) (2007) 1885–1906.
- [20] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, *Pattern Recognit.* 39 (9) (2006) 1725–1745.
- [21] W. Zhao, R. Chellappa, *Face Processing: Advanced Modeling and Methods*, Academic Press, 2011.
- [22] S.K. Zhou, R. Chellappa, W. Zhao, *Unconstrained Face Recognition*, vol. 5, Springer Science & Business Media, 2006.
- [23] S. Gong, S. McKenna, A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, World Scientific, 2000.
- [24] S. Li, A. Jain, *Handbook of Face Recognition*, Springer, 2011.
- [25] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *TPAMI* 19 (7) (1997) 711–720.
- [26] K. Messer, J. Matas, J. Kittler, J. Luetttin, G. Maitre, Xm2vtsdb: The extended m2vts database, in: *Second International Conference on Audio and Video-based Biometric Person Authentication*, vol. 964, 1999, pp. 965–966.
- [27] A.S. Georgiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *TPAMI* 23 (6) (2001) 643–660.
- [28] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression (pie) database, in: *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, IEEE, 2002, pp. 53–58.
- [29] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, *Image Vis. Comput.* 28 (5) (2010) 807–813.
- [30] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, IEEE, 1994, pp. 138–142.
- [31] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, University of Massachusetts, 2007.
- [32] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, *arXiv preprint arXiv:1411.7923* (2014).
- [33] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: *CVPR*, 2014, pp. 1891–1898.
- [34] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, Ms-celeb-1m: a dataset and benchmark for large-scale face recognition, in: *ECCV*, Springer, 2016, pp. 87–102.
- [35] I. Kemelmacher-Shlizerman, S.M. Seitz, D. Miller, E. Brossard, The megaface benchmark: 1 million faces for recognition at scale, in: *CVPR*, 2016, pp. 4873–4882.
- [36] K. Ricanek, T. Tesafaye, Morph: A longitudinal image database of normal adult age-progression, in: *International Conference on Automatic Face and Gesture Recognition*, IEEE, 2006, pp. 341–345.
- [37] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: *CVPR*, 2011, pp. 529–534.
- [38] D. Chen, X. Cao, L. Wang, F. Wen, J. Sun, Bayesian face revisited: a joint formulation, in: *ECCV*, Springer, 2012, pp. 566–579.
- [39] H.-W. Ng, S. Winkler, A data-driven approach to cleaning large face datasets, in: *ICIP*, 2014, pp. 343–347.
- [40] B.F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, A.K. Jain, Pushing the frontiers of unconstrained face detection and recognition: larpa janus benchmark a, in: *CVPR*, 2015, pp. 1931–1939.
- [41] A. Bansal, A. Nanduri, C. Castillo, R. Ranjan, R. Chellappa, Umdfaces: an annotated face dataset for training deep networks, (2016).
- [42] A. Bansal, C. Castillo, R. Ranjan, R. Chellappa, The do's and don'ts for cnn-based face verification, in: *ICCVW*, 2017, pp. 2545–2554.
- [43] C. Whitlam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A.K. Jain, J.A. Duncan, K. Allen, et al., larpa janus benchmark-b face dataset, in: *CVPRW*, 2017, pp. 90–98.
- [44] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, Vggface2: A dataset for recognising faces across pose and age, in: *IEEE International Conference on Automatic Face & Gesture Recognition*, IEEE, 2018, pp. 67–74.
- [45] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: *CVPR*, vol. 1, 2005, pp. 947–954.
- [46] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The cas-peal large-scale chinese face database and baseline evaluations, *IEEE Trans. Syst. Man Cybern.-Part A* 38 (1) (2007) 149–161.
- [47] J.R. Beveridge, P.J. Phillips, D.S. Bolme, B.A. Draper, G.H. Givens, Y.M. Lui, M.N. Teli, H. Zhang, W.T. Scruggs, K.W. Bowyer, et al., The challenge of face recognition from digital point-and-shoot cameras, in: *BTAS*, IEEE, 2013, pp. 1–8.
- [48] M. Grgic, K. Delac, S. Grgic, Sface-surveillance cameras face database, *Multimed. Tools Appl.* 51 (3) (2011) 863–879.
- [49] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, X. Chen, A benchmark and comparative study of video-based face recognition on cox face database, *TIP*, 24 (12) (2015) 5967–5981.
- [50] P. Li, J. Brogan, P.J. Flynn, Toward facial re-identification: experiments with data from an operational surveillance camera plant, in: *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, 2016, pp. 1–8.
- [51] S. Gupta, N. Gupta, S. Ghosh, M. Singh, S. Nagpal, M. Vatsa, R. Singh, Facesurv: a benchmark video dataset for face detection and recognition across spectra and resolutions, in: *FG*, IEEE, 2019, pp. 1–7.
- [52] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, *TPAMI* 28 (12) (2006) 2037–2041.
- [53] X. Cao, D. Wipf, F. Wen, G. Duan, J. Sun, A practical transfer learning algorithm for face verification, in: *ICCV*, 2013, pp. 3208–3215.
- [54] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition, *TIP* 16 (1) (2006) 57–68.
- [55] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, *TIP*, 11 (4) (2002) 467–476.
- [56] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: *CVPR*, 2014, pp. 1701–1708.
- [57] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: *CVPR*, 2015, pp. 815–823.
- [58] Y. Li, G. Wang, L. Nie, Q. Wang, W. Tan, Distance metric optimization driven convolutional neural network for age invariant face recognition, *Pattern Recognit.* 75 (2018) 51–62.
- [59] M. He, J. Zhang, S. Shan, M. Kan, X. Chen, Deformable face net for pose invariant face recognition, *Pattern Recognit.* 100 (2020) 107113.
- [60] X. Wei, H. Wang, B. Scotney, H. Wan, Minimum margin loss for deep face recognition, *Pattern Recognit.* 97 (2020) 107012.
- [61] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: *NeurIPS*, 2012, pp. 1097–1105.
- [62] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [63] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *CVPR*, 2015, pp. 1–9.
- [64] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *CVPR*, 2016, pp. 770–778.
- [65] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010).
- [66] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes, R.M. Mersereau, Eigenface-domain super-resolution for face recognition, *TIP* 12 (5) (2003) 597–606.
- [67] X. Wang, X. Tang, Face hallucination and recognition, in: *International Conference on Audio-and Video-Based Biometric Person Authentication*, Springer, 2003, pp. 486–494.
- [68] P.H. Hennings-Yeomans, S. Baker, B.V. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: *CVPR*, 2008, pp. 1–8.
- [69] W.W. Zou, P.C. Yuen, Very low resolution face recognition problem, *TIP* 21 (1) (2011) 327–340.
- [70] Z. Wang, S. Chang, Y. Yang, D. Liu, T.S. Huang, Studying very low resolution recognition using deep networks, in: *CVPR*, 2016, pp. 4792–4800.
- [71] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Color face recognition for degraded face images, *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* 39 (5) (2009) 1217–1230.
- [72] Z. Lei, T. Ahonen, M. Pietikainen, S.Z. Li, Local frequency descriptor for low-resolution face recognition, in: *Face and Gesture*, IEEE, 2011, pp. 161–166.
- [73] Y. Wong, C. Sanderson, S. Mau, B.C. Lovell, Dynamic amelioration of resolution mismatches for local feature based identity inference, in: *ICPR*, IEEE, 2010, pp. 1200–1203.
- [74] C.-X. Ren, D.-Q. Dai, H. Yan, Coupled kernel embedding for low-resolution face image recognition, *TIP* 21 (8) (2012) 3770–3783.
- [75] S. Shekhar, V.M. Patel, R. Chellappa, Synthesis-based recognition of low resolution faces, in: *IJCB*, 2011, pp. 1–6.
- [76] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: *CVPR*, 2017, pp. 4681–4690.
- [77] X. Yu, F. Porikli, Ultra-resolving face images by discriminative generative networks, in: *ECCV*, Springer, 2016, pp. 318–333.
- [78] A. Chakrabarti, A. Rajagopalan, R. Chellappa, Super-resolution of face images using kernel pca-based prior, *IEEE Trans. Multimed.* 9 (4) (2007) 888–892.
- [79] C. Liu, H.-Y. Shum, W.T. Freeman, Face hallucination: theory and practice, *IJCV* 75 (1) (2007) 115–134.
- [80] S. Zhu, S. Liu, C.C. Loy, X. Tang, Deep cascaded bi-network for face hallucination, in: *ECCV*, Springer, 2016, pp. 614–630.

- [81] Q. Cao, L. Lin, Y. Shi, X. Liang, G. Li, Attention-aware face hallucination via deep reinforcement learning, in: CVPR, 2017, pp. 690–698.
- [82] X. Yu, F. Porikli, Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders, in: CVPR, 2017, pp. 3760–3768.
- [83] C. Liu, S. Gong, C.C. Loy, On-the-fly feature importance mining for person re-identification, Pattern Recognit. 47 (4) (2014) 1602–1615.
- [84] J. Meng, A. Wu, W.-S. Zheng, Deep asymmetric video-based person re-identification, Pattern Recognit. 93 (2019) 430–441.
- [85] T. Wang, S. Gong, X. Zhu, S. Wang, Person re-identification by video ranking, in: ECCV, Springer, 2014, pp. 688–703.
- [86] J. Wang, X. Zhu, S. Gong, W. Li, Transferable joint attribute-identity deep learning for unsupervised person re-identification, in: CVPR, 2018, pp. 2275–2284.
- [87] Z. Cheng, Q. Dong, S. Gong, X. Zhu, Inter-task association critic for cross-resolution person re-identification, in: CVPR, 2020.
- [88] A. Dantcheva, J.-L. Dugelay, Frontal-to-side face re-identification based on hair, skin and clothes patches, in: AVSS, IEEE, 2011, pp. 309–313.
- [89] G.M. Farinella, G. Farioli, S. Battiato, S. Leonardi, G. Gallo, Face re-identification for digital signage applications, in: International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage, Springer, 2014, pp. 40–52.
- [90] Y. Wang, J. Shen, S. Petridis, M. Pantic, A real-time and unsupervised face re-identification system for human-robot interaction, Pattern Recognit. Lett. 128 (2019) 559–568.
- [91] Y. Kawanishi, Y. Wu, M. Mukunoki, M. Minoh, Shinpuhan2014: a multi-camera pedestrian dataset for tracking people across multiple cameras, in: 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision, vol. 5, 2014.
- [92] N. Martinel, C. Micheloni, Re-identify people in wide area camera network, in: CVPRW, 2012, pp. 31–36.
- [93] A. Das, A. Chakraborty, A.K. Roy-Chowdhury, Consistent re-identification in a camera network, in: ECCV, Springer, 2014, pp. 330–345.
- [94] D.S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, in: BMVC, vol. 1, 2011, p. 6.
- [95] D. Baltieri, R. Vezzani, R. Cucchiara, Sarc3d: a new 3d body model for people tracking and re-identification, in: ICIA, Springer, 2011, pp. 197–206.
- [96] W.R. Schwartz, L.S. Davis, Learning discriminative appearance-based models using partial least squares, in: 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing, IEEE, 2009, pp. 322–329.
- [97] D. Baltieri, R. Vezzani, R. Cucchiara, 3dpc: 3d people dataset for surveillance and forensics, in: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, 2011, pp. 59–64.
- [98] C.C. Loy, T. Xiang, S. Gong, Multi-camera activity correlation analysis, in: CVPR, 2009, pp. 1988–1995.
- [99] K. Liu, B. Ma, W. Zhang, R. Huang, A spatio-temporal appearance representation for video-based pedestrian re-identification, in: ICCV, 2015, pp. 3810–3818.
- [100] P.M. Roth, M. Hirzer, M. Köstinger, C. Belezni, H. Bischof, Mahalanobis Distance Learning for Person Re-Identification, in: Person Re-Identification, Springer, 2014, pp. 247–267.
- [101] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: ECCV, Springer, 2008, pp. 262–275.
- [102] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: deep filter pairing neural network for person re-identification, in: CVPR, 2014, pp. 152–159.
- [103] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: ICCV, 2015, pp. 1116–1124.
- [104] M. Gou, S. Karanam, W. Liu, O. Camps, R.J. Radke, Dukemtmc4reid: a large-scale multi-camera person re-identification dataset, in: CVPRW, 2017, pp. 10–19.
- [105] T. Xiao, S. Li, B. Wang, L. Lin, X. Wang, Joint detection and identification feature learning for person search, in: CVPR, 2017, pp. 3415–3424.
- [106] G. Song, B. Leng, Y. Liu, C. Hetang, S. Cai, Region-based quality estimation network for large-scale person re-identification, in: AAAI, 2018.
- [107] P. Hu, D. Ramanan, Finding tiny faces, in: CVPR, 2017, pp. 951–959.
- [108] P. Grother, M. Ngan, Face recognition vendor test (frvt): performance of face identification algorithms, NIST Interagency Report 8009 (5) (2014).
- [109] S. Liao, Z. Lei, D. Yi, S.Z. Li, A benchmark study of large-scale unconstrained face recognition, in: IJCB, IEEE, 2014, pp. 1–8.
- [110] P.J. Phillips, P. Grother, R. Micheals, Evaluation Methods in Face Recognition, in: Handbook of Face Recognition, Springer, 2011, pp. 551–574.
- [111] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local Nash equilibrium, in: NeurIPS, 2017, pp. 6626–6637.
- [112] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (10) (2016) 1499–1503.
- [113] E. Tzeng, J. Hoffman, T. Darrell, K. Saenko, Simultaneous deep transfer across domains and tasks, in: ICCV, 2015, pp. 4068–4076.
- [114] K. Saenko, B. Kulic, M. Fritz, T. Darrell, Adapting visual category models to new domains, in: ECCV, Springer, 2010, pp. 213–226.
- [115] Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, arXiv preprint arXiv:1409.7495(2014).
- [116] M. Ghifary, W.B. Kleijn, M. Zhang, D. Balduzzi, W. Li, Deep reconstruction-classification networks for unsupervised domain adaptation, in: ECCV, Springer, 2016, pp. 597–613.
- [117] L.A. Gatys, A.S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in: CVPR, 2016, pp. 2414–2423.
- [118] C. Li, M. Wand, Precomputed real-time texture synthesis with Markovian generative adversarial networks, in: ECCV, Springer, 2016, pp. 702–716.
- [119] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, M.-H. Yang, Diversified texture synthesis with feed-forward networks, in: CVPR, 2017, pp. 3920–3928.
- [120] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: ICCV, 2017, pp. 2223–2232.
- [121] C.-Y. Yang, C. Ma, M.-H. Yang, Single-image super-resolution: a benchmark, in: ECCV, Springer, 2014, pp. 372–386.
- [122] S. Zhu, C. Li, C. Change Loy, X. Tang, Face alignment by coarse-to-fine shape searching, in: CVPR, 2015, pp. 4998–5006.
- [123] Z. Zhang, P. Luo, C.C. Loy, X. Tang, Facial landmark detection by deep multi-task learning, in: ECCV, 2014, pp. 94–108.
- [124] X. Cao, Y. Wei, F. Wen, J. Sun, Face alignment by explicit shape regression, IJCV, 107 (2) (2014) 177–190.
- [125] A.C. Gallagher, T. Chen, Understanding images of groups of people, in: CVPR, 2009, pp. 256–263.
- [126] W.-S. Zheng, S. Gong, T. Xiang, Associating groups of people, in: BMVC, vol. 2, 2009.
- [127] D. Helbing, P. Molnar, Social force model for pedestrian dynamics, Phys. Rev. E 51 (5) (1995) 4282.
- [128] R.L. Hughes, A continuum theory for the flow of pedestrians, Transport. Res. Part B 36 (6) (2002) 507–535.
- [129] A. Bendale, T. Boulton, Towards open world recognition, in: CVPR, 2015, pp. 1893–1902.
- [130] A. Bendale, T.E. Boulton, Towards open set deep networks, in: CVPR, 2016, pp. 1563–1572.
- [131] W.-S. Zheng, S. Gong, T. Xiang, Towards open-world person re-identification by one-shot group-based verification, TPAMI 38 (3) (2016) 591–606.

Zhiyi Cheng received the BE degree from Beijing University of Posts and Telecommunications in 2014, and the MPhil degree from Chinese University of Hong Kong in 2016. She is currently working toward the PhD degree at Queen Mary University of London. Her research interests include computer vision and machine learning.

Xiatian Zhu is a Computer Vision Researcher at Vision Semantics Limited, London, UK. He received his Ph.D. from Queen Mary University of London. He won Sullivan Doctoral Thesis Prize 2016, an annual award representing the best doctoral thesis submitted to a UK University in computer vision. His research interests include computer vision and machine learning.

Shaogang Gong is Professor of Visual Computation at Queen Mary University of London (since 2001), a Fellow of the Institution of Electrical Engineers and the British Computer Society. He received his D.Phil (1989) in computer vision from Keble College, Oxford University. His research interests include computer vision, machine learning and video analysis.