MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World

Conference Paper · February 2016

CITATIONS
5

READS
6,265

5 authors, including:

Yandong Guo
Microsoft
21 PUBLICATIONS **145** CITATIONS
SEE PROFILE

Yuxiao Hu
Microsoft
44 PUBLICATIONS **4,839** CITATIONS
SEE PROFILE

Xiaodong He
Microsoft Research & University of Washington
151 PUBLICATIONS **3,840** CITATIONS
SEE PROFILE

MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World

Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao
Microsoft; Redmond, WA 98052

Abstract

Face recognition, as one of the most well-studied problems in computer vision, consists of two subproblems, verification and identification. Face verification is to determine whether two given face images belong to the same person, while face identification is typically to fetch the most similar faces in a gallery image set for any given query image. In this paper, we define our face recognition task as to determine the identity of a person from this individual's face image by using all the possibly collected face images of this individual as training data. More specifically, our task is to recognize the face image and link the face to a corresponding entity key in a knowledge base. With the unique key and the associated rich information provided by the knowledge base, our face recognition is an end-to-end simulation of the human behavior in face recognition. For this purpose, we design a benchmark task, which is to recognize one million celebrities in the world from their face images, which probably lead to one of the largest classification problems in computer vision. We describe and provide both training and measurement datasets to facilitate research in this area. Our datasets are larger than any existing datasets which are publicly available, and can help close the gap to the scale of the datasets used privately in industry.

Introduction

The recent breakthrough in computer vision benefits greatly from large scale training datasets and clearly defined tasks with rigorous metrics to inspire the community. A typical example is the large scale visual recognition challenge (ILSVRC) [1], which provides both instrumental training data and clearly defined tasks including image classification, object detection and localization, has inspired phenomenal great progresses in the area, including [2, 3, 4].

In this paper, we design a new benchmark task, and provide the corresponding large scale training and measurement datasets for face recognition, which has been a critical and special problem in computer vision. There are two types of tasks for face recognition. One is very well-studied, called face verification, which is to determine whether two given face images belong to the same person. Recently, the interest in the other type of face recognition task, face identification, has greatly increased [5, 6, 7, 8]. For typical face identification problems, two sets of face images are given, called gallery set and query set. Then the task is, for a given face image in the query set, to find the most similar faces in the gallery image set. When the gallery image set only has a very limited number (say, less than five) of face images for each individual, the most effective solution is still to learn a generic feature which can tell whether or not two face images are the same person, which is essentially still the problem of face verification.

Despite all the efforts and progresses in face recognition, we still observe gaps in the following two aspects. First, we do not see enough effort in determining the identity of a person from a face image with disambiguation, especially in web-scale. The current face identification task mainly focuses on finding similar images (in terms of certain types of distance metric) for the input image, rather than answering questions such as "who is in the image?" and "if it is George in the image, which George?". This lacks the important last step of "recognizing". Second, the publicly available datasets are much smaller than that being used privately in industry, such as Facebook [9, 8] and Google [10], as summarized in Table 1. Though the research in face recognition highly desires large datasets consisting of many distinct people, such large dataset is not easily or publicly accessible to most researchers. This greatly limits the contributions from research groups, especially in academia.

This paper is mainly to close the above two gaps. Our first contribution is that we define our face recognition as to determine the identity of a person from his/her face images. More specifically, we introduce a knowledge base into face recognition, since in the knowledge base, each person (typically celebrity) is identified by a unique entity key, which is associated with rich descriptions/properties, including date of birth, professions, place of birth, representative images, etc. Then our face recognition task is to recognize the face image and link the face to a unique entity key in the knowledge base. For example, in Figure 1, given an image on the left, our task is to recognize the person in the image (who is "George W. Bush"), and link this image with the entity key associated with the following properties: named "George W. Bush", who was born on July 6, 1946, and served as the 43rd president of the United States of American from 2001 to 2009.

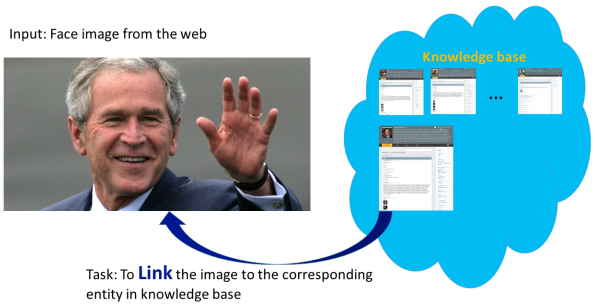


Figure 1. *An example of our face recognition task. Input is an image from the web, the task is to recognize the face in the image and then link this face with the corresponding entity key in the knowledge base.*

Our face recognition task has the following advantages. First, our face recognition task is to link the image with an entity key in the knowledge base, rather than an isolated string for a person's name. This setup naturally solves the disambiguation issue in the traditional face recognition task. Moreover, the linked entity key is associated with rich and comprehensive information in the knowledge base, which makes our task more similar to human behavior compared with traditional face identification, since retrieving the individual's name as well as the associated information naturally take place when humans are viewing a face image. Our face recognition task completes the last step of "recognizing" a face image. A lot of real applications could benefit from our face recognition task, including image search/retrieval, ranking, caption generation, image deep understanding, etc.

Our second contribution is that we design a large scale benchmark task for our face recognition problem described above and construct a measurement set for this task. More specifically, we organize a list for one million celebrities from a knowledge base and encourage researchers to build recognizers to recognize (from face images) as many celebrities in this one-million list as possible with high accuracy. We focus on one million popular celebrities because only with popular celebrities, we can build up publicly available dataset consisting of large scale of images of many distinct people, which is highly desired by the research in face recognition area. We leverage a knowledge base to generate our one-million celebrity list since recently knowledge bases can provide accurate identifiers and rich properties for celebrities. Examples include Satori knowledge graph in Microsoft and "freebase" in [11]. We chose to leverage freebase to provide entity key for the one million celebrities in our list due to its good coverage, public availability, and free licensing. More details are provided in the dataset construction section.

In order to evaluate the recognition performance on our one-million celebrity list, we construct a measurement set by blending a set of carefully labeled images and a set of distractor images. For the labeled images, due to limited resource, we sample 1000 celebrities from the one-million celebrity list and provide one image per celebrity. We will have more celebrities carefully labeled to expand our measurement set in the near future. The distractor images are images of other celebrities or ordinary people on the web.

Note that we do not expose the selected celebrities in our measurement set. Though we do not require to exclude celebrities in our measurement from the training data set, our task still requires the recognition model to have generalization ability, due to the following reasons. First, there are one million celebrities to be recognized in our task, and there are millions of images for some popular celebrities on the web. Therefore the chance to include the measurement images in the training set is relatively low, as long as the celebrity list in the measurement set is hidden. Second, it is too expensive to manually label all the images for every celebrity in our one-million list. This is different from most of the existing face recognition benchmark tasks, in which the measurement set is published and targeted on a small group of people. For these traditional benchmark tasks, the images of all the persons in the measurement set need to be removed from the training set.

With the above setup, our benchmark task has the following attractive challenges. First, researchers need to build a recognizer which could robustly distinguish one million people faces, rather than focusing merely on a small group of people. To the best of our knowledge, this may lead to the largest classification problem in the face recognition area. With the increase of the number of classes, the inter-variation between classes tends to decrease. There are celebrities look very similar to each other (or even twins) in our one-million list. Second, some celebrities have millions of images available on the web and he/she may look very different in different images due to different lighting conditions, poses, makeups, growth of age, or even sex reassignment surgery. This introduces a large intra-class variation.

The above challenges triggers our third contribution: we provide a very large training dataset to facilitate the above face recognition task. In order to prepare this training dataset, we select the top 100K celebrities from our one-million celebrity list in terms of their web appearance frequency, and provide approximately 100 images per entity. Note these images are obtained by scraping popular search engines and may contain noises. This is because to prepare training data of this size is beyond the scale of manually labeling, and we observe that, even with noise, more data could still be able to benefit the face recognition model, which is also reported in [12]. Moreover, we believe that to remove the noisy label as preprocessing and/or make the model training more robust to noisy labels are valuable research topics too, so we leave this problem open. For this reason, we do not limit the use of outside training data that can be collected on the web. This is similar to the unrestricted, labeled outside data setting in Labeled Faces in the Wild (LFW) [13] benchmark.

The 100K celebrities in the training set covers about 75% of celebrities in our measurement set, And we encourage people to bring in more outside data and evaluate experimental results in a separate track. Especially, we encourage people label their data with entity keys in the freebase snapshot provided together with the training and measurement datasets, so that different datasets could be easily merged to facilitate collaboration.

Related works

There are two types of tasks for face recognition. One is face verification, which is to determine whether two given face images belong to the same person. Face verification has been heavily investigated. One of the most widely used measurement sets for verification is Labeled Faces in the Wild (LFW) in [13][14], which provides 3000 matched face image pairs and 3000 mismatched face image pairs, and allow researchers to report verification accuracy with different settings. The best performance on LFW datasets has been frequently updated in the past several years. Especially, with the "unrestricted, labeled outside data" setting, multiple research groups have claimed higher accuracy than human performance for verification task on LFW [10][5].

Recently, the interest in the other type of face recognition task, face identification, has increased [5] [6] [7] [8]. For typical face identification problem, two sets of face images are given, called gallery set and query set. Then the task is, for a given face image in the query set, to find the most similar faces in the gallery image set. Currently, the MegaFace in [7] might be one of the most difficult face identification benchmarks. The difficulty of MegaFace mainly comes from the up to one million distractors blended in the gallery image set. Note that the query set in

MegaFace are selected from images from FaceScrub [15] and FG-NET [16], which contains 530 and 82 persons respectively.

Several datasets have been published to facilitate the training for the face verification and identification tasks. Examples include LFW [13, 14], Youtube Face Database (YFD) [17], CelebFaces+ [18], and CASIA-WebFace [19]. In LFW, 13000 images of faces were collected from the web, and then carefully labeled with celebrities' names. The YFD contains 3425 videos of 1595 different people. The CelebFace+ dataset contains 202,599 face images of 10,177 celebrities. People in CelebFaces+ and LFW are mutually exclusive. A quick summary is shown in Table 1.

Table 1. Face recognition datasets

Dataset	Available	people	images
LFW	public	5K	13K
YFD	public	1595	3425 videos
CelebFaces	public	10K	202K
CASIA-WebFace	public	10K	500K
Ours	**public**	100**K**	about 10**M**
Facebook	private	4K	4400K
Google	private	8M	100-200M

As shown in Table 1, our training dataset is considerably larger than the publicly available datasets. Another important difference between our training dataset and other datasets is that our dataset is determined to facilitate our celebrity recognition task, so our dataset needs to cover as many popular celebrities as possible, while other datasets are mainly used to train a generalizable face feature, and celebrity coverage is not a concern for these datasets. That is, if a people name corresponds to multiple celebrities, for example, Mike Smith, and will lead to ambiguous image search result, these celebrities are normally removed from these datasets to ensure the precision of the collected training data.

Dataset construction

Our face recognition task is to recognize one million celebrities from their face images. This section is to describe the steps to obtain the one-million celebrity list, and steps to construct the measurement and the training set.

One million celebrity list

We select one million celebrities, who are real persons in the world and have/had public attentions. The steps for selection are described in details in the following paragraphs.

First, we select a subset of entities from a knowledge base called freebase [11] based on the information within freebase. In freebase, each entity is identified by a unique key (called machine identifier, mid in [11]), and associated with rich properties. More specifically, we select the entities of which the properties satisfy all the three following conditions.

- The object type of the entity is defined as "people.person" in freebase.
  This condition means that we select entities which are claimed (by freebase) to be real persons in the world. We don't include movie characters since their appearance is not strictly defined, especially when a classic movie is retaken.
- The entities are required to have at least one of the properties unique for human beings, such as "person's name", "place of birth", "date of birth", "person's professions".
  This condition removes the entities which have too sparse information for us to collect and label images. This condition also helps us to remove some of the entities of which the object type are mislabeled as "people.person" in freebase.
- If the date of birth is available for a given entity in freebase, this entity can not be selected if he/she was born before the mid-nineteenth century.
  The reason for this condition is as follows. The first roll-film specialized camera "Kodak" was invented in 1888 [20] and started to get popular in late nineteenth century. We can not rely on drawings or sculptures to recognize people's faces, since whether they are visually similar to the actual person could be subjective and arguable. An interesting example is that the sculpture of John Harvard in Harvard university is claimed to be inspired by a Harvard student Sherman Hoar rather than Harvard himself, since no one knew what John Harvard had looked like [21].

In the second step, we rank all the entities in the above subset according to the frequency of their occurrence on the web. Then, we select the top one million entities to form our one million celebrity list and provide their entity keys (mid) in freebase. The occurrence frequency for a given entity is obtained by counting how many documents contain this entity in a large corpus with billions of documents from the web.

Measurement dataset

The measurement set is to evaluate the performance of recognizing the one million celebrities described in the last subsection. The measurement set is constructed by blending a set of carefully labeled images and a set of distractor images.

It is not feasible to include all the one million celebrities in the measurement set. Therefore, for the labeled images, we sample 1000 celebrities [1] from the one-million celebrity list and provide one image for each of the celebrity. The correctness of our measurement dataset is ensured by multiple iterations of careful review and rigorous consensus verification. In order to represent the popularity distribution of celebrities on the web, we applied the weighted random sampling.

Let f_i denote the number of documents mentioned the i^{th} celebrity on the web. Though setting the sampling weights to be proportional to f_i seems to be a natural solution, we don't choose this option since this option will make our measurement set barely contain any celebrities from the bottom 90% in our one-million list (ordered by f_i). The reason is that the distribution of f is too skewed. For example, Justin Bieber (entity key m.06w2sn5 in [11]) has been mentioned by millions of distinct documents in our experiment, while most of the professors have been mentioned by less than 10 documents.

In order to make our task more challenge, in the random sampling procedure, we set the probability for the i^{th} celebrity to get selected to be proportional to f_i' defined in the following equation.

[1] We will increase the number of celebrities in our measurement set in the future.

$$f_i' = f_i^{\frac{1}{\sqrt{5}}}. \quad (1)$$

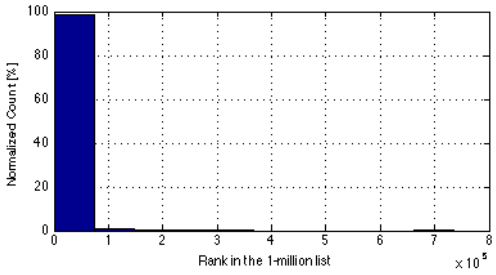
According to Equation 1, on one hand, celebrities who are more popular (with larger f_i) have larger chances to be selected. On the other hand, the exponent $1/\sqrt{5}$, which is obtained empirically, is used to penalize the probability of celebrities with large f to be selected to include some celebrities with small f.

According to our experiments, with the adjustment in Equation 1, though the measurement set is still mainly focused on the most popular celebrities, about 25% of the celebrities in our measurement set are ranked in the bottom 90% in our one-million celebrity list (ordered by f), as shown in Figure 2 (b). Since the list of the celebrities in our measurement set is not exposed, researchers need to include as many celebrities as possible from our one-million list to improve the recall. This may make our task one of the largest classification problems in the area.

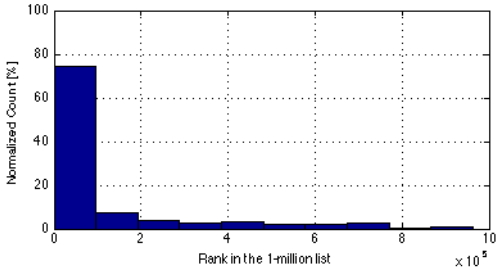
In our measurement set, the 1000 images for these 1000 celebrities are blended with images from other celebrities or ordinary people online. The precision and recall are used to evaluate the model performance.

Training dataset

In order to facilitate the above face recognition task we provide a large training dataset. This training dataset is prepared by the following two steps. First, we select the top 100K entities from our one-million celebrity list in terms of their web appearance frequency. Then, we leverage popular search engines to provide approximately 100 images per celebrity.



(a) With sampling weight f



(b) With sampling weight f' defined in (1)

Figure 2. *Distribution of the celebrities obtained by sampling with different weights. With the sampling weights f' defined in (1), about 25% of our measurement set are ranked in the bottom 90% in our one-million list, which encourages researchers to explore the entire celebrity list.*

Two examples are shown in Figure 3 and Figure 4. As shown in the figures, same celebrity may look very differently in different images. In Figure 3, Lady Gaga (m.0478_m) looks visually different due to different lighting, different poses, and heavy makeups. In Figure 4, we see images for Steve Jobs (m.06y3r) when he was about 20/30 years old, as well as images when he was about 50 years old. The image at row 9, column 4 in Figure 4 is claimed to be Steve Jobs when he was in high school. Notice that the image in the right corner in Figure 4, marked with red rectangle is

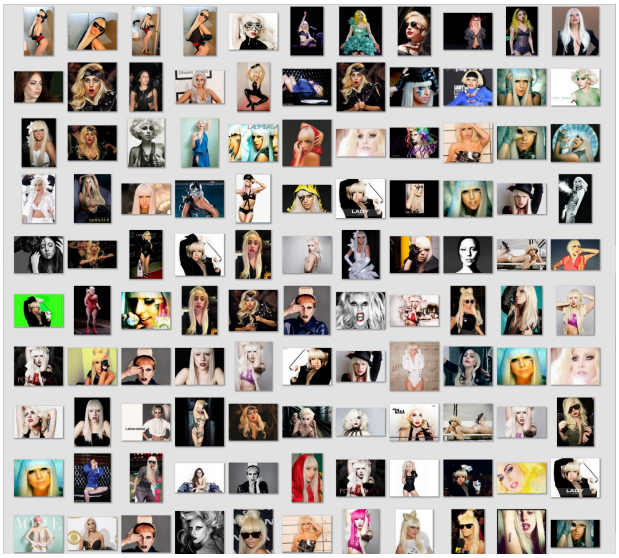


Figure 3. *Examples of the training images we provided for the celebrity with entity key m.0478_m (Lady Gaga)*

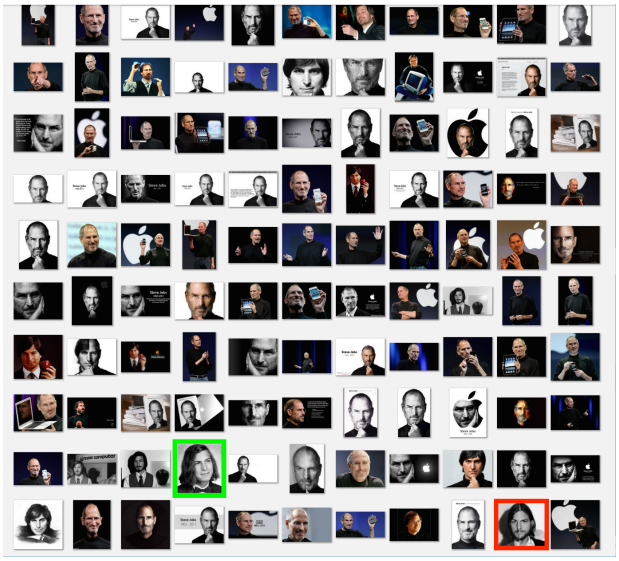


Figure 4. *Examples of the training images we provided for the celebrity with entity key m.06y3r (Steve Jobs). The image marked with a green rectangle (at row 9, column 4) is claimed to be Steve Jobs when he was in high school. The image in the right corner, marked with a red rectangle is considered as a noise sample in our dataset, since it is synthesized by combining one image of Steve Jobs and one image of Ashton Kutcher, who is the actor in the movie "Jobs".*

considered as a noise sample in our dataset, since this image was synthesized by combining one image of Steve Jobs and one image of Ashton Kutcher, who is the actor in the movie "Jobs".

As we have mentioned in the introduction section, we do not manually remove the noise in this training data set. This is partially because to prepare training data of this size is beyond the scale of manually labeling. In addition, we have observed that the state-of-the-art deep neural network learning algorithm can tolerate a certain level of noise in the training data. Though for a small percentage of celebrities their image search result is far from perfect, more data especially more individuals covered by the training data could still be of great value to the face recognition research, which is also reported in [12]. Moreover, we believe that data cleaning, noisy label removal, and learning with noisy data are all good and real problems that are worth of dedicated research efforts. Therefore, we leave this problem open and even do not limit the use of outside training data.

As mentioned in the previous subsection, our 100K-list in the training set only covers about 75% of celebrities in our measurement set, so that to encourage researchers to bring in outside data to get higher recognition recall rate and compare experimental results in a separate track. Especially, we encourage people label their data with entity keys in the freebase snapshot we provided and publish, so that different dataset could be easily merged to facilitate collaboration.

Conclusions and Discussions

In this paper, we have defined our face recognition task as to determine the identity of a person from the face image. More specifically, our task is to recognize the face image and link the face to a corresponding entity key in a knowledge base. With the unique key and the associated rich information provided by the knowledge base, our face identification is an end-to-end simulation of the human behavior in face recognition. Moreover, we design a benchmark task, with the target to recognize one million celebrities in the world from their face images, which probably lead to one of the largest classification problems in the area. We describe and provide both training and measurement datasets to facilitate research in the area. Our dataset are larger than the datasets which are now publicly available, and closes the gap to the scale of the datasets used privately in industry.

Beyond face recognition, our datasets could inspire other research topics. For example, people could adopt one of the cutting-edge unsupervised/semi-supervised clustering algorithms [22, 23, 24, 25] on our training dataset, and/or develop new algorithms which can accurately locate and remove outliers in a large, real dataset. Another interesting topic is the to build estimators to predict a person's properties from his/her face images. For example, the images in our training dataset are associated with entity keys in knowledge base, of which the gender information (or other properties) could be easily retrieved. People could train a robust gender classifier for the face images in the wild based on this large scale training data. We look forward to exciting research inspired by our training dataset and benchmark task.

References

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012, MIT Press, pp. 1097–1105.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint abXiv:1409.1556*, 2014.

[4] S. R. Kaiming He, Xiangyu Zhang and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[5] Y. Sun, X. Wang, and X. Tang, "DeepID3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2014.

[6] H. Fan, M. Yang, Z. Cao, Y. Jiang, and Q. Yin, "Learning compact face representation: Packing a face into an int32," *Proc. of ACM Int'l Conf. on Multimedia*, 2014, ACM, pp. 933–936.

[7] I. Kemelmacher-Shlizerman, S. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," *ArXiv e-prints*, 2015.

[8] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification.," *Proc. of IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, IEEE, pp. 2746–2754.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *Proc. of IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[10] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *Proc. of IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[11] Google, "Freebase data dumps." https://developers.google.com/freebase/data, 2015.

[12] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," *Proceedings of the British Machine Vision Conference (BMVC)*, 2015.

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[14] G. B. H. E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Tech. Rep. UM-CS-2014-003, University of Massachusetts, Amherst, May 2014.

[15] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," *Proc. of IEEE Int'l Conf. on Image Proc. (ICIP)*, Oct 2014.

[16] G. Panis and A. Lanitis, "An overview of research activities in facial age estimation using the FG-NET aging database," *ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part II*, 2014.

[17] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," *Proc. of IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[18] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," *Proc. of IEEE Computer Soc. Conf.*

on Computer Vision and Pattern Recognition (CVPR), June 2014.

[19] S. L. Dong Yi, Zhen Lei and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[20] G. Eastman, "Camera," *US Patent 388850 A*, 1888.

[21] R. S. John T. Bethell, Richard M. Hunt, *Harvard A to Z.* Harvard University Press, 2004.

[22] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in Neural Information Processing Systems*, 2001, MIT Press, pp. 849–856.

[23] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," *Journal of Machine Learning*, vol. 56, no. 1-3, pp. 209–239, June 2004.

[24] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," *Proc. of Int'l Conf. on Machine Learning*, 2003, pp. 912–919.

[25] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schlkopf, "Learning with local and global consistency," *Advances in Neural Information Processing Systems*, 2004, MIT Press, pp. 321–328.

Author Biography

Yandong Guo earned a Ph.D. in electrical engineering at Purdue University at West Lafayette, under the supervision of Prof. Charles Bouman and Prof. Jan Allebach. Before that, he received his B.S. and M.S. degree in electrical engineering from Beijing University of Posts and Telecommunications, China, in 2005 and 2008, receptively. He has been working as a researcher at Microsoft since January 2014. His research has focused on statistical image models, machine learning, and computer vision.

Lei Zhang received his PhD degree in computer science from Tsinghua University in 2001. He is now a Senior Researcher with Microsoft Research. Prior to this, he worked with Microsoft Research Asia, Beijing, China for 12 years on large-scale multimedia content analysis, visual recognition, and machine learning, and moved to Bing image search as a Principal Development Manager for two years working on knowledge mining from web-scale image data. He holds 40+ U.S patents and has served as Program Area Chairs or Committee Members for many conferences in multimedia and computer vision.

Yuxiao Hu got his CS Master degree from Tsinghua University (China, 2001). He worked on face related projects in Microsoft Research Asia during 2001 and 2004 as an Assistant Researcher and then joined Microsoft Bing on 2008, working as a senior developer in multimedia and web platform team, after getting his PhD in ECE from UIUC. Currently, he is working in Microsoft Research ,Cloud Computing and Storage group, doing research and development on machine learning, image understanding and distributed systems.

Xiaodong He is a Senior Researcher at Microsoft Research. He is also an Affiliate Professor in Electrical Engineering at the University of Washington, Seattle. His research interests are mainly in the machine intelligence areas, including deep learning, speech, natural language, computer vision, information retrieval, and knowledge representation. He and colleagues won several Challenges in NLP and Computer Vision including the MS COCO Image Captioning Challenge 2015. He is an elected member of the IEEE SLTC.

Jianfeng Gao is a Principal Researcher of Microsoft Research, USA. Gao received his PhD in computer science from Shanghai Jiaotong Univerity (1999). His research interests include Web search and information retrieval, natural language processing and machine learning. He was Associate Editor of ACM Trans on Asian Language Information Processing, and was Member of the editorial board of Computational Linguistics. He received an outstanding paper award from ACL-2015.