A Survey on Deep Learning based Face Recognition

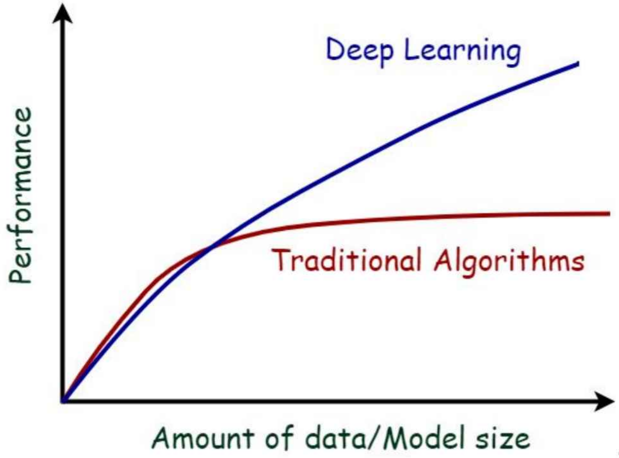
Na Zhang

Part V: Databases

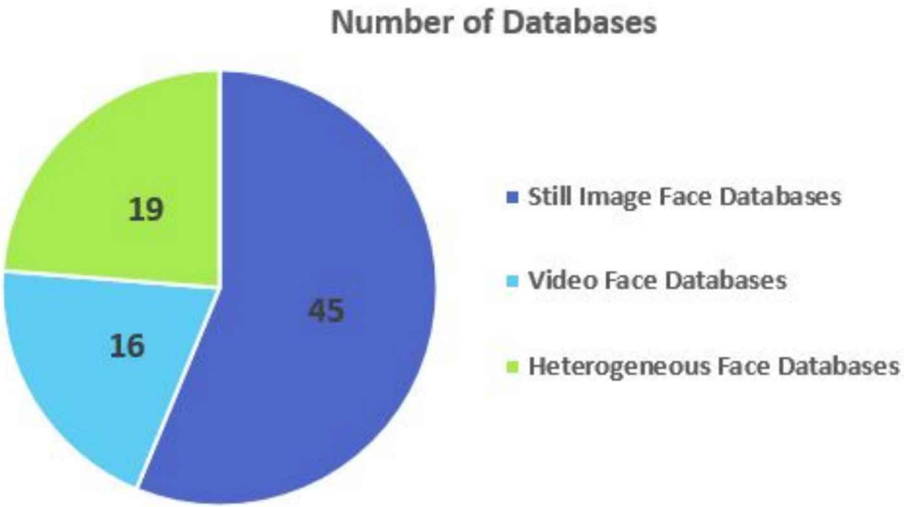
- ✓ Introduction
- ✓ Deep Learning Methods
- ✓ Some Specific Face Recognition Problems
- ⏻ Databases

- Data and Algorithm are two essential components for FR
- In some sense, the FR research is driven by face data
- With the wider use of deep neural networks in FR
  - the requirement of a huge amount of training data becomes more urgent
  - the deep learning methods are expected to learn a complex data distribution from large-scale training datasets containing many identities

- Experiments have demonstrated that:
  - large amount of labeled data can help the network learn better deep models



- An overview of face datasets
  - Still Image Faces
  - Video Faces
  - Heterogeneous Faces



Still Image Face Databases

- Early face datasets were almost collected under pre-defined or controlled environments, such as PIE, Yale, CMU PIE, FERET, etc.

- Along with the practical requirement, more attentions are paid to uncontrolled or unconstrained scenarios. i.e., face recognition in the wild.

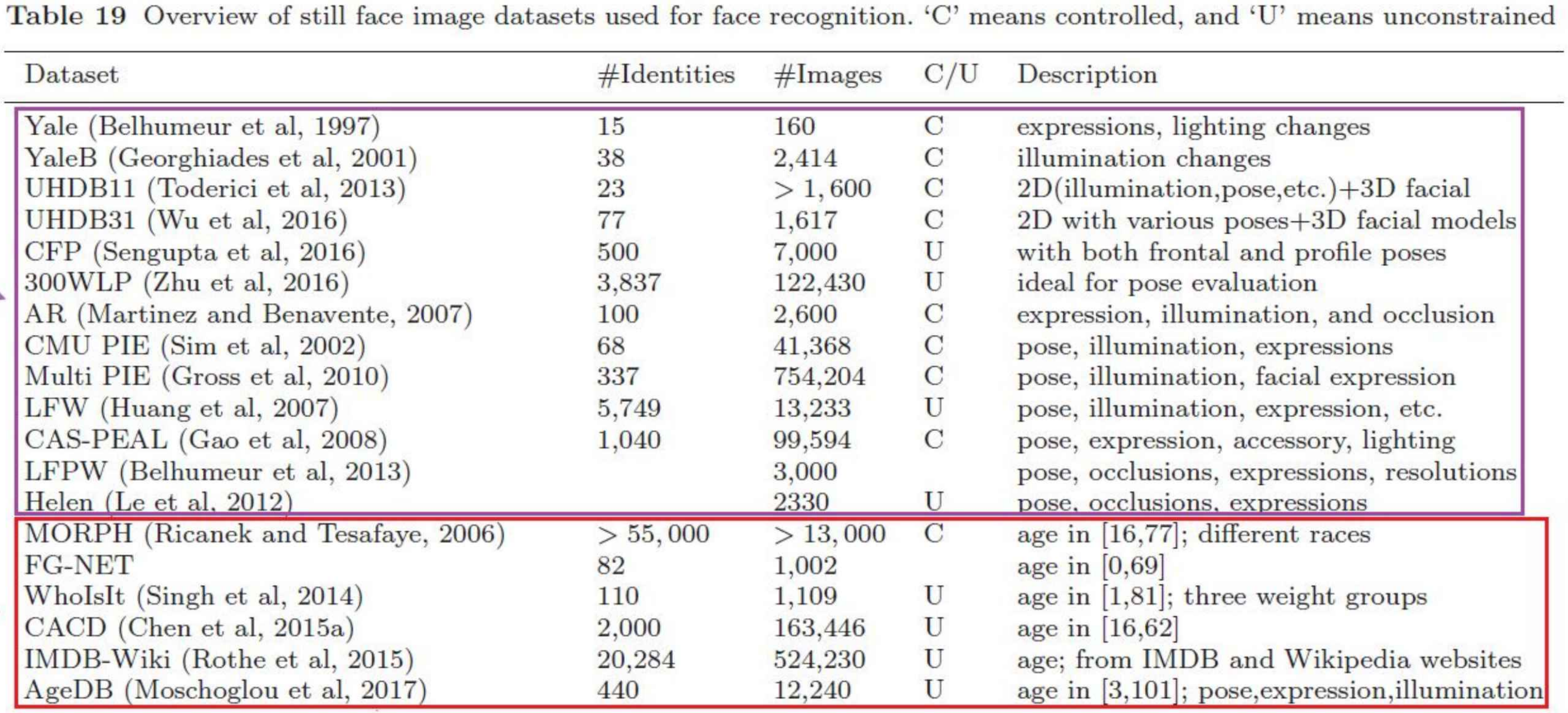
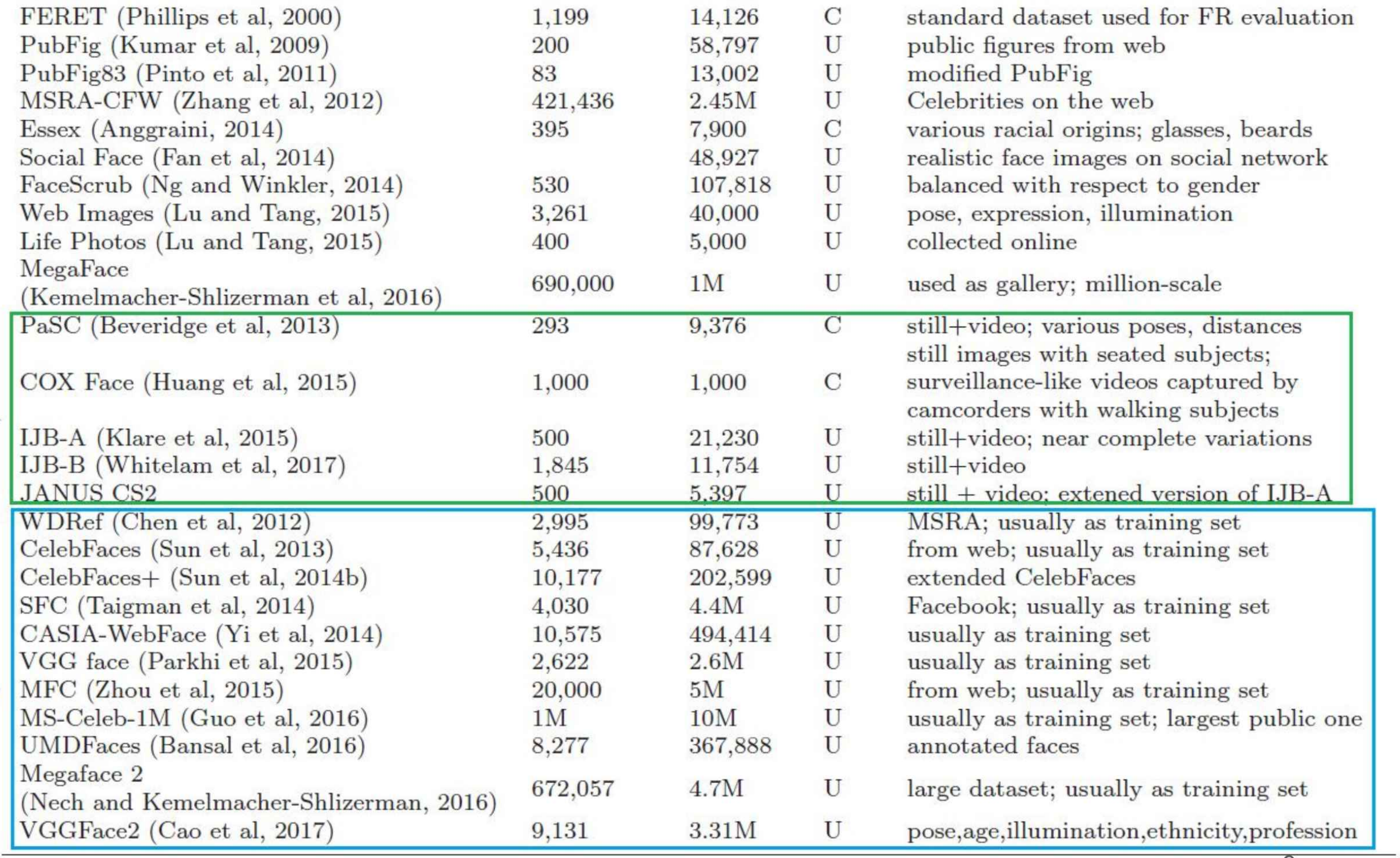


Table 19 Overview of still face image datasets used for face recognition. 'C' means controlled, and 'U' means unconstrained

Dataset	#Identities	#Images	C/U	Description
Yale (Belhumeur et al, 1997)	15	160	C	expressions, lighting changes
YaleB (Georghiades et al, 2001)	38	2,414	C	illumination changes
UHDB11 (Toderici et al, 2013)	23	$> 1,600$	C	2D(illumination,pose,etc.)+3D facial
UHDB31 (Wu et al, 2016)	77	1,617	C	2D with various poses+3D facial models
CFP (Sengupta et al, 2016)	500	7,000	U	with both frontal and profile poses
300WLP (Zhu et al, 2016)	3,837	122,430	U	ideal for pose evaluation
AR (Martinez and Benavente, 2007)	100	2,600	C	expression, illumination, and occlusion
CMU PIE (Sim et al, 2002)	68	41,368	C	pose, illumination, expressions
Multi PIE (Gross et al, 2010)	337	754,204	C	pose, illumination, facial expression
LFW (Huang et al, 2007)	5,749	13,233	U	pose, illumination, expression, etc.
CAS-PEAL (Gao et al, 2008)	1,040	99,594	C	pose, expression, accessory, lighting
LFPW (Belhumeur et al, 2013)		3,000		pose, occlusions, expressions, resolutions
Helen (Le et al, 2012)		2330	U	pose, occlusions, expressions
MORPH (Ricanek and Tesafaye, 2006)	$> 55,000$	$> 13,000$	C	age in [16,77]; different races
FG-NET	82	1,002		age in [0,69]
WhoIsIt (Singh et al, 2014)	110	1,109	U	age in [1,81]; three weight groups
CACD (Chen et al, 2015a)	2,000	163,446	U	age in [16,62]
IMDB-Wiki (Rothe et al, 2015)	20,284	524,230	U	age; from IMDB and Wikipedia websites
AgeDB (Moschoglou et al, 2017)	440	12,240	U	age in [3,101]; pose,expression,illumination

illumination, pose, etc.

Age



FERET (Phillips et al, 2000)	1,199	14,126	C	standard dataset used for FR evaluation
PubFig (Kumar et al, 2009)	200	58,797	U	public figures from web
PubFig83 (Pinto et al, 2011)	83	13,002	U	modified PubFig
MSRA-CFW (Zhang et al, 2012)	421,436	2.45M	U	Celebrities on the web
Essex (Anggraini, 2014)	395	7,900	C	various racial origins; glasses, beards
Social Face (Fan et al, 2014)		48,927	U	realistic face images on social network
FaceScrub (Ng and Winkler, 2014)	530	107,818	U	balanced with respect to gender
Web Images (Lu and Tang, 2015)	3,261	40,000	U	pose, expression, illumination
Life Photos (Lu and Tang, 2015)	400	5,000	U	collected online
MegaFace (Kemelmacher-Shlizerman et al, 2016)	690,000	1M	U	used as gallery; million-scale
PaSC (Beveridge et al, 2013)	293	9,376	C	still+video; various poses, distances
COX Face (Huang et al, 2015)	1,000	1,000	C	still images with seated subjects; surveillance-like videos captured by camcorders with walking subjects
IJB-A (Klare et al, 2015)	500	21,230	U	still+video; near complete variations
IJB-B (Whitelam et al, 2017)	1,845	11,754	U	still+video
JANUS CS2	500	5,397	U	still + video; extened version of IJB-A
WDRef (Chen et al, 2012)	2,995	99,773	U	MSRA; usually as training set
CelebFaces (Sun et al, 2013)	5,436	87,628	U	from web; usually as training set
CelebFaces+ (Sun et al, 2014b)	10,177	202,599	U	extended CelebFaces
SFC (Taigman et al, 2014)	4,030	4.4M	U	Facebook; usually as training set
CASIA-WebFace (Yi et al, 2014)	10,575	494,414	U	usually as training set
VGG face (Parkhi et al, 2015)	2,622	2.6M	U	usually as training set
MFC (Zhou et al, 2015)	20,000	5M	U	from web; usually as training set
MS-Celeb-1M (Guo et al, 2016)	1M	10M	U	usually as training set; largest public one
UMDFaces (Bansal et al, 2016)	8,277	367,888	U	annotated faces
Megaface 2 (Nech and Kemelmacher-Shlizerman, 2016)	672,057	4.7M	U	large dataset; usually as training set
VGGFace2 (Cao et al, 2017)	9,131	3.31M	U	pose,age,illumination,ethnicity,profession

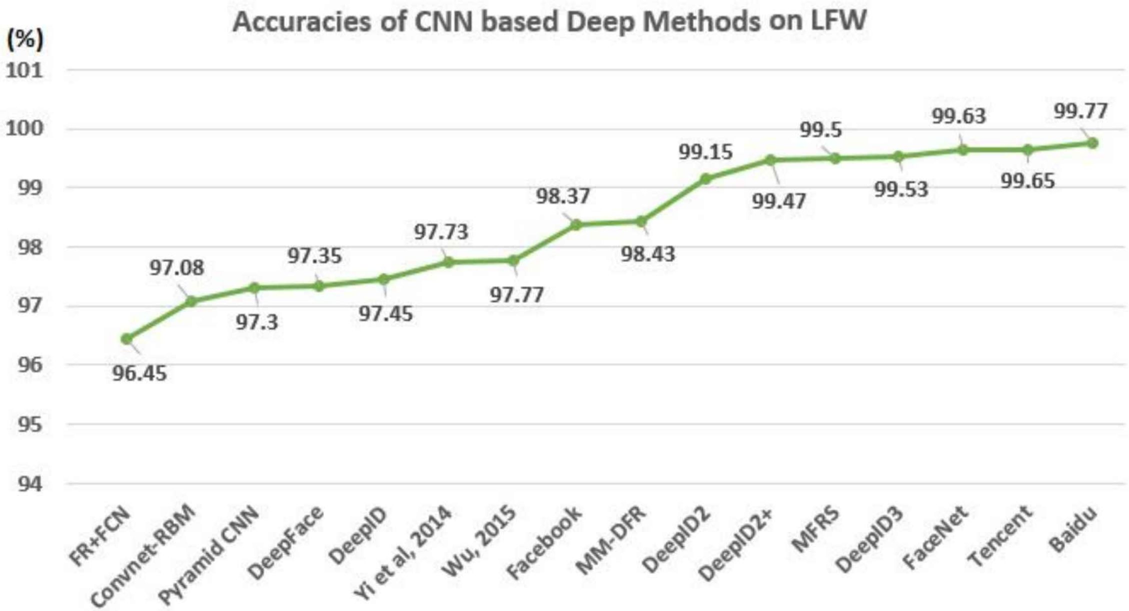
still+video →

training →

- Most datasets are public
- Some of them provided links for researchers to download from the Internet

- Several big datasets are private:
    - MSRA’s WDRef
    - Facebook’s SFC
    - MFC (Megvii Face Classification)

- LFW
  - can be viewed as a milestone dataset in which images are crawled from the Internet containing variations in pose, illumination, expression, resolution, etc.
  - Many research works have been focused on improving the performance on LFW
  - Recent advances, especially the CNN based face recognition, enabled close to 100% accuracy in LFW



However, the face recognition problem is far from being solved

- IJB-A
    - The performance of state-of-the-art face recognition systems are far less than satisfactory on a newly released dataset, IJB-A.
    - This benchmark is considered more challenging than LFW and has become more popular

- Several large and public available Training Datasets

  □ CASIA-WebFace

  http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html

  □ MS-Celeb-1M

  http://www.msceleb.org/

  □ CelebFaces

  http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

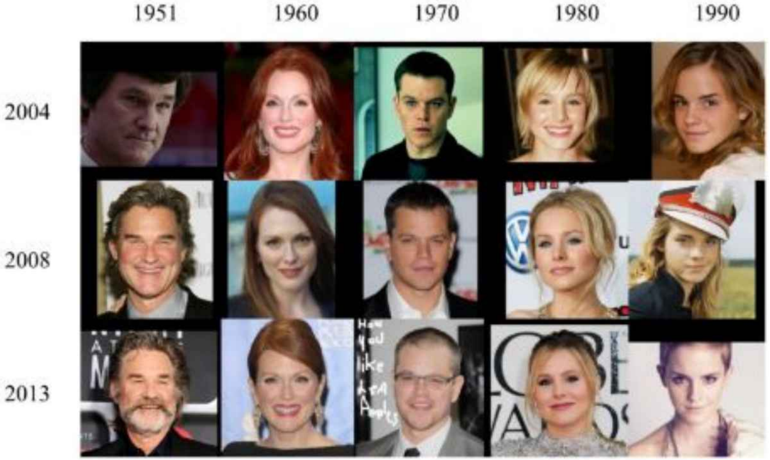
  □ VGGFace2

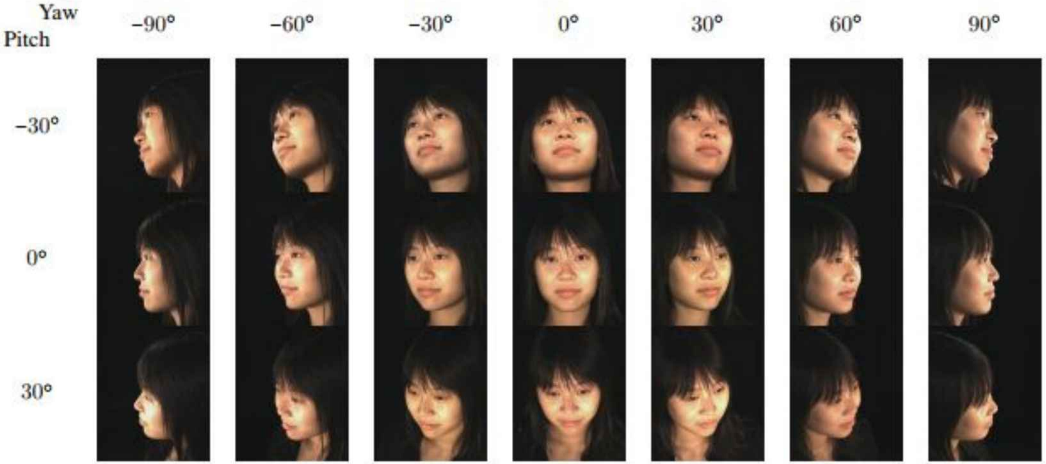
  https://www.robots.ox.ac.uk/~vgg/data/vgg_face2/vggface2.pdf

  □ Megaface 2, etc.

  http://megaface.cs.washington.edu/

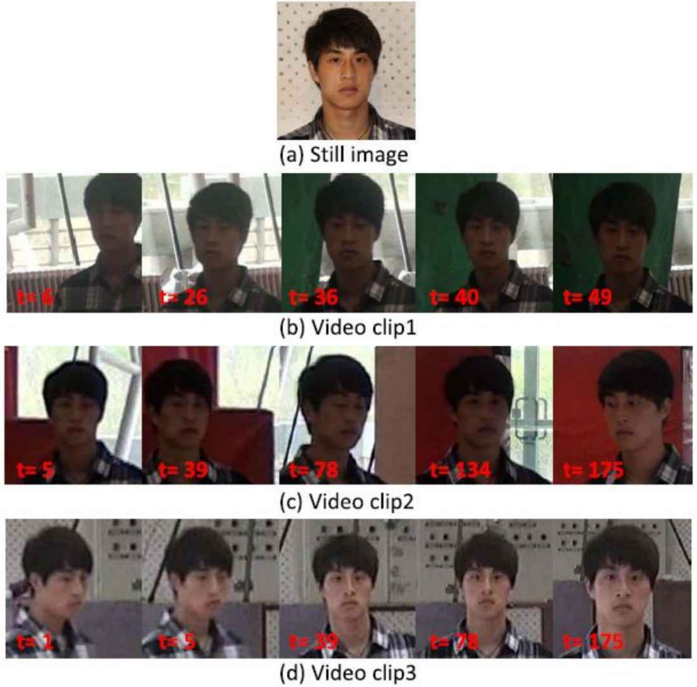
- Some datasets are used for specific tasks, e.g., age, pose, illumination, expression, and so on.





Video Face Databases

- Video based face recognition has also gained much attention
- Most are public available
- Still + Video faces: COX Face, PaSC, JANUS CS2 and IJB-A



(a) Still image

(b) Video clip1

(c) Video clip2

(d) Video clip3

- YouTube Faces (YTF) and YouTube Celebrities (YTC) are often used to test the recognition performance of various deep models.

YouTube Faces
Same / Not Same ?

still+video

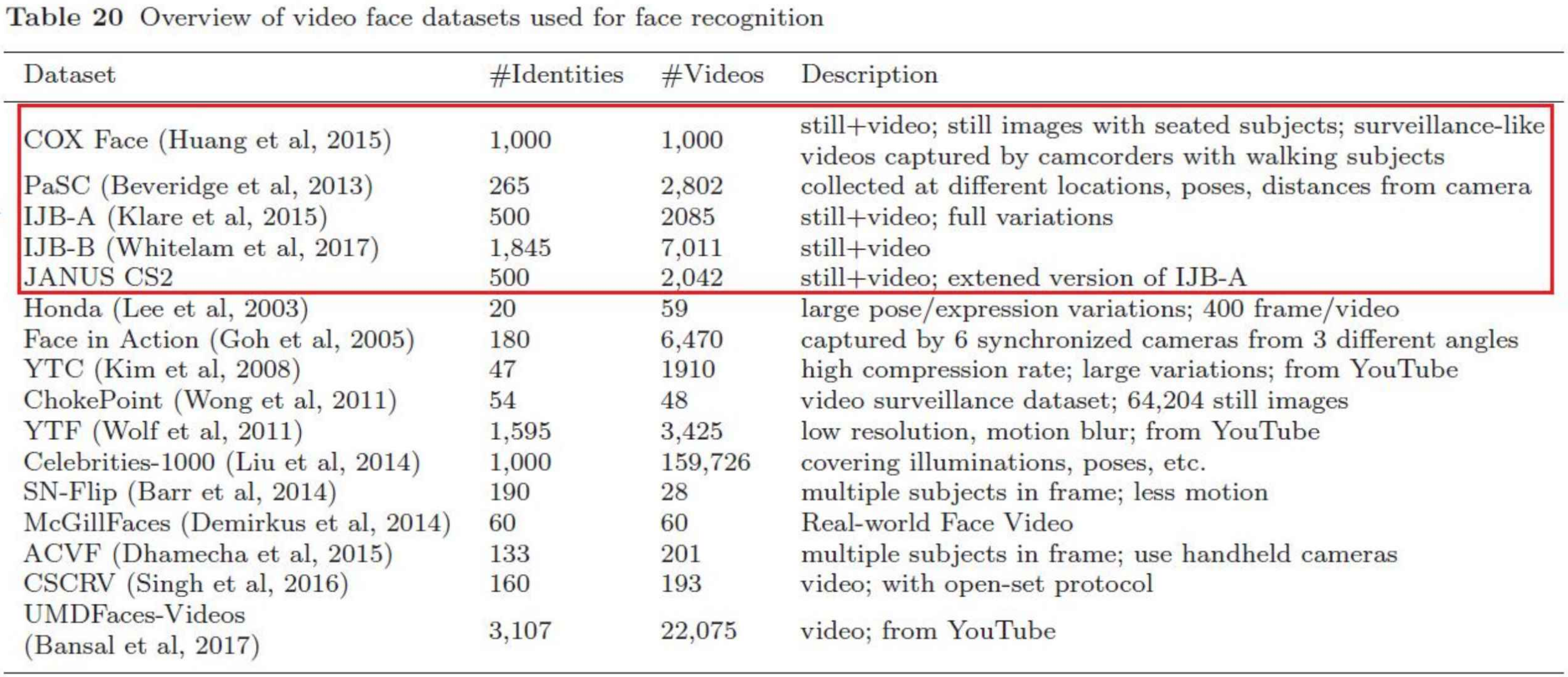


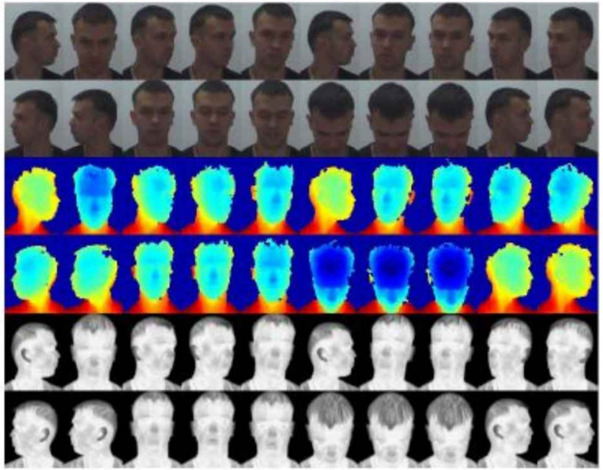
Table 20 Overview of video face datasets used for face recognition

Dataset	#Identities	#Videos	Description
COX Face (Huang et al, 2015)	1,000	1,000	still+video; still images with seated subjects; surveillance-like videos captured by camcorders with walking subjects
PaSC (Beveridge et al, 2013)	265	2,802	collected at different locations, poses, distances from camera
IJB-A (Klare et al, 2015)	500	2085	still+video; full variations
IJB-B (Whitelam et al, 2017)	1,845	7,011	still+video
JANUS CS2	500	2,042	still+video; extened version of IJB-A
Honda (Lee et al, 2003)	20	59	large pose/expression variations; 400 frame/video
Face in Action (Goh et al, 2005)	180	6,470	captured by 6 synchronized cameras from 3 different angles
YTC (Kim et al, 2008)	47	1910	high compression rate; large variations; from YouTube
ChokePoint (Wong et al, 2011)	54	48	video surveillance dataset; 64,204 still images
YTF (Wolf et al, 2011)	1,595	3,425	low resolution, motion blur; from YouTube
Celebrities-1000 (Liu et al, 2014)	1,000	159,726	covering illuminations, poses, etc.
SN-Flip (Barr et al, 2014)	190	28	multiple subjects in frame; less motion
McGillFaces (Demirkus et al, 2014)	60	60	Real-world Face Video
ACVF (Dhamecha et al, 2015)	133	201	multiple subjects in frame; use handheld cameras
CSCRV (Singh et al, 2016)	160	193	video; with open-set protocol
UMDFaces-Videos (Bansal et al, 2017)	3,107	22,075	video; from YouTube

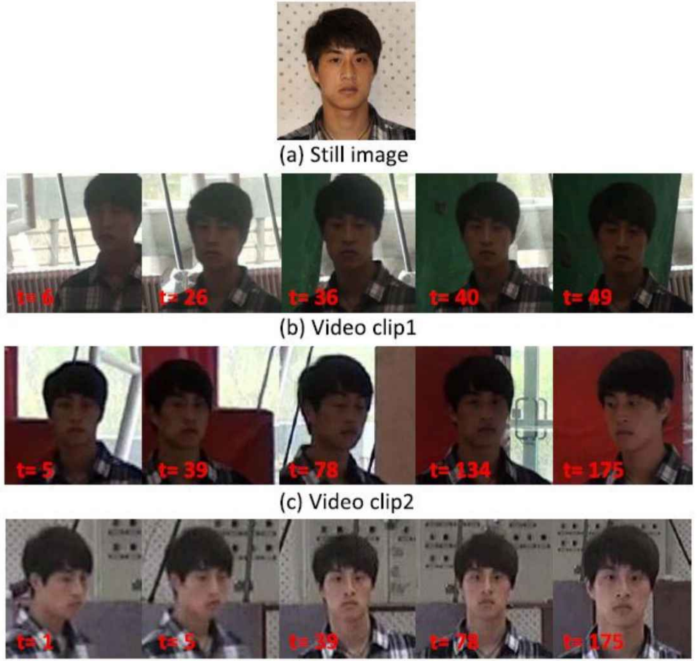
Heterogeneous Face Databases

For HFR, multi-modal data are needed

- Visible and Thermal
  - ✔ WSRI
  - ✔ Oulu-CASIA
  - ✔ CASIA NIR-VIS 2.0
  - ✔ Night Vision (NVESD)
- 3D
  - ✔ Curtin-Faces, LS3DFace, RGB-D-T



- Still and Video
  - ✔ COX-S2V
- photo and sketch
  - ✔ CUHK Face Sketch (CUFS)
  - ✔ CUHK Face Sketch FERET (CUFSF)



(a) Still image

(b) Video clip1

(c) Video clip2

(d) Video clip3

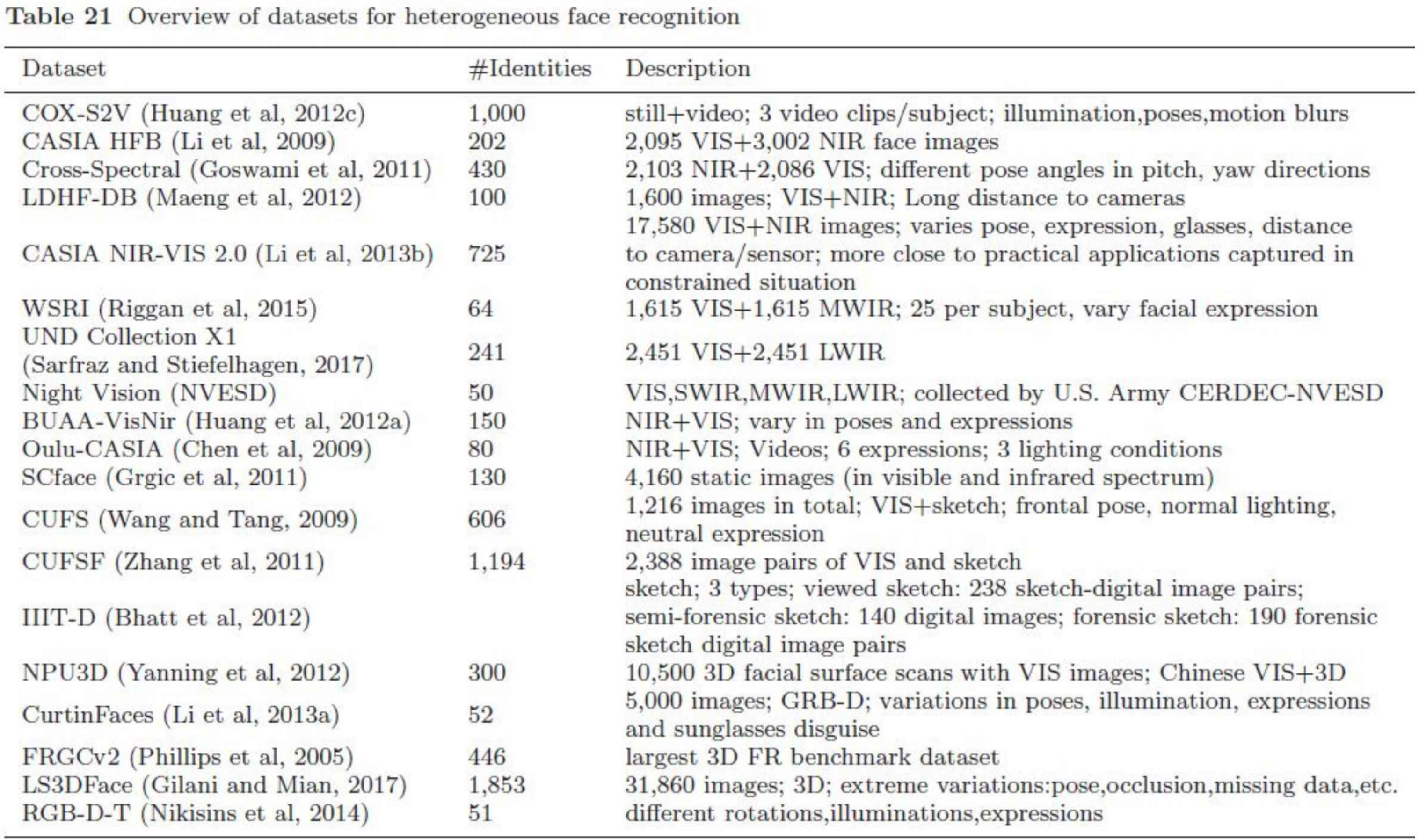


Table 21 Overview of datasets for heterogeneous face recognition

Dataset	#Identities	Description
COX-S2V (Huang et al, 2012c)	1,000	still+video; 3 video clips/subject; illumination,poses,motion blurs
CASIA HFB (Li et al, 2009)	202	2,095 VIS+3,002 NIR face images
Cross-Spectral (Goswami et al, 2011)	430	2,103 NIR+2,086 VIS; different pose angles in pitch, yaw directions
LDHF-DB (Maeng et al, 2012)	100	1,600 images; VIS+NIR; Long distance to cameras
CASIA NIR-VIS 2.0 (Li et al, 2013b)	725	17,580 VIS+NIR images; varies pose, expression, glasses, distance to camera/sensor; more close to practical applications captured in constrained situation
WSRI (Riggan et al, 2015)	64	1,615 VIS+1,615 MWIR; 25 per subject, vary facial expression
UND Collection X1 (Sarfraz and Stiefelhagen, 2017)	241	2,451 VIS+2,451 LWIR
Night Vision (NVESD)	50	VIS,SWIR,MWIR,LWIR; collected by U.S. Army CERDEC-NVESD
BUAA-VisNir (Huang et al, 2012a)	150	NIR+VIS; vary in poses and expressions
Oulu-CASIA (Chen et al, 2009)	80	NIR+VIS; Videos; 6 expressions; 3 lighting conditions
SCface (Grgic et al, 2011)	130	4,160 static images (in visible and infrared spectrum)
CUFS (Wang and Tang, 2009)	606	1,216 images in total; VIS+sketch; frontal pose, normal lighting, neutral expression
CUFSF (Zhang et al, 2011)	1,194	2,388 image pairs of VIS and sketch
IIIT-D (Bhatt et al, 2012)		sketch; 3 types; viewed sketch: 238 sketch-digital image pairs; semi-forensic sketch: 140 digital images; forensic sketch: 190 forensic sketch digital image pairs
NPU3D (Yanning et al, 2012)	300	10,500 3D facial surface scans with VIS images; Chinese VIS+3D
CurtinFaces (Li et al, 2013a)	52	5,000 images; GRB-D; variations in poses, illumination, expressions and sunglasses disguise
FRGCv2 (Phillips et al, 2005)	446	largest 3D FR benchmark dataset
LS3DFace (Gilani and Mian, 2017)	1,853	31,860 images; 3D; extreme variations:pose,occlusion,missing data,etc.
RGB-D-T (Nikisins et al, 2014)	51	different rotations,illuminations,expressions

Discussion of Challenges in Face Data

It is not trivial to get a huge amount of labeled face data

- Some strategies have been developed to address this issue
  - Minimize the need of data
    - Peng et al (2016)
    - used a modeling method to minimize the need of huge amount of data
  - Data synthesis
    - Lv et al (2017)
    - provided five data augmentation methods for face images, such as landmark perturbation, hairstyles, glasses, poses and illuminations synthesis
  - Instead of directly manipulating the input images, Leng et al (2017) performed virtual sample generation at the feature level for handling unbalanced training set

- For both still and video FR, large-scale datasets are important

- However, large-scale datasets often contain massive noisy labels, especially when automatically collected from the Internet

- Web-collected data could be unbalanced, where some subjects have much more faces than some others

- Unlike 2D images, 3D facial scans are not easy to crawl from the web

- With the progress in sensor technology, low cost 3D sensors may pave the way for multimodal systems, such as color and depth (RGB-D)

□Gilani and Mian (2017) proposed:

✔A method for generating a large corpus of labeled 3D face identities and their multiple instances for training the models

✔A protocol for merging the most challenging existing 3D datasets for testing

- In heterogeneous face recognition, the datasets are typically small

- Developing deep models is likely to overfit or underfit due to the small training set for HFR

- Exploring optimal methods to fit deep models for small-scale HFR datasets remains a critical problem

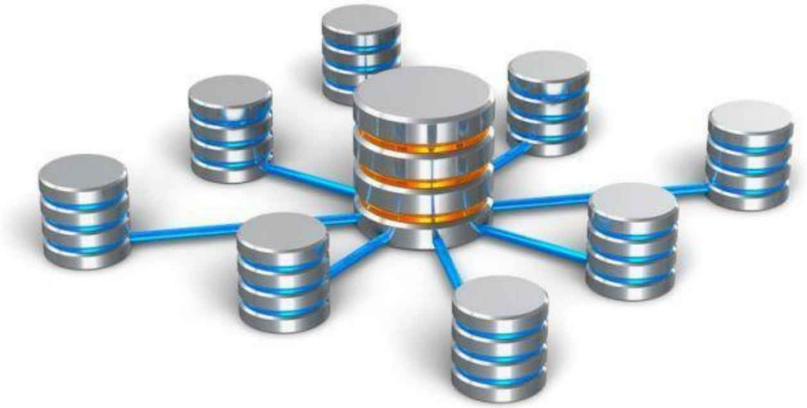
Conclusion

- We have presented a complete, comprehensive survey of face recognition methods based on deep learning
- Mainly focus on:
  - □ deep architectures
  - □ some specific recognition problems

- Deep learning techniques have been fully used for face recognition
- Have played important roles in addressing or circumventing challenges in FR
  - pose variations
  - illumination changes
  - facial expression, etc.

- Deep methods have also shown good performance to handle
  - RGB-D
  - Video
  - Heterogeneous face recognition

- A review of related face databases is given as well
    - Still Images
    - Videos Faces
    - Heterogeneous face



- Although the face recognition accuracies have been improved on many existing still image face datasets, here still exists a lot of challenges
  - For example, video face recognition is still a challenge
  - How to adapt a generic recognition system into a different domain is another open problem
  - the face image quality issue is a challenge for deep learning as well

Thank You and Questions