V-FACE: A Large-Scale Vietnamese Face Image Database in Unconstrained Environments

Minh Hieu Duong*, Thom Tran, Viet Anh Dao, Viet-Bac Nguyen, and Hoang Anh The Nguyen

***Abstract*— Researches have been conducted to exploit and collect details in every part of the human body to understand ourselves and improve many aspects of science and technology. Especially, the face is one of the most informative parts that is used commonly in computer vision such as face identification, face recognition, and landmarks detection, etc. A number of face image databases have been published and used worldwide in different technologies. In this paper, we introduce a new large-scale face image database for Vietnamese and describe a capturing system specifically designed to obtain the data. The V-FACE database contains more than 3 million high-quality images of more than 300 subjects and has a balanced ratio of genders. The database includes a variety of attributes, including different angles, lighting conditions, facial expressions, and occlusions with the combination of four types of accessories. Therefore, the V-FACE database can be exploited to be used in various tasks such as unconstrained face recognition, 3D face model reconstruction, face image frontalization, etc. In a variety of unconstrained environments, V-FACE provides a sufficient amount of data to improve the accuracy of face recognition technologies in practical scenarios. Additionally, with the information on different poses, we can reconstruct the human face in the most accurate details. There are several other Vietnamese face image datasets such as COMASK20 and VN-Celeb. V-FACE is a large-scale Vietnamese face database with a balance ratio in genders and age groups and different unconstrained environments such as illumination, facial expressions, poses and occlusions. We hope this database will improve many types of research in the world and especially in Vietnam.**

***Index Terms*—Face image database, cameras, capturing device, poses, illumination and occlusions.**

I. Introduction

Nowadays, there are cameras everywhere to capture moments of daily life or events which they can observe. There is a wide variety of poses, illuminations and occlusions, etc. that could occur in a face image. Studies about the human face have shown positive results in obtaining personal information or recognizing identities. With the development of deep learning and face image databases, deep learning algorithms' performance has improved significantly in technologies such as face recognition [1], age estimation [2], 3D face modeling [7], etc. A high-quality large-scale face image database is one important factor for improving deep learning models. A good face image database should have features such as quantity, quality, and accurate annotation. This paper introduces a large-scale Vietnamese face image database that assures the diversity and uniformity of data, correct annotation, and a capturing device that was systematically constructed the database. With V-FACE database, we can analyze factors that cause performance defects such as illumination/lighting changes, poses and accessories to technologies like face recognition.

V-FACE database can be used in many applications. For example, V-FACE database provides all the necessary elements to improve the quality of 3D face model reconstruction. Constructing a 3D face model needs a number of images of different poses of a human face. Illumination is also an important factor to create the best texture for the 3D face model. Furthermore, V-FACE database includes face images with light intensity, lighting source direction and accessories which are common in real-world scenarios. These factors are great contributions to improving the performance of face recognition system in unconstrained environments. In Vietnam, there has not been any large-scale face image database to analyze the characteristics of Vietnamese in unconstraned environments. There is a Vietnamese face image datasets denoted as COMASK20 dataset [16]. This dataset is only useful for face-masked detection. Therefore, this work will provide a Vietnamese unconstrained face image database denoted as V-FACE. V-FACE is inspired by K-FACE which is a face image database introduced by KIST [6]. V-FACE is a large-scale, diverse and well-balanced face image database with high-quality images. The rest of this paper is organized as followed: section II describes some novel face image databases, section III describes V-Face face image database in detail. The paper is concluded in section III.

II. Related works

A. *CASIA-WebFace*

CASIA-WebFace [15], a novel database, was introduced by Yi et al. This face image database consists of 494,414 images from 10,575 individuals collected from different sources in the Internet. The CASIA-WebFace face database is close to real-world scenarios because of the face images of this database

This paper was submitted on June 12, 2022.
This work is supported by Korea Institute of Science and Technology, Korea and Vietnam – Korea Institute of Science and Technology, Vietnam. (Corresponding author: Minh Hieu Duong).
Minh Hieu Duong currently is with Information Technology Division at Vietnam – Korea Institute of Science and Technology, Hoa Lac High-Tech Park, Thang Long Highway Km 29, Hanoi (e-mail: dmhieu@most.gov.vn).
Thom Tran currently is with Information Technology Division at Vietnam – Korea Institute of Science and Technology, Hoa Lac High-Tech Park, Thang Long Highway Km 29, Hanoi (e-mail: ttthom@most.gov.vn).
Viet Anh Dao currently is with Information Technology Division at Vietnam – Korea Institute of Science and Technology, Hoa Lac High-Tech Park, Thang Long Highway Km 29, Hanoi (e-mail: dvanh@most.gov.vn).
Viet-Bac Nguyen currently is with Information Technology Division at Vietnam – Korea Institute of Science and Technology, Hoa Lac High-Tech Park, Thang Long Highway Km 29, Hanoi (e-mail: nvbac@most.gov.vn).
Hoang Anh Nguyen The currently is with Information Technology Division at Vietnam – Korea Institute of Science and Technology, Hoa Lac High-Tech Park, Thang Long Highway Km 29, Hanoi (e-mail: anhnth@most.gov.vn).

are captured and collected in the unconstrained environments. Therefore, it has been widely used to train face recognition models, particularly with models based on deep learning methods. However, it does not provide enough information on the data such as pose direction, illumination condition, and correct annotation so it is difficult to analyze the factors that affect the recognition performance.

A. CelebA

CelebFaces Attributes Dataset (CelebA) is another large-scale face image dataset introduced for face representation and attribute prediction in the wild [11]. It consists of 202,599 images from 10,177 identities with five facial landmarks and 40 facial attributes that include hair color, gender, age and occlusion such as wearing sunglasses or not. This database has been widely used not only for face recognition but also for facial landmark detection and attribute prediction. Additionally, the CelebA database is used more for generative adversarial network-based methods and style learning.

B. VGGFace2

The previous version of VGGFace2 is the VGGFace database, which was released by Parkhi in 2015 [4]. The purpose of this database is to improve the accuracy of face recognition systems. Studies show that the number of identities and images of each subject in the training set is one of the most important factors which affects face recognition technologies's performance. To address this issue, VGGFace consists of 2.6 million images acquired from 2,622 people crawled on the web. The VGGFace2 database, which was extended from VGGFace, contains 3.31 million images of 9131 subjects, with an average of 362.6 images for each subject. Images are downloaded from Google Image Search and have a diversity in pose, age, illumination, ethnicity and even profession (e.g. actors, athletes, politicians, …). For ethnic balancing, it includes more Asian people than the previous version, VGGFace, although it is still limited.

C. K-FACE

K-FACE [6] is a large-scale database introduced by Yeji Choi et al.. from Korean Institute of Science and Technology (KIST). It contains a total of 17,550,000 images of 1,000 subjects with 17,550 images per person with various attributes. All images are captured using digital single-len reflex (DSLR) cameras, which provides high-quality images with resolution of 2592×1728 pixels. Images are captured with a hemisphere multi-camera system of 27 different cameras, 35 lighting conditions and multiple occlusions. These occlusions are five combinations of different types of accessories such as glasses, cap and mask. The K-FACE database has an equivalent ratio of age distribution, gender and a number of images per subject to overcome the limitations of the existing public databases with data imbalance.

D. Multi-PIE

Gross et al. releases a face image database denoted as Multi-PIE acquired under various illumination conditions and camera directions to consider the poses, lighting conditions and expressions [8]. It consists of over 750,000 images of 337 identities with 15 camera directions and 19 light directions. This database is an expanded version of the PIE database with high-resolution still images and geometrical calibration images. It also has more expressions, cameras, and recording sessions than PIE database. The Multi-PIE face image database has been improved compared to the PIE database by using a uniform, static background and live monitors allowing for constant control of the head position. This database has been used to improve the quality of researches for face recognition across pose, illumination and expression.

E. Labelled Faces in the Wild

Labelled Faces in the Wild (LFW) [10] was created and maintained by researchers at the University of Massachusetts, Amherst. This is a database of face images designed for researching in the problem of unconstrained face recognition. LFW contains 13,233 images of 5749 identities detected and centered by the Viola Jones face detector. These images are collected from different sources from the web. 1,680 subjects have two or more distinct photos in the dataset. The first version of the database consists of four different sets of LFW images and three different types of aligned images.

III. V-FACE DATABASE

A. Capturing Device

This elaborate hemispherical system is designed by KIST [6] to build the K-FACE database, which is shown in Fig. 1. This system was transferred to VKIST in a cooperation project between Korea and Vietnam. The system consists of 27 Full HD digital single-lens reflex cameras and 10 lighting devices installed at regular angular intervals. The type of camera we use was Canon EOS 1500D. All cameras acquire data at once by using a shutter button. The resolution of the output images is 2976 × 1984. As shown in Table I, we use 11 different lighting conditions by adjusting the intensities from 0 to 1,000 lux and changing the lighting direction to include partial illumination for shadows on the face. The lighting directions are included in Table I. The database provided labels such as: poses, lighting intensities and directions. These labels are systematically structured in folders, which is shown in Fig. 2.

B. Data Collection and Annotating

Using the device described in Section III.A, we develop a capturing process that lasts approximately an hour. All cameras capture images of a subject with and withoutaccessoriessuch as cap, sunglasses, wig, and mask, in various lighting environments. Subjects' age ranges from 20 to 50 years. They are students, workers, officers, researchers, etc. V-FACE database has been constructed since 2021. A subject joins a

TABLE I
THE CONFIGURATION OF V-FACE DATABASE

Label	Accessories
S001	Neutral (None)
S002	Neutral (None)
S003	Neutral (None)
S004	Sunglasses
S005	Mask
S006	(E01) Sunglasses + Mask (E02) Cap + Mask (E03) Wig (E04) Cap

Label	Expression
E01	Neutral
E02	Happy
E03	Sad
E04	Afraid
E05	Angry
E06	Surprised
E07	Disgust

Illumination			
Label	Lux	Direction	
		Vertical	Horizontal
L1	1000	All -Vertical: +30, -15 -Horizontal: +90, -90 (interval: 45)	
L2	400		
L3	200		
L4	150		
L5	100		
L16	400	All	+90
L17	200		
L24	400		0
L25	200		
L32	400		-90
L33	200		

Direction	+	-
Vertical	Up	Down
Horizontal	Right	Left

Pose				
Label	Direction		Label	Direction
	Vertical	Horizontal		Vertical	Horizontal
C1	0	+90	C14	+30	+45
C2		+75	C15		+30
C3		+60	C16		+15
C4		+45	C17		0
C5		+30	C18		-15
C6		+15	C19		-30
C7		0	C20		-45
C8		-15	C21	-15	+45
C9		-30	C22		+30
C10		-45	C23		+15
C11		-60	C24		0
C12		-75	C25		-15
C13		-90	C26		-30
			C27		-45

capturing process of six sessions with a total of 11,004 images per subject including 27 poses and seven facial expressions.

Total number of captured images has been 3,183,300 images from 300 subjects. Images are checked immediately after they were collected. This inspection process is manually carried out. Afterward, if there is any false data existed (e.g., blur images, closed eyes, moving subjects, not matching the lighting conditions due to synchronization delays), the data are recaptured to maintain the uniformity of the data distribution for all subjects and conditions.

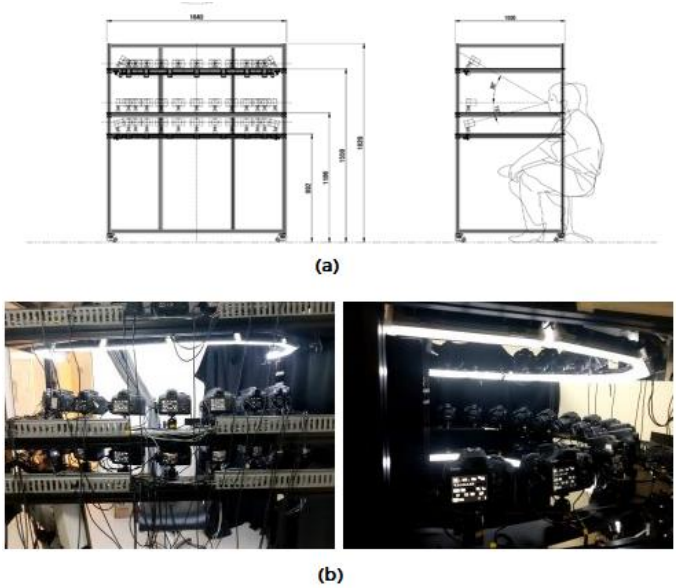

Fig. 1. Capturing device. (a) Concept design and (b) actual appearance.

C. Database Statistics

V-FACE was systematically constructed and maintained the balanced ratio of gender and an equal number of images per each subject. Therefore, this database overcome limitations faced with existing public databases because of imbalance in data quantity and attributes. Moreover, this is the first large-scale face image database whose subjects are Vietnamese people. It has a variety of attributes so researchers can study more about the feature of Vietnamese face. Futhermore, it contributes to researches on face image data of Asians. The data are acquired from subjects with age ranges from 20 to 50 but mostly in their 20 years old. The ratio of gender is almost the same that are 48% males and 52% females. Moreover, the number of images are the same in every subject and in each lighting condition to have an accurate statistical analysis of diversity in illumination, pose, facial expressions and occlusions (accessories).

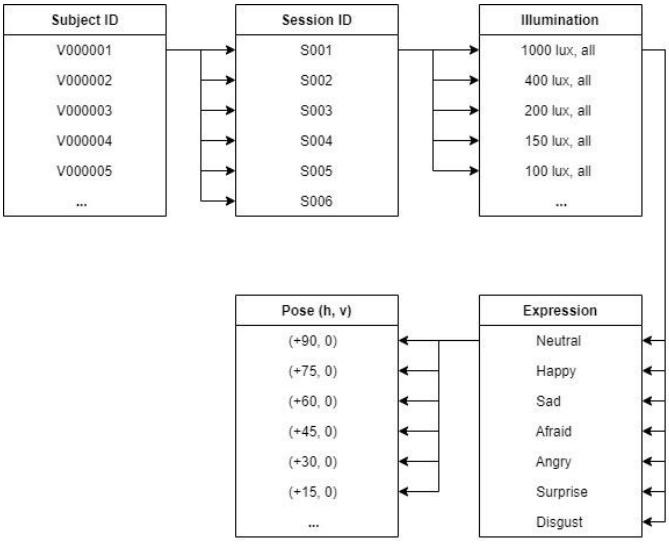

Fig. 2. Structure of the V-FACE database.

As shown in Fig. 2, the structure of the V-FACE image database consists of a total of five layers: subject identity, sessions identity, illumination, facial expression and pose. There are a total of six sessions. In the first three sessions, there is no occlusion (neutral). In the fourth session, subjects wear a sunglasses. In the fifth session, due to the situation that face mask are mandatory due to the coronavirus pandemic since 2019 (COVID-19), subjects wear a mask (of any colors). In th

Fig. 3. Sample images of accessories and expressions (neutral, happy, sad, afraid, angry, surprise, disgust).

sixth session, subjects wear different combinations of mask, hat, sunglasses and wig. The first fifth sessions consisted of full light conditions and facial expressions. The sixth session is unique and different from the capturing process of K-FACE. It only has two light conditions with four expressions representing where subjects wear combinations of accessories. There are a total of eleven lighting conditions and each lighting condition has seven facial expressions: neutral, happy, sad, afraid, angry, surprise and disgust. The last session is conducted in the maner that has only one facial expression that is neutral. Finally, there are 27 poses corresponding to 27 angles per facial expression. In each expression, there is also an image acquired from webcam which is located in front of the face of subjects to monitor the position of subjects during the capturing process. V-FACE is different from K-FACE in several ways. K-FACE has 35 lighting conditions and three facial expressions. Meanwhile V-FACE has 11 lighting conditions and seven expressions. V-FACE and K-FACE use different combinations of accessories. Sample images of all expressions, all poses and all lighting changes are shown in Fig. 3,. Fig. 4 and Fig. 5, respectively.

In conclusion, V-FACE image face database is a uniformly structured database of six capturing sessions, eleven lighting conditions, seven facial expressions and 27 poses. Therefore, the total number of images per subject is 10,611: 5 (capturing sessions) × 11 (lighting conditions) × 7 (facial expressions) × 27 (poses) + 4 (combinations of accessory conditions) × 2 (lighting conditions) × 1 (facial expression) × 27 (poses). Labels for all attributes are also provided, such as the first capturing session, which are denoted as S001, facial expressions, lighting conditions and poses, for all images.

IV. Applicable Areas of Research

A. 3D Face Reconstruction

Recovering the 3D shapes of human faces has been a common application in many studies such as photogrammetry [13], 3D modeling, etc. The usage of 3D face data in face analysis applications has received considerable attention recently. Reconstructing the 3D face model of a person from unconstrained 2D images has been a challenging task and has a variety of applications such as face recognition [14], face animation [3], etc. There is a surge of interest in 3D face reconstruction from a single image using Deep Convolutional Neural Networks (CNN). Because of this reason, 2D face images of different angles are needed to improve the performance of these methods. Multiple 2D images in different poses are good materials for training a Deep CNN model in 3D face reconstructions with multiple images. With each pose and correct annotations, researchers can analyze the difference between angles and occlusions when reconstructing the 3D face model. V-FACE provides a sufficiency of poses and several occlusions to improve the accuracy of the models in real-life scenarios, especially during the time of the Coronavirus pandemic. In animation, we can replace the face of the virtual characters with anyone's face generated by 3D face model reconstruction techniques. The more images and different angles of the subject are presented, the more accurate the 3D face model is. Therefore, in this case, V-FACE with face images acquired from various angles is very useful for many types of researches and applications.

B. Unconstrained Face Recognition

With the development and evolution of Deep Convolutional Neural Network, deep-learning-based image classification technologies and large-scale face image databases, the accuracy of face recognition has been significantly improved. It can surpass the human recognition performance in some benchmarks. However, there are still difficulties remaining in the real-life scenarios for unconstrained face recognition such as changing poses, illumination, facial expressions and occlusions [9]. Accordingly, V-FACE face image database contains face images with highly balanced distribution and

Fig. 4. Sample images of poses.

accurate labels for all information such as identity, pose, lighting conditions (intensity and direction), facial expressions (neutral, happy, sad, afraid, angry, surprise, confused) and different face accessories.

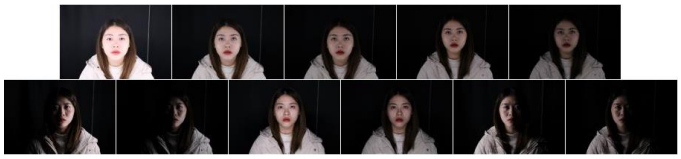

Fig. 5. Sample images of lighting changes.

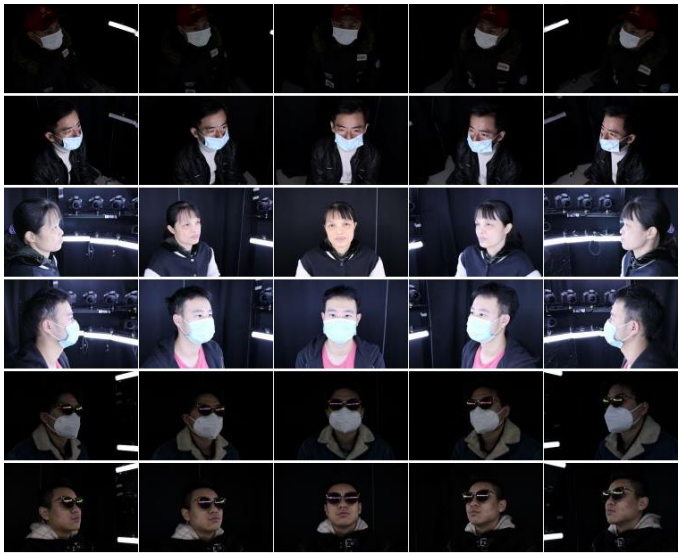

Fig. 6. Sample images of unconstrained face recognition.

These attributes are useful to improve the face recognition accuracy and develop a robust face recognition system in the real world. Different poses of a face image is one of the most important factors that can degrade the performance of face recognition system. To find a solution to this problem, major studies have been carried out based on pose-invariant face recognition and face frontalization. Face frontalization is the technique that converts the face images from different views to the frontal view. V-FACE not only includes images from different angles in the horizontal directions but also in the vertical direction. With respect to the changes in illuminations, the accuracy of face recognition can be degraded due to shadows or reflections caused by partial illumination or an extremely high/low light intensity. V-FACE contains the light intensity so we can study image-based lighting normalization under various lighting conditions. Additionally, V-FACE enables researches such as analyzing face recognition in the cases that face in the face image is deformed while facial expression changes. These changes are neutral, happy, sad, afraid, angry, surprised, confused, and occlusions. Since the outbreak of the COVID-19 pandemic, the demand for masked face recognition has increased significantly. Medical services need to know the trace of people who wear or don't wear a mask. The V-FACE database will help researchers to develop and improve the face recognition algorithms that exploit the information of the upper part of the head such as the forehead, eyebrow, or ear. Sample images in unconstrained environments, including different facial expressions, wearing accessories such as caps, sunglasses, and masks and with angles under different lighting conditions, are shown in Fig. 6. There are images acquired in extreme lighting conditions such as the intensity of illumination is close to zero, wearing different accessories and in the upper/lower angles.

C. Face Age Estimation and Aging Simulation

Age estimation and age simulation from face images, also known as age synthesis or age progression, have been applied in various domains. These domains include age-invariant face recognition, missing relative (specifically children) seeking, entertainment, etc. Age is a factor that constantly and permanently causes variations in facial appearance. Aging is an inevitable stochastic process. Facial aging adversely affects the performance of face recognition, face verification, and authentication. Age estimation is the technology that we label a face image with the exact real age or age group. Many works have been done to develop these techniques by either using good databases or bettering the models. Impressive aging results for natural face images and numerous face estimation and aging models have been provided [5, 12]. Nevertheless, face estimation and aging can be still improved in real-life scenarios where many different angles, illuminations and occlusions occur. These changes significantly affect the facial appearance and shape of the face. When a person becomes aging, the identity of that person changes and leads to the loss of the original person's identity. However, aging does not change the identity much according to small periods. In this regard, the V-FACE face image database enables researchers to estimate the age of identities in real life due to the fact that it consists of face images with ages of 20 to 55 years acquired at various lighting conditions, poses, facial expressions, and occlusions. With this database. we can exploit information such as face shape, skin texture and wrinkles that represent the information of age.

V. CONCLUSION

This paper introduces V-FACE, a large-scale Vietnamese face image database that consists of 3,183,300 images of 300 subjects. This database is constructed uniformly and systematically by considering each attribute: gender ratio and

data distribution per subject. It also contains images acquired in different unconstrained conditions such as different light conditions, occlusions and different poses. It includes accurate labels for all attributes listed above. So, this database is very useful in various researches and applications that include unconstrained face recognition, 3D face reconstruction, age estimation and age simulation. V-FACE is the first large-scale database of Vietnamese, so it is very valuable for researchers to study more about Vietnamese and South-Asian features and characteristics. In the future, we will expand the capturing device to be a sphere so the information on the back of a human is able to be acquired. Furthermore, we will balance the ratio of each age group and use more accessories and lighting conditions to expand the variety of unconstrained environments to be closer to real-life scenarios.

REFERENCES

[1] Patricia S. Abril and Robert Plant. 2007. The Patent Holder's Dilemma: Buy, Sell, or Troll? Commun. ACM 50, 1, pp. 36–44, January 2007. https://doi.org/10.1145/1188913.1188915.

[2] Sten Andler. Predicate Path Expressions. In Proceedings of the 6th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (San Antonio, Texas) (POPL '79). Association for Computing Machinery, New York, NY, USA, pp. 226–236, 1979. https://doi.org/10.1145/567752.567774.

[3] Chen Cao, Yanlin Weng, Stephen Lin, and Kun Zhou. 3D Shape Regression for Real-Time Facial Animation. ACM Trans. Graph. 32, 4, Article 41, July 2013. https://doi.org/10.1145/2461912.2462012.

[4] Qiong Cao, Li Shen, Weidi Xie, Omkar Parkhi, and Andrew Zisserman. VGGFace2: A Dataset for Recognising Faces across Pose and Age. pp. 67–74, 2018. https://doi.org/10.1109/FG.2018.00020.

[5] Bor-Chun Chen, Chu-Song Chen, and Winston H. Hsu. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. In Computer Vision – ECCV 2014, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, pp. 768–783, 2014.

[6] Yeji Choi, Hyunjung Park, Gi Pyo Nam, Haksub Kim, Hee seung Choi, Junghyun Cho, and Ig-Jae Kim. 2021. K-FACE: A Large-Scale KIST Face Database in Consideration with Unconstrained Environments. ArXiv abs/2103.02211 (2021).

[7] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3D Face Reconstruction With Weakly-Supervised Learning: From Single Image to Image Set. pp. 285–295, 2019. https://doi.org/10.1109/CVPRW.2019.00038.

[8] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. In 2008 8th IEEE International Conference on Automatic Face Gesture Recognition.pp. 1–8, 2008. https://doi.org/10.1109/AFGR.2008.4813399.

[9] Gary Huang, Manjunath Narayana, and Erik Learned-Miller. Towards unconstrained face recognition, pp. 1–8, 2008. https://doi.org/10.1109/CVPRW.2008.4562973.

[10] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, 2007. University of Massachusetts, Amherst.

[11] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2014. Deep Learning Face Attributes in the Wild., November 2014. https://doi.org/10.1109/ICCV.2015.425.

[12] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal Regression With Multiple Output CNN for Age Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[13] P. Schrott, A. Detrekoi, and K. Fekete. Photogrammetric network for evaluation of human faces for face reconstruction purpose. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives 39 (08 2012), 549–552, 2012. https://doi.org/10.5194/isprsarchives-XXXIX-B3-549-2012.

[14] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 1701–1708, 2014.

[15] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Li. 2014. Learning Face Representation from Scratch., November 2014.

[16] Vu, H.N., Nguyen, M.H. & Pham, C. Masked face recognition with convolutional neural networks and local binary patterns. Appl Intell 52, 5497–5512 (2022). https://doi.org/10.1007/s10489-021-02728-1.