Studies of Deep-Learning Face Recognition
7824519
Lin Ming Tai

1. Introduction

Deep-learning face recognition (FR) has been a major focus of research in recent years because of the stunningly high accuracy and possibility of wide real-world applications. FR provides a non-interacting method in verifying a person's identification compared to other methods of biosecurity (such as fingerprints, voice, footprint, or retina), which can speed up the progress and expand the scope. Nowadays, face recognition technology has become an indispensable part of our lives and widely applied to public security, ID verification, social media. On the other side, dissenting voice and controversy is appearing due to overly strong ability in facial recognition that may violate human rights and privacy. In this paper, we will explore the history of FR, the basic principle behind the deep-learning FR and its important innovation, and also the technical and social challenges.

2.History and development of face recognition

Face recognition can be largely divided into shallow(traditional) and deep methods. Former had been being the focus of people since the very beginning of development, but it had never been reliable and practical. It was thereby replaced by deep-learning method in recent years. We will step into the history of FR by looking at the development and characteristics of shallow method to explore why it is replaced by deep-learning method.

The concept of face recognition was firstly raised in the field of psychology in the 1950s, which is used to understand patient's mind by observing face expression, interpretation of emotion or perception of gestures (Bruner and Tagiuri, 1954). Computer scientists and engineers did not conduct any research on face recognition until 1960's. During 1964 and 1965, Woodrow Bledsoe, along with Helen Chan and Charles Bisson, was the first group of researchers worked on using computers to recognize human faces (Bledsoe, 1968). However, only few related publications were accessible to the public because they were funded by unnamed agency (Bledsoe, 1968). Bledsoe continues later his researches at Stanford Research Institute and successfully designed and implemented a semi-automatic FR system that can recognize face with the selection of face coordinates by humans in 1968 (Bledsoe, 1968). Other researches proceeded with improving the FR system and raising the accuracy with different approaches.

The first automated face recognition system was invented in 1973 by Kenade using a dedicated computer system and algorithms that could automatically extract sixteen facial parameters with

correct identification of 45-75% (Sakai et al., 1973). Eigenfaces, a dominant image processing technique in subsequent years, was proposed by L. Sirovich and M. Kirby in 1986 (Sirovich and Kirby, 1987). It represented an image in a lower dimension without losing much information and then reconstructing it based on Principal Component Analysis (Kirby and Sirovich, 1987). This approach established the foundation of many state-of-art approaches and the first industrial applications in 1990s (Carrera, 2010). Since the 1990s, face recognition area has received a lot of attention, with a noticeable increase in the number of publications.

Despite the hard work of face recognition community, FR was still infeasible in real life because of the technical limitation of shallow method against uncontrolled facial variations and scenes. Generally, the traditional and shallow method utilized only one- or two-layer representation, such as the distribution of the dictionary atoms or histogram of feature codes, to recognize face (Zhong, Ling and Wang, 2018). It is far not enough to process the complex nonlinear facial appearance variations. Different attempts of community were made in order to fix this problem. Holistic approaches FR community developed obtain low-dimensional representation and pre-calculated assumption through certain distribution assumptions, such as sparse representation (Moghaddam, Wahid and Pentland, 1998), or linear subspace (Wright et al., 2009) to predict the facial expression. But accuracy drops significantly once it is applied in an unexpected facial change (Zhong, Ling and Wang, 2018).

The research community also endeavored to separately improve the preprocessing, local descriptors, and feature transformation, but these the outcome and efficiency of these approaches are unacceptable (Zhong, Ling and Wang, 2018). Besides, most methods aimed to address one aspect of unconstrained facial changes only, such as lighting, pose, expression, or disguise. There was no integrated technique to address these unconstrained challenges integrally (Zhong, Ling and Wang, 2018). Consequently, shallow method of FR only achieved at most 95% of accuracy in Labeled Faces in the Wild (LFW) benchmark that is unusable in real world to deal with uncontrolled environment. (Chen, Cao and Wen, 2013). Inaccuracy and instability cause the impracticality of real-world application of shallow face recognition. Due to the weakness of shallow face recognition, scientists had been striving to look for other direction to overcome the obstacles. It gave a rise of deep-learning method FR.

3.1 The deep-learning face recognition system

Computer scientists and engineers were inspired to use deep learning methods to create a whole new generation of facial recognition technology after the huge success of AlexNet in ImageNet competition (Krizhevsky, Sutskever and Hinton, 2017).

According to "Deep Face Recognition: A Survey" by Wang and Deng (2018), deep learning methods in face recognition, such as convolutional neural networks, using a multi-layer cascade to substitute handmade features with powerful algorithms for unsupervised or semi-supervised learning features and extraction of hierarchical features. They learn multiple levels of representations that suit various abstraction levels. The levels form a hierarchy of concepts without the change of the face, pose, lighting, and expression changes. More specific texture features are acquired by second layer. Third layer starts identifying some special facial structures, such as a big mouth or high-bridged nose. In the fourth, the network output obtains enough information to explain certain facial attributes, which can make a special response to some clear abstract concepts such as smile, roar, and even blue eye. The initial layers of Deep convolutional neural networks (CNN) automatically learn the features extraction (e.g., Gabor Filter Based Feature Extraction) designed for years or even decades. And the later layers further learn higher-level abstraction. Finally, the combination of these higher-level abstraction represents facial identity with unprecedented stability.

Given this new technology, FR has a dramatic improvement in accuracy, compactness, and speed. It took less than ten years to advance from the impossibility of real-world application to wide application. The trend of deep-learning face recognition is unstoppable and posing great changes to our lives. There are four milestone systems on deep learning for face recognition that drove further innovations; they are DeepFace, DeepID, VGGFace, and FaceNet. In the following, characteristic and breakthrough of these methods would be examined only because they generally share the same concept of the utilization of CNN.

3.2 DeepFace

Commented [1]: Put this to last paragraph of 1.1 . what d oyou mean by shallow method?

Commented [MTL2R1]:

The following information is obtained from "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" by Taigman, Yang and Marc'Aurello (2014) unless specifically cited.

DeepFace is a deep learning face recognition system created by a research group formed by Yaniv Taigman, Ming Yang and Marc'Aurelio Ranzato in 2014. It plays a crucial role in the development of face recognition system because it is the first efficient and compact FR approach human performance on the unconstrained condition for the first time (DeepFace: 97.35% vs. Human: 97.53%) by using the deep learning method. It trains a nine-layer neural net with over 120 million connection weights, organized as a siamese network, and was trained on four million images uploaded by Facebook users. This system adopts Alexnet as the architecture, Softmax as the loss function.

DeepFace reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance. Benefit from the direct application toward RGB pixel values, the system is extremely compact in sheer contrast to the shift toward ten of thousands of appearance features in other systems. Besides, efficiency is also an advantage of it, the DeepFace runs at 0.33 seconds per image, accounting for image decoding, face detection and alignment, the feedforward network, and the final classification output.

In modern face recognition, the conventional pipeline consists of four stages: 1. Detect, 2. Align, 3. Represent, 4. Classify. DeepFace reprograms the step of face alignment and representation to achieve the unprecedentedly high accuracy and establish the foundation of future deep-learning FR.
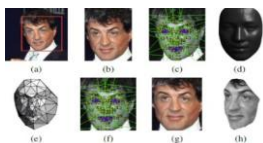
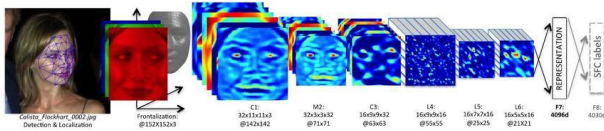**3.2.1 Face Alignment**

Fig.1

Research group of DeepFace has readopted 3D model, which has fallen out of favor in the FR community for the face alignment. They build a system that includes analytical 3D modeling of the face based on fiducial points, that is used to warp a detected facial crop to a 3D frontal mode (Frontalization). In general, the alignment process can be divided into three steps: 1. 2D alignment, 2. 3D alignment, 3. Frontalization.

2D Alignment: Firstly, the alignment process by detecting six fiducial points inside the detection crop, centered at the center of the eyes, tip of the nose and mouth as illustrated in Fig. 1(a). And apply it in several iterations to refine its output. At each iteration, fiducial points are extracted by a Support Vector Regressor (SVR) trained to predict point configurations from an image descriptor. Our image de-scripter is based on LBP Histograms. Iterating on the new warped image until there is no substantial change, eventually composing the final 2D similarity. This aggregated transformation generates a 2D aligned crop, as shown in Fig. 1(b).

3D Alignment: In order to align faces undergoing out-of-plane rotations, researcher uses a generic 3D shape model and registers a 3D affine camera, which is used to warp the 2D-aligned crop to the image plane of the 3D shape. After finding 67 fiducial points on the result, generate a 2D-aligned crop with their corresponding Delaunay triangulation, then transformed to the 2D-aligned crop image-plane, as shown in Fig. 1(c,d,e). Then mapping each part of the 2D image to a polygon in the mesh Fig. 1(g).

Frontalization: From the 3D mesh, frontaliztion is achieved by generating a 2D image using piecewise affine transformations on each part of the image. (Fig. 1(g)).

3.2.2 Representation

Researchers choose learning-based method rather than engineered features in representation step because the former offers an advantage that they can discover and optimize features for the specific task at hand and the performance exceeds the latter with the help of the availability of

rich data. A deep neural network is trained on a multi-class face recognition task to classify the identity of a face image. Also, the network architecture is customized by researchers and assumed that once the alignment is completed, the location of each facial region is fixed at the pixel level. In general, the whole process can be broken into preprocessing and main step.

Preprocessing step: A 3D-aligned 3-channels (RGB) face image of size 152 by 152 pixels is input to the network, then being passed through a convolutional layer with 32 filters of size 11*11*3, and the resulting 32 features maps are fed to a max-pooling layer (M2) which takes the max over 3x3 spatial neighborhoods with a stride of 2, separately for each channel. This is followed by another convolutional layer (C3) that has 16 filters of size 9x9x16. The purpose of these three layers is to extract low-level features, like simple edges and texture.

Main Step: The subsequent layers (L4, L5, and L6) are instead locally connected with different sets of filters at every location in the feature map. Eventually, the top two layers (F7 and F8) are fully connected that each output unit is connected to all inputs. Output of layer F7 will be used as a raw face representation feature vector while the output of the layer F8 is fed to a K-way softmax which a distribution over the K different classes (where each class corresponds to a person).

3.3 DeepID

The following information is obtained from "Deep Learning Face Representation by Joint Identification-Verification" by Yi, Wang and Tang (2014) unless specifically cited.

The DeepID, a series of systems (e.g. DeepID, DeepID2, etc.), found by Yi, Wang and Tang in 2014. Their system was first described much like DeepFace, the biggest difference is researcher of DeepID develop a better feature representation for reducing intra-personal variations while enlarging inter-personal variations by using both face identification and verification signals as supervision. It further improves the accuracy and even exceeds human with an accuracy of 99.15% (Human: 97.53%) on the LFW dataset. Compared with the best previous deep learning result on LFW, the error rate has been greatly reduced by 67%. This system adopts Alexnet as the architecture, contrastive as the loss function.

3.3.1 More effective representation

Face verification represents the computation of one-to-one similarity between the gallery and probe to determine whether the two images are of the same subject, whereas face identification computes one-to-many similarity to determine the specific identity of a probe face.

In the training stage, given an input face image with the identification signal, its DeepID2 features are extracted in the top hidden layer of the learned hierarchical nonlinear feature representation, and then mapped to one of a large number of identities through another function g. In the testing stage, the learned DeepID2 features can be generalized to other tasks (such as face verification) and new identities unseen in the training data. The supervisory identification signal tends to pull apart the DeepID2 features of different identities since they have to be classified into different classes. Therefore, the learned features would have rich identity-related or inter-personal variations. However, the identification signal has a relatively weak constraint on DeepID2 features extracted from the same identity, since dissimilar DeepID2 features could be mapped to the same identity through function g. This leads to problems when DeepID2 features are generalized to new tasks and new identities in a test where g is not applicable anymore.

Researchers solve this by using an additional face verification signal, which requires that every two DeepID2 feature vectors extracted from the same identity are close to each other while those extracted from different identities are kept away. The strong per-element constraint on DeepID2 features can effectively reduce the intra-personal variations. On the other hand, using the verification signal alone (i.e. only distinguishing a pair of DeepID2 feature vectors at a time) is not as effective in extracting identity-related features as using the identification signal (i.e. distinguishing thousands of identities at a time). Therefore, the two supervisory signals emphasize different aspects of feature learning and should be employed together. To characterize faces from different aspects, complementary DeepID2 features are extracted from various face regions and resolutions and are concatenated to form the final feature representation after PCA dimension reduction. Since the learned DeepID2 features are diverse among different identities while consistent within the same identity, it makes the following face recognition easier. Using the learned feature representation and a recently proposed face verification model, DeepID achieved the highest 99.15% face verification accuracy on the LFW dataset.

3.4 VGGFace

The following information is obtained from "Deep Face Recognition" by Parkhi, Vedaldi and Zisserman (2015) unless specifically cited.

The VGGFace was developed by Omkar Parkhi, et al. from the Visual Geometry Group (VGG) at the University of Oxford in 2015. The availability of large quantities of training data is critical for deep-learning facial recognition. However, this resource is usually owned by those Internet giants only. For example, research group of Facebook develops FaceDeep with 4.4 million images and 4.4 thousand identities (Taigman, Yang and Marc'Aurelio, 2014). Building a dataset this large is beyond the capabilities of most international research groups, particularly in academia. Therefore, VGGFace focus on building an algorithm which can collect large face dataset containing hundreds of example images for thousands of unique identities without the support of rich resources, and use this to train a very deep CNN model for face recognition that allowed them to achieve then state-of-the-art results on standard datasets. VGGFace's accuracy reaches 98.95% in the LFW dataset suggests that deep FR is possible for public without the support of great resources.

3.4.1 New method of data set collection

Researchers propose a multi-stage strategy to collect a large face dataset which contains hundreds of celebrities and public figures from the internet, then go through different filter to ensure the precision and purity of the image.

In order to use limited image resources from internet to generate a competitive dataset, more processes and filters are required to ensure the precision and purity of image. The whole work can be largely divided into six stages. In stage 1, researchers bootstrap and filter a list of candidate identity names and photos. Researchers collect photos and names of celebrities from IMDB to avoid the privacy issue and portrait rights, then passed to human annotators to sieve the candidate. More photos of those celebrities are gathered from Google and Bing in stage 2. After that, more than five million of image and two thousand unpurified images per person is acquired, filtering is thereby necessary. In stage 3 and 4, researchers use automatic filter to improve the purity and remove erroneous in each set using classifier, then remove the duplicate images using descriptor. After all those steps, stage 5 progress the final manual filtering. Human annotation is required to increase the purity (precision) of the data, by filtering 1.6 million of a total number of images to less than 1 million. The whole progress takes around 10 days.

3.5 FaceNet

The following information is obtained from "FaceNet: A Unified Embedding for Face Recognition and Clustering." by Schroff, Kalenichenko and Philbin (2015) unless specifically cited.

FaceNet was found by Florian Schroff, et al. at Google in 2015. The focus of this work is to raise the efficiency of face representation by presenting a new innovative unified system that allowed images to be encoded efficiently as feature vectors that allowed rapid similarity calculation and matching via distance calculations. The system achieved the state-of-the-art results with a new record accuracy of 99.63% in LFW dataset, lower the error rate in comparison to the best-publicized result by 30%.

Researchers adopt a new method for face verification, recognition and clustering that based on learning Euclidean embedding per image using a deep convolutional network. The network is trained such that the squared L2 distances in the embedding space directly correspond to face similarity: faces of the same person have small distances and faces of distinct people have large distances.

Besides, previous approaches based on deep network use a classification layer trained over a set of known face identities and then take an intermediate bottleneck layer as a representation used to generalize recognition beyond the set of identities used in training (Wang and Deng, 2018). In contrast, FaceNet directly trains output to be a compact 128-D embedding using a triplet-based loss function based on LMNN. This approach largely improves the directness and raise efficiency.

4. Technical challenge of deep learning face recognition

With the significant advancement in face recognition system, it successfully applied in our real world and became an indispensable part of our lives that assisting and helping us. However, we still have a long way to go to produce a "perfect" face recognition that we can wholly rely on it.

First, there is an insufficient dataset for the general application of FR. In order to increase the adaption of different scenes, diversity of images is needed for training the network. However, it is difficult to collect and label enough samples for innumerable scenes in real world, because

those images are collected on internet so bias and preference toward certain scenes and variables are unavoidable (Singh and Prasad, 2018). Even though deep domain adaption is found to reduce the algorithm bias against different races and scenes, there is still no general solution for all variables to minimize the likeliness of error (Wang and Deng, 2018).

Besides, researchers are still not able to have a grasp of deep learning and the neural network. Although neural network and deep learning are developed by scientist, they can't fully understand the causal chain and reason of outcome behind (Gong and Jain, 2017). The complexity and self-learning characteristics made accountability became impossible in deep FR. Many fundamental questions, such as what the identity capacity is and why the systems are easily fooled by adversarial samples, are still open(Wang and Deng, 2018). Once there is any unexpected bug in the system, in-depth knowledge can help researchers to fix and adjust according to the situation. Deep understanding of is helpful in improving FR.

Last but not least, security issue is also a matter of concern in developing FR system. Deep FR is not inevitable, hackers trying to develop and fool the system in different ways, such as presentation way, adversarial attack and template attack (Wang and Deng, 2018). Presentation attack subverts the face recognition system by presenting a facial biometric artifact. For instance, the 3D silicone mask with skin-like appearance and facial motion challenges current anti-spoofing methods (Patricia and de Vries, 2017). Future challenge of FR community is how to develop resolution and prevention of consistently upgraded attack.

5.Social challenge of face recognition

As mentioned above, face recognition has been applied widely and became an important part of our lives. But human still not get used to this new technology, many dissenting voices and controversies appeared.

Firstly, strong face recognition system may violate citizen's human rights and privacy. The appearance of robust face recognition system becomes a new trend of law enforcement to fight crime. For example, all Canadian international airports use face recognition as part of the Primary Inspection Kiosk program that compares traveler's faces to their photos stored on the ePassport to avoid illegal immigrants and criminals (Patel, 2018). On the other side, it may become the tool of government to surveil and control people. Over the last decade, the Chinese

Communist Party has built an unprecedented surveillant assemblage in the Xinjiang Uyghur Autonomous Region (XUAR) with the region's Uyghur Muslim minority as the chief target of augmented Party-state controls. The Chinese Party-state employs surveillance to delineate 'correct' thought and behavior among its citizens, and then persuade self-alignment with Party and Han-defined norms (Leibold, 2019); In Hong Kong, pro-democracy protesters are trying to topple the face recognition devices that they are afraid to be recognized and arrested due to their political orientations (SCMP, 2019). This clearly shows the white fear among citizens if this robust technology is used by the autocratic government.

Secondly, the face recognition technology might be abused. With the rapid advancement of face recognition in recent years, some new technologies are derived from it and could be utilized with bad intentions. For instance, Deepfake takes a person in an existing image or video and replace them with someone else's likeness using artificial neural networks (Güera and Delp, 2016). However, when this technology became available for the community in 2017, it became a tool for making altered photos and realistic fake videos which created pornography of people, often female celebrities whose likeness is typically used without their consent, or misrepresent well-known politicians in videos to create political distress (Shao, 2019). The stunning likeliness misleads the public and causes person involved distressed. But due to the defect of law among different countries, this trend became difficult to be terminated.

Thirdly, the system showed bias against people of color. Even though accuracy of face recognition system has exceeded human (>99%), the application on real world is not 100% guaranteed, especially on women and minorities. In 2018, Gender Shades, a seminal study led by MIT Media Lab researcher Joy Buolamwini, found that gender classification systems sold by IBM, Microsoft, and Face++ had an error rate as much as 34.4 percentage points higher for darker-skinned females than lighter-skinned males (Buolamwini and Gebru, 2018). This largely due to the dataset for training is collected randomly from the internet and far fewer images of women and people with dark skin than they are images of men and people with light skin (Buolamwini and Gebru, 2018). And while many of them are supposedly tested for fairness, those tests don't check performance on a wide enough range of faces (Benuwa et al., 2016). As mentioned above, face recognition became one of the important tools for law enforcement in

recent years, a minor mistake made by the system may cause unnecessary and serious trouble for a citizen.

6. Discussion

Without any doubt, the advancement of deep-learning face recognition in recent years is beyond our imagination. Its accuracy was developed from barely approaching human to exceed human and make the real-world application become possible in less than ten years. The trend of deep-learning FR is inevitable, and more resources will be input to this field. Rather than technical challenge, I believe we should prioritize solving social challenges to achieve sustainable development of this field.

In the light of FR system development, a comprehensive law system should be established. The government should legislate and amend the laws to prevent people from abusing this new technology to do illegal or immoral things. Meanwhile, the public should monitor the government to uses the technology properly and transparently, without compromising human rights and privacy. General education is needed to increase the awareness of general public to their interest and ethics of users.

Real-world application provides some important samples for researchers to improve the technology further. But it is not guaranteed to operate ideally, and it may cause big trouble especially when it is used in law enforcement. The way to balance the degree of utilization of AI and real-person should be considered, that would be a waste of resource and meaningless if there is too many real-person assistance with AI, but this may be insecure if AI replace human especially inaccuracy and bias still exists in FR. This may irreversibly mistake and trouble to citizens.

Easing public's concern and balancing the implication of FR is vital in gaining people's support. I argue that building a healthy environment for researchers without much dissenting voice is essential for helping them to get more supports and resources to overcome the technical challenges and obstacles we are facing now.

Reference list

Ballantyne, M. and Boyer, R. (1996). Woody bledsoe: His life and legacy. *AI Magazine*, 17(1), pp.7–20.

Benuwa, B.B., Zhan, Y.Z., Ghansah, B., Wornyo, D.K. and Banaseka Kataka, F. (2016). A Review of Deep Machine Learning. *International Journal of Engineering Research in Africa*, 24, pp.124–136.

Bledsoe, W. (1968). Semiautomatic facial recognition. *Technical report sri project 6693*.

Bruner, J. and Tagiuri, R. (1954). The percepton of people. *Handbook of Social Psycology*, 2(17).

Buolamwini, J. and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, pp.77–91.

Chen, W., Cao, X. and Wen, S. (2013). Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. *CVPR*, pp.3025–3032.

Daniel, N. (2019). Deep Learning Techniques for Face Recognition: A Review. *International Journal for Research in Applied Science and Engineering Technology*, 7(6), pp.1604–1607.

Deng, W., Hu, J., Zhang, N., Chen, B. and Guo, J. (2017). Fine-grained face verification: FGLFW database, baselines, and human-DCMN partnership. *Pattern Recognition*, 66, pp.63–73.

Goldstein, A.J., Harmon, L.D. and Lesk, A.B. (1971). Identification of human faces. *Proceedings of the IEEE*, 59(5), pp.748–760.

Goswami, G., Agarwal, A., Ratha, N., Singh, R. and Vatsa, M. (2019). Detecting and Mitigating Adversarial Perturbations for Robust Face Recognition. *International Journal of Computer Vision*, 127(6–7), pp.719–742.

Güera, D. and Delp, E. (n.d.). Deepfake Video Detection Using Recurrent Neural Networks. *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*.

Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), pp.103–108.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), pp.84–90.

Leibold, J. (2019). Surveillance in China's Xinjiang Region: Ethnic Sorting, Coercion, and Inducement. *Journal of Contemporary China*, pp.1–15.

Liu, C. and Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 11(4), pp.467–476.

Meder, R. (2016). Near Infrared Spectroscopy: Seeing the Wood in the Trees. *NIR news*, 27(1), pp.26–28.

Moghaddam, B., Wahid, W. and Pentland, A. (1998). Beyond eigenfaces: probabilistic matching for face recognition. *Automatic Face and Gesture Recognition*, pp.30–35.

Mohan, D., Sankaran, N., Tulyakov, S., Setlur, S. and Govindaraju, V. (2019). Significant Feature Based Representation for Template Protection.

Pan, L. and Peng, X. (2012). RETRACTED: Face Recognition with Learning-based Descriptor. *IERI Procedia*, 2(17), pp.114–119.

Parkhi, O., Vedaldi, A. and Zisserman, A. (2015). Deep Face Recognition.

Patel, V. (2018). *Airport Passenger Processing Technology: A Biometric Airport Journey*. Embry riddle aeronautical university.

Patricia and de Vries (2017). Dazzles, Decoys, and Deities: The Janus Face of Anti-Facial Recognition Masks. *Journal of Media & Communication*, 8(1), pp.72–86.

Preprint repository arXiv achieves milestone million uploads. (2014). *Physics Today*.

Ramachandra, R. and Busch, C. (2017). Presentation Attack Detection Methods for Face Recognition Systems. *ACM Computing Surveys*, [online] 50(1), pp.1–37. Available at: https://dl.acm.org/citation.cfm?id=3038924 [Accessed 31 May 2019].

Sakai, T., Kanade, T., Nagao, M. and Ohta, Y. (1973). Picture processing system using a computer complex. *Computer Graphics and Image Processing*, 2(3–4), pp.207–215.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering.

SCMP. (2019). *Why are Hong Kong protesters targeting lamp posts?* [online] Available at: https://www.scmp.com/tech/big-tech/article/3024997/why-are-hong-kong-protesters-targeting-lamp-posts.

Shao, G. (2019). *What "deepfakes" are and how they may be dangerous*. [online] CNBC. Available at: https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html.

Singh, S. and Prasad, S.V.A.V. (2018). Techniques and Challenges of Face Recognition: A Critical Review. *Procedia Computer Science*, 143, pp.536–543.

Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, [online] 4(3), p.519. Available at: https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-4-3-519 [Accessed 31 May 2019].

Taigman, Y., Yang, M. and Marc'Aurelio, R. (2014). *DeepFace: Closing the Gap to Human-Level Performance in Face Verification*. Menlo Park, CA, USA: Facebook AI Research.

Turk, M. and Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1), pp.71–86.

Wang, M. and Deng, W. (2018a). Deep Face Recognition: A Survey.

Wang, M. and Deng, W. (2018b). Deep visual domain adaptation: A survey. *Neurocomputing*, [online] 312, pp.135–153. Available at: https://arxiv.org/pdf/1802.03601.pdf [Accessed 7 Sep. 2019].

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Yi Ma (2009). Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), pp.210–227.

Yi, S., Chen, Y., Wang, X. and Tang, X. (2014). *Deep Learning Face Representation by Joint Identification-Verification*.

Zhen, L., Pietikainen, M. and Li, S.Z. (2014). Learning Discriminant Face Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2), pp.289–302.

Zhong, G., Ling, X. and Wang, L. (2018). From shallow feature learning to deep learning: Benefits from the width and depth of deep architectures. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(1).