Deepfake Forensics Using Recurrent Neural Networks

RAHUL U[1], RAGUL M[2],RAJA VIGNESH K[3], TEJESWINEE K[4]
[1] Student, Dept of CSE, Rajalakshmi Engineering College, Chennai, TN, India
[2] Student, Dept of CSE, Rajalakshmi Engineering College, Chennai, TN, India
[3] Student, Dept of CSE, Rajalakshmi Engineering College, Chennai, TN, India
[4] Assistant Professor, Dept of CSE, Rajalakshmi Engineering College, Chennai, TN, India

Abstract

As of late an AI based free programming device has made it simple to make authentic face swaps in recordings that leaves barely any hints of control, in what are known as "deepfake" recordings. Situations where these genuine istic counterfeit recordings are utilized to make political pain, extort somebody or phony fear based oppression occasions are effectively imagined. This paper proposes a transient mindful pipeline to automat-ically recognize deepfake recordings. Our framework utilizes a convolutional neural system (CNN) to remove outline level highlights. These highlights are then used to prepare a repetitive neural net-work (RNN) that figures out how to characterize if a video has been sub-ject to control or not. We assess our technique against a huge arrangement of deepfake recordings gathered from different video sites. We show how our framework can accomplish aggressive outcomes in this assignment while utilizing a basic design.

1. Introduction

The principal known endeavor at attempting to swap somebody's face, around 1865, can be found in one of the notorious por-characteristics of U.S. President Abraham Lincoln. The lithography, as found in Figure 1, blends Lincoln's head in with the collection of Southern government official John Calhoun. After Lincoln's assassination, interest for lithographies of him was incredible to the point that inscriptions of his head on different bodies showed up practically medium-term [27].

Ongoing advances [21, 42] have drastically changed the playing field of picture and video control. The democratization of present day devices, for example, Tensorflow [6] or Keras [12] combined with the open openness of the re-penny specialized writing and modest access to figure infras-tructure have moved this change in perspective. Convolutional autoencoders [38, 37] and generative ill-disposed system (GAN) [17, 7] models have made altering pictures and recordings, which used to be held to exceptionally prepared professional fessionals, an extensively open activity inside reach of practically any person with a PC. Cell phone and work area applications like FaceApp [1] and FakeApp [2] are based upon this advancement.

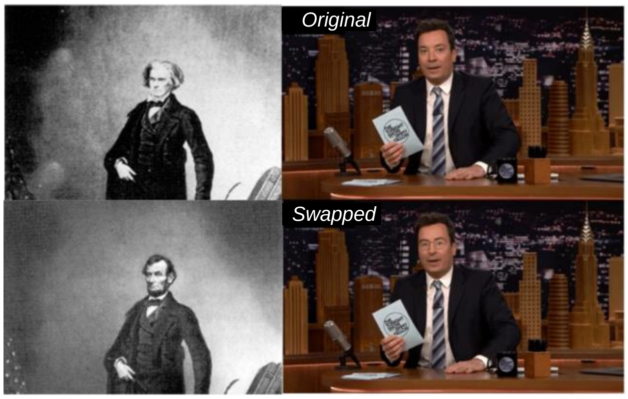

Figure 1. Face swapping isn't new. Models, for example, the swap of U.S. President Lincoln's head with legislator John Calhoun's body were created in mid-nineteenth century (left). Present day devices like FakeApp [2] have made it simple for anybody to create "deepfakes, for example, the one swapping the heads recently night TV has Jimmy Fallon and John Oliver (right).

FaceApp naturally produces exceptionally practical trans-arrangements of countenances in photos. It enables one to change face hairdo, sex, age and different qualities utilizing a cell phone. FakeApp is a work area application that enables one to make what are currently known as "deepfakes" recordings. Deepfake recordings are controlled videoclips which were first made by a Reddit client, deepfake, who utilized Ten-sorFlow, picture web indexes, web based life sites and open video film to embed another person's face onto prior recordings outline by outline.

Albeit some kind deepfake recordings exist, they stay a minority. Up until now, the discharged instruments [2] that produce deepfake recordings have been extensively used to make counterfeit superstar explicit recordings or retribution pornography [5]. This sort of erotic entertainment has just been restricted by locales including Reddit, Twitter, and Pornhub. The reasonable idea of deepfake recordings likewise makes them an objective for age of pedopornographic material, counterfeit news, counterfeit reconnaissance recordings, and pernicious lies. These phony recordings have just been utilized to make political strains and they are being considered by administrative substances [4].

As exhibited in the Malicious AI report [11], specialists in man-made brainpower ought to consistently think about the double use nature of their work, permitting abuse contemplations to impact investigate needs and standards. Given the seriousness of the malevolent assault vectors that deepfakes have caused, in this paper we present a novel answer for the recognition of this sort of video.

The primary commitments of this work are outlined as pursues. To start with, we propose a two-organize investigation made out of a CNN to separate highlights at the casing level pursued by a transiently mindful RNN system to catch fleeting incon-sistencies between outlines presented by the face-swapping process. Second, we have utilized an assortment of 600 recordings to assess the proposed strategy, with half of the recordings being deepfakes gathered from various video facilitating sites. Third, we show tentatively the viability of the de-scribed approach, which enables use to distinguish if a speculate video is a deepfake control with 94% more exactness than an arbitrary finder gauge in a decent setting.

Related Work

Digital Media Forensics. The field of digital media Advanced Media Forensics. The field of computerized media legal sciences expects to create advancements for the mechanized as-sessment of the honesty of a picture or video. Both component based [35, 16] and CNN-based [18, 19] uprightness investigation techniques have been investigated in the writing. For video-based advanced criminology, most of the proposed so-lutions attempt to identify computationally modest controls, for example, dropped or copied outlines [40] or duplicate move controls [9]. Systems that recognize face-based mama nipulations incorporate techniques that recognize PC gen-erated faces from common ones, for example, Conotter et al. [13] or Rahmouni et al. [33]. In biometry, Raghavendra et al. [32] as of late proposed to identify transformed countenances with two pre-prepared profound CNNs and Zhou et al. [41] proposed identification of two diverse face swapping controls utilizing a two-stream arrange. Of uncommon enthusiasm to professionals is another dataset by Ro¨ssler et al. [34], which has about a large portion of a million altered pictures that have been produced with include based face altering [38].

Face-based Video Manipulation Methods. Multi-ple approaches that target face controls in video se-quences have been proposed since the 1990s [10, 14]. Thies et al. shown the primary continuous appearance move for faces and later proposed Face2Face [38], a constant fa-cial reenactment framework, equipped for changing facial move-ments in various kinds of video streams. Options to Face2Face have additionally been proposed [8].

A few face picture blend methods utilizing profound learning have additionally been investigated as overviewed by Lu et al. [29]. Generative antagonistic systems (GANs) are utilized for maturing changes to faces [7], or to adjust face properties, for example, skin shading [28]. Profound element introduction [39] appears amazing outcomes in changing face characteristics, for example, age, fa-cial hair or mouth looks. Comparable consequences of characteristic introductions are accomplished by Lample et al. [24]. A large portion of these profound learning based picture combination strategies experience the ill effects of low picture goals. Karras et al. [22] show great combination of appearances, improving the picture quality us-ing dynamic GANs.

Recurrent Neural Networks. – Long Short Term Mem-ory (LSTM) systems are a specific sort of Recurrent Neural Network (RNN), first presented by Hochreiter and Schmidhuber [20] to adapt long haul conditions in information successions. At the point when a profound learning engineering is furnished with a LSTM joined with a CNN, it is ordinarily con-sidered as "somewhere down in space" and "somewhere down in time" individually, which can be viewed as two particular framework modalities. CNNs have made enormous progress in visual acknowledgment errands, while LSTMs are broadly utilized for long succession process-ing issues. As a result of the characteristic properties (rich vi-sual portrayal, long haul worldly memory and start to finish preparing) of a convolutional LSTM design, it has been completely read for other PC vision undertakings in-volving groupings (for example action acknowledgment [15] or human re-ID in recordings [30]) and has lead to noteworthy upgrades.

2. Deepfake Videos Exposed

Because of the way that FakeApp [2] produces the manipulated deepfake video, intra-outline irregularities and temporal irregularities between outlines are made. These video abnormalities can be misused to identify if a video under examination is a deepfake control or not. Let us quickly clarify how a deepfake video is produced to comprehend why these abnormalities are presented in the recordings and how we can misuse them.

2.1. Creating Deepfake Videos

It is notable that profound learning systems have been effectively used to improve the presentation of picture pressure. Particularly, the autoencoder has been applied for dimensionality decrease, smaller portrayals of pictures, and generative models learning [26]. In this manner, autoencoders can remove increasingly packed portrayals of pictures with a limited misfortune work and are required to accomplish preferable pressure execution over existing picture pressure benchmarks. The packed represen-tations or inactive vectors that current convolutional autoen-coders learn are the principal foundation behind the faceswap-ping abilities of [2]. The subsequent knowledge is the utilization of two arrangements of encoder-decoders with shared loads for the encoder systems. Figure 2 shows how these thoughts are utilized in the preparation and age stages that occur during the making of a deepfake video.

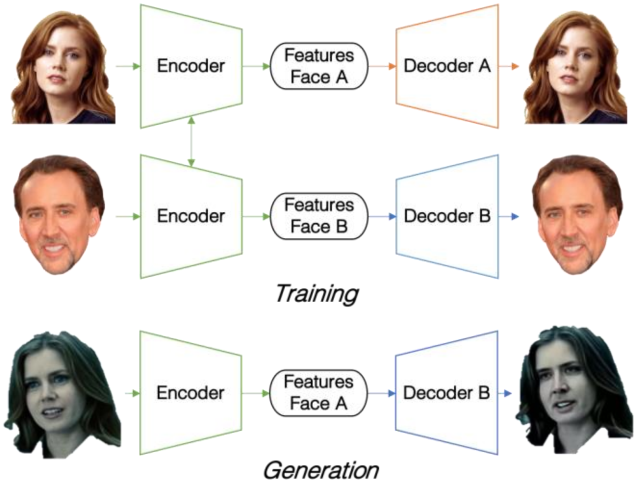

Figure 2. What makes deepfakes conceivable is figuring out how to drive both inactive countenances to be encoded on similar highlights. This is tackled by having two systems having the equivalent encoder, yet utilizing two distinct decoders (top). At the point when we need to do another faceswapp, we encode the information confront and disentangle it utilizing the objective face decoder (base).

2.1.1 Training

Two arrangements of preparing pictures are required. The principal set just has tests of the first face that will be supplanted, which can be separated from the objective video that will be manipu-lated. This initially set of pictures can be additionally reached out with pictures from different hotspots for increasingly practical outcomes. The second arrangement of pictures contains the ideal face that will be swapped in the objective video. To facilitate the preparation procedure of the autoencoders, the most effortless face swap would have both the first face and target face under comparable survey and light conditions. In any case, this is normally not the situation. Various camera sees, contrasts in lightning con-ditions or essentially the utilization of various video codecs makes it hard for autencoders to deliver reasonable faces under all conditions. This generally prompts swapped faces that are outwardly conflicting with the remainder of the scene. This edge level scene irregularity will be the primary component that we will abuse with our methodology.

It is additionally essential to take note of that on the off chance that we train two autoen-coders independently, they will be contradictory with one another. In the event that two autoencoders are prepared independently on various arrangements of countenances, their dormant spaces and portrayals will be dif-ferent. This implies every decoder is just ready to interpret a solitary sort of idle portrayals which it has picked up during the preparation stage. This can be overwhelmed by forc-ing the two arrangement of autoencoders to share the loads for the encoder systems, yet utilizing two distinct decoders. In this design, during the preparation stage these two systems are dealt with independently and every decoder is just prepared with faces from one of the subjects. Be that as it may, every single dormant face are created by the equivalent encoder which powers the encoder it-self to distinguish regular highlights in the two appearances. This can be effectively cultivated because of the regular arrangement of shared characteristics of every single human face (for example number and position of eyes, nose,

. . .).

2.1.2 Video Generation

At the point when the preparation procedure is finished, we can pass a dormant portrayal of a face created from the first subject present in the video to the decoder organize prepared on countenances of the subject we need to embed in the video. As appeared in Figure 2, the decoder will attempt to reproduce a face from the new subject, from the data comparative with the orig-inal subject face present in the video. This procedure is re-peated for each casing in the video where we need to do a faceswapping activity. It is essential to call attention to that for doing this casing level activity, initial a face finder is utilized to remove just the face district that will be passed to the prepared autoencoder. This is typically a second wellspring of scene irregularity between the swapped face and the re-set of the scene. Since the encoder doesn't know about the skin or other scene data it is extremely basic to have limit impacts because of a seamed combination between the new face and the remainder of the casing.

The third significant shortcoming that we misuse is inborn to the age procedure of the last video itself. Since the autoencoder is utilized casing by-outline, it is totally ignorant of any past produced face that it might have made. This absence of fleeting mindfulness is the wellspring of different inconsistencies. The most unmistakable is an inconsis-tent selection of illuminants between scenes with outlines, with prompts a flashing wonder in the face district com-mon to most of phony recordings. In spite of the fact that this phenomenon can be difficult to acknowledge to the unaided eye in the best physically tuned deepfake controls, it is effectively caught by a pixel-level CNN include extractor. The phe-nomenon of inaccurate shading steadiness in CNN-created recordings is a notable and still open research issue in the PC vision field [31]. Consequently, it isn't amazing that an autoencoder prepared with compelled information neglects to render illuminants accurately.

3. Recurrent Network for Deepfake Detection

In this segment, we present our start to finish trainable recurrent deepfake video recognition framework (Figure 3). The proposed framework is created by a convolutional LSTM structure for handling outline groupings. There are two basic parts in a convolutional LSTM:

1. CNN for outline highlight extraction.

2. LSTM for transient grouping investigation.

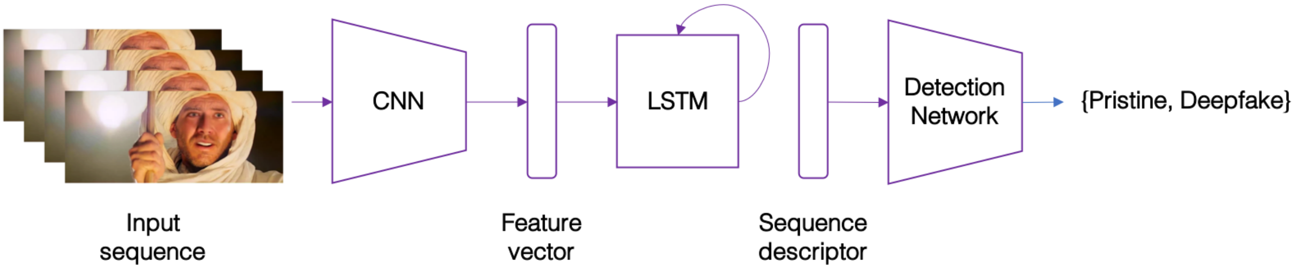

Figure 3. Review of our location framework. The framework learns and induces in a start to finish way and, given a video grouping, yields a likelihood of it being a deepfake or a perfect video. It has a convolutional LSTM subnetwork, for handling the info worldly grouping.

Given a concealed test succession, we get a lot of fea-tures for each casing that are produced by the CNN. After-wards, we link the highlights of different back to back casings and pass them to the LSTM for examination. We at last produce a gauge of the probability of the grouping being either a deepfake or a nonmaninpulated video.

4.1. Convolutional LSTM

Given a picture arrangement (see Figure 3), a convolutional LSTM is utilized to create a transient grouping descrip-tor for picture control of the shot casing. Focusing on start to finish learning, a reconciliation of completely associated lay-ers is utilized to delineate high-dimensional LSTM descrip-tor to a last recognition likelihood. In particular, our shal-low system comprises of two completely associated layers and one dropout layer to limit preparing over-fitting. The convo-lutional LSTM can be partitioned into a CNN and a LSTM, which we will depict independently in the accompanying para-diagrams.

CNN for Feature Extraction. Propelled by its accomplishment in the IEEE Signal Processing Society Camera Model Identi-fication Challenge, we embrace the InceptionV3 [36] with the completely associated layer at the highest point of the system expelled to straightforwardly yield a profound portrayal of each casing utilizing the ImageNet pre-prepared model. Following [3], we don't tweak the system. The 2048-dimensional element vec-tors after the last pooling layers are then utilized as the sequential LSTM input.

LSTM for Sequence Processing. Let us expect a se-quence of CNN highlight vectors of info outlines as information and a 2-hub neural system with the probabilities of the sequence being a piece of a deepfake video or an untampered video. The key test that we have to address is the deindication of a model to recursively process a succession in a mean-ingful way. For this issue, we resort to the utilization of a 2048-wide LSTM unit with 0.5 possibility of dropout, which is competent to do precisely what we need. All the more especially, during preparing, our LSTM model takes an arrangement of 2048-dimensional ImageNet include vectors. The LSTM is fol-lowed by a 512 completely associated layer with 0.5 possibility of

dropout. At last, we utilize a softmax layer to process the probabilities of the casing succession being either immaculate or deepfake. Note that the LSTM module is a middle of the road unit in our pipeline, which is prepared completely start to finish without the need of assistant misfortune capacities.

4. Experiments

In this segment we report the insights concerning our experiments. To begin with, we depict our dataset. At that point, we give subtleties of the trial settings to guarantee reproducibility and end up by breaking down the detailed outcomes.

4.1. Dataset

For this work, we have gathered 300 deepfake recordings from different video-facilitating sites. We further incorporate 300 additional recordings haphazardly chose from the HOHA dataset [25], which prompts a last dataset with 600 recordings. We chose the HOHA dataset as our wellspring of immaculate recordings since it contains a practical arrangement of grouping tests from renowned motion pictures with an accentuation on human activities. Given that an impressive number of the deepfake recordings are produced utilizing cuts from significant movies, utilizing recordings from the HOHA dataset further guarantees that the general framework figures out how to spot control highlights present in the deepfake recordings, rather than remembering semantic substance from the two classes of recordings present in the last dataset.

4.2. Parameter Settings

In the first place, we have utilized an irregular 70/15/15 split to create three disjoints sets, utilized for preparing, approval and test re-spectively. We do a reasonable parting, i.e., we do the split-ting first for the 300 deepfake recordings and afterward we rehash the procedure for the 300 nonmanipulated recordings. This guar-antees that every last set has precisely half recordings of each class, which enables use to report our outcomes as far as air conditioning curacy without considering predispositions because of the appearance recurrence of each class or the need of utilizing reg-ularizing terms during the preparation stage. As far as information preprocessing of the video successions, we do:

- Subtracting channel mean from each channel.
- Resizing of each edge to 299×299.
- Sub-arrangement inspecting of length N controlling the length of info grouping – N = 20, 40, 80 edges. This enables use to perceive what number of casings are fundamental per video to have a precise recognition.
- The analyzer is set to Adam [23] for start to finish train-ing of the total model with a learning pace of 1e−5 and rot of 1e−6.

4.3. Results

It is not unusual to find deepfake videos where the manipulation is only present in a small portion of the video (i.e. the target face only appears briefly on the video, hence the deepfake manipulation is short in time). To account for this, for every video in the training, validation and test splits, we extract continuous subsequences of fixed frame length that serve as the input of our system.

In Table 1 we present the performance of our system in terms of detection accuracy using sub-sequences of length $N = 20, 40, 80$ frames. These frame sequences are extracted sequentially (without frame skips) from each video. The entire pipeline is trained end-to-end until we reach a 10-epoch loss plateau in the validation set.

Model	Training acc. (%)	Validation acc. (%)	Test acc. (%)
Conv-LSTM, 20 frames	99.5	96.9	96.7
Conv-LSTM, 40 frames	99.3	97.1	97.1
Conv-LSTM(80 frames),	99.7	97.2	97.1

Table 1. Grouping consequences of our dataset parts utilizing video sub-arrangements with various lengths.

As we can see in our outcomes, with under 2 seconds of video (40 casings for recordings examined at 24 edges for each second) our framework can precisely foresee if the part being broke down originates from a deepfake video or not with an exactness more noteworthy than 97%.

5. Conclusion

In this paper we have presented a temporal-aware system to automatically detect deepfake videos. Our experimental results using a large collection of manipulated videos have shown that using a simple convolutional LSTM structure we can accurately predict if a video has been subject to manipulation or not with as few as 2 seconds of video data. We accept that our work offers a ground-breaking first line of guard to spot counterfeit media made utilizing the apparatuses portrayed in the paper. We show how our framework can accomplish compet-itive outcomes in this errand while utilizing a basic pipeline archi-tecture. In future work, we intend to investigate how to build the heartiness of our framework against controlled recordings us-ing concealed systems during preparing.

Affirmations: This material depends on look into supported by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under understanding number FA8750-16-2-0173. The U.S.Government is approved to imitate and appropriate reprints for Governmental purposes despite any copyright documentation subsequently. The perspectives and ends con-tained in this are those of the creators and ought not be deciphered as fundamentally speaking to the official arrangements or supports, either communicated or inferred, of DARPA, AFRL or the U.S. Government.

References

[1] Faceapp. https://www.faceapp.com/. (Accessed on 05/29/2018). 1

[2] Fakeapp. https://www.fakeapp.org/. (Accessed on 05/29/2018). 1, 2

[3] IEEE's Signal Processing Society - Camera Model Identification — Kaggle. https://www.kaggle.com/c/ sp-society-camera-model-identification/discussion/49299. (Accessed on 05/29/2018). 4

[4] The Outline: Experts fear face swapping tech could start an international showdown. https://theoutline.com/post/3179/deepfake-videos-are-freaking-experts-out?zd=1&zi=hbmf4svs. (Accessed on 05/29/2018). 1

[5] What are deepfakes & why the future of porn is terrifying. https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/. (Accessed on 05/29/2018). 1

[6] M. Abadi et al. Tensorflow: A system for large-scale machine learning. *Proceedings of the USENIX Conference on Operating Systems Design and Implementation*, 16:265–283, Nov. 2016. Savannah, GA. 1

[7] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. *arXiv:1702.01983*, Feb. 2017. 1, 2

[8] H. Averbuch-Elor et al. Bringing portraits to life. *ACM Transactions on Graphics*, 36(6):196:1–196:13, Nov. 2017. 2

[9] P. Bestagini et al. Local tampering detection in video sequences. *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pages 488–493, Sept. 2013. Pula, Italy. 2

[10] C. Bregler, M. Covell, and M. Slaney. Video rewrite: Driving visual speech with audio. *Proceedings of the ACM Annual*

Conference on Computer Graphics And Interactive Techniques, pages 353–360, Aug. 1997. Los Angeles, CA. 2

[11] M. Brundage et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv:1802.07228*, Feb. 2018. 2

[12] F. Chollet et al. Keras. https://keras.io, 2015. 1

[13] V. Conotter, E. Bodnari, G. Boato, and H. Farid. Physiologically-based detection of computer generated faces in video. *Proceedings of the IEEE International Conference on Image Processing*, pages 248–252, Oct. 2014. Paris, France. 2

[14] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister. Video face replacement. *ACM Transactions on Graphics*, 30(6):1–130, Dec. 2011. 2

[15] J. Donahue et al. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):677–691, Apr. 2017. 2

[16] H. Farid. *Photo Forensics*. MIT Press Ltd, 2016. 2

[17] I. Goodfellow et al. Generative adversarial nets. *Advances in Neural Information Processing Systems*, pages 2672–2680, Dec. 2014. Montréal, Canada. 1

[18] D. Güera, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, and E. J. Delp. A counter-forensic method for CNN-based camera model identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1840–1847, July 2017. Honolulu, HI. 2

[19] D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, and E. J. Delp. Reliability map estimation for cnn-based camera model attribution. *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, Mar. 2018. Lake Tahoe, NV. 2

[20] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, Nov. 1997. 2

[21] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5967–5976, July 2017. Honolulu, HI. 1

[22] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv:1710.10196*, Oct. 2017. 2

[23] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, Dec. 2014. 5

[24] G. Lample et al. Fader networks: Manipulating images by sliding attributes. *Advances in Neural Information Processing Systems*, pages 5967–5976, Dec. 2017. Long Beach, CA. 2

[25] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008. Anchorage, AK. 4

[26] Y. Liao, Y. Wang, and Y. Liu. Graph regularized autoencoders for image representation. *IEEE Transactions on Image Processing*, 26(6):2839–2852, June 2017. 2

[27] S. Lorant. *Lincoln; a picture story of his life*. Norton, 1969. 1

[28] Y. Lu, Y.-W. Tai, and C.-K. Tang. Conditional cyclegan for attribute guided face image generation. *arXiv:1705.09966*, May 2017. 2

[29] Z. Lu, Z. Li, J. Cao, R. He, and Z. Sun. Recent progress of face image synthesis. *arXiv:1706.04717*, June 2017. 2

[30] N. McLaughlin, J. M. d. Rincon, and P. Miller. Recurrent convolutional network for video-based person reidentification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1325–1334, June 2016. Las Vegas, NV. 2

[31] Y. Qian et al. Recurrent color constancy. *Proceedings of the IEEE International Conference on Computer Vision*, pages 5459–5467, Oct. 2017. Venice, Italy. 3

[32] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1822–1830, July 2017. Honolulu, HI. 2

[33] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen. Distinguishing computer graphics from natural images using convolution neural networks. *Proceedings of the IEEE Workshop on Information Forensics and Security*, pages 1–6, Dec. 2017. Rennes, France. 2

[34] A. Rössler et al. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv:1803.09179*, Mar. 2018. 2

[35] H. T. Sencar and N. Memon, editors. *Digital Image Forensics*. Springer New York, 2013. 2

[36] C. Szegedy et al. Rethinking the inception architecture for computer vision. *Proeedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, June 2016. Las Vegas, NV. 4

[37] A. Tewari et al. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1274–1283, Oct. 2017. Venice, Italy. 1

[38] J. Thies et al. Face2Face: Real-time face capture and reenactment of rgb videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395, June 2016. Las Vegas, NV. 1, 2

[39] P. Upchurch et al. Deep feature interpolation for image content changes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6090–6099, July 2017. Honolulu, HI. 2

[40] W. Wang and H. Farid. Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Transactions on Information Forensics and Security*, 2(3), 2007. 2

[41] P. Zhou et al. Two-stream neural networks for tampered face detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1831–1839, July 2017. Honolulu, HI. 2

[42] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2242–2251, Oct. 2017. Venice, Italy. 1