Predicting Heart Rate Variations of Deepfake Videos using Neural ODE

Steven Fernandes[1], Sunny Raj[1], Eddy Ortiz[2], Iustina Vintila[2], Margaret Salter[2], Gordana Urosevic[2], Sumit K. Jha[1]
[1]University of Central Florida, USA
[2]Royal Bank of Canada
{steven, sraj, jha}@cs.ucf.edu
{eddy.ortiz, iustina.vintila, margaret.salter, gordana.urosevic}@rbc.com

Abstract

Deepfake is a technique used to manipulate videos using computer code. It involves replacing the face of a person in a video with the face of another person. The automation of video manipulation means that deepfakes are becoming more prevalent and easier to implement. This can be credited to the emergence of apps like FaceApp and FakeApp, which allow users to create their own deepfake videos using their smartphones. It has hence become essential to detect fake videos, to avoid the spread of false information. A recent study shows that the heart rate of fake videos can be used to distinguish original and fake videos. In the study presented, we obtained the heart rate of original videos and trained the state-of-the-art Neural Ordinary Differential Equations (Neural-ODE) model. We then created deepfake videos using commercial software. The average loss obtained for ten original videos is 0.010927, and ten donor videos are 0.010041. The trained Neural-ODE was able to predict the heart rate of our 10 deepfake videos generated using commercial software and 320 deepfake videos of deepfakeTIMI database. To best of our knowledge, this is the first attempt to train a Neural-ODE on original videos to predict the heart rate of fake videos.

1. Introduction

Deepfake is an artificial intelligence method of video manipulation, which involves replacing the face of a person in a video with another person's face [1]. To create deepfake videos, an auto-encoder is trained using an input of a large collection of photos and condensing the photos into specific data points. A second auto-encoder performs the same condensing on stills of the face to be replaced in the video. The data points of the input photos are superimposed onto the data points from the video to replace the heads, based on each specific feature [2]. Deepfakes are becoming more prevalent and easier to implement, with the emergence of apps like FaceApp [3] and FakeApp [4] applications.

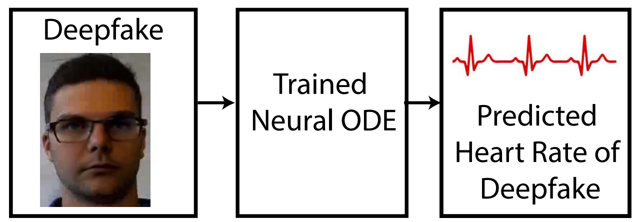


Figure 1: Neural-ODE prediction of heart rate variations

There are some valid concerns raised about the potentially damaging consequences of deepfake videos. First, while some applications of deepfake are harmless enough, it is increasingly being used to overlay unrelated faces onto the actors of pornographic videos [5]. Another danger of deepfake videos is the spread of false information and manipulated news through the Internet and specifically, social media [2]. With the large amount of false news already being spread via social media, it's important to identify further manipulations of the truth. Finally, scenes captured on video have long been accepted forms of evidence in legal proceedings; the increasing prevalence of deepfake videos has called into question the suitability of depending on video evidence to make legal verdicts [1]. To enable us to sort through original and fake videos, we must focus on developing sophisticated artificial intelligence methods for detecting deepfakes.

A recent study shows that the heart rate of fake videos can be used to distinguish original and fake videos [6]. However, obtaining a heart rate directly from fake videos is a time-consuming task. In this paper, we use the state-of-the-art Neural Ordinary Differential Equation (Neural-ODE) [7] solver to predict the heart rate of fake videos trained on original videos.

To the best of our knowledge, this is the first attempt to use state-of-the-art Neural-ODE solver to predict the heart rates variations of fake video obtained from commercial website [8].

2. Related Work

Forgeries and image spoofing have been traditionally studied by analyzing the pixels and frequencies of visual artifacts. With the advent of deep learning and generative adversarial networks, it has become easy to create dystopian situations related to fake images and videos.

2.1. Manipulating Faces in Videos

Face manipulation in videos was first introduced in the 90's. Video Rewrite was the first facial-animation based system, introduced by Bregler et al [9] in the late 1990's. It had the capability of automating the labeling and assembling tasks required to resync current footage with a new audio track. Koopman et al. [1] extracted the video frames containing the subject's face and split them uniformly into eight groups. An average value of photo response non uniformity (PRNU) [10] patterns were calculated for each of the eight groups. The PRNU patterns of the first group were compared with the PRNU patterns of the other seven groups, and the normalized cross-correlation scores were calculated for each video. The authors tested their methods on a small dataset, noting that the approach must be validated on a wider dataset before being widely accepted.

Recently, several techniques have been proposed using Generative Adversarial Networks (GANs) for generating fake faces in videos [11]. GANs are also used to alter the age [12]and skin color of a face [13], and facial hair and mouth expressions can be altered using feature interpolation [14]. GANs have also been used in several image synthesis techniques [15, 16, 17], and in the synthesis of high-quality images from low-resolution images [18]. The recent advancements in GANs have contributed to the development of deepfakes.

2.2. Image and Video based Digital Forensics

Traditionally, image inconsistencies are detected by finding compression artifacts [19] and distortions [20], as well as assessing image quality [21]. The color and noise distributions in original images can be investigated using specific networks [35, 48]. However, it is hard to find distortion, compression artifacts, and noises in synthetic images due to non-linearity [22]. Hence, feature-based techniques [23, 24] and convolutional neural networks (CNNs) [25] are used to find the authenticity of digital images. CNNs have been used to detect morphed facial images [26]. Recently, feature-based face detection was proposed by Thies et al. [27] on the dataset created by Rssler et al. [28], which contains around half a million edited images. Video manipulations are usually detected by finding duplicated or dropped frames [29], or copy and move manipulations [30].

2.3. Biological Signals

Subtle motion and color variations within videos can be observed [31, 32],enabling remote photoplethysmography (rPPG) [33, 34] and ballistocardiogram (BCG) [35] techniques for heart rate detection from facial videos. rPPG has proven to be more robust compared to BCG. There are several proposed methods for using rPPG, including using optical properties [36], Kalman filters [37], and extracting signal information from different facial areas [38, 39, 40, 33].

2.4. Recurrent Neural Network

Long Short Term Memory (LSTM) networks are the most popular recurrent neural networks (RNNs), introduced by Schmidhuber et al. [41]. They have been used for temporal analysis by Gera et al. [42], and in CNNs extracting frame features. Pre-processing is done by subtracting the mean and then resizing the frame to 299x299. The features from multiple frames are concatenated and given to LSTM for temporal analysis. However, missing data is a major issue in time series analysis. Typically it is addressed using generative models [43, 44], concatenating time stamp information of the input to RNN [45, 46, 47], or data imputation [48].

In this paper, we used Neural ODE, which is a recent generative approach for modeling time series. In this model, each time series is represented by a latent trajectory. The latent model was trained using a variational autoencoder [49, 50] by considering sequence-valued observations. We first created deepfake videos using the commercial deepfake video generating website, deepfakeweb.com [8]. The average loss obtained using deepfakeweb.com for ten original videos was 0.010925, and ten donor videos were 0.010041. Heart rate from original videos were extracted using three well known approaches: facial skin color variation [51], average optical intensity in the forehead [52], and Eulerian video magnification [31]. The Neural-ODE was trained using the heart rate obtained from the original videos. It was then used to predict the heart rate of deepfake videos obtained from deepfakeweb.com [8] and a publicly available DeepfakeTIMI database [53].

The key contributions of the paper are listed below.

- A new deepfake database was created using the commercial website [8] by considering ten original videos and ten donor videos from the COHFACE database.
- Predicting heart rate variations of deepfake videos using Neural-ODE trained on original videos from the COHFACE and publicly available VidTIMIT database.

To the best of our knowledge, this is the first attempt to use state-of-the-art Neural-ODE solver to predict the heart rates variations from deepfake videos [8].

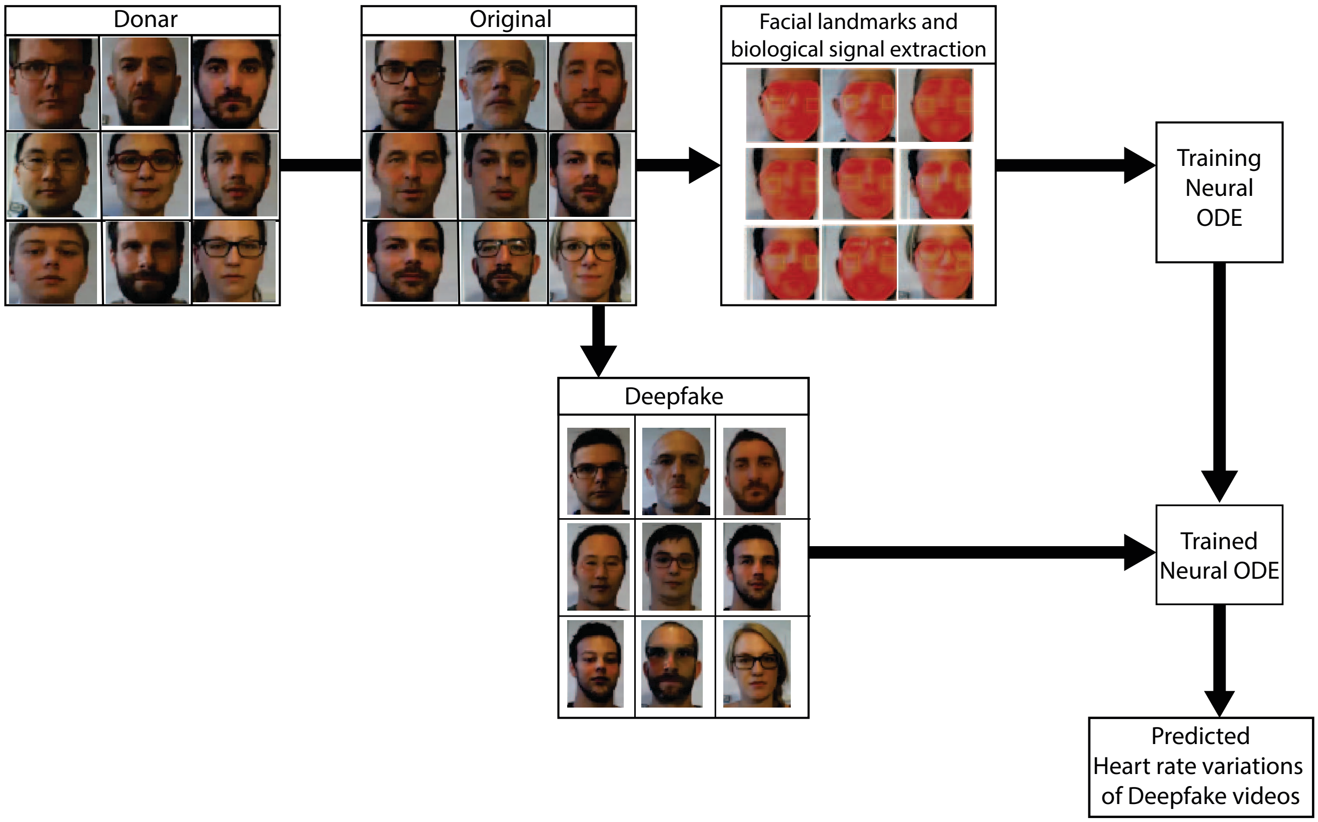


Figure 2: Block diagram of the proposed system used to predict the heart rate of deepfake videos using Neural-ODE

3. Proposed System

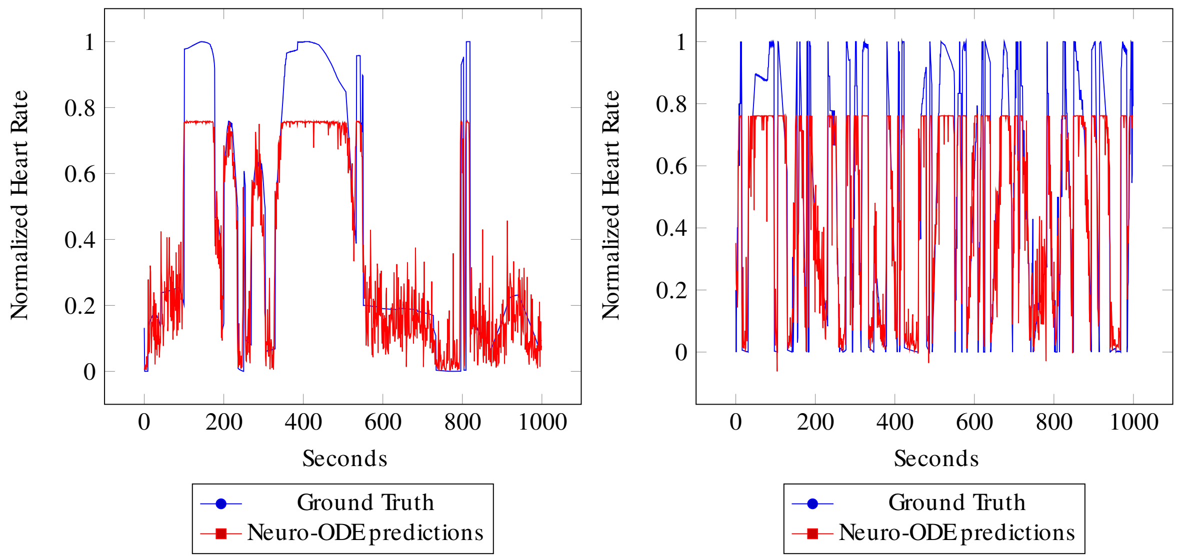
The four main steps are (i) Creating a deepfake dataset using a commercial website [8] (ii) Extracting heart rate from facial videos (iii) Training Neural-ODE using heart rate from original videos (iv) Predicting heart rates of deepfake videos using trained Neural-ODE.

3.1. Deepfake Databases

We considered twenty videos from the COHFACE database and uploaded them to commercial website [8] to create ten deepfake videos. Each deepfake video was created by considering an original video and a donor video. The resolution of the original video was 640x480 pixels, and the frame rate was 20Hz. The original and donor inputs and deepfake output from deepfakesweb.com are shown in Fig. 6. We also considered DeepfakeTIMI database [53], containing 320 videos from 32 subjects (10 videos per subject). The image resolution for videos was 128x128. The corresponding 320 original videos were constructed by concatenating the frames from VidTIMIT database.

3.2. Detecting Heart Rate from Facial Videos

We extracted the heart rates from the original videos using three known approaches: (i) measuring facial skin color variation caused by blood flow [51]; (ii) measuring average optical intensity in the forehead [52]; and (iii) magnifying and processing temporal changes in the color using Eulerian method [31]. In the facial skin color variation approach, the facial landmarks were detected using dlib [54], and ROI was obtained. Average RGB values for all the frames containing the ROI was obtained. Fast Fourier Transform was applied to obtain the heart rate [51]. In the optical intensity approach, the forehead region was isolated from the face, and the average optical intensity [52] in this region was used to detect the heart rate. With standard lighting conditions and considerably less noise caused by motion, a stable heart rate was obtained after 15 seconds. After the stable heart rate was obtained, phase variation with respect to frequency was computed. In Eulerian method, the color values at a given spatial location is amplified within a specific range of temporal frequency band. The amplification indicates that the changes in the redness are more significant as the blood flows into the facial region.



(a) (b)

Figure 3: Min-max normalized heart rate obtained from skin color variation (ground truth) and Neural-ODE (predictions) on: (a) Our deepfake videos (b) DeepfakeTIMI database videos

3.3. Training Neural-ODE

The heart rates obtained from the original videos using the three approaches discussed in section 3.2 were normalized using min-max normalization and applied to Neural-ODE for training. The Neural-ODE was trained separately using the ten original videos from COHFACE and the 320 original videos from VidTIMIT. Among the 320 original videos from VidTIMIT database, the videos of poor quality were automatically discarded. The steps involved in training the NeuralODE are listed below:

- The training data was split using the sliding window approach with time steps of 5.
- The RNN encoder was executed over the time series data of the heart rate, obtained from optical intensity and skin color variation approaches.
- For the posterior, validate the parameters over.

$$q\left(\boldsymbol{z}_{t_0} | \{\boldsymbol{y}_{t_i}, t_i\}_i, \phi\right) = \mathcal{N}\left(\boldsymbol{z}_{t_0} | \mu_{\boldsymbol{z}_{t_0}}, \sigma_{z_0}\right) \quad (1)$$

  where, $\mu_{\boldsymbol{z}_0}, \sigma_{z_0}$ are from the hidden states of $RNN\left(\{y_{t_i}, t_i\}, \phi\right)$.

- The isotropic unit Gaussia was sampled using the reparameterization trick.
- The variational autoencoder was built with Adam optimizer having learning rate of 0.00001.

3.4. Predicting the Heart Rate of Deepfake Videos using NeuralODE

For prediction of heart rate from deepfake videos, the trained Neural-ODE is given our 10 deepfake videos and 320 deepfake vidoes from the DeepfakeTIMI database. Among the 320 deepfake videos from the DeepfakeTIMI database, the videos of poor quality were automatically discarded. The layers in the encoder, decoder models of LSTM and variational autoencoder (VAE) of our prediction network are tabulated in Table 1.

Model	Layers
Encoder	Input Layer
	LSTM-1
	Dense Layer
Decoder	Input Layer
	RepeatVector
	LSTM-1
	LSTM-2
VAE	Input Layer
	LSTM-1
	Dense Layer-1
	Dense Layer-2
	LSTM-3
	LSTM-4

Table 1: Encoder, Decoder Layers of LSTM and VAE

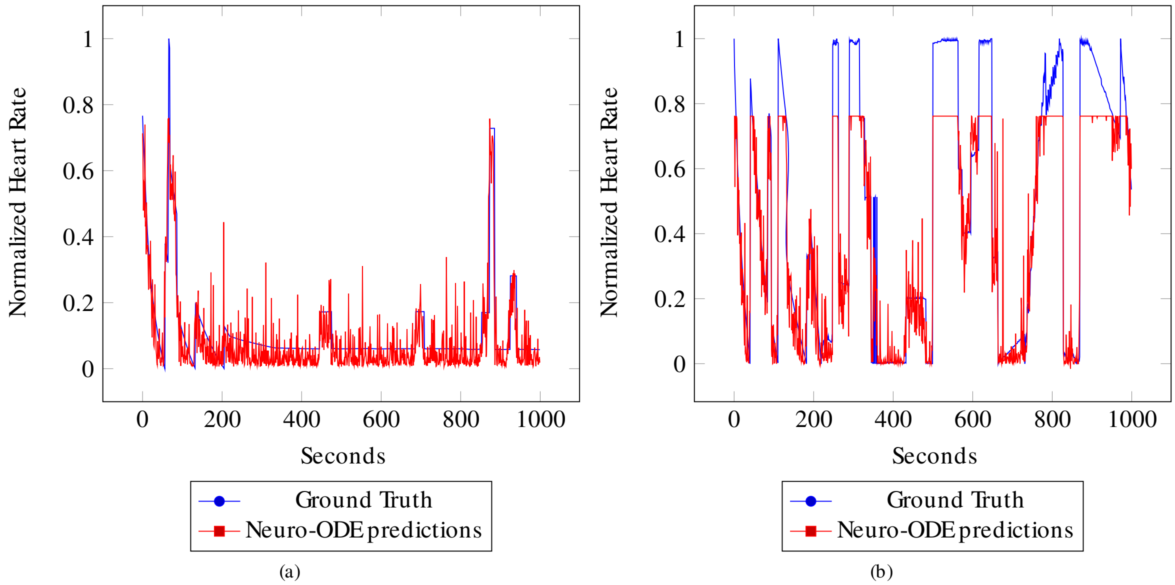


Figure 4: Min-max normalized heart rate obtained from optical intensity (ground truth) and Neural-ODE (predictions) on: (a) Our deepfake videos (b) DeepfakeTIMI database videos

4. Results and Discussions

4.1. Detecting Heart Rate from Facial Videos

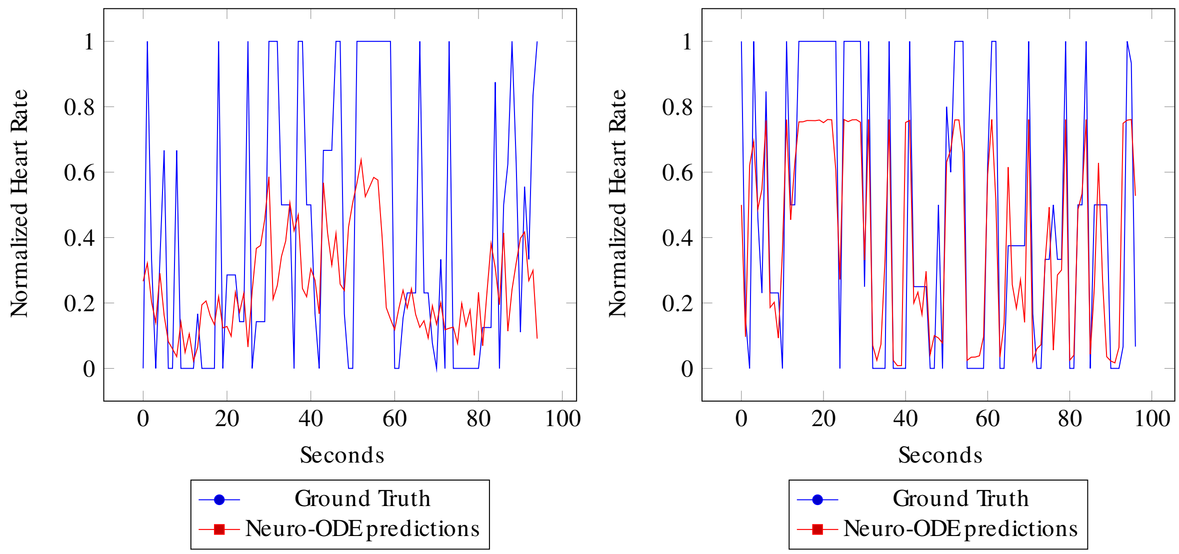
We considered 20 videos from the COHFACE dataset. This included ten original videos and ten donor videos. The 20 videos were uploaded to commercial deepfake video generating website [8]. To obtain the deepfake videos, 4 hours of GPU cloud usage was purchased. The losses obtained for the ten original and ten donar videos are tabulated in Table 2. The average loss obtained for ten original videos was 0.010927, and for the ten donor videos, 0.010041. Three well known approaches: facial skin color variation, average optical intensity and Eulerian video magnification are used to extract the heart rate of original videos. The Neural-ODE is trained using the min-max normalized heart rate obtained from the three approaches for 10 original videos from COHFACE database and 320 original videos of VidTIMIT database. To obtain the ground truth, the min-max normalized heart rate is again obtained using the three approaches for our 10 deepfake videos and 320 deepfake videos from the DeepfakeTIMI database. Among the 320 original videos from VidTIMIT database and 320 deepfake videos from DeepfakeTIMI database, the videos of poor quality were automatically discarded. The training loss values for Neural-ODE on original videos using the three approaches: facial skin color variation, average optical intensity and Eulerian video magnification on original videos from COHFACE database and VidTIMIT database are tabulated in Table 3.

Subject	Original video loss	Donor video loss
1	0.01179	0.00979
2	0.0113	0.0097
3	0.0085	0.00981
4	0.00962	0.01132
5	0.00924	0.01118
6	0.01056	0.0097
7	0.01071	0.00745
8	0.01433	0.00986
9	0.01153	0.01268
10	0.01169	0.00892

Table 2: Loss values for 10 original and donor videos obtained from commercial deepfake website [8]

Heart rate extraction techniques	Database	Loss
Skin color variation	COHFACE	0.0189
	VidTIMIT	0.0401
Optical intensity	COHFACE	0.0166
	VidTIMIT	0.0261
Eulerian magnification	COHFACE	0.1254
	VidTIMIT	0.0727

Table 3: Training loss values for Neural-ODE on original videos from COHFACE and VidTIMIT databases.



(a) (b)

Figure 5: Min-max normalized heart rate obtained from Eulerian magnification (ground truth) and Neural-ODE (predictions) on: (a) Our deepfake videos (b) DeepfakeTIMI database videos

4.2. Predicting the Heart Rate of Deepfake Videos using Neural-ODE

The ground truth heart rate prediction obtained from the three approaches: facial skin color variation, average optical intensity and Eulerian video magnification are min-max normalized and compared with Neural-ODE. The loss values for Neural-ODE on fake videos are tabulated in Table 4.

Heart rate extraction techniques	Database	Loss
Skin color variation	Our fake videos	0.0215
	DeepfakeTIMI	0.0327
Optical intensity	Our fake videos	0.0154
	DeepfakeTIMI	0.0252
Eulerian magnification	Our fake videos	0.0353
	DeepfakeTIMI	0.0565

Table 4: Loss values for Neural-ODE on Deepfake videos

The Neural-ODE is trained for 5000 epochs using the min-max normalized heart rate obtained using skin color variation, and optical intensity and 10000 epochs using Eulerian video magnification. The learning rate is 0.0001 and Adam optimizer is used. The prediction result obtained from Neural-ODE using skin color variation, average optical intensity, and Eulerian video magnification on our 10 fake videos and 320 fake videos from DeepfakeTIMI database are shown in Fig. 3, Fig. 4 and Fig. 5 respectively.

5. Conclusion and Future Work

Neural-ODEs have lead a revolution in the area of deep neural networks. They combine traditional theory of differential equations and numerically stable forward simulation [7]. In this paper, we have developed a novel approach to predict the heart rates of deepfake videos using state-of-the-art Neural-ODE. The Neural-ODE is trained using min-max normalized heart rate obtained from original face videos using three well-known approaches: skin color variation, average optical intensity, and Eulerian video magnification.

The significant contributions of our paper are listed.

- Created a new fake video database using commercial deepfake video generating website [8].
- Predict the heart rate of the deepfake videos generated using commercial websites and other fake datasets.

To the best of our knowledge, this is the first attempt to detect heart rate of deepfake videos using Neural-ODE. In the future, we will optimize the network to implement it on low-power/cost single-board computer [55, 56, 57].

Acknowledgements. We acknowledge support from NSF Awards #1822976 and #1422257, an award from the Florida Cybersecurity Center, Royal Bank of Canada, and the Air Force Young Investigator Award to Sumit Jha. We thank Sharma Valliyil Thankachan for providing us his valuable insights on deepfakes.

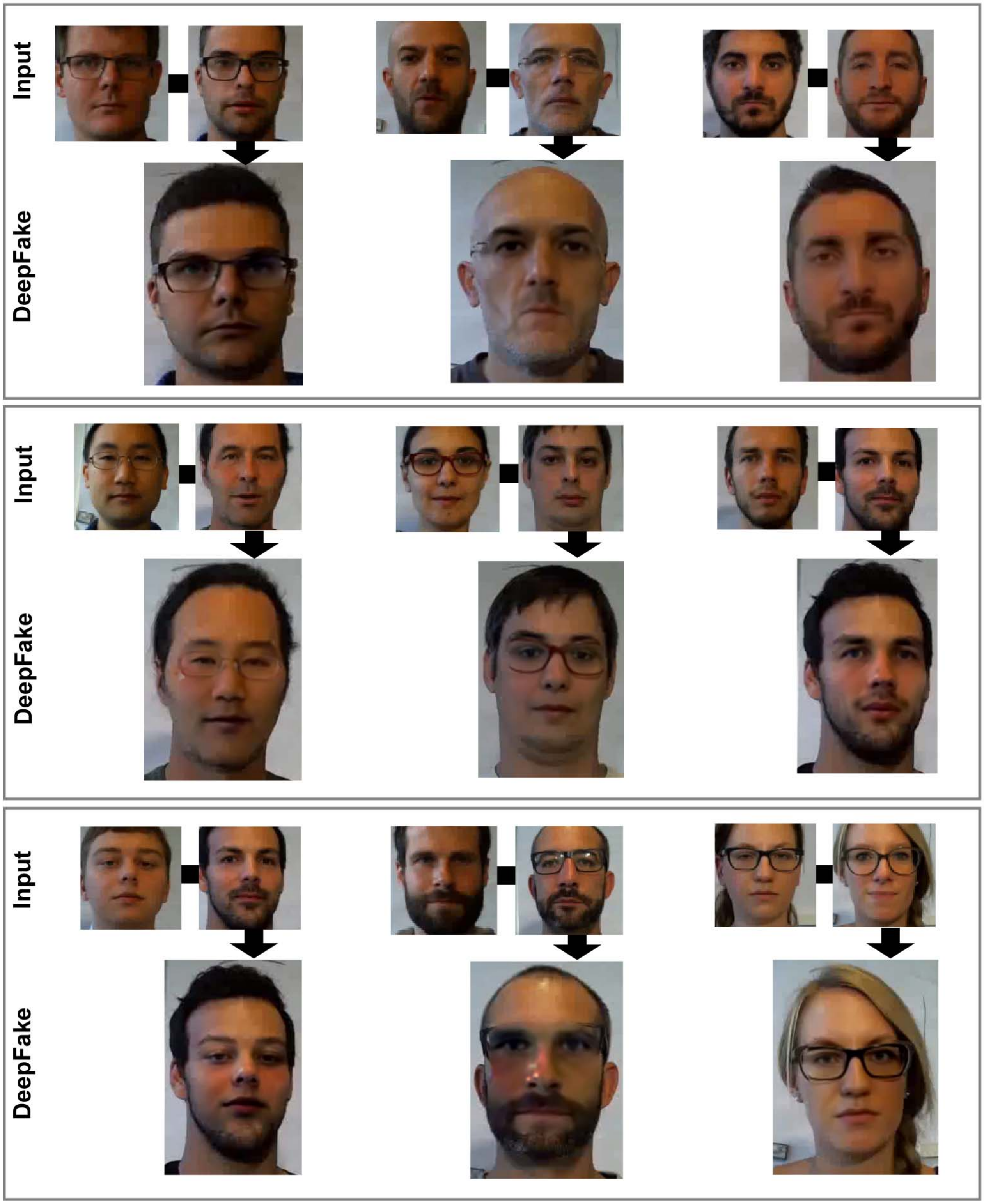


Figure 6: Original, donar, and deepfake video frame generated using commercial deepfake website [8]

References

[1] Marissa Koopman, Andrea Macarulla Rodriguez, and Zeno Geradts. Detection of deepfake video manipulation. 08 2018. 1, 2

[2] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network. *CoRR*, abs/1809.00888, 2018. 1

[3] FaceApp. https://www.faceapp.com/. (Accessed on 07/15/2019). 1

[4] FakeApp. https://www.fakeapp.org/. (Accessed on 07/03/2019). 1

[5] Marie-Helen Maras and Alex Alexandrou. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. *The International Journal of Evidence & Proof*, 23(3):255–262, 2019. 1

[6] Umur Aybars Ciftci and Ilke Demir. Fakecatcher: Detection of synthetic portrait videos using biological signals. *CoRR*, abs/1901.02212, 2019. 1

[7] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *CoRR*, abs/1806.07366, 2018. 1, 6

[8] Deepfakes web. https://deepfakesweb.com/. (Accessed on 07/30/2019). 1, 2, 3, 5, 6, 7

[9] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: Driving visual speech with audio. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 353–360, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 2

[10] Alan J. Cooper. Improved photo response non-uniformity (prnu) based source camera identification. *Forensic Science International*, 226(1):132 – 141, 2013. 2

[11] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2089–2093. IEEE, 2017. 2

[12] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 2089–2093. IEEE, 2017. 2

[13] Supasorn Suwajanakorn, Ira Kemelmacher-Shlizerman, and Steven M Seitz. Total moving face reconstruction. In *European Conference on Computer Vision*, pages 796–812. Springer, 2014. 2

[14] Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation for image content changes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7064–7073, 2017. 2

[15] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017. 2

[16] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015. 2

[17] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 2

[18] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 2

[19] Mauro Barni, Luca Bondi, Nicolò Bonettini, Paolo Bestagini, Andrea Costanzo, Marco Maggini, Benedetta Tondi, and Stefano Tubaro. Aligned and non-aligned double jpeg detection using convolutional neural networks. *Journal of Visual Communication and Image Representation*, 49:153–163, 2017. 2

[20] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016. 2

[21] Javier Galbally and Sébastien Marcel. Face anti-spoofing based on general image quality assessment. In *2014 22nd International Conference on Pattern Recognition*, pages 1173–1178. IEEE, 2014. 2

[22] Vineet Kushwaha, Maneet Singh, Richa Singh, Mayank Vatsa, Nalini Ratha, and Rama Chellappa. Disguised faces in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–9, 2018. 2

[23] Jessica Fridrich. Digital image forensics. *IEEE Signal Processing Magazine*, 26(2):26–37, 2009. 2

[24] Hany Farid. *Photo forensics*. MIT Press, 2016. 2

[25] David Güera, Yu Wang, Luca Bondi, Paolo Bestagini, Stefano Tubaro, and Edward J Delp. A counter-forensic method for cnn-based camera model identification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1840–1847. IEEE, 2017. 2

[26] David Güera, Fengqing Zhu, Sri Kalyan Yarlagadda, Stefano Tubaro, Paolo Bestagini, and Edward J Delp. Reliability map estimation for cnn-based camera model attribution. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 964–973. IEEE, 2018. 2

[27] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395, 2016. 2

[28] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv preprint arXiv:1803.09179*, 2018. 2

[29] Weihong Wang and Hany Farid. Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Transactions on Information Forensics and Security*, 2(3):438–449, 2007. 2

[30] Paolo Bestagini, Simone Milani, Marco Tagliasacchi, and Stefano Tubaro. Local tampering detection in video sequences. In *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pages 488–493. IEEE, 2013. 2

[31] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012. 2, 3

[32] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. 2

[33] Philipp V Rouast, Marc TP Adam, Raymond Chiong, David Cornforth, and Ewa Lux. Remote heart rate measurement using low-cost rgb face video: a technical literature review. *Frontiers of Computer Science*, 12(5):858–872, 2018. 2

[34] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010. 2

[35] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2013. 2

[36] Litong Feng, Lai-Man Po, Xuyuan Xu, Yuming Li, and Ruiyi Ma. Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):879–891, 2014. 2

[37] Sakthi Kumar Arul Prakash and Conrad S Tucker. Bounded kalman filter method for motion-robust, non-contact heart rate estimation. *Biomedical optics express*, 9(2):873–897, 2018. 2

[38] Changchen Zhao, Chun-Liang Lin, Weihai Chen, and Zhengguo Li. A novel framework for remote photoplethysmography pulse extraction on compressed videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1299–1308, 2018. 2

[39] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F Cohn, and Nicu Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2396–2404, 2016. 2

[40] Litong Feng, Lai-Man Po, Xuyuan Xu, Yuming Li, and Ruiyi Ma. Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):879–891, 2014. 2

[41] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 2

[42] D. Gera and E. J. Delp. Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, Nov 2018. 2

[43] Mauricio A Álvarez and Neil D Lawrence. Computationally efficient convolved multiple output gaussian processes. *Journal of Machine Learning Research*, 12(May):1459–1500, 2011. 2

[44] Joseph Futoma, Sanjay Hariharan, and Katherine Heller. Learning to detect sepsis with a multitask gaussian process rnn classifier. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1174–1182. JMLR. org, 2017. 2

[45] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference*, pages 301–318, 2016. 2

[46] Zachary C Lipton, David Kale, and Randall Wetzel. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. In *Machine Learning for Healthcare Conference*, pages 253–270, 2016. 2

[47] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1555–1564. ACM, 2016. 2

[48] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018. 2

[49] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[50] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014. 2

[51] Hamidur Rahman, Mobyen Uddin Ahmed, Shahina Begum, and Peter Funk. Real time heart rate monitoring from facial rgb color video using webcam. In *The 29th Annual Workshop of the Swedish Artificial Intelligence Society*, June 2016. 2, 3

[52] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 2, 3

[53] Pavel Korshunov and Sébastien Marcel. Deepfakes: a new threat to face recognition? assessment and detection. *CoRR*, abs/1812.08685, 2018. 2, 3

[54] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014. 3

[55] Raspberry Pi 4. https://www.raspberrypi.org/products/raspberry-pi-4-model-b/. (Accessed on 08/01/2019). 6

[56] ODROID-N2. https://www.hardkernel.com/shop/odroid-n2-with-4gbyte-ram/. (Accessed on 08/04/2019). 6

[57] Tinker-Board. https://www.asus.com/us/Single-Board-Computer/Tinker-Board/. (Accessed on 07/15/2019). 6