



REFERENCES

- Afchar, D., Nozick, V., Yamagishi, J., Echizen, I., 2018. MesoNet: a Compact Facial Video Forgery Detection Network, in: 2018 IEEE International Workshop on Information Forensics and Security (WIFS). pp. 1–7. <https://doi.org/10.1109/WIFS.2018.8630761>
- Altuncu, E., Franqueira, V.N.L., Li, S., 2022. Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review. <https://doi.org/10.48550/arXiv.2208.10913>
- Anonymous., Deepfakes, 2017. deepfakes_faceswap [WWW Document]. URL <https://github.com/deepfakes/faceswap> (accessed 12.15.22).
- Carlini, N., Wagner, D., 2017. Towards Evaluating the Robustness of Neural Networks. <https://doi.org/10.48550/arXiv.1608.04644>
- Goodfellow, I.J., Shlens, J., Szegedy, C., 2015. Explaining and Harnessing Adversarial Examples. <https://doi.org/10.48550/arXiv.1412.6572>
- Gu, S., Rigazio, L., 2015. Towards Deep Neural Network Architectures Robust to Adversarial Examples. <https://doi.org/10.48550/arXiv.1412.5068>
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Hussain, S., Neekhara, P., Jere, M., Koushanfar, F., McAuley, J., 2020. Adversarial Deepfakes: Evaluating Vulnerability of Deepfake Detectors to Adversarial Examples. <https://doi.org/10.48550/arXiv.2002.12749>
- Kowalski, M., 2016. FaceSwap [WWW Document]. URL <https://github.com/MarekKowalski/FaceSwap> (accessed 11.29.23).
- Kurakin, A., Goodfellow, I., Bengio, S., 2017. Adversarial Machine Learning at Scale. <https://doi.org/10.48550/arXiv.1611.01236>
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. <https://doi.org/10.1109/5.726791>



- Oster, M., Douglas, R., Liu, S.-C., 2009. Computation with Spikes in a Winner-Take-All Network. *Neural Computation* 21, 2437–2465. <https://doi.org/10.1162/neco.2009.07-08-829>
- Papernot, N., McDaniel, P., Wu, X., Jha, S., Swami, A., 2016. Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks. <https://doi.org/10.48550/arXiv.1511.04508>
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. <https://doi.org/10.48550/arXiv.1506.02640>
- Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M., 2019. FaceForensics++: Learning to Detect Manipulated Facial Images. <https://doi.org/10.48550/arXiv.1901.08971>
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R., 2014. Intriguing properties of neural networks. <https://doi.org/10.48550/arXiv.1312.6199>
- Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708. <https://doi.org/10.1109/CVPR.2014.220>
- Thies, J., Zollhöfer, M., Nießner, M., 2019. Deferred Neural Rendering: Image Synthesis using Neural Textures. <https://doi.org/10.48550/arXiv.1904.12356>
- Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., Nießner, M., 2020. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. <https://doi.org/10.48550/arXiv.2007.14808>
- Xiao, C., Zhong, P., Zheng, C., 2019. Enhancing Adversarial Defense by k-Winners-Take-All. <https://doi.org/10.48550/arXiv.1905.10510>
- Xiang, J., Zhu, G., 2017. Joint Face Detection and Facial Expression Recognition with MTCNN, in: 2017 4th International Conference on Information Science and Control Engineering (ICISCE). Presented at the 2017 4th International Conference on



UNIVERSITAS
GADJAH MADA

Assessing the Effectiveness of K-Winners-Takes-All Activation Function on Defending Deepfake Detector Model from Whitebox Adversarial Attack
ALLEN NATHAEL ARDY, Faizal Makhrus, S.Kom., M.Sc., Ph.D
Universitas Gadjah Mada, 2024 | Diunduh dari <http://etd.repository.ugm.ac.id/>

Information Science and Control Engineering (ICISCE), pp. 424–427.
<https://doi.org/10.1109/ICISCE.2017.95>

Xu, W., Evans, D., Qi, Y., 2018. Feature Squeezing: Detecting Adversarial Examples in Deep Neural Networks, in: Proceedings 2018 Network and Distributed System Security Symposium. <https://doi.org/10.14722/ndss.2018.23198>