PROTEKSI CITRA FOTO KPM MAHASISWA FASILKOM UNSRI DARI *DEEPFAKE* DENGAN CMUA-*WATERMARK*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika

Oleh :

Renaldi Budi Setiawan
NIM : 09021281823066

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN SKRIPSI

PROTEKSI CITRA FOTO KPM MAHASISWA FASILKOM UNSRI DARI *DEEPFAKE* DENGAN CMUA-*WATERMARK*

Oleh :

Renaldi Budi Setiawan
NIM : 09021281823066

Indralaya, 2 Juli 2023

Pembimbing I

Samsuryadi, S.Si., M.Kom., Ph.D.
NIP. 197102041997021003

Pembimbing II,

Muhammad Qurhanul Rizqie, S.Kom,. M.T., Ph.D.
NIP. 198712052022031006

Mengetahui,
Ketua Jurusan Teknik Informatka

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari **Senin** tanggal **27 Juni 2023** telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Renaldi Budi Setiawan
NIM : 09021281823066
Judul : Proteksi Citra Foto KPM Mahasiswa Fasilkom UNSRI dari DeepFake dengan CMUA-Watermark.

Dan dinyatakan **LULUS**.

1. Ketua Penguji

Mastura Diana Marieska, M.T.
NIP. 198603212018032001

2. Penguji

Julian Supardi , M.T., Ph.D.
NIP. 197207102010121001

3. Pembimbing I

Samsuryadi, S.Si., M.Kom., Ph.D.
NIP. 197102041997021003

4. Pembimbing II

Muhammad Qurhanul Rizqie, PH.D.
NIP. 198712032022031006

Mengetahui,
Ketua Jurusan Teknik Informatika

Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama	: Renaldi Budi Setiawan
NIM	: 09021281823066
Program Studi	: Teknik Informatika Reguler
Judul Skripsi	: Proteksi Citra Foto KPM Mahasiswa Fasilkom UNSRI dari DeepFake dengan CMUA-Watermark.

Hasil pengecekan Software iThenticate/Turnitin : 5%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya dan Ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 10 Agustus 2023

Renaldi Budi Setiawan
NIM. 09021281823066

MOTTO DAN PERSEMBAHAN

"Sesungguhnya sesudah kesulitan itu ada kemudahan, maka apabila kamu telah selesai dari suatu urusan kerjakanlah dengan sungguh-sungguh urusan yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap."(Al-Insyirah, 6-8)

"Yang Sudah Yasudah"

Ku persembahkan karya tulis ini kepada:

- Ayah, Ibu, Kakak, dan Adik
- Teman-teman Seperjuangan
- Dosen Pembimbing
- Fakultas Ilmu Komputer Universitas Sriwijaya

ABSTRACT

The development of technology and information provides great benefits for information exchange, but also opens up opportunities for misuse of information and the spread of harmful hoaxes. The deepfake phenomenon, which is the manipulation of images and videos using Artificial Intelligence (AI), poses a serious threat to security and privacy. This research focuses on deepfakes of human face images, especially those obtained from the Sriwijaya University (UNSRI) website, which can be exploited without verification. Commonly used passive defense approaches to detect deepfakes have limitations because the modified images remain at risk of being widely spread on the internet. Therefore, this research proposes an active defense approach using Adversarial Watermarking techniques. The proposed method, Cross-Model Universal Adversarial Watermark (CMUA-Watermark), protects thousands of face images from various deepfake models simultaneously. The results show that CMUA-Watermark provides a high protection success rate with satisfactory performance on various datasets. However, the HiSD deepfake model shows a decrease in performance against the KPM photo image dataset of Fasilkom UNSRI students. Therefore, the use of CMUA-Watermark on older versions of UNSRI websites is recommended to prevent deepfake exploitation and maintain the security and privacy of student facial images.

Keywords: *Images, KPM photo Image, deepfake, CMUA-Watermark*

ABSTRAK

Perkembangan teknologi dan informasi memberikan manfaat besar untuk pertukaran informasi, namun juga membuka peluang penyalahgunaan informasi dan penyebaran *hoaks* yang merugikan. Fenomena *deepfake*, yaitu manipulasi gambar dan video menggunakan *Artificial Intelligence* (AI), menjadi ancaman serius terhadap keamanan dan privasi. Penelitian ini fokus pada *deepfake* pada citra gambar wajah manusia, terutama yang diperoleh dari situs web Universitas Sriwijaya (UNSRI), yang dapat dieksploitasi tanpa verifikasi. Pendekatan pertahanan pasif yang umum digunakan untuk mendeteksi *deepfake* memiliki keterbatasan karena citra yang sudah dimodifikasi tetap berisiko tersebar luas di internet. Oleh karena itu, penelitian ini mengusulkan pendekatan pertahanan aktif dengan menggunakan teknik *Adversarial Watermarking*. Metode yang diusulkan, yaitu *Cross-Model Universal Adversarial Watermark* (CMUA-Watermark), melindungi ribuan citra wajah dari berbagai model *deepfake* secara bersamaan. Hasil penelitian menunjukkan bahwa CMUA-Watermark memberikan tingkat keberhasilan proteksi yang tinggi dengan kinerja memuaskan pada berbagai *dataset*. Namun, model *deepfake* HiSD menunjukkan penurunan performa terhadap *dataset* citra foto KPM mahasiswa Fasilkom UNSRI. Oleh karena itu, penggunaan CMUA-Watermark di situs web UNSRI versi lama dianjurkan untuk mencegah eksploitasi d*eepfake* dan menjaga keamanan serta privasi citra wajah mahasiswa.

Kata Kunci: *Citra, Foto KPM, deepfake, CMUA-Watermark*

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT atas segala nikmat, rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **"Proteksi Citra Foto KPM Mahasiswa Fasilkom Unsri dari Deepfake dengan CMUA-Watermark"** Tugas Akhir ini disusun untuk memenuhi salah satu persyaratan kelulusan tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah membantu penulis dalam menyelesaikan Tugas Akhir ini, yaitu kepada:

1. Allah Subhanahu Wa Ta'ala yang telah memberikan hamba keimanan, kesehatan, kecerdasan, kemudahan dan kelancaran sehingga hamba dapat menyelesaikan tugas-tugas hamba sebagai seorang mahasiswa.
2. Kedua Orang Tua penulis tercinta Ayah Husdi dan Ibu Nyimas Nurhasanah yang telah memberikan doa dan restu serta dukungan yang sangat besar selama mengikuti dan melaksanakan perkuliahan di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, memberikan semangat sehingga penulis dapat menyelesaikan Tugas Akhir ini.
3. Universitas Sriwijaya yang telah memberikan saya kesempatan dan berbagai fasilitas dalam perkuliahan.
4. Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika.
5. Bapak Samsuryadi, S.Si., M.Kom., Ph.D. dan Bapak Muhammad Qurhanul Rizqie, S.Kom., M.T., Ph.D. sebagai pembimbing Tugas Akhir yang mengarahkan dan memberi masukan dalam proses pengerjaannya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik.

6. Ibu Mastura Diana Marieska, M.T. dan Bapak Julian Supardi, M.T., Ph.D. selaku dosen penguji, yang telah memberikan masukkan sehingga Tugas akhir ini menjadi lebih baik lagi.
7. Kak Ricy selaku admin Jurusan Teknik Informatika Reguler yang telah membantu mengurus seluruh berkas yang diperlukan.
8. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Para teman-teman seperjuangan Tasya, Sholeh, Roni, Reza, Nanda, Nabila, Nadia, dan Nikea yang telah membantu penulis saat kesulitan dalam mengerjakan Tugas Akhir, memberikan motivasi dan semangat.
10. Serta teman-teman seperjuangan angkatan 2018 yang tidak tertuliskan dalam kata pengantar ini namun turut membantu dalam proses untuk mencapai gelar sarjana ini.

Penulis menyadari bahwa laporan tugas akhir ini masih banyak kekurangan dan masih jauh dari kata sempurna karena keterbatasan ilmu yang dimiliki penulis. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk kesempurnaan tugas akhir ini. Semoga tugas akhir ini dapat memberikan manfaat bagi orang banyak.

Palembang, 10 Agustus 2023

Renaldi Budi Setiawan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan dibahas berkenaan dengan garis besar pokok-pokok pikiran dalam penelitian ini. Pokok pikiran yang akan dibahas antara lain latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian. Pokok-pokok pikiran yang diuraikan akan dijadikan acuan dalam kajian penelitian ini.

1.2 Latar Belakang Masalah

Perkembangan teknologi dan informasi saat ini maju dengan sangat pesat sehingga mempermudah manusia dalam bertukar informasi. Seiring dengan kemajuannya, banyak pula terjadi pelanggaran di dalamnya, salah satunya merupakan penyebaran informasi palsu (Borges et al., 2019; Qayyum et al., 2019). *Deepfakes* adalah produk dari aplikasi kecerdasan buatan (AI) hasil penggabungan dari proses, menggabungkan, mengganti, dan melapiskan gambar dan penjepit video untuk membuat video palsu yang tampak otentik (Maras & Alexandrou, 2019). *Deepfake* telah menghadirkan bahaya serius terkait privasi dan keamanan pada citra gambar wajah seseorang (Tolosana et al., 2020).

Citra gambar wajah merupakan target utama dari pembuatan *deepfake* karena hal ini berkaitan dengan ketersediaan citra wajah yang mudah didapatkan dari berbagai sumber, seperti media sosial dan situs-situs yang menyediakan data gambar wajah lainnya. Diperlukan sebuah proteksi yang ditanamkan pada citra gambar wajah untuk mencegah generasi *deepfake*. Permasalahan yang sama juga

dapat ditemukan di situs resmi Universitas Sriwijaya (UNSRI) versi lama, yang dapat diakses oleh siapa saja tanpa perlu verifikasi terlebih dahulu. Situs ini mengandung informasi tentang mahasiswa, termasuk foto kartu pengenal mahasiswa (KPM) mereka. Kondisi ini memberikan peluang bagi pihak yang tidak bertanggung jawab untuk mengeksploitasi informasi pribadi mahasiswa dan melakukan penyalahgunaan melalui *deepfake*. Dalam konteks ini, *deepfake* dapat digunakan untuk mengatasnamakan identitas mahasiswa dan menggunakan foto-foto mereka dalam konten palsu yang dapat merugikan individu tersebut secara serius. Oleh karenanya, dibutuhkan adanya sebuah Proteksi yang dapat mengatasi permasalahan tersebut.

Penelitian terkait mengenai proteksi deepfake umumnya hanya melatih pendeteksi deepfake untuk mendeteksi konten yang telah dimodifikasi (Afchar et al., 2018; Tariq, Lee, & Woo, 2021; Zhao et al., 2021; Sun et al., 2021; Chen et al., 2021). Detektor semacam itu pada dasarnya adalah pengklasifikasi biner, yang memprediksi apakah sebuah gambar dipalsukan oleh model deepfake atau tidak. Metode pendeteksian gambar seperti ini dapat disebut dengan pendekatan pertahan pasif (Huang et al., 2022).

Meskipun metode pendekatan pasif dapat membantu mengidentifikasi *deepfake*, efek dan bahaya yang ditimbulkan tidak dapat dicegah sepenuhnya karena citra atau gambar tersebut dapat dengan mudah menyebar di antara banyak gambar lainnya yang ada di internet. Hal ini menyebabkan penyebaran *deepfake* yang belum terdeteksi dan mempertahankan risikonya yang tetap ada. Oleh karena itu, diperlukan penggunaan pendekatan pertahanan aktif seperti penggunaan

Adversarial Watermark untuk memerangi model *deepfake*. *Adversarial Watermark* memungkinkan penambahan watermark pada gambar wajah sebelum diunggah ke internet, watermark dapat memberikan lapisan perlindungan tambahan pada gambar wajah, membuatnya tampak tidak nyata saat dimodifikasi menggunakan model *deepfake.* Sehingga dapat menghindari risiko serangan deepfake yang berbahaya di masa mendatang. Salah satu metode *Adversarial Watermark* yang sangat efektif adalah *Cross-Model Universal Adversarial Watermark* (CMUA-Watermark), yang mampu melindungi ribuan citra gambar wajah dari berbagai model deepfake secara bersamaan (Huang et al., 2022).

Dengan adanya referensi penelitian terkait yang telah dijelaskan, metode *Cross-Model Universal Adversarial Watermark* (CMUA-Watermark) akan menjadi metode yang dipergunakan dalam penelitian proteksi citra gambar. Diharapkan dengan menggunakan metode ini, citra gambar dapat diproteksi oleh sistem dengan tingkat keberhasilan proteksi yang baik.

1.3 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang yang telah diuraikan sebelumnya didapatkan rumusan masalah yaitu maraknya kasus pemalsuan foto dengan *deepfake* menggunakan konten yang tersebar di internet. Konten foto KPM mahasiswa Fasilkom Unsri pada situs lama Unsri juga dapat diakses secara bebas sehingga berpotensi mendapat serangan *deepfake* oleh karena itu didapatkan bebarapa pertanyaan yang harus diselesaikan dalam penelitian ini yaitu :

1. Bagaimana cara membangun perangkat lunak yang dapat memproteksi citra gambar foto KPM mahasiswa Fasilkom UNSRI dari *deepfake* menggunakan metode CMUA-*Watermark.*
2. Bagaimana tingkat keberhasilan metode CMUA-*Watermark* dalam memproteksi citra foto KPM mahasiswa Fasilkom UNSRI dari *deepfakes*?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Menghasilkan perangkat lunak yang dapat memproteksi citra gambar foto KPM mahasiswa Fasilkom UNSRI dari *deepfake* menggunakan metode CMUA-*Watermark.*
2. Mengetahui tingkat keberhasilan penggunaan metode CMUA-*Watermark* dalam memproteksi citra foto KPM mahasiswa Fasilkom UNSRI dari *deepfake*.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Sistem yang dibuat dapat memproteksi citra gambar foto KPM mahasiswa Fasilkom UNSRI dari *deepfake* menggunakan metode CMUA-*Watermark.*
2. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian terkait di masa mendatang.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dari *dataset* Celeb-a digunakan sebagai data latih dan data verifikasi, *dataset* ini didapatkan penelitian *Deep Learning Face Attributes in the Wild* (Liu et al., 2015) .

2. Data uji yang digunakan merupakan *dataset* foto mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya Angkatan 2018.
3. Format citra yang digunakan JPG.
4. Penitian tidak membuat algortima generasi *deepfake* sendiri dan hanya menggunakan yang sudah ada.
5. Penelitian ini tidak melakukan pencarian *Auto-step size* dan hanya menggunakan dari *auto-step size* dari penelitian sebelumnya untuk melakukan training pada model CMUA-*watermark*.

1.7 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah. Pokok-pokok pikiran ini akan menjadi dasar dan acuan pengembangan penelitian pada bab selanjutnya. **BAB II. KAJIAN LITERATUR**

Pada bab ini membahas literatur pada penelitian, seperti pengertian Citra, *Deepfake*, CMUA-*Watermark* dan penelitian yang relevan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dibahas proses pengumpulan data dan tahapan-tahapan di dalam penelitian. Tahapan penelitian dibahas lebih rinci berdasarkan kerangka kerja tertentu. Di bagian akhir bab ini akan dimuat rancangan manajemen proyek penelitian

BAB IV. PENGEMBANGAN PERANGKAT LUNAK.

Pada bab ini menjelaskan analisa dan proses pengembangan sistem, seperti analisis kebutuhan sistem dan konstruksi sistem. Pada akhir bab dilakukan analisa pengujian.

BAB V. HASIL DAN ANALISIS PENELITIAN.

Pada bab ini berisi hasil yang dilakukan pada sistem serta keakuratan sistem. Pada akhir bab dijelaskan analisis dari penelitian.

BAB VI. KESIMPULAN DAN SARAN.

Pada bab ini menjelaskan kesimpulan dari penelitian dan saran ke depannya dari penelitian ini

1.8 Kesimpulan

Pada Bab ini telah menjelaskan dasar dan patokan pada penelitian , seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

Anwar, A. (2014). A Review of RUP (Rational Unified Process). *International Journal of Software Engineering*, *5*(2), 8–24. http://www.cscjournals.org/library/manuscriptinfo.php?mc=IJSE-142

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, 1–9.

Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality*, *11*(3). https://doi.org/10.1145/3287763

Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8789–8797. https://doi.org/10.1109/CVPR.2018.00916

Day, C. (2019). The Future of Misinformation. *Computing in Science and Engineering*, *21*(1), 108. https://doi.org/10.1109/MCSE.2018.2874117

Fletcher, J. (2018). Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, *70*(4), 455–471. https://doi.org/10.1353/tj.2018.0097

Gonzalez, R. C., & Woods, R. E. (2018). *4TH EDITION Digital image processing*.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–11.

He, Z., Zuo, W., Kan, M., Shan, S., & Chen, X. (2019). AttGAN: Facial Attribute Editing by only Changing What You Want. *IEEE Transactions on Image Processing*, *28*(11), 5464–5478. https://doi.org/10.1109/TIP.2019.2916751

Huang, H., Wang, Y., Chen, Z., Li, Y., Tang, Z., Chu, W., Chen, J., Lin, W., & Ma, K.-K. (2021). *CMUA-Watermark: A Cross-Model Universal Adversarial Watermark for Combating Deepfakes*. http://arxiv.org/abs/2105.10872

Huang, H., Wang, Y., Chen, Z., Zhang, Y., Li, Y., Tang, Z., Chu, W., Chen, J., Lin, W., & Ma, K. (2022). CMUA-Watermark: A Cross-Model Universal Adversarial Watermark for Combating Deepfakes. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(1), 989–997. https://doi.org/10.1609/aaai.v36i1.19982

Jain, P., Dave, M., & Patel, V. M. (2020). A Comprehensive Review on Steganography Techniques in Digital Images. *Journal of King Saud University-Computer and Information Sciences*, *32*(4), 395–408.

Li, X., Zhang, S., Hu, J., Cao, L., Hong, X., Mao, X., Huang, F., Wu, Y., & Ji, R. (2021). Image-to-image Translation via Hierarchical Style Disentanglement. *Proceedings of the IEEE Computer Society Conference on Computer Vision*

and Pattern Recognition, *i*, 8635–8644. https://doi.org/10.1109/CVPR46437.2021.00853

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 3730–3738. https://doi.org/10.1109/ICCV.2015.425

M. Sonka, V. H. and R. B. (2014). Image processing, analysis, and machine vision. Cengage Learning. In *IEEE Aerospace and Electronic Systems Magazine*.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 1–28.

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226

Metzen, J. H., Kumar, M. C., Brox, T., & Fischer, V. (2017). Universal Adversarial Perturbations Against Semantic Image Segmentation. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2774–2783. https://doi.org/10.1109/ICCV.2017.300

Moosavi-Dezfooli, S. M., Fawzi, A., Fawzi, O., & Frossard, P. (2017). Universal adversarial perturbations. *Proceedings - 30th IEEE Conference on Computer*

Vision and Pattern Recognition, CVPR 2017, *2017-Janua*, 86–94. https://doi.org/10.1109/CVPR.2017.17

Qayyum, A., Qadir, J., Janjua, M. U., & Sher, F. (2019). Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News. *IT Professional*, *21*(4), 16–24. https://doi.org/10.1109/MITP.2019.2910503

Rafique, M. A., Younus, S., & Bhatti, M. A. (2018). A comprehensive review on digital image representation and its applications. *Digital Communications and Networks*, *4*(1), 1–14.

Ruiz, N., Bargal, S. A., & Sclaroff, S. (2020). Disrupting Deepfakes: Adversarial Attacks Against Conditional Image Translation Networks and Facial Manipulation Systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12538 LNCS*, 236–251. https://doi.org/10.1007/978-3-030-66823-5_14

Tang, H., Xu, D., Sebe, N., & Yan, Y. (2019). Attention-Guided Generative Adversarial Networks for Unsupervised Image-to-Image Translation. *Proceedings of the International Joint Conference on Neural Networks*, *2019-July*. https://doi.org/10.1109/IJCNN.2019.8851881

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Wang, L., Cho, W., & Yoon, K. J. (2020). Deceiving Image-to-Image Translation Networks for Autonomous Driving with Adversarial Perturbations. *IEEE Robotics and Automation Letters*, *5*(2), 1421–1428. https://doi.org/10.1109/LRA.2020.2967289