Proteksi Citra Foto KPM Mahasiswa Fasilkom UNSRI dari *DeepFake* dengan *CMUA-Watermark*

Diajukan untuk Menyusun Skripsi
di jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI

Oleh :

Renaldi Budi Setiawan
NIM : 09021281823066

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN PROPOSAL SKRIPSI

Proteksi Citra Foto KPM Mahasiswa Fasilkom UNSRI dari *DeepFake* dengan *CMUA-Watermark*

Oleh :

Renaldi Budi Setiawan

NIM : 09021281823066

Indralaya, 26 mei 2023

Pembimbing I	Pembimbing II,
Syamsuryadi, S.Si., M.Kom., Ph.D.	Muhammad Qurhanul Rizqie, S.KOM,. M.T., Ph.D.
NIP 197102041997021003	NIP 1671060312870008

Mengetahui,
Ketua Jurusan

Alvi Syahrini Utami, M.Kom.
NIP. 19781222200642003

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

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

Berita palsu telah menjadi isu yang merupakan ancaman bagi kepentingan masyarakat umum (Borges et al., 2019; Qayyum et al., 2019). Berita palsu mengacu pada konten gaya berita fiktif yang dibuat untuk menipu publik (Aldwairi & Alwahedi, 2018; Jang & Kim, 2018). Informasi palsu menyebar dengan cepat melalui media sosial, yang dapat berdampak pada jutaan pengguna (Figueira & Oliveira, 2017). Saat ini, satu dari lima pengguna internet mendapatkan berita melalui *YouTube*, kedua setelah Facebook (Anderson, 2018). Peningkatan popularitas video ini menyoroti perlunya alat untuk mengonfirmasi keaslian konten media dan berita, karena teknologi baru memungkinkan manipulasi video yang meyakinkan (Anderson, 2018). Mengingat kemudahan dalam memperoleh dan menyebarkan informasi yang salah melalui platform media sosial, semakin sulit untuk mengetahui apa yang harus dipercaya, yang mengakibatkan konsekuensi berbahaya bagi pengambilan keputusan yang

terinformasi (Borges et al., 2019; Britt et al., 2019). Memang, hari ini kita hidup di apa yang oleh beberapa orang disebut era “pasca-kebenaran”, yang ditandai dengan disinformasi digital dan perang informasi yang dipimpin oleh aktor jahat yang menjalankan kampanye informasi palsu untuk memanipulasi opini publik (Anderson, 2018; Qayyum et al., 2019; Zannettou et al., 2019).

Situs resmi milik UNSRI versi lama[1] hingga saat ini dapat dengan mudah diakses oleh siapapun tanpa memerlukan verifikasi terlebih dahulu. Adanya laman ini juga memberikan informasi terkait mahasiswa, termasuk didalamnya foto diri mahasiswa. Setiap foto yang diunggah oleh pihak UNSRI di situs tersebut akan muncul dalam hasil pencarian gambar Google dan dapat dengan mudah diunduh. Informasi pribadi mahasiswa yang ada di dalam situs ini sangat memungkinkan dapat dieksploitasi oleh seseorang untuk melakukan tindakan kejahatan, seperti mengatasnamakan identitas dan menggunakan wajah mahasiswa tersebut dengan menggunakan *deepfake*.

Adversarial watermark dapat digunakan untuk memerangi *deepfake model*, *adversarial watermark* dapat menghasilkan citra gambar yang terdistorsi (Ruiz et al., 2020). Namun metode ini masih kurang efisien karena memerlukan proses pelatihan individu untuk setiap citra gambar wajah, untuk menghasilkan *adversarial attack model* terhadap *deepfake model* tertentu (Huang et al., 2021). Untuk mengatasi masalah ini, penelitian ini menggunakan metode *universal adversarial attack model* pada *deepfake model*, untuk menghasilkan *Cross-Model*

[1] https://old.UNSRI.ac.id/?act=daftar_mahasiswa

Universal Adversarial Watermark (*CMUA-Watermark*) yang dapat melindungi ribuan citra gambar wajah dari beberapa model *deepfake* (Huang et al., 2021).

1.3 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang yang telah diuraikan maka rumusan masalah dari penelitian ini adalah

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

1.4 Tujuan Masalah

Tujuan penelitian ini adalah:

1. Menghasilkan kerangka kerja yang sesuai untuk penelitian proteksi citra gambar foto KPM mahasiswa Fasilkom UNSRI dari *deepfake* dengan *CMUA-Watermark.*
2. Menghasilkan perangkat lunak yang dapat memproteksi citra gambar foto KPM mahasiswa Fasilkom UNSRI dari *deepfake* menggunakan metode *CMUA-Watermark.*

3. 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. Dataset yang digunakan pelatihan merupakan dataset Celeb-a, dari penelitian *Deep Learning Face Attributes in the Wild* (2015).
2. Data uji yang digunakan merupakan dataset foto mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya Angkatan 2018.
3. Ekstensi citra yang didukung oleh perangkat lunak adalah .jpg.
4. Penelitian hanya fokus dalam proteksi citra dari *deepfake*.

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yaitu sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan membahas landasan dari penelitian, seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

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 menjelaskan pelaksanaan alur penelitian, yakni pengumpulan data dan perancangan pembangunan perangkat lunak. Serta tahapan dijelaskan secara detail berdasarkan kerangka yang dibuat.

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.

BAB II

KAJIAN LITERATUR

2.1 Pendahuluan

pada bab ini akan dijelaskan mengenai dasar-dasar teori digunakan pada penelitian ini. Serta penjelasan hasil dari penelitian-penelitian terkait mengenai citra, *Deepfakes*, *CMUA-Watermark* dan RUP. Pada bab ini pula dibahas mengenai penelitian terkait lainnya yang relevan.

2.2 Landasan Teori

2.2.1 Citra

Citra merupakan representasi visual dari objek atau scene yang dibentuk oleh kumpulan piksel-piksel (Rafique et al., 2018). Citra digital terdiri dari matriks piksel, di mana setiap piksel mewakili intensitas cahaya atau warna pada posisi tertentu dalam citra (Gonzalez & Woods, 2018). Citra dapat berupa citra grayscale yang hanya memiliki tingkat keabuan (*grayscale*) atau citra berwarna yang terdiri dari tiga komponen warna dasar (merah, hijau, biru) yang membentuk citra dalam model warna RGB (*Red-Green-Blue*) (M. Sonka, 2014).

2.2.1.1 Ekstraksi Fitur Citra

Ekstraksi fitur citra adalah proses pengambilan informasi yang relevan atau karakteristik khusus dari citra (Rafique et al., 2018). Fitur-fitur ini dapat digunakan untuk analisis, pengenalan pola, atau pemrosesan selanjutnya. Beberapa metode umum untuk ekstraksi fitur citra termasuk Transformasi Wavelet, Deteksi Tepi, dan Deskriptor Kejadian Tertentu (M. Sonka, 2014).

2.2.1.2 Segmentasi Citra

Segmentasi citra adalah proses pemisahan citra menjadi beberapa bagian yang memiliki kesamaan atribut atau ciri tertentu (Gonzalez & Woods, 2018). Tujuan dari segmentasi citra adalah untuk memisahkan objek dari latar belakang atau mengidentifikasi bagian-bagian yang memiliki makna atau karakteristik tertentu dalam citra. Metode segmentasi citra dapat mencakup teknik berbasis intensitas, berbasis tepi, atau berbasis region (Rafique et al., 2018).

2.2.1.3 Peningkatan dan Restorasi Citra

Peningkatan citra adalah proses memperbaiki atau meningkatkan kualitas citra dengan memperjelas detail, meningkatkan kontras, atau mengurangi *noise* (M. Sonka, 2014). Restorasi citra adalah proses pemulihan citra yang terdegradasi akibat gangguan atau kerusakan, seperti *blur* atau *noise* (Gonzalez & Woods, 2018). Teknik umum yang digunakan dalam peningkatan dan restorasi citra termasuk filter spasial, filter frekuensi, atau metode restorasi berbasis statistik.

2.2.1.4 Kompresi Citra

Kompresi citra adalah proses pengurangan ukuran file citra dengan mempertahankan kualitas citra seoptimal mungkin (Rafique et al., 2018). Kompresi citra digunakan untuk mengurangi ruang penyimpanan dan memungkinkan transmisi data citra yang efisien. Teknik kompresi citra meliputi kompresi berbasis piksel seperti JPEG (*Joint Photographic Experts Group*) dan kompresi berbasis *wavelet* seperti JPEG2000 (M. Sonka, 2014).

2.2.1.5 Steganografi pada Citra

Steganografi pada citra adalah teknik menyisipkan pesan rahasia ke dalam citra dengan cara yang tidak terlihat oleh mata manusia secara kasat mata (Yang et al., 2019). Pesan rahasia ini dapat disembunyikan dalam domain spasial atau domain frekuensi citra dengan memanfaatkan properti citra yang dapat ditoleransi perubahan kecil. Beberapa metode steganografi pada citra termasuk *Least Significant Bit* (LSB), *Discrete Cosine Transform* (DCT), dan *Pixel Value Differencing* (PVD) (Jain et al., 2020).

2.2.2 *DeepFake*

Deepfake sendiri baru dipopulerkan di tahun 2017, berawal dari pengguna *Reddit* mengunggah video porno hasil editan. Pengguna *Reddit* ini mengembangkan GAN menggunakan *TensorFlow*. Teknologi *Deepfake* dapat berupa video lucu, pornografi, atau politik seseorang yang mengatakan apa pun, tanpa persetujuan orang yang citra gambar dan suaranya terlibat (Day, 2019; Fletcher, 2018).

Deepfake adalah Kombinasi dari "pembelajaran mendalam" dan "palsu", *deepfake* adalah video *hiper*-realistis yang dimanipulasi secara digital untuk menggambarkan orang-orang yang mengatakan dan melakukan hal-hal yang tidak pernah benar-benar terjadi (Metz, 2019; Metz & O'Sullivan, 2019). *Deepfake* mengandalkan jaringan saraf yang menganalisis kumpulan besar sampel data untuk belajar meniru ekspresi wajah, tingkah laku, suara, dan infleksi seseorang (Dickson, 2018). Prosesnya melibatkan memasukkan rekaman dua orang ke dalam algoritme pembelajaran mendalam untuk melatihnya bertukar

wajah(Rubenking & Eddy, 2019). Dengan kata lain, *deepfake* menggunakan teknologi pemetaan wajah dan AI yang menukar wajah seseorang di video menjadi wajah orang lain (Horowitz, 2019; Wallace, 2019). *Deepfake* muncul ke publisitas pada tahun 2017 ketika pengguna *Reddit* mem-*posting* video yang menunjukkan selebriti dalam situasi seksual yang membahayakan (Brown, 2019; Leetaru, 2019; Marr, 2019). *Deepfake* sulit dideteksi, karena mereka menggunakan rekaman nyata, dapat memiliki audio yang terdengar otentik, dan dioptimalkan untuk menyebar di media sosial dengan cepat. Dengan demikian, banyak pemirsa menganggap bahwa video yang mereka lihat adalah asli.

2.2.2.1 *Photo Deepfake*

2.2.2.1.1 *Face and Body Swapping*

Dalam hal ini, perubahan dilakukan pada wajah dan tubuh dengan mengganti atau memadukan tubuh dan wajah dengan wajah atau tubuh orang lain. Hasilnya adalah orang yang sama sekali berbeda dalam citra gambar aslinya. Contoh pendekatan ini dapat dilihat di banyak aplikasi menggunakan *Aging filter*. Ini dapat berguna bagi pelanggan untuk mencoba pakaian, kosmetik, atau gaya rambut secara virtual.

2.2.2.2 *Deepfake Creation*

Video *Deepfake* sangat sempurna sehingga dapat membodohi siapa pun. Berbagai alat dan aplikasi digunakan untuk mengembangkan video *deepfake* ini. Aplikasi ini sebagian besar menggunakan teknik pembelajaran mendalam untuk mengembangkan video ini. Video *deepfake* pertama dibuat menggunakan *FakeApp* yang dikembangkan oleh pengguna *Reddit.* Untuk memahaminya lebih

jelas, mari kita ambil contoh gambar diam ini dari film "*Man of Steel*" di mana wajah aktris Amy Adams diganti dengan aktor lain Nicolas Cage seperti yang ditunjukkan pada Gambar II-1.

Gambar II-1, *Frame* dari klip *deepfake* film "*Man of Steel*".

Gambar II-1 menunjukkan gambar asli dari film *Man of Steel* dengan wajah aktris Amy Adams di sebelah kiri, dan di sebelah kanan adalah bingkai *deepfake* yang menggantikan wajah dengan Nicolas Cage. Contoh ini menunjukkan bagaimana wajah perempuan diganti dengan wajah laki-laki. Beginilah cara melakukannya:

1. Wilayah gambar yang menunjukkan wajah Amy Adams diambil dari video aslinya.
2. Gambar yang diekstrak ini digunakan sebagai *input* untuk di proses *deep learning*, teknik AI ini digunakan untuk secara otomatis menghasilkan gambar yang cocok, Nicolas Cage.
3. Gambar yang dihasilkan sekarang ditukar dengan wajah asli di dalam video asli dan menghasilkan video deepfake.

2.2.3 *CMUA-Watermark*

CMUA-Watermark adalah sebuah teknik *watermarking* yang digunakan untuk memberikan tanda air (*watermark*) pada data multi-media secara rahasia dan tangguh. Metode ini juga memiliki kemampuan untuk mengatasi masalah kehilangan atau perubahan data yang disebabkan oleh proses kompresi, *cropping*, rotasi, dll (Li et al., 2019).

2.2.3.1 Metode *CMUA-Watermark*

Metode *CMUA-Watermark* menggabungkan teknik *deep learning* dan *adversarial learning* untuk memberikan *watermark* pada data multimedia. Pada tahap pelatihan, sebuah jaringan saraf berbasis CNN (*Convolutional Neural Network*) dilatih menggunakan pasangan data multimedia yang memiliki *watermark* dan tanpa *watermark*. Setelah pelatihan selesai, jaringan saraf akan menghasilkan *watermark* yang dapat diterapkan pada berbagai jenis data multimedia.

Selama proses *embedding*, data multimedia diubah menjadi representasi vektor dalam ruang fitur oleh jaringan saraf yang telah dilatih sebelumnya. Kemudian, *watermark* yang dihasilkan oleh jaringan saraf dimasukkan ke dalam representasi vektor tersebut. Setelah itu, representasi vektor yang telah dimodifikasi digunakan untuk menghasilkan data multimedia yang telah ditandai dengan *watermark*.

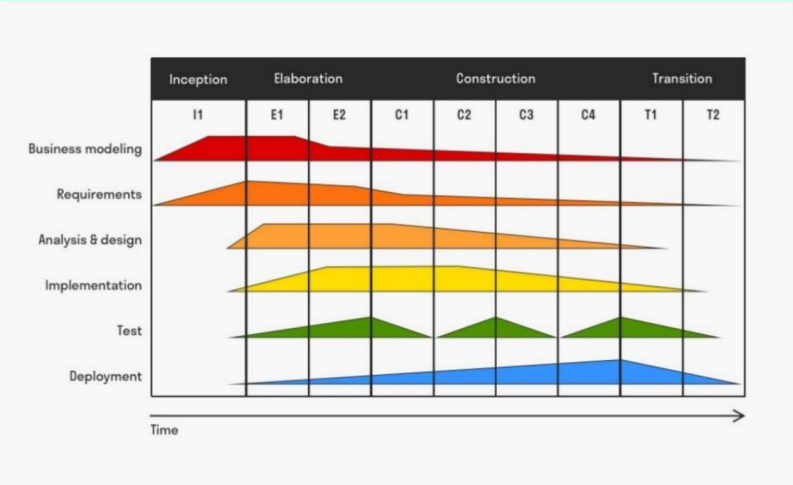
Proses pencarian *watermark* dilakukan dengan menggunakan sebuah jaringan saraf yang dibuat khusus untuk memisahkan *watermark* dari data

multimedia. Jaringan saraf ini juga dilatih menggunakan pasangan data multimedia yang memiliki *watermark* dan tanpa *watermark*.

Metode *CMUA-Watermark* memiliki beberapa keunggulan dibandingkan dengan metode *watermarking* lainnya. Salah satu keunggulannya adalah kemampuan untuk menghasilkan *watermark* yang dapat diterapkan pada berbagai jenis data multimedia dengan keandalan yang tinggi. Selain itu, metode ini juga dapat mengatasi masalah kehilangan atau perubahan data yang disebabkan oleh proses kompresi, cropping, rotasi, dll dengan baik (Li et al., 2019).

2.2.4 Rational Unified Process

Rational Unified Process (RUP) adalah metode rekayasa pengembangan perangkat lunak yang digunakan untuk kedisiplinan dalam penetapan tugas dan tanggung jawab. Tujuan RUP adalah memastikan bahwa produk perangkat lunak yang dihasilkan akan berkualitas dan sesuai kebutuhan pengguna akhir (end-users) (Anwar, 2014).



Gambar II-2,Arsitektur Rasional Unified Process

RUP yang baik akan tercipta lewat hasil kerja sama antara pengembang perangkat lunak, mitra dan pengguna. Salah satu perspektif dalam RUP merupakan *Dynamic Perspective* & *Lifecycle Phases* yang penggunaannya digambarkan dalam bidang dua dimensi. Bidang horizontal menyatakan lamanya waktu pengembangan dan aspek dinamis lainnya, sedangkan bidang vertikal menyatakan aspek statis dalam rekayasa pengembangan perangkat lunak. Perspektif RUP model ini dinyatakan seperti dalam Gambar II-2.

Dalam bidang horizontal, terdapat fase atau tahap dalam proses rekayasa perangkat lunak yang memaparkan peran dari tiap unit. Fase dalam bidang ini terbagi ke dalam fase insepsi, elaborasi, konstruksi dan transisi.

1. Fase insepsi merupakan fase yang berfokus pada pendefinisian ruang lingkup atau batasan dalam proyek pengembangan dengan cara melakukan analisis desain berorientasi objek (*Object Oriented Analysis Design*). Tujuan dari fase ini adalah untuk mendapatkan seluruh pemahaman dari pihak yang berkaitan agar sistem yang diajukan sesuai dengan keinginan dan kebutuhan.
2. Fase elaborasi merupakan fase yang akan membuat arsitektur dasar sistem lewat hasil analisis sebelumnya. Fase ini juga akan menentukan perencanaan proyek serta spesifikasi dari fitur yang akan dimuat dalam sistem. Hasil dari fase ini merupakan dokumen arsitektur yang berguna untuk fase selanjutnya.
3. Fase Konstruksi merupakan fase menerjemahkan spesifikasi fitur dari dokumen rancangan sebelumnya ke dalam bentuk program/sistem sesuai

dengan arsitekturnya. Fase ini berfokus pada peningkatan fungsi serta implementasi yang lebih mendalam terhadap spesifikasi sistem.

4. Fase Transisi merupakan fase pengujian sistem ke pengguna akhir dimana sistem yang dibuat harus memenuhi kebutuhan perangkat lunak dan kebutuhan penggunanya. Kendali dalam fase ini mulai dipindah kepada tim pemeliharaan perangkat lunak.

2.3 Penelitian Lain yang Relevan

2.3.1 *Landmark Breaker: Obstructing DeepFake By Disturbing Landmark Extraction*

Penelitian yang telah dilakukan mengenai *Landmark Breaker: Obstructing DeepFake By Disturbing Landmark Extraction* (Sun et al., 2020). Tulisan ini menjelaskan metode baru, yaitu *Landmark Breaker*, untuk menghalangi generasi *DeepFake* dengan melanggar langkah prasyarat ekstraksi *landmark* wajah. Dengan menciptakan *adversarial perturbations* untuk mengganggu ekstraksi *landmark* wajah, sehingga wajah input ke model *DeepFake* tidak dapat disejajarkan dengan baik. *Landmark Breaker* divalidasi pada himpunan data Celeb-DF, yang menunjukkan kemanjuran *Landmark Breaker* pada ekstraksi *landmark* wajah yang mengganggu.

2.3.2 Penelitian-Penelitian tentang *Face Modification*

Adapun penelitian tentang *face Modification* Dalam beberapa tahun terakhir, akses gratis ke gambar wajah berskala besar dan kemajuan luar biasa dari model generatif telah membuat jaringan modifikasi wajah menghasilkan gambar wajah yang lebih realistis dengan target orang atau atribut. StarGAN (Choi et al.,

2018) mengusulkan pendekatan baru dan terukur untuk melakukan penerjemahan gambar-ke-gambar di berbagai domain, mencapai kualitas visual yang lebih baik pada gambar yang dihasilkan. Kemudian, AttGAN (He et al., 2019) menggunakan batasan klasifikasi atribut untuk memberikan gambar wajah yang lebih alami pada manipulasi atribut wajah. Selain itu, AGGAN (Tang et al., 2019) memperkenalkan *attention mask* melalui mekanisme perhatian bawaan untuk mendapatkan gambar target dengan kualitas tinggi. Baru-baru ini, (Li et al., 2021) mengusulkan HiSD yang merupakan metode penerjemahan gambar-ke-gambar yang canggih untuk skalabilitas beberapa label dan keragaman yang dapat dikontrol dengan pelepasan yang mengesankan. Meskipun model-model ini mengadopsi beragam arsitektur dan kerugian, *watermark* CMUA kami berhasil mencegah gambar wajah dimodifikasi dengan benar oleh semuanya.

2.4 Kesimpulan

Pada bab ini telah dibahas teori yang akan digunakan sebagai dasar penelitian ini. Pada bab ini juga telah dibahas mengenai penelitian terkait yang mendukung literatur penelitian ini. Mekanisme pelaksanaan penelitian selengkapnya akan dibahas dalam bab selanjutnya.

BAB III

METODE PENELITIAN

3.1 Pendahuluan

Pada bab ini akan dijelaskan mengenai tahapan penelitian, metode penelitian serta manajemen proyek penelitian. Tahapan penelitian dijadikan sebagai acuan pada setiap fase pengembangan perangkat lunak agar dapat memberikan solusi untuk rumusan masalah dan tercapainya tujuan penelitian.

3.2 Unit Penelitian

Penelitian ini dilaksanakan di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

3.3 Pengumpulan Data

Pada bagian ini akan dijelaskan tahapan pengumpulan data meliputi jenis dan sumber data dan metode pengumpulan data yang digunakan dalam penelitian.

3.3.1 Jenis Data

Jenis data yang digunakan sebagai objek penelitian ini ada dua yaitu, data primer dan sekunder.

3.3.2 Sumber Data

1. Data primer berupa kumpulan data citra foto KPM mahasiswa Fasilkom UNSRI Angkatan 2018.

2. data sekunder berupa dataset Celeb-A dari peneletian *Deep Learning Face Attributes in the Wild* (Liu et al., 2015).seperti pada Gambar III-1

Gambar III-1. Contoh data yang dari *dataset* celebA

3.3.3 Metode pengumpulan Data

Ada dua metode pengumpulan data yang digunakan dalam penelitian ini:

1. *Crawling* data citra foto KPM mahasiswa Fasilkom Unsri Angkatan 2018 dari laman situs UNSRI lama[2].
2. mengunduh dataset Celeb-A dari di unduh dari halaman Kaggle[3].

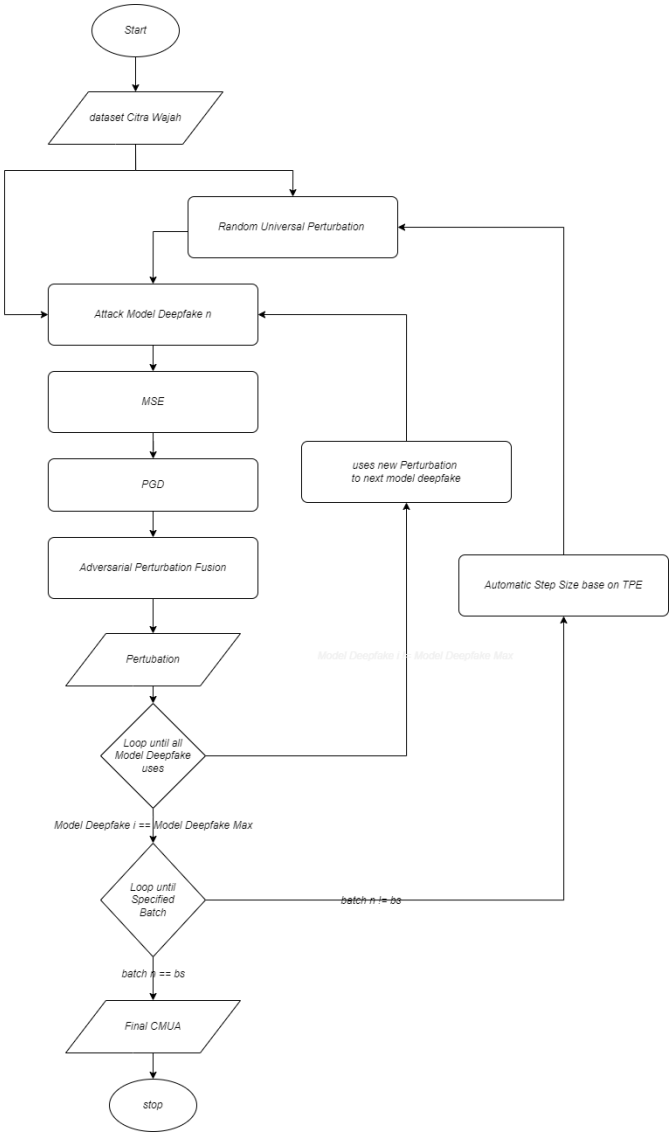
3.4 Tahapan Penelitian

Tahapan penelitian adalah rincian proses yang akan dilakukan pada penelitian. Tahapan penelitian yang akan dilakukan pada penelitian ini adalah sebagai berikut.

[2] https://old.unsri.ac.id/?act=daftar_mahasiswa
[3] https://www.kaggle.com/datasets/nikhilbartwal001/celeba

3.4.1 Kerangka Kerja

Kerangka kerja pada penelitian ini adalah sebagai berikut:



Gambar III-2. Diagram Alir Sistem Proteksi Citra dari *deepfake* dengan metode CMUA

Berdasarkan kerangka kerja penelitian pada Gambar III-2, sistem proteksi dari *deepfake* dengan metode CMUA memiliki alur sistem sebagai berikut:

1. Input Dataset Citra Wajah

Proses penginputan dataset yang digunakan dalam sistem ini, yaitu berupa citra gambar wajah dari dataset Celeb-A yang digunakan sebagai data latih.

2. Menyerang Dataset *Random Universal Perturbation*

Memasangkan proteksi awal dengan perturbation awal yang bernilai acak

3. Attack Model Deepfake

Melakukan penyerangan terhadap dataset yang belum diproteksi dan dataset yang yang telah di proteksi

4. *Mean Square Error* (MSE)

Pada tahapan ini gambar yang sebelumnya sudah diserang oleh model deepfake akan di uji dengan metode MSE untuk mengukur perbedaan antara $G(I_1)$... $G(I_n)$(gambar yang deepfake tanpa proteksi) dan $G(I_1 + W)$... $G(I_n + W)$(gambar yang telah diproteksi terlebih dahulu sebelum di deepfake), di mana E adalah nilai batas atas dari watermark lawan W.

$$\max_{W} \sum_{i=1}^{n} MSE(G(I_i), G(I_i + +W)), s.t. ||W||_\infty \leq \epsilon,$$

5. PGD

menggunakan PGD (Madry et al., 2018) sebagai metode serangan dasar untuk memperbarui *adversarial perturbations* pada setiap iterasi serangan,

$$I_{adv}^0 = I + W,$$

$$I_{adv}^{r+1} = clip_{I,\in}\{I_{adv}^{r} + a\ sign(\nabla_I L(G(I_{adv}^{r}), (G(I)))\}$$

6. Adversarial Perturbation Fusion

Konflik di antara watermark yang berlawanan yang dihasilkan dari gambar dan model yang berbeda akan mengurangi kemampuan transferabilitas *CMUA-Watermark* yang diusulkan. Untuk melemahkan konflik ini, digunakan strategi fusi gangguan dua tingkat selama proses serangan. Secara khusus, ketika menyerang satu model *deepfake* tertentu, akan melakukan **fusi tingkat gambar** untuk merata-rata gradien yang di *sign* dari sekumpulan gambar wajah,

$$G_{avg} = \frac{\sum_{j}^{bs} sign(\nabla_{I_j} L\left(G\left(I_j^{adv}\right), G\left(I_j\right)\right))}{bs}$$

di mana *bs* adalah ukuran kumpulan gambar wajah, dan $I^{adv}{}_j$ adalah *adversarial image* ke-j dari sebuah *batch*. Operasi ini akan menyebabkan G_{avg} lebih berkonsentrasi pada atribut umum wajah manusia daripada atribut wajah tertentu. Kemudian, menggunakan PGD untuk menghasilkan adversarial perturbation P_{avg} melalui $G_{avg.}$

Setelah mendapatkan P_{avg} dari satu model, melakukan **fusi tingkat model**, yang secara iteratif menggabungkan P_{avg} yang dihasilkan dari model tertentu ke W_{CMUA} dalam pelatihan, dan W_{CMUA} awal hanyalah P_{avg} yang dihitung dari model deepfake pertama,

$$W_{CMUA}^{o} = p_{avg}^{0},$$

$$W_{CMUA}^{t+1} = \alpha\ .\ W_{CMUA}^{t} + (1-\alpha).\ P_{avg}^{t},$$

di mana α adalah faktor peluruhan, P^t_{avg} adalah rata-rata gangguan yang dihasilkan dari model *deepfake* yang diserang ke-*t*, dan W^t_{CMUA} adalah *CMUA-Watermark* pelatihan setelah model *deepfake* yang diserang ke-*t*.

7. Automatic Step Size Tuning based on TPE

Selain fusi dua tingkat yang disebutkan di atas, ditemukan bahwa ukuran langkah serangan untuk model yang berbeda juga penting untuk transferabilitas CMUA-Watermark yang dihasilkan. Oleh karena itu, mengeksploitasi pendekatan heuristik untuk secara otomatis menemukan ukuran langkah serangan yang sesuai.

Metode serangan dasar yang dipilih (PGD) termasuk ke dalam keluarga FGSM (Goodfellow et al., 2015), dan gradien $\nabla_x L$ dinormalisasi oleh fungsi *sign*:

$$sign\, x = \begin{cases} -1\ , & x < 0, \\ 0\ , & x = 0, \\ 1\ , & x > 0. \end{cases}$$

Dalam perhitungan nyata, elemen-elemen dalam $\nabla_x L$ hampir tidak pernah mencapai 0, sehingga $||\mathrm{sign}(\nabla_x L)||2 \approx 1$ adalah tetap untuk setiap gradien. Perturbasi ΔP yang diperbarui dalam iterasi metode serangan berbasis *sign* dirumuskan sebagai:

$$\Delta P = a.sign(\nabla_X L).$$

Dengan kata lain, hanya ukuran langkah *a* yang menentukan tingkat pembaruan selama serangan, sehingga pemilihan *a* memiliki pengaruh yang besar terhadap performa serangan. Kesimpulan ini juga berlaku untuk serangan universal lintas

model; perturbasi yang diperbarui ΔP^u dalam sebuah iterasi serangan universal lintas model dibentuk dengan menggabungkan ΔP^i dari beberapa model G_1,..., G_m:

$$\Delta P^u = \sum_{i=1}^{m} \alpha^{(m-i)} \Delta P_i = \sum_{i}^{m} \alpha^{(m-i)} a_i . sign(\nabla_X L_i) .$$

Dalam rumus di atas, *m* adalah jumlah model, faktor peluruhan α adalah sebuah konstanta, dan *sign* ($\nabla_X L_i$) memberikan arah optimasi untuk G_i . Oleh karena itu, arah optimasi secara keseluruhan sangat dipengaruhi oleh a_1, ..., a_m, dan memilih a_1, ..., a_m yang sesuai di berbagai model untuk menemukan arah keseluruhan yang ideal adalah masalah utama untuk serangan lintas model.

Menggunakan algoritma TPE ((Bergstra et al., 2011)) untuk memecahkan masalah ini, yang secara otomatis mencari a_1, ..., a_m yang sesuai untuk menyeimbangkan arah yang berbeda yang dihitung dari berbagai model. TPE adalah metode optimasi hiper-parameter berdasarkan *Sequential Model-Based Optimization* (SMBO), yang secara berurutan membangun model untuk memperkirakan kinerja hiper-parameter berdasarkan pengukuran historis, dan kemudian memilih hiperparameter baru untuk diuji berdasarkan model ini. Dalam penelitian ini, menganggap ukuran langkah a_1, ..., a_m sebagai hyperparameter input *x* dan tingkat keberhasilan serangan sebagai nilai kualitas y dari TPE. TPE menggunakan $P(x|y)$ dan $P(y)$ untuk memodelkan $P(y|x)$, dan $p(x|y)$ diberikan oleh:

$$p(x|y) = \begin{cases} l(x), & if\ y < y^*, \\ g(x), & if\ y \geq y^*, \end{cases}$$

di mana y^* ditentukan oleh pengamatan terbaik secara historis, e(x) adalah densitas yang dibentuk dengan pengamatan {$x(i)$} sedemikian rupa sehingga kerugian yang sesuai lebih rendah dari y^* , dan $g(x)$ adalah densitas yang dibentuk dengan pengamatan yang tersisa. Setelah memodelkan $P(y|x)$, lalu terus mencari ukuran langkah yang lebih baik dengan mengoptimalkan kriteria *Expected Improvement* (EI) di setiap iterasi pencarian, yang diberikan oleh,

$$EI_{y^*}(x) = \frac{\gamma y^* l(x) - l(x)\int_{-\infty}^{y^*} p(x)dy}{\gamma l(x) + (1-\gamma)g(x)}$$
$$\propto \left(y + \frac{g(x)}{l(x)}(1-\gamma)\right)^{-1},$$

di mana γ = p (y < y^*). Dibandingkan dengan kriteria lainnya, EI bersifat intuitif dan telah terbukti memiliki kinerja yang sangat baik. Untuk detail lebih lanjut mengenai TPE, lihat (Bergstra et al., 2011).

8. Final CMUA-Watermark

Setelah semua iterasi *batch max* yang diinginkan selesai didapatkan *CMUA-watermark* final yang siap digunakan dan diimplementasikan keberbagai jenis gambar wajah untuk proteksi citra dari deepfake.khusunya foto KPM mahasiswa Unsri Fasilkom 2018.

3.4.2 Kriteria Pengujian

Kriteria pengujian menggunakan matriks *Mask*, yang lebih berkonsentrasi pada area yang dimodifikasi,

$$Mask_{(i,j)} = \begin{cases} 1, & if \left|\left|G(I)_{(i,j)} - I_{(i,j)}\right|\right| > 0.5, \\ 0, & else \end{cases}$$

di mana (*i*, *j*) adalah koordinat piksel dalam gambar. Dengan cara ini, ketika menghitung L^2_{mask}, hanya piksel dengan perubahan besar yang akan dihitung dan area lainnya akan ditinggalkan,

$$L^2{}_{mask} = \frac{\sum_i \sum_j Mask_{(i,j)} . \left|\left| G(I)_{(i,j)} - G(I + C_{CMUA})_{(i,j)} \right|\right|}{\sum_i \sum_j Mask_{(i,j)}}.$$

Dalam eksperimen kami, jika $L^2_{mask} > 0{,}05$, kami menentukan bahwa gambar berhasil dilindungi, dan menggunakan SR_{mask} untuk merepresentasikan tingkat keberhasilan melindungi gambar wajah.

3.4.3 Format data Pengujian

Format data pengujian yang digunakan berupa table

- L^2_{mask} adalah perbandingan antara *original image* dengan *distorted* image
- SR_{mask} untuk merepresentasikan tingkat keberhasilan melindungi gambar wajah

dataset	$L^2_{mask}\uparrow$	$SR_{mask}\uparrow$
CelebA Training		
CelebA verifikasi		

Foto KPM mahasiswa Fasilkom Unsri 2018		

Tabel III-1. Tabel akurasi dan tingkat keberhasilan CMUA pada tiap Dataset

3.4.4 Alat yang digunakan dalam Pelaksanaan Penelitian

Alat bantu penelitian Proteksi Citra Foto KPM Mahasiswa Fasilkom UNSRI dari *DeepFake* dengan *CMUA-Watermark* yang akan digunakan dalam penelitian ini adalah sebagai berikut.

1. Perangkat Keras

   Proccessor : intel® core™ i7 8th *generation*

   RAM : 16GB RAM

   HDD : 1TB HDD strorage.

   VRAM : Nvidia Geforce GTX 1050 4GB RAM

2. Perangkat Lunak

   Sistem Operasi : Windows 10 64-bit

   Teks Editor : Visual Studio Code

   Basaha Pemprograman : Python

Selain perangkat lunak diatas diatas.terdapat tambahan perangkat lunak ketiga

3.4.5 Pengujian Penelitian

Pengujian penelitian akan dijabarkan pada bab V.

3.4.6 Analisis dan Kesimpulan Hasil Pengujian Penetian

Analisi hasil pengujian dilakukan dengan memperhatikan nilai $L^2_{mask\ untuk}$ menentukan bahwa gambar berhasil dilindungi dan SR_{mask} untuk merepresentasikan tingkat keberhasilan melindungi gambar wajah. Serta perbandingan tingkat keberhasilan dari masing masing dataset yang diujikan.

3.5 Metode Pengembangan Perangkat Lunak

Metode pengembangan yang digunakan dalam penelitian ini adalah metode Rational Unified Process (RUP). Pengembangan sistem deteksi kemiripan kode sumber dibagi ke dalam empat tahap, yaitu fase insepsi, fase elaborasi, fase konstruksi dan fase transisi. Berikut merupakan tahapan pengembangan perangkat lunak yang akan dilakukan dalam tiap fasenya.

3.5.1 Face Insepsi

Tahapan yang akan dilakukan dalam fase ini adalah sebagai berikut.

1. Pemodelan Sistem : Menentukan ruang lingkup dan batasan masalah.
2. Kebutuhan : Mendefinisikan spesifikasi perangkat lunak.
3. Analisis dan Perancangan : Melakukan analisis terhadap kebutuhan perangkat lunak termasuk di dalamnya kebutuhan fungsional dan non fungsional dari spesifikasi perangkat lunak.
4. Implementasi : Membuat seluruh rancangan sistem ke dalam bentuk diagram use-case.

3.5.2 Fase Elaborasi

Tahapan yang akan dilakukan dalam fase ini adalah sebagai berikut.

1. Pemodelan Sistem: Membuat rancangan antarmuka (interface) sistem.
2. Kebutuhan: Menentukan spesifikasi dari sistem.
3. Analisis dan Perancangan: Membangun model *activity diagram* dan *sequence diagram* dari rancangan sistem.
4. Impelementasi: Membuat program berdasarkan diagram yang ditentukan sebelumnya.

3.5.3 Fase Konstruksi

Tahapan yang akan dilakukan dalam fase ini adalah sebagai berikut.

1. Pemodelan Bisnis : Menentukan bahasa pemrograman yang akan membangun sistem.
2. Kebutuhan : Menentukan kebutuhan sistem sesuai dengan fungsi yang telah ditentukan.
3. Analisis dan Perancangan : Membangun tampilan antar-muka sistem.
4. Implementasi: Membangun sistem dengan membuat program menggunakan bahasa pemrograman yang telah ditentukan.

3.5.4 Fase Transisi

Tahapan yang akan dilakukan dalam fase ini adalah sebagai berikut.

1. Pemodelan Sistem : Menentukan pengujian terhadap sistem.
2. Kebutuhan : Menentukan alat bantu pengujian terhadap sistem.
3. Analisis dan Perancangan : Merancang kasus penggunaan selama pengujian sistem.

4. Implementasi : Melaksanakan pengujian terhadap sistem menggunakan kasus penggunaan yang telah ditentukan.

3.6 Manajemen Proyek Perangkat Lunak

Rencana manajemen proyek penelitian merupakan perencanaan aktivitas penelitian dari tahap awal hingga selesai. Perencanaan aktivitas pada penelitian ini akan menggunakan *Gantt Chart* seperti pada Tabel III-4.

No.	Uraian Kegiatana	Tahun 2023 bulan ke-					
		1	2	3	4	5	6
1	**Melakukan Pengumpulan Data**						
a	Mengumpulkan data						
b	Melakukan pra-pengolahan data						
c	Melakukan modifikasi data sesuai skenario						
d	Tersedia dokumen hasil tahapan penelitian						
2	**Rekayasa Perangkat Lunak**						
2.1	**Insepsi**						
a	Menentukan pemodelan bisnis						
b	Menentukan kebutuhan pengguna						
c	Menentukan kebutuhan sistem						
2.2	**Elaborasi**						
a	Menentukan spesifikasi sistem						
b	Membangun model *acitvity diagram* dan *sequence diagram*						
c	Membangun rancangan tampilan antar-muka						
2.3	**Konstruksi**						
a	Membangun Model *Class Diagram*						
b	Membangun Sistem						

	(impelementasi kode)						
c	Perbaikan Sistem						
2.4	**Transisi**						
a	Melakukan pengujian awal terhadap sistem						
b	Tersedia dokumen hasil tahapan penelitian						
3	**Melakukan Pengujian Penelitian Terhadap Sistem**						
a	Membuat rancangan hasil pengujian dalam penelitian						
b	Melakukan pengujian final terhadap sistem						
c	Tersedia dokumen hasil penelitian						
4	**Melakukan Analisis dan Kesimpulan dari Hasil Pengujian**						
a	Melakukan analisis terhadap hasil pengujian penelitian						
b	Membuat kesimpulan dan saran terhadap hasil pengujian penelitian						
c	Tersedia dokumen hasil penelitian						

Tabel III-2. Tabel Rencana Manajemen Proyek Penelitian

3.7 Kesimpulan

Pada bab ini telah dibahas tentang proses pengumpulan data yang digunakan sebagai bahan uji perangkat lunak, tahapan penelitian, metode pengembangan perangkat lunak yang akan digunakan serta kriteria pengujian penelitian yang akan dilakukan terhadap sistem.

DAFTAR PUSTAKA

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, *141*, 215–222. https://doi.org/10.1016/j.procs.2018.10.171

Anderson, K. E. (2018). Getting acquainted with social networks and apps: combating fake news on social media. *Library Hi Tech News*, *35*(3), 1–6. https://doi.org/10.1108/LHTN-02-2018-0010

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

Britt, M. A., Rouet, J. F., Blaum, D., & Millis, K. (2019). A Reasoned Approach to Dealing With Fake News. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 94–101. https://doi.org/10.1177/2372732218814855

Brown, D. (2019). *Wait, is that video real? The race against deepfakes and dangers of manipulated recordings*.

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

Dickson, B. (2018). When AI Blurs the Line Between Reality and Fiction. *PCMag*. https://www.pcmag.com/news/when-ai-blurs-the-line-between-reality-and-fiction

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: Challenges and opportunities. *Procedia Computer Science*, *121*, 817–825. https://doi.org/10.1016/j.procs.2017.11.106

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

Horowitz, B. T. (2019). *AI and Machine Learning Exploit, Deepfakes, Now Harder to Detect*.

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

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.

Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, *80*, 295–302. https://doi.org/10.1016/j.chb.2017.11.034

Leetaru, K. (2019). *DeepFakes: The Media Talks Politics While The Public Is Interested In Pornography*.

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.

Marr, B. (2019). The Best (And Scariest) Examples Of AI-Enabled Deepfakes. *Forbes*.

Metz, R. (2019, June 12). *The fight to stay ahead of deepfake videos before the 2020 US election*. CNN BUSINESS. https://edition.cnn.com/2019/06/12/tech/deepfake-2020-detection/index.html

Metz, R., & O'Sullivan, D. (2019). A deepfake video of Mark Zuckerberg presents a new challenge for Facebook. *CNN BUSINESS*. https://edition.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html

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.

Rubenking, N. J., & Eddy, M. (2019). Detecting Deepfakes May Mean Reading Lips. *PCMag*. https://www.pcmag.com/news/detecting-deepfakes-may-mean-reading-lips

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

Sun, P., Li, Y., Qi, H., & Lyu, S. (2020). Landmark Breaker: Obstructing DeepFake by Disturbing Landmark Extraction. *2020 IEEE International Workshop on Information Forensics and Security, WIFS 2020*, 6–11. https://doi.org/10.1109/WIFS49906.2020.9360910

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

Wallace, D. (2019). "Deepfake" videos, other tech could threaten 2020 election, Dems fear. *FOX NEWS*. https://www.foxnews.com/tech/house-lawmakers-concerned-deepfake-tech-could-pose-national-security-threat-before-2020-election

Yang, C., Liu, X., & Sun, X. (2019). Steganography Algorithms for Digital Image Hiding: A Review. *IEEE Access*, *7*, 134764–134781.

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other

zzshenanigans. *Journal of Data and Information Quality*, *11*(3). https://doi.org/10.1145/3309699