**Proteksi Citra 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 dari DeepFake dengan CMUA-Watermark**

Oleh :

Renaldi Budi Setiawan

NIM : 09021281823066

Indralaya,Oktober 2022

Pembimbing I Pembimbing II,

Syamsuryadi, S.Si., M.Kom., Ph.D. Muhammad Qurhanul Rizqie  
NIP 197102041997021003 NIP

Mengetahui,  
ketua Jurusan

Alvi Syahrini Utami, M.Kom.  
NIP. 19781222200642003

DAFTAR ISI

Halaman

HALAMAN JUDUL i

HALAMAN PENGESAHAN ii

HALAMAN PERTUJUAN KOMISI PENGUJI iii

HALAMAN MOTTO DAN PERSEMBAHAN iv

ABSTRACT v

ABSTRAKSI vi

KATA PENGANTAR vii

DAFTAR TABEL viii

DAFTAR GAMBAR ix

DAFTAR LAMPIRAN x

DAFTAR ISTILAH, SINGKATAN DAN LAMBANG xi

BAB I PENDULUAN

* 1. Pendahuluan 1
  2. Latar Belakang Masalah 1
  3. Rumusan Masalah 2
  4. Tujuan Penelitian 2
  5. Manfaat Penelitian 2
  6. Batasan Masalah
  7. Sistematika penulisan
  8. Kesimpulan

BAB II KAJIAN LITERATUR

* 1. Pendahuluan
  2. Landasan Teori
  3. Penelitian Lain yang Relevan
  4. Kesimpulan

BAB III METODE PENELITIAN

* 1. Pendahuluan
  2. Landasan Teori
     1. Jenis dan Sumber Data
     2. Metode Pengumpulan Data
  3. Tahapan penelitian
     1. Mengumpulkan Data
     2. Menentukan Kerangka Kerja Penelitian
     3. Menentukan Kriteria Pengujian \
     4. Menentukan Format Data Pengujian
     5. Menentukan Alat Bantu Penelitian
     6. Melakukan Pengujian Penelitian
     7. Mengevaluasi Hasil Penelitian dan Membuat Kesimpulan
  4. Metode pengembang Perangkat Lunak
  5. Kesimpulan

# **BAB I PENDAHULUAN**

## **Pendahuluan**

Pada bab ini akan dibahas berkenaan dengan garis besar pokok-pokok pikirandalam 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.

## **Latar Belakang Masalah**

Dalam beberapa tahun terakhir, berita palsu telah menjadi isu yang merupakan ancaman bagi wacana publik, masyarakat manusia, dan demokrasi (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, di mana hal itu dapat berdampak pada jutaan pengguna (Figueira & Oliveira, 2017). Saat ini, satu dari lima pengguna internet mendapatkan berita mereka melalui YouTube, kedua setelah Facebook (Anderson, 2018). Peningkatan popularitas video ini menyoroti perlunya alat untuk mengkonfirmasi keaslian konten media dan berita, karena teknologi baru memungkinkan manipulasi video yang meyakinkan (Anderson, 2018).

Kemajuan teknologi baru-baru ini telah membuatnya mudah untuk menciptakan apa yang sekarang disebut "deepfakes", video hiper-realistis menggunakan face swap yang meninggalkan sedikit jejak manipulasi (Chawla, 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, 2018). Teknologi Deepfake dapat menghasilkan, misalnya, video lucu, pornografi, atau politik seseorang yang mengatakan apa pun, tanpa persetujuan orang yang gambar dan suaranya terlibat (Day, 2019; Fletcher, 2018). Faktor deepfakes yang mengubah permainan adalah ruang lingkup, skala, dan kecanggihan teknologi yang terlibat, karena hampir semua orang dengan komputer dapat membuat video palsu yang praktis tidak dapat dibedakan dari media otentik (Fletcher, 2018).

Untungnya, adversarial watermark dapat digunakan untuk memerangi model deepfake, adversarial watermark dapat menghasilkan gambar yang terdistorsi. Metode yang ada memerlukan proses pelatihan individu untuk setiap gambar wajah, untuk menghasilkan adversarial attack model terhadap model deepfake tertentu, yang sangat tidak efisien. Untuk mengatasi masalah ini, penelitian ini menggunakan metode universal adversarial attack model pada model deepfake, untuk menghasilkan Cross-Model Universal Adversarial Watermark (CMUA-Watermark) yang dapat melindungi ribuan gambar wajah dari beberapa model deepfake. Secara khusus, mengusulkan pipeline universal adversarial cross model dengan menyerang beberapa model deepfake dan menggabungkan gradien dari model-model ini secara berulang. Kemudian memperkenalkan metode berbasis batch untuk mengurangi konflik adversarial watermark yang dihasilkan oleh gambar wajah yang berbeda. Akhirnya, merancang metode evaluasi yang lebih masuk akal dan *comprehensive* untuk mengevaluasi efektivitas adversarial Watermark.

## **Rumusan Masalah**

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

1. Bagaimana cara memproteksi citra gambar dengan metode CMUA-Watermark?
2. Bagaimana tingkat akurasi metode CMUA-Watermark dalam memproteksi citra dari deepfakes?

## **Tujuan Masalah**

Tujuan dari penelitian ini adalah:

1. Membangun perangkat lunak yang dapat memproteksi citra gambar menggunakan metode CMUA-Watermark.
2. Mengetahui tingkat akurasi penggunaan metode CMUA-Watermark dalam memproteksi citra dari deepfakes.

## **Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini adalah:

1. Sistem yang dibuat dapat membantu pengguna untuk memproteksi citra gambar menggunakan metode CMUA-Watermark.
2. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian terkait di masa mendatang.

## **Batasan Masalah**

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan merupakan dataset yang dipakai dalam penelitian milik dw (202x).
2. Data uji yang digunakan merupakan dataset foto mahasiswa jurusan Teknik Informatika Universitas Sriwijaya Angkatan 2018.
3. Ekstensi citra yang didukung oleh perangkat lunak adalah .jpg.

## **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, CMUAI-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.

## **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**

## **Pendahuluan**

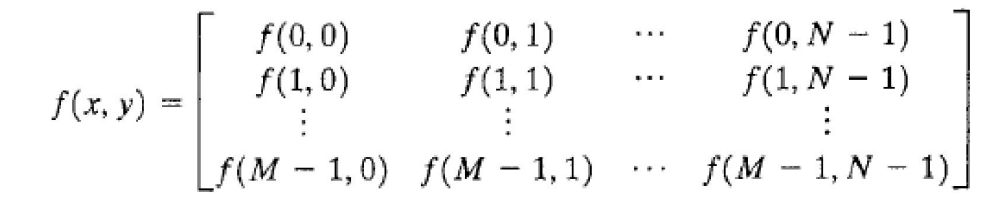
pada bab ini ini akan dijelaskan menganai dasar-dasar teori digunakan pada penelitian ini. Serta penjelasan hasil dari penelitian-penitian terkait pengenai citra, *Deepfakes*, CMUAI-Watermark, RUP, dan penelitian yang relevan*.*

## **Landasan Teori**

### **Citra**

Citra didefinisikan sebagai fungsi dari dua variable misalnya *a*(*x,y*) dimana *a* sendiri sebagai amplitudo (misalnya kecerahan) citra pada koordinat (x,y). Citra digital *a*[*m,n*] merupakan citra dalam ruang diskrit 2D yang berasal dari citral Analog *a*(*x,y*) di ruang kontinyu 2D melalui proses sampling yaitu yang biasa disebut sebagai digitalisasi (Young et al., 2006).

Menurut McAdrew citra digital adalah citra *f*(*x,y*) yang telah diskritkan pada kordinat spasial dan kecerahan. Citra digital direpresentasikan oleh *array* dua dimensi dimana setiap *array* merepresentasikan satu kanal warna. Nilai warna kecerahan yang didigitalkan ini dinamakan nilai tingkat keabuan . Setiap elemen array tersebut dinamakan piksel atau pel yang diambil dari istilah ‘picture element’. Dimensi pada citra ditulis dengan format panjang x tinggi. Namun pada citra digital didefinisikan dengan ukuran tinggi *M* dan panjang *N.*



Koordinat citra dimulai dari pojok kiri atas, secara Sistematis di dimulai dari (0,0) dan berakhir di (M-1,N-1)(McAndrew, 2014).

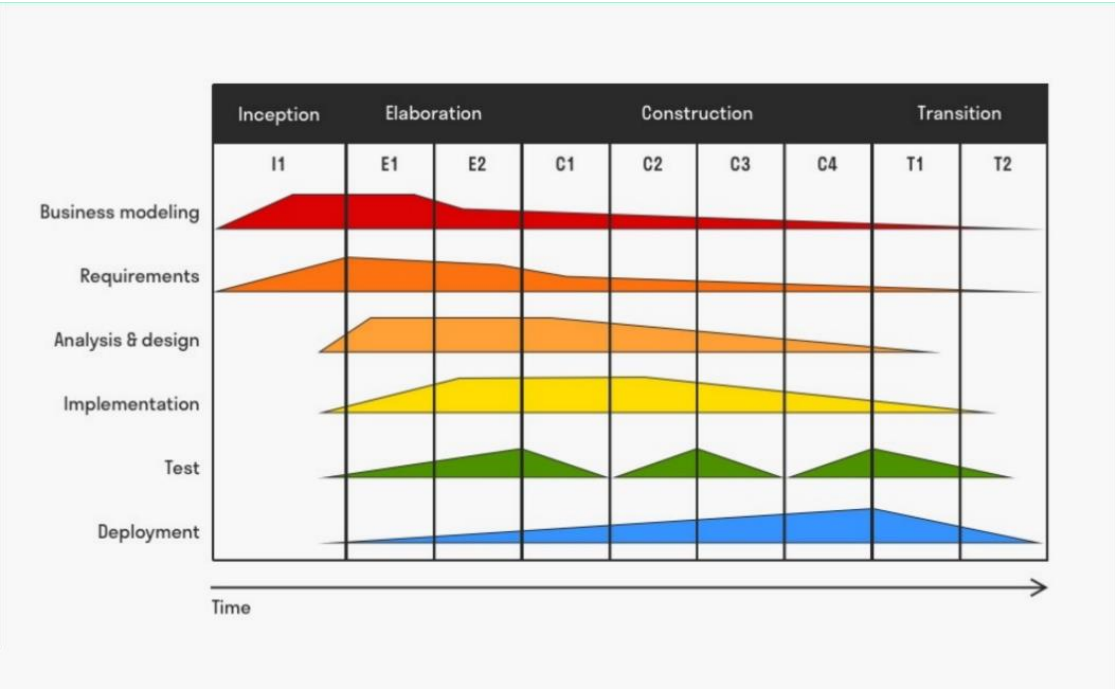
### **DeepFakes**

Deepfake merupakan teknik sintetis citra manusia yang berdasarkan pada kecerdasan buatan/AI. Ini digunakan untuk menggabungkan serta menempatkan gambar dan video yang ada ke sumber gambar atau video menggunakan teknik mesin belajar yang dikenal sebagai jaringan generatif adversarial (generative adversarial network) atau GAN(Hidayatul Khusna Sri Pangestuti, 2019).

### **CMUAI-Watermark**

### **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). 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.

  
Gambar \*\*-\*,Arsitektur Rasional Unified Process

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.

## **Penelitian Lain yang Relevan**

Penelitian yang telah dilakukan mengenai *adversarial machine learning at scale (A. Kurakin, I. Goodfellow, and S. Bengio, in ICLR, 2017.).* Dalam makalah ini mempelajari cara meningkatkan ketahanan terhadap contoh adversarial model besar (*Inception v3*) yang dilatih pada set data besar (*ImageNe*t). Menunjukkan bahwa pelatihan adversarial memberikan ketahanan terhadap contoh adversarial yang dihasilkan menggunakan metode satu langkah. Meskipun pelatihan adversarial tidak banyak membantu melawan metode iteratif yang diamati bahwa contoh adversarial yang dihasilkan oleh metode iteratif lebih kecil kemungkinannya untuk ditransfer di antara jaringan, yang memberikan ketahanan tidak langsung terhadap serangan adversarial kotak hitam. Selain itu kami mengamati bahwa peningkatan kapasitas model juga dapat membantu meningkatkan ketahanan terhadap contoh adversarial terutama ketika digunakan bersamaan dengan pelatihan adversarial. Akhirnya menemukan efek kebocoran label yang menghasilkan akurasi yang lebih tinggi pada contoh adversarial FGSM dibandingkan dengan contoh bersih ketika jaringan dilatih secara adversarial.

## **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**

## **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.

## **Pengumpulan Data**

Pada bagian ini akan dijelaskan mengenai data yang digunakan dalam

penelitian.

### Jenis dan Sumber Data

Jenis data yang digunakan sebagai objek penelitian adalah data Sekuder berupa data Kumpulan data dari *dataset* data CelebA (Liu et al., in Deep learning face attributes in the wild, 2015.) Gambar \*\*\*-\* contoh data yang terdapat dalam dataset CelebA.

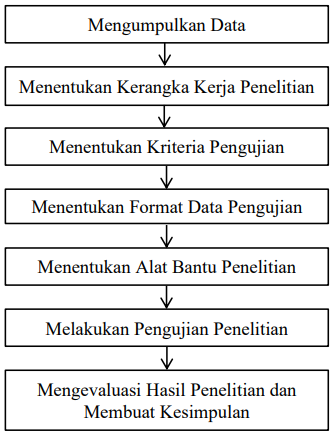
Gambar III-1. Contoh data yang dari *dataset* celebA

### Metode pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data berupa video DeepFake dengan intensitas tinggi dari 59 wajah selebiriti dari berbagai umur yang terdapat dalam *dataset* Celeb-df. Untuk menunjukkan keefektifan *Landmark Breaker* dalam menghambat generasi *DeepFake*. Setiap video berisi satu subjek dengan berbagai pose kepala dan ekspresi wajah.

## **Tahapan Penelitian**

tahapan penelitian yang dilakukan pada tugas akhir ini adalah:



Gambar III-2.alur Tahapan penelitian

Mengumpulkan data

data berupa video DeepFake dengan intensitas tinggi dari 59 wajah selebiriti dari berbagai umur yang terdapat dalam *dataset* Celeb-df. Untuk menunjukkan keefektifan *Landmark Breaker* dalam menghambat generasi *DeepFake*. Setiap video berisi satu subjek dengan berbagai pose kepala dan ekspresi wajah.

### Menentukan Kerangka Kerja Penelitian

Kerangka kerja yang dilakukan pada penelitian ini adalah sebagai berikut :

### Menetukan Kriteria Pengujian

Kriteria pengujian yang akan dilakukan pada penelitian:

Pengujian dilakukan dengan menggunakan metode Regresi Logistik Ordinal.

### Penarikan Hipotesa

Format data pengujian menggunakan pendekatan Confusion Matrix.

### Menentukan Sumber Data

Dalam melaksanakan penelitian Menentukan tingkat keparahan luka pada kecelakaan lalu lintas menggunakan metode Regresi Logistik Ordinal dibutuhkan alat dalam menunjang proses penelitian. Adapun alat yang digunakan sebagai berikut

1. Perangkat Keras

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

RAM : 16GB RAM

HDD : 1TB HDD strorage.

1. Perangkat Lunak

Sistem Operasi : Windows 10 64 bit

Teks Editor : Spyder

### Melakukan Pengujian Penelitian

Pengujian penelitian dimulai dengan tahap pra-pengulahan data masukkan, kemudian akan dilakukan beberapa tahap pengujian terhadap data tersebut.

### Mengevalusi Hasil penelitian dan Membuat kesimpulan

## **Metode Pengembangan Perangkat Lunak**

Metodologi pengembangan perangkat lunak yang digunakan dalam penelitian ini menggunakan pemograman berorientasi objek. Metode yang digunakan adalah Rational Unified Process (RUP). Berikut adalah fase-fase pada RUP yang akan dilakukan pada penelitian.

### Face Inception

Hal-hal yang dilakukan pada fase ini adalah :

1. Pemodelan Bisnis : Menentukan ruang lingkup masalah.
2. Kebutuhan : Menentukan kebutuhan fungsional dan non-fungsional serta fitur-fitur pada perangkat lunak.
3. Analisis dan Perancangan : Menganalisis kebutuhan fungsional dan non-fungsional serta menganalisis skenario alur proses pada perangkat lunak.
4. Implementasi : Merancang use case berdasarkan kebutuhan fungsional.

Tahapan yang dilakukan dalam fase ini antara lain:

### Fase Elaboration

Hal-hal yang dilakukan pada fase ini adalah :

1. Pemodelan Bisnis : Merancang antarmuka perangkat lunak.
2. Kebutuhan : Menentukan kebutuhan untuk membangun perangkat lunak.
3. Analisis dan Perancangan : Membuat model diagram aktivitas dan diagram alur.

### Fase Construction

Hal-hal yang dilakukan pada fase ini adalah :

1. Kebutuhan : Membuat model diagram kelas.
2. Implementasi : Membuat kode program menggunakan bahasa pemograman yang telah ditentukan.

### Fase Transition

Tahapan yang akan dilakukan pada fase ini adalah:

1. Pemodelan Bisnis : Menentukan rencana pengujian perangkat lunak.
2. Kebutuhan : Menentukan alat yang digunakan untuk melakukan pengujian.
3. Analisis dan Perancangan : Membuat skenario pengujian use case.
4. Implementasi : Melakukan pengujian use case.

## **Kesimpulan**

Bab ini telah membahas mengenai tahapan penelitian yang akan dilakukan serta metode pengembangan perangkat lunak yang akan digunakan. Selain itu juga menjelaskan mengenai teknik pengumpulan data sebagai bahan uji dalam penelitian