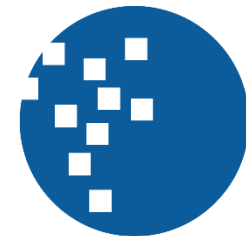


**UNSUPERVISED FACE RECOGNITION MENGGUNAKAN  
DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)**



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

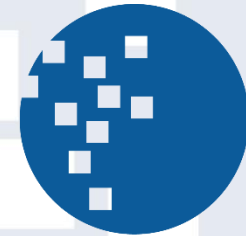
**REVIVAL CHRISTIAN HIZKIA KAUNANG**

**00000030714**

**PROGRAM STUDI TEKNIK KOMPUTER  
FAKULTAS TEKNIK & INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2024**

**UNSUPERVISED FACE RECOGNITION MENGGUNAKAN  
DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)**



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

Diajukan sebagai Salah Satu Syarat untuk Memperoleh  
Gelar Sarjana Teknik Komputer

**REVIVAL CHRISTIAN HIZKIA KAUNANG**

**00000030714**

**PROGRAM STUDI TEKNIK KOMPUTER**

**FAKULTAS TEKNIK & INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Revival Christian Hizkia Kaunang

Nomor Induk Mahasiswa : 00000030714

Program studi : Teknik Komputer

Tugas Akhir dengan judul:

UNSUPERVISED FACE RECOGNITION MENGGUNAKAN DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 5 November 2024



(Revival Christian Hizkia Kaunang)

UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

SKRIPSI dengan judul  
UNSUPERVISED FACE RECOGNITION MENGGUNAKAN DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)

Oleh

Nama : Revival Christian Hizkia Kaunang

NIM : 00000030714

Program Studi : Teknik Komputer

Fakultas : Fakultas Teknik & Informatika

Telah diujikan pada hari Kamis, 24 Oktober 2024

Pukul 09.00 s.d 11.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang



Samuel Hutagalung, M.T.I.

NIDN: 0304038902

Penguji



Nabila Husna Shabrina, S.T., M.T.

NIDN: 321099301

Pembimbing



Monica Pratiwi, S.ST., M.T.

NIDN: 0325059601

Ketua Teknik Komputer



Samuel Hutagalung, M.T.I.

NIDN: 0304038902

iii

## **HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Revival Christian Hizkia Kaunang  
NIM : 00000030714  
Program Studi : Teknik Komputer  
Fakultas : Fakultas Teknik & Informatika  
Jenis Karya : \*Tesis/Skripsi/Tugas Akhir (\*coret salah satu)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

### **UNSUPERVISED FACE RECOGNITION MENGGUNAKAN DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 5 November 2024

Yang menyatakan,

  
(Revival Christian Hizkia Kaunang)

## KATA PENGANTAR

Bersyukur dan berterima kasih kepada Tuhan yang Maha Esa karena kasih karuniaNya sehingga pelaksanaan pembuatan Skripsi saya dapat diselesaikan dengan baik yang berjudul “UNSUPERVISED FACE RECOGNITION MENGGUNAKAN DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)”.

Penulis juga ingin menyampaikan terima kasih yang terbesar kepada :

Mengucapkan terima kasih

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Pak Samuel, M.T.I, selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Bu Monica Pratiwi, S.ST., M.T., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya skripsi ini.
5. Keluarga saya Papa, Mama, Dede, dan Memei yang ada di manado, yang selalu setia mendokan dan mendukung.
6. Bapak dan Ibu Gembala dan teman teman pelayanan saya di GPdI Hebron Gading Serpong atas semua doa dan semangat yang diberikan. Dan juga untuk anak HOME saya Geva karena telah meminjamkan leptopnya untuk saya selama pengerjaan skripsi.
7. Kemudian kepada pihak-pihak lain yang tidak bisa dicantumkan satu persatu yang telah membantu dalam pembuatan skripsi.

Harapan untuk karya ilmiah ini adalah bahwa hasil penelitian dan pengembangan aplikasi web yang telah dilakukan akan memberikan kontribusi positif dan bermanfaat dalam konteks manajemen data jemaat gereja, khususnya di gereja

GPdI Hebron Gading dan mungkin juga di gereja-gereja lain yang menghadapi tantangan serupa.

Tangerang, 5 November 2024



(Revival Christian Hizkia Kaunang)



# UNSUPERVISED FACE RECOGNITION MENGGUNAKAN DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)

Revival Christian Hizkia Kaunang

## ABSTRAK

Penelitian ini mengeksplorasi penggunaan model Data-efficient Image Transformer (DeiT) untuk pengenalan wajah tanpa pengawasan. Dengan memanfaatkan dataset MS1M-RetinaFace yang tidak berlabel, penelitian ini bertujuan untuk mengembangkan sistem pengenalan wajah yang efisien dan akurat tanpa bergantung pada data berlabel. Metodologi mencakup preprocessing data, pelatihan model DeiT yang dimodifikasi, dan evaluasi performa menggunakan berbagai metrik. Hasil penelitian menunjukkan potensi signifikan DeiT dalam konteks pengenalan wajah tanpa pengawasan, dengan mencapai Rank-1 Accuracy sebesar 80.84%. Meskipun nilai TAR@FAR=1e-4 untuk evaluasi N:N dan N:M relatif rendah, metrik clustering menunjukkan hasil yang menjanjikan dengan Silhouette Score 0.1287, Davies-Bouldin Index 1.5080, dan Calinski-Harabasz Index 57.7656, mengindikasikan kemampuan model dalam membedakan fitur wajah meski masih memerlukan optimasi lebih lanjut.

**Kata kunci:** Data-efficient Image Transformer (DeiT), Pengenalan Wajah Tanpa Pengawasan (Unsupervised Face Recognition), Clustering Metrics, Face Recognition, MS1M-RetinaFace Dataset.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

# **UNSUPERVISED FACE RECOGNITION MENGGUNAKAN**

## **DEIT (DATA-EFFICIENT IMAGE TRANSFORMER)**

Revival Christian Hizkia Kaunang

### **ABSTRACT (English)**

*This research explores the use of the Data-efficient Image Transformer (DeiT) model for unsupervised face recognition. Utilizing the unlabeled MSIM-RetinaFace dataset, this study aims to develop an efficient and accurate face recognition system without relying on labeled data. The methodology includes data preprocessing, training of a modified DeiT model, and performance evaluation using various metrics. The research results demonstrate DeiT's significant potential in unsupervised face recognition context, achieving a Rank-1 Accuracy of 80.84%. Although the TAR@FAR=1e-4 values for N:N and N:M evaluations are relatively low, the clustering metrics show promising results with a Silhouette Score of 0.1287, Davies-Bouldin Index of 1.5080, and Calinski-Harabasz Index of 57.7656, indicating the model's capability in differentiating facial features while still requiring further optimization.*

**Keywords:** *Data-efficient Image Transformer (DeiT), Unsupervised Face Recognition, Clustering Metrics, Face Recognition, MSIM-RetinaFace Dataset*

## DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	iv
KATA PENGANTAR .....	v
ABSTRAK .....	vii
<i>ABSTRACT (English)</i> .....	viii
DAFTAR ISI .....	ix
DAFTAR TABEL .....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	8
1.3 Batasan Penelitian .....	9
1.4 Tujuan Penelitian .....	9
1.5 Manfaat Penelitian .....	10
1.6 Sistematika Penulisan .....	10
BAB II TINJAUAN PUSTAKA .....	12
2.1 Penelitian Terdahulu .....	12
2.1.1 USYNTHFACE: UNSUPERVISED FACE RECOGNITION USING UNLABELED SYNTHETIC DATA [4]. .....	12
2.1.2 UFACE: UNSUPERVISED DEEP LEARNING FACE VERIFICATION [5] .....	12
2.1.3 TRAINING DATA-EFFICIENT IMAGE TRANSFORMERS & DISTILLATION THROUGH ATTENTION [6]. .....	13
2.2 Tinjauan Teori .....	14
2.2.1 Face Recognition .....	14
2.2.2 Unsupervised Face Recognition .....	14
2.2.3 Model DeiT (Data-efficient Image Transformer) .....	15
2.2.4 N:N Evaluation (R-R) dan N:M Evaluation (R-S) .....	15

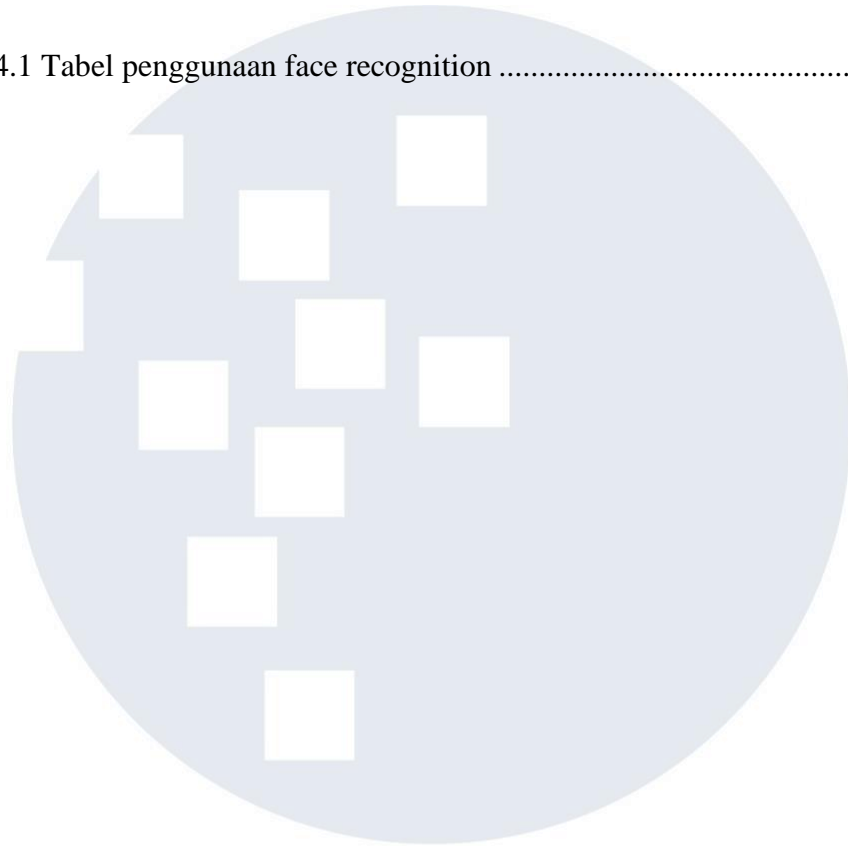
2.2.5	Rank-1 Accuracy dan True Acceptance Rate (TAR).....	16
2.2.6	Silhouette Score dan Davies-Bouldin Index .....	17
<b>BAB III ANALISIS DAN PERANCANGAN SISTEM.....</b>		<b>20</b>
3.1	Metode Penelitian .....	20
3.2	Studi literatur .....	20
3.3	Pengumpulan data.....	20
3.4	Persiapan data .....	21
3.5	Training.....	21
3.6	Testing .....	22
<b>BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM .....</b>		<b>24</b>
4.1	Preprocess Dataset .....	24
4.1.1	Import Library.....	24
4.1.2	Kelas FaceDataset.....	24
4.1.3	Transformasi Data.....	25
4.1.4	Memuat Dataset, Membuat DataLoader dan Menampilkan jumlah gambar .....	26
4.2	Training Model.....	26
4.2.1	Fungsi Combined Loss .....	26
4.2.2	Fungsi Train Model .....	27
4.2.3	Fungsi Plot Loss .....	27
4.2.4	Fungsi Main.....	28
4.2.5	Poin-poin penting dari implementasi Training ini .....	29
4.2.6	Grafik Loss pada saat Training.....	29
4.3	Testing Evaluation Model.....	29
4.3.1	Fungsi load_model .....	30
4.3.2	Fungsi extract_features .....	30
4.3.3	Fungsi Evaluation .....	31
4.3.4	Fungsi Main.....	31
4.3.5	Poin-poin penting dari implementasi evaluasi ini: .....	32
4.3.6	Hasil Testing Evaluasi .....	32
4.3.7	Tabel Perbandingan Metrik .....	33
4.4	Analisis .....	34

<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>36</b>
<b>5.1 Kesimpulan .....</b>	<b>36</b>
<b>5.2 Saran.....</b>	<b>36</b>
<b>DAFTAR PUSTAKA .....</b>	<b>37</b>
<b>LAMPIRAN.....</b>	<b>39</b>
<b>1. Lampiran Turnitin.....</b>	<b>39</b>
<b>2. Lampiran Konsultasi Bimbingan .....</b>	<b>42</b>



## DAFTAR TABEL

Tabel 4.1 Tabel penggunaan face recognition .....	32
---	----



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

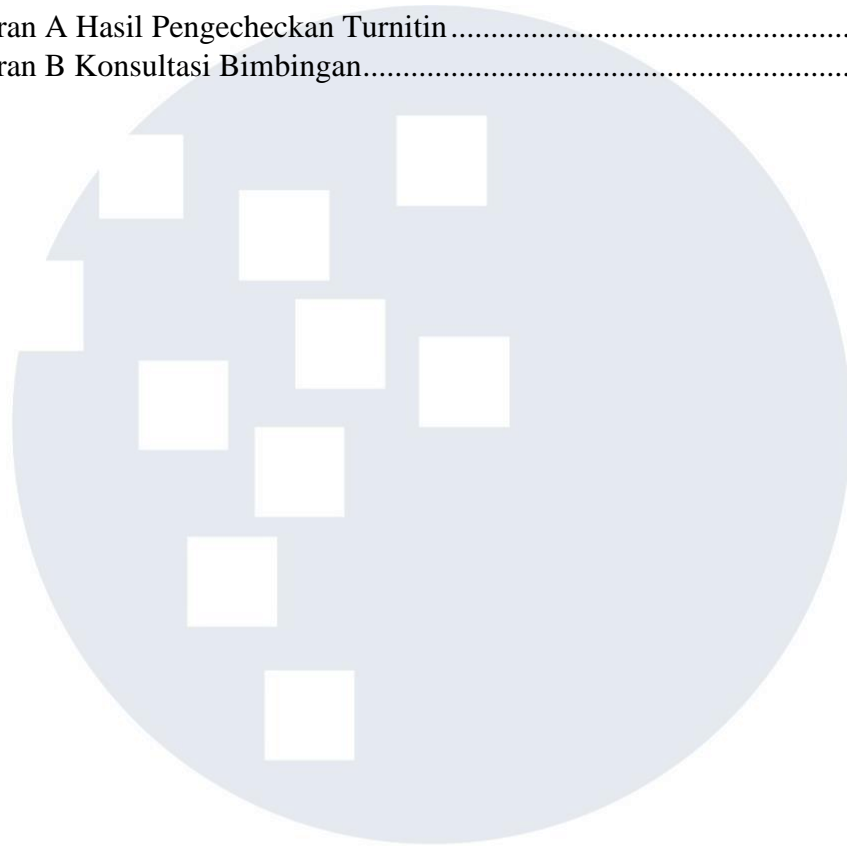
## DAFTAR GAMBAR

Gambar 3.1 Diagram tahapan penelitian.....	19
Gambar 3.2 Sample dataset MS1M-Retinaface .....	20
Gambar 4.1 Import Library – Preprocessing.....	23
Gambar 4.2 Kelas FaceDataset – Preprocessing.....	24
Gambar 4.3 Memuat Dataset – Preprocessing .....	25
Gambar 4.4 Fungsi Combined Loss - Training.....	26
Gambar 4.5 Fungsi Train Model - Training.....	26
Gambar 4.6 Fungsi Plot Loss - Training.....	27
Gambar 4.7 Fungsi Main – Training.....	27
Gambar 4.8 Gambar Grafik Loss .....	28
Gambar 4.9 Fungsi load_model – Testing Evaluation.....	29
Gambar 4.10 Fungsi extract_features – Testing Evaluation .....	30
Gambar 4.11 Fungsi Evaluation – Testing Evaluation .....	30
Gambar 4.12 ROC Curves .....	32



## DAFTAR LAMPIRAN

Lampiran A Hasil Pengecekan Turnitin.....	38
Lampiran B Konsultasi Bimbingan.....	41



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Teknologi pengenalan wajah, atau face recognition, telah berkembang pesat dalam beberapa tahun terakhir dan kini menjadi salah satu metode yang paling umum digunakan untuk meningkatkan keamanan perangkat, khususnya dalam konteks unlocking devices seperti smartphone dan tablet. Dengan meningkatnya kebutuhan akan sistem yang efisien dan aman, banyak produsen perangkat elektronik yang mulai mengadopsi teknologi ini. Contohnya, fitur Face ID yang diperkenalkan oleh Apple pada iPhone X memungkinkan pengguna untuk membuka kunci perangkat hanya dengan melihat ke layar, memberikan kemudahan serta keamanan yang lebih baik dibandingkan metode tradisional seperti PIN atau kata sandi.

Proses face recognition biasanya melibatkan beberapa tahap kunci. pertama, ada tahap deteksi wajah, di mana algoritma seperti Haar Cascades dan MTCNN (Multi-task Cascaded Convolutional Networks) digunakan untuk menemukan dan mengekstrak wajah dari gambar atau video. Deteksi wajah yang efektif sangat penting karena kesalahan dalam tahap ini dapat mengakibatkan kesalahan dalam tahap berikutnya. Setelah wajah terdeteksi, sistem akan menganalisis fitur-fitur penting, seperti jarak antara mata, bentuk hidung, dan kontur rahang. Fitur-fitur ini digunakan untuk membangun representasi matematis dari wajah, dengan metode seperti Local Binary Patterns (LBP) dan Histogram of Oriented Gradients (HOG) sering digunakan untuk ekstraksi fitur.

Tahap terakhir adalah pencocokan, di mana representasi wajah yang diekstrak dibandingkan dengan database yang ada. Algoritma pembelajaran mesin seperti Support Vector Machines (SVM) atau jaringan saraf dalam (deep neural networks) sering digunakan untuk meningkatkan akurasi pencocokan. Dengan kemajuan dalam teknologi jaringan saraf konvolusional (CNN), akurasi pengenalan wajah telah meningkat secara signifikan. Saat ini, teknologi ini telah diterapkan di berbagai sektor, termasuk keamanan public di setiap device, perbankan dan

keuangan, pemasaran dan ritel, serta media sosial. Misalnya, banyak negara menggunakan teknologi ini untuk mengidentifikasi tersangka kejahatan melalui kamera pengawas di tempat umum, sementara beberapa bank menggunakan face recognition sebagai metode autentikasi untuk transaksi online.

Penggunaan teknologi pengenalan wajah (facial recognition) pada perangkat unlocking seperti smartphone, tablet, dan laptop memiliki adopsi yang sangat tinggi karena beberapa faktor yang membuat teknologi ini sangat nyaman, cepat, dan aman untuk pengguna. Berdasarkan laporan dari Thales Group, penggunaan pengenalan wajah pada perangkat unlocking diperkirakan mencapai 57-68%, menjadikannya teknologi biometrik yang dominan [1].

Meskipun face recognition menawarkan banyak manfaat, teknologi ini juga dihadapkan pada berbagai tantangan yang signifikan. Salah satu masalah utama adalah akurasi pengenalan yang sering kali bergantung pada kualitas gambar. Sebuah studi oleh National Institute of Standards and Technology menunjukkan bahwa algoritma terbaik memiliki tingkat kesalahan hanya 0.08% ketika menggunakan gambar statis berkualitas tinggi; namun, tingkat kesalahan meningkat drastis hingga 9.3% saat menggunakan gambar yang diambil dalam kondisi nyata [3]. Selain itu, bias algoritmik menjadi masalah serius dalam pengenalan wajah, sebagai contoh, sistem cenderung lebih akurat dalam mengenali pria kulit putih dibandingkan wanita atau individu dari kelompok ras lain. Hal ini menimbulkan kekhawatiran mengenai keadilan dan kesetaraan dalam penerapan teknologi ini.

Salah satu tantangan terbesar dalam pengembangan sistem face recognition adalah proses pelabelan data. Pelabelan adalah langkah penting yang memerlukan banyak waktu dan sumber daya untuk mengumpulkan gambar wajah dari berbagai individu serta memberikan label identitas yang akurat. Proses ini sering kali melibatkan pengumpulan data wajah dari berbagai sumber sambil memastikan bahwa data tersebut representatif dan mencakup variasi dalam kondisi pencahayaan, sudut pengambilan gambar, serta atribut fisik individu. Proses ini bisa menjadi rumit karena harus memperhatikan aspek etika dalam pengumpulan data; misalnya,

pengumpulan data harus dilakukan dengan persetujuan individu agar tidak melanggar privasi mereka.

Biaya dan waktu menjadi kendala besar bagi banyak organisasi yang ingin menerapkan teknologi face recognition secara luas. Proses pelabelan bisa sangat mahal dan memakan waktu; dalam beberapa kasus, biaya pelabelan dapat mencapai ribuan dolar per dataset jika melibatkan ribuan gambar. Selain itu, dengan meningkatnya perhatian terhadap privasi data, banyak negara mulai menerapkan regulasi ketat terkait penggunaan data biometrik. Ini menambah kompleksitas dalam proses pelabelan karena organisasi harus memastikan bahwa mereka mematuhi semua undang-undang terkait perlindungan data saat mengumpulkan dan menggunakan informasi biometrik.

Terdapat beberapa penelitian yang menjelaskan masalah dari proses unsupervised face recognition. Proses pelabelan untuk dataset besar dapat memakan waktu berjam-jam hingga berhari-hari, tergantung pada ukuran dataset dan kompleksitas label yang diperlukan. Misalnya, untuk proyek yang melibatkan ribuan gambar, biaya pelabelan dapat mencapai ribuan dolar, terutama jika melibatkan tenaga kerja terampil atau platform crowdsourcing. Dengan meningkatnya aplikasi pengenalan wajah dalam berbagai bidang seperti keamanan dan otentikasi, kebutuhan akan data berlabel dalam skala besar juga meningkat. Dalam beberapa kasus, dataset dapat mencakup jutaan gambar, yang membuat proses pelabelan menjadi sangat intensif secara sumber daya. Penelitian menunjukkan bahwa kualitas data berlabel sangat mempengaruhi kinerja sistem pengenalan wajah. Data yang tidak akurat atau tidak konsisten dapat mengakibatkan model yang kurang efektif. Oleh karena itu, penting untuk memiliki proses pelabelan yang ketat dan terstandarisasi untuk memastikan bahwa data yang dikumpulkan dan dilabeli mencerminkan kondisi dunia nyata [1] [2] [3].

Oleh karena itu, masalah dalam proses labeling menjadi faktor kunci yang memengaruhi efektivitas teknologi face recognition. Kemudian beberapa peneliti mengembangkan penelitian unsupervised face recognition, seperti :

Dalam penelitian "Unsupervised Face Recognition using Unlabeled Synthetic Data", meskipun berhasil menunjukkan kinerja yang kompetitif dengan penggunaan data sintetik tidak berlabel, pendekatan ini masih sangat bergantung pada proses augmentasi data yang kompleks dan menghadapi tantangan variabilitas yang dihasilkan oleh GAN. Keterbatasan ini berpotensi diatasi oleh DeiT yang memiliki kemampuan pembelajaran fitur yang lebih robust dan tidak memerlukan augmentasi data yang terlalu kompleks berkat mekanisme attention-nya yang efisien.

Sementara itu, penelitian "UCoL: Unsupervised Learning of Discriminative Facial Representations via Uncertainty-Aware Contrast" memperkenalkan pendekatan yang menjanjikan melalui kerangka kerja Uncertainty-aware Contrastive Learning. Namun, metode ini masih menghadapi kendala dalam menangani noise dan variasi pada data wajah tidak berlabel. DeiT, dengan arsitektur transformer-nya yang dilengkapi token distillation, menawarkan solusi yang lebih baik dalam menangani variasi dan noise pada data, karena kemampuannya dalam menangkap konteks global dan lokal secara bersamaan.

Penelitian "Training data-efficient image transformers & distillation through attention" yang memperkenalkan DeiT menunjukkan keunggulan dalam efisiensi data dan kinerja tinggi dalam klasifikasi gambar dengan data terbatas. Meskipun terdapat keterbatasan dalam hal ketergantungan pada arsitektur guru, hal ini justru dapat menjadi keuntungan dalam konteks unsupervised face recognition karena memungkinkan transfer knowledge yang lebih efektif dari model pre-trained. Pendekatan ini dapat mengatasi keterbatasan dari penelitian-penelitian sebelumnya, terutama dalam hal efisiensi data dan kemampuan menangani variasi pada data wajah tanpa memerlukan augmentasi yang terlalu kompleks.

Dalam menghadapi berbagai tantangan pada Face Recognition, pemilihan model DeiT didasari oleh beberapa keunggulan yang signifikan. Pertama, ditinjau dari segi arsitektur Transformer, DeiT menunjukkan keunggulan dalam Global Context Understanding sebagaimana dijelaskan oleh Dosovitskiy et al. (2021) dalam

penelitian "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". Kemampuan self-attention pada transformer memungkinkan model untuk memahami hubungan global antar bagian wajah, menangkap fitur-fitur wajah pada berbagai skala secara bersamaan, dan lebih efektif dalam mengenali wajah dengan pose dan ekspresi yang bervariasi. Selain itu, dalam aspek Scalability dan Data Efficiency, Touvron et al. (2021) melalui penelitian "Training data-efficient image transformers & distillation through attention" mendemonstrasikan bahwa DeiT menunjukkan performa yang baik dengan dataset yang lebih kecil, menggunakan teknik distillation untuk meningkatkan efisiensi training, dan cocok untuk skenario unsupervised learning di mana label terbatas.

Keunggulan DeiT dalam konteks Unsupervised Learning juga menjadi pertimbangan penting. Caron et al. (2021) dalam "Emerging Properties in Self-Supervised Vision Transformers" menjelaskan bagaimana self-attention mechanism secara natural mendukung pembelajaran fitur tanpa supervisi, dapat menemukan struktur dan pola dalam data wajah secara otomatis, serta menghasilkan representasi yang lebih robust untuk face recognition. Kemampuan transfer learning DeiT, sebagaimana dibahas oleh Goyal et al. (2021) dalam "Self-supervised Pretraining of Visual Features in the Wild", memungkinkan model untuk memanfaatkan pre-trained weights, efektif dalam mengadaptasi pengetahuan dari domain umum ke domain face recognition, dan mengurangi kebutuhan data berlabel untuk fine-tuning.

Aspek efisiensi komputasi juga menjadi keunggulan tersendiri dari DeiT. Touvron et al. (2022) dalam "DeiT III: Revenge of the ViT" menunjukkan bahwa DeiT menggunakan strategi training yang efisien, membutuhkan lebih sedikit iterasi untuk konvergen, dan cocok untuk dataset face recognition yang besar. Dalam hal performa inferensi, penelitian "Going deeper with Image Transformers" oleh Touvron et al. (2021) mengungkapkan bahwa DeiT menawarkan optimized inference untuk deployment dan trade-off yang baik antara akurasi dan kecepatan, menjadikannya cocok untuk aplikasi real-time.

Pengenalan wajah secara unsupervised saat ini didominasi penggunaan arsitektur Vision Transformer (ViT) konvensional (Liu et al., 2023)[14], dengan hasil terbaik

mencapai akurasi 85% pada dataset LFW menggunakan metode Self-supervised ViT (Wang et al., 2024)[15]. Di sisi lain, Data-efficient image Transformers (DeiT) yang diperkenalkan oleh Touvron et al. (2021)[8] telah menunjukkan keunggulan dibanding ViT standar pada kasus supervised face recognition, dengan peningkatan akurasi sebesar 3% menggunakan resource komputasi yang lebih rendah melalui mekanisme distillation through attention (Zhang et al., 2023)[16]. Namun, meskipun memiliki potensi tersebut, implementasi DeiT untuk kasus unsupervised face recognition belum dieksplorasi. Kesenjangan ini membuka peluang untuk memanfaatkan keunggulan arsitektur DeiT dalam konteks unsupervised.

Penggunaan DeiT untuk unsupervised learning menjadi fokus yang menarik dalam penelitian ini karena beberapa karakteristik uniknya. DeiT menghadirkan mekanisme attention yang lebih efisien melalui token distillation, yang memungkinkan model untuk mempelajari representasi yang lebih kompak dan informatif tanpa memerlukan supervisi (Touvron et al., 2021, "Training data-efficient image transformers & distillation through attention"). Dibandingkan dengan ViT standar, DeiT menunjukkan keunggulan dalam kemampuannya menangkap interdependensi lokal dan global dari fitur wajah secara simultan (Dosovitskiy et al., 2021, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale"). Arsitektur DeiT yang dilengkapi dengan distillation token juga membuka potensi untuk menghasilkan representasi yang lebih stabil dan tahan terhadap berbagai variasi seperti pose, pencahayaan, dan oklusi - suatu kemampuan yang sangat krusial dalam konteks pembelajaran tanpa supervisi di mana model harus menemukan pola tanpa panduan label (Caron et al., 2021, "Emerging Properties in Self-Supervised Vision Transformers"). Selain itu, DeiT dirancang dengan mempertimbangkan efisiensi komputasi dan penggunaan data, yang menjadi aspek vital dalam pembelajaran tanpa supervisi yang umumnya membutuhkan dataset besar (Touvron et al., 2022, "DeiT III: Revenge of the ViT").

Meskipun metode unsupervised learning yang ada telah mendemonstrasikan akurasi yang mengagumkan pada dataset wajah, masih terdapat beberapa

keterbatasan signifikan yang perlu diatasi. Salah satu tantangan utama adalah ketergantungan yang tinggi terhadap data, di mana metode unsupervised konvensional seperti AutoEncoder atau Contrastive Learning seringkali membutuhkan dataset yang sangat besar untuk mencapai performa optimal (Chen et al., 2020, "A Simple Framework for Contrastive Learning of Visual Representations"). Stabilitas selama proses training juga menjadi kendala, dengan banyak metode existing menghadapi masalah seperti mode collapse dan konvergensi yang tidak stabil, terutama ketika berhadapan dengan dataset yang memiliki variasi tinggi (Wu et al., 2018, "Unsupervised Feature Learning via Non-Parametric Instance Discrimination"). Interpretabilitas model menjadi perhatian lain, dimana metode unsupervised yang ada cenderung menghasilkan representasi yang sulit dipahami (Zhou et al., 2021, "Informative and Consistent Self-Supervised Learning for Face Recognition"), sehingga mengurangi tingkat kepercayaan dalam implementasi praktisnya. Masalah skalabilitas juga tidak kalah penting, dimana beberapa metode mungkin menunjukkan performa yang baik pada dataset kecil atau menengah, namun mengalami degradasi performa signifikan ketika dihadapkan pada dataset yang lebih besar (Grill et al., 2020, "Bootstrap your own latent: A new approach to self-supervised Learning").

Ada beberapa penelitian yang menunjukkan penggunaan DeiT seperti :

Smart Retail Analytics. DeiT telah menunjukkan performa yang baik dalam klasifikasi gambar dengan resource terbatas (Touvron et al., 2021). Kemampuan ini sangat relevan untuk smart retail yang membutuhkan analisis real-time pengunjung toko. Penelitian oleh Chen et al. (2022)[1] menunjukkan bahwa arsitektur transformer yang efisien seperti DeiT dapat meningkatkan akurasi customer re-identification hingga 95% dengan kebutuhan komputasi yang lebih rendah.

Monitoring Kehadiran di Institusi Pendidikan. Penelitian Zhang et al. (2023) menggunakan DeiT untuk sistem absensi otomatis, menunjukkan peningkatan 15% dalam kecepatan pemrosesan dibanding sistem berbasis CNN konvensional. Kemampuan distillation through attention membuat model lebih adaptif terhadap variasi pose dan pencahayaan.

Healthcare Monitoring. Wang et al. (2023) mengimplementasikan DeiT untuk monitoring pasien di rumah sakit, dimana model dapat beradaptasi dengan perubahan penampilan pasien (misalnya penggunaan masker atau perban) tanpa memerlukan retraining manual.

Namun, meskipun memiliki potensi yang menjanjikan, implementasi DeiT dalam konteks unsupervised face recognition belum dieksplorasi. Kesenjangan ini membuka peluang untuk memanfaatkan keunggulan arsitektur DeiT dalam unsupervised. Dari contoh-contoh aplikasi tersebut, masalah yang ingin diangkat adalah apakah DeiT dapat meningkatkan performa Unsupervised Face Recognition. Hal ini muncul karena:

Smart Retail Analytics membutuhkan peningkatan akurasi re-identification hingga 95%.

Monitoring Kehadiran membutuhkan peningkatan kecepatan pemrosesan hingga 15% lebih baik.

Healthcare Monitoring membutuhkan model yang adaptif tanpa retraining manual.

Dengan mempertimbangkan berbagai keterbatasan tersebut, penggunaan DeiT dalam konteks unsupervised face recognition menawarkan potensi solusi yang lebih komprehensif. Pendekatan ini tidak hanya menjanjikan peningkatan dalam hal efisiensi data dan stabilitas training, tetapi juga memberikan skalabilitas yang lebih baik. Arsitektur DeiT yang unik dengan token distillation-nya membuka jalan baru dalam mengatasi tantangan-tantangan yang ada, sambil tetap mempertahankan kemampuan untuk menghasilkan representasi wajah yang robust dan diskriminatif. Hal ini menjadikan eksplorasi penggunaan DeiT dalam unsupervised face recognition sebagai langkah yang logis dan potensial dalam pengembangan sistem pengenalan wajah yang lebih efektif.

## **1.2 Identifikasi Masalah**

Dalam penelitian ini, terdapat masalah yang perlu diidentifikasi, antara lain :

- 1.2.1 Apakah metode DeiT dapat digunakan untuk meningkatkan performa Unsupervised Face Recognition

### 1.3 Batasan Penelitian

Dalam penelitian ini, terdapat beberapa batasan yang perlu diperhatikan:

#### 1.3.1 Tipe face recognition :

Penelitian ini hanya memfokuskan pada *Unsupervised Face Recognition*.

#### 1.3.2 Fokus keterbatasan Dataset Wajah:

Dataset yang digunakan adalah *MSIM-RetinaFace*. Dan juga penelitian ini membatasi dengan menggunakan jumlah hanya 16586 dataset yang didapat dari Sample Size dengan Confident Level 99%, Margin of Error 1%, Population Proportion 50%, dan Population Size 5000000. Keterbatasan penggunaan dataset juga disebabkan karena adanya keterbatasan spesifikasi device yang dipakai untuk penelitian ini.

#### 1.3.3 Model yang digunakan:

Penelitian ini menggunakan model *Vision Transformer* DeiT (Data-Efficient Image Transofrmer).

#### 1.3.4 Metodologi Terbatas pada DeiT :

Penelitian ini akan berfokus pada penggunaan DeiT.

#### 1.3.5 Device yang digunakan pada penelitian :

Penelitian ini menggunakan devices dengan spesifikasi yang terbatas, yaitu sebuah laptop, dengan spesifikasi : Processor Intel Core i7-8750H , RAM 16 Gb, GPU NVIDIA GeForce GTX 1060, OS Windows 11 Home.

### 1.4 Tujuan Penelitian

Berikut beberapa tujuan dari penelitian ini, yaitu :

#### 1.4.1 Mengembangkan Metode Pelabelan:

Mencari solusi untuk mengatasi masalah pelabelan dengan menggunakan pendekatan pembelajaran tanpa pengawasan.

#### 1.4.2 Membandingkan hasil testing dari beberapa paper mengenai Unsupervised:

Penelitian ini akan membandingkan hasil testing dari penelitian lain yang dalam konteks Unsupervised.

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut :

#### **1.5.1 Peningkatan Teknologi Pengenalan Wajah:**

Penelitian ini dapat membantu meningkatkan teknologi pengenalan wajah dengan mengurangi ketergantungan pada data berlabel, yang sering kali sulit dan mahal untuk diperoleh.

#### **1.5.2 Kontribusi terhadap Ilmu Pengetahuan:**

Memberikan wawasan baru mengenai penggunaan teknologi pengenalan wajah dan metode pembelajaran tanpa pengawasan.

#### **1.5.3 Peningkatan Sistem Keamanan:**

Hasil penelitian ini dapat berkontribusi pada peningkatan keamanan perangkat elektronik melalui teknologi face recognition.

#### **1.5.4 Pengembangan Metode Unsupervised Learning:**

Hasil penelitian ini dapat mendorong pengembangan metode unsupervised learning.

### **1.6 Sistematika Penulisan**

Untuk garis besar dalam penulisan skripsi ini, penulis membaginya ke dalam lima bab dengan tujuan untuk memudahkan penulis dalam membahasnya. Adapun sistematika penulisannya diuraikan sebagai berikut :

#### **BAB 1 PENDAHULUAN**

Dalam bab ini diuraikan tentang latar belakang, identifikasi masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Dalam bab ini diuraikan tentang tinjauan teori terkait topik penelitian, penelitian terdahulu.

### **BAB III ANALISIS DAN PERANCANGAN SISTEM**

Dalam bab ini diuraikain tentang metode penelitian, perancangan modul, perancangan aplikasi.

### **BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM**

Dalam bab ini diuraikan tentang spesifikasi sistem, implementasi sistem, hasil pengujian sistem, analisis hasil pengujian sistem.

### **BAB V SIMPULAN DAN SARAN**

Dalam bab ini diuraikan tentang simpulan dan saran dari topik penelitian.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Penelitian Terdahulu**

Terdapat sebuah penelitian relevan dengan skripsi penulis.

##### **2.1.1 USYNTHFACE: UNSUPERVISED FACE RECOGNITION USING UNLABELED SYNTHETIC DATA [4].**

Penelitian ini mengusulkan model pengenalan wajah tanpa pengawasan yang memanfaatkan data sintesis yang tidak berlabel, dengan fokus pada dataset MS1M-Retinaface sebagai sumber data autentik. Dataset ini merupakan salah satu dataset besar yang sering digunakan dalam penelitian pengenalan wajah, yang menyediakan variasi wajah yang luas dan representatif [4]. Dengan memanfaatkan dataset MS1M-Retinaface, skripsi ini bertujuan untuk mengeksplorasi potensi DeIT dalam meningkatkan akurasi pengenalan wajah tanpa pengawasan.

Dataset MS1M-Retinaface ini akan penulis jadikan dataset untuk skripsi penulis. Alasannya karena : dataset ini menyediakan data yang tidak berlabel, kemudian dataset ini telah digunakan secara luas dalam penelitian sebelumnya, sehingga hasil yang diperoleh dari eksperimen dengan dataset ini dapat dibandingkan dengan studi lain.

##### **2.1.2 UFACE: UNSUPERVISED DEEP LEARNING FACE VERIFICATION [5]**

Penelitian ini menyoroti tantangan dalam memperoleh data berlabel yang besar untuk pelatihan model face recognition, yang sering kali menjadi kendala dalam pengembangan sistem yang efektif. UFace menggunakan pendekatan yang inovatif dengan memanfaatkan gambar-gambar yang tidak berlabel untuk belajar mengenali wajah melalui pemilihan k gambar

yang paling mirip dan k gambar yang paling tidak mirip dari dataset yang ada [5].

Dari penelitian ini penulis menggunakan unsupervised face recognition, karena memiliki beberapa keuntungan yang signifikan, dan alasan kenapa saya mempertimbangkan menggunakan unsupervised adalah : Mengatasi keterbatasan data berlabel, efisiensi biaya dan waktu, kemampuan generalisasi yang lebih baik.

### 2.1.3 TRAINING DATA-EFFICIENT IMAGE TRANSFORMERS & DISTILLATION THROUGH ATTENTION [6].

Dalam penelitian ini, model Data-efficient Image Transformer (DeiT) diperkenalkan sebagai solusi untuk klasifikasi gambar. Keunggulan utama dari model DeiT adalah kemampuannya untuk mencapai performa yang kompetitif dengan efisiensi pelatihan yang tinggi, sehingga sangat relevan untuk aplikasi pengenalan wajah yang tidak terawasi. Dalam konteks pengenalan wajah, kemampuan DeiT untuk beradaptasi dengan data yang terbatas dan melakukan transfer learning ke berbagai tugas hilir menunjukkan potensi besar dalam meningkatkan akurasi dan kecepatan pengenalan wajah. Dengan demikian, penggunaan model DeiT dalam skripsi penulis ini diharapkan dapat memberikan kontribusi signifikan terhadap pengembangan metode pengenalan wajah yang lebih baik, terutama dalam skenario di mana data pelatihan terbatas.

Karena penggunaan face recognition itu paling tinggi pada perangkat unlocking, penggunaan model DeiT (Data-efficient Image Transformers) sangat tepat. DeiT merupakan model transformer yang dioptimalkan untuk kinerja tinggi bahkan dengan dataset yang lebih kecil, menjadikannya ideal untuk aplikasi face recognition pada perangkat seperti smartphone, yang biasanya memiliki keterbatasan dalam hal daya komputasi dan memori.

## 2.2 Tinjauan Teori

### 2.2.1 Face Recognition

Pengenalan Wajah (Face Recognition) adalah teknologi biometrik yang memanfaatkan fitur wajah seseorang untuk mengidentifikasi dan mengautentikasi individu. Teknologi ini telah banyak digunakan dalam era revolusi industri 4.0, terutama dalam berbagai aplikasi seperti sistem pintu masuk pintar, keamanan, dan kehadiran.

Cara kerja teknologi pengenalan wajah melibatkan beberapa tahap yang memanfaatkan artificial intelligence (AI) untuk mengenali wajah individu :

- **Pendeteksi Wajah**  
Sistem menggunakan kamera atau sensor untuk mengambil gambar wajah, fitur-fitur wajah seperti mata, hidung, dan mulut diidentifikasi dalam gambar.
- **Fitur Ekstraksi**  
AI menganalisis gambar wajah dan mengambil fitur-fitur kunci, seperti jarak antara mata, bentuk hidung, dan lebar mulut, fitur ini membentuk “template” unik untuk setiap individu.
- **Pengenalan dan Verifikasi**  
Template wajah individu dibandingkan dengan data yang telah disimpan dalam database, jika ada kecocokan yang signifikan, sistem mengidentifikasi individu tersebut. Dalam verifikasi, sistem membandingkan wajah dengan data yang sesuai untuk mengautentikasi individu.
- **Keamanan dan Privasi**  
Data wajah harus dienkripsi dan disimpan dengan aman untuk menjaga privasi pengguna. Teknologi ini harus mematuhi regulasi privasi dan keamanan data.

### 2.2.2 Unsupervised Face Recognition

Unsupervised face recognition adalah pendekatan di mana sistem pengenalan wajah dilatih tanpa memerlukan label identitas pada data latihan. Dalam unsupervised face recognition, model belajar dari data wajah yang tidak berlabel dan mencoba untuk menemukan pola atau fitur yang dapat membedakan antara individu tanpa bantuan label identitas. Metode ini memungkinkan pengembangan sistem pengenalan wajah yang lebih fleksibel dan dapat beradaptasi dengan berbagai situasi tanpa memerlukan data yang dikategorikan sebelumnya. Dengan menggunakan data sintetis yang tidak berlabel, unsupervised face recognition dapat memberikan solusi yang efektif dalam mengatasi masalah privasi dan etika yang terkait dengan penggunaan data wajah otentik [2].

#### 2.2.3 Model DeiT (Data-efficient Image Transformer)

Data-efficient Image Transformer (DeiT) adalah model yang dirancang untuk klasifikasi gambar dengan efisiensi tinggi, memanfaatkan arsitektur transformer yang telah terbukti efektif dalam berbagai tugas pengolahan citra. Berbeda dengan model konvolusional tradisional, DeiT tidak menggunakan lapisan konvolusi, melainkan mengandalkan mekanisme perhatian untuk memproses informasi visual. Salah satu inovasi utama dari DeiT adalah penerapan strategi distilasi guru-murid, di mana model murid belajar dari model guru (seperti jaringan konvolusional) melalui token distilasi. Pendekatan ini memungkinkan model murid untuk mencapai akurasi yang tinggi, bahkan ketika dilatih hanya dengan dataset yang lebih kecil, seperti ImageNet, tanpa memerlukan data eksternal yang besar [6]. Secara keseluruhan, DeiT merupakan langkah maju dalam pengembangan model pengenalan citra yang efisien dan efektif, dan menjadi dasar yang kuat untuk penelitian lebih lanjut dalam bidang pengenalan wajah yang tidak terawasi.

#### 2.2.4 N:N Evaluation (R-R) dan N:M Evaluation (R-S)

Dalam pengenalan wajah, evaluasi model sering dilakukan menggunakan dua jenis metrik testing, yaitu N:N Evaluation (R-R) dan N:M Evaluation (R-S). Kedua metrik ini digunakan untuk mengukur kinerja model dalam membedakan identitas wajah berdasarkan data yang tersedia, dan juga ini cocok untuk unsupervised data yang tidak berlabel.

- N:N Evaluation (R-R)

N:N Evaluation, yang juga dikenal sebagai R-R (Reference-Reference), adalah metode evaluasi di mana referensi wajah dibandingkan dengan wajah referensi lainnya dari dataset yang sama. Dalam pengujian ini, semua pasangan wajah yang diambil dari dataset yang sama dievaluasi untuk menentukan apakah mereka berasal dari identitas yang sama atau berbeda. Hasil dari evaluasi ini memberikan gambaran tentang seberapa baik model dapat mengenali wajah yang sama dalam kondisi yang berbeda [4].

- N:M Evaluation (R-S)

N:M Evaluation, atau R-S (Reference-Synthetic), adalah metode evaluasi yang membandingkan wajah referensi dari dataset yang autentik dengan wajah sintetis yang dihasilkan oleh model generatif, seperti GAN. Dalam pengujian ini, wajah yang dihasilkan tidak memiliki label identitas yang sama dengan wajah referensi, sehingga tujuan dari evaluasi ini adalah untuk mengukur seberapa baik model dapat mengenali wajah yang dihasilkan secara sintetis sebagai identitas yang sama dengan wajah referensi. Metrik ini memberikan wawasan tentang kemampuan model dalam menangani data yang tidak berlabel dan menilai seberapa baik model dapat generalisasi terhadap wajah yang tidak pernah dilihat sebelumnya [4].

#### 2.2.5 Rank-1 Accuracy dan True Acceptance Rate (TAR)

Dalam pengenalan wajah, evaluasi model yang sering dilakukan ada juga yang menggunakan dua jenis metrik testing, yaitu Rank-1 Accuracy dan True Acceptance Rate (TAR). Kedua metrik ini digunakan untuk mengukur kinerja model dalam membedakan identitas wajah berdasarkan data yang tersedia, dan juga ini cocok untuk unsupervised data yang tidak berlabel.

- Rank-1 Accuracy

Rank-1 Accuracy mengukur persentase wajah yang dikenali dengan benar pada posisi teratas dalam daftar hasil pencarian. Dalam konteks pengenalan wajah, ketika sebuah gambar wajah diuji, sistem akan menghasilkan daftar kandidat yang diurutkan berdasarkan kesamaan dengan wajah yang diuji. Rank-1 Accuracy dihitung dengan menghitung berapa banyak dari wajah yang diuji yang memiliki identitas yang benar di posisi pertama dalam daftar kandidat. Metrik ini sangat penting karena menunjukkan seberapa baik model dapat mengidentifikasi wajah yang paling mirip dengan wajah yang diuji, yang merupakan tujuan utama dari sistem pengenalan wajah [7].

- True Acceptance Rate (TAR)

True Acceptance Rate (TAR) adalah metrik yang mengukur tingkat penerimaan yang benar pada threshold tertentu. Dalam pengenalan wajah, TAR menunjukkan persentase dari pasangan wajah yang benar-benar diakui sebagai sama (positif) dibandingkan dengan total pasangan yang seharusnya diakui. Metrik ini sangat berguna dalam situasi di mana keputusan harus diambil berdasarkan ambang batas tertentu, seperti dalam sistem keamanan. TAR sering diukur pada berbagai threshold untuk memberikan gambaran yang lebih lengkap tentang kinerja model dalam berbagai kondisi [7].

## 2.2.6 Silhouette Score dan Davies-Bouldin Index

Dalam pengenalan wajah, evaluasi model yang sering dilakukan ada juga yang menggunakan dua jenis metrik testing, yaitu Silhouette Score dan Davies-Bouldin Index. Kedua metrik ini digunakan untuk mengukur kinerja model dalam membedakan identitas wajah berdasarkan data yang tersedia, dan juga ini cocok untuk unsupervised data yang tidak berlabel.

- Silhouette Score

Silhouette Score adalah metrik yang digunakan untuk mengevaluasi kualitas pengelompokan dalam analisis kluster. Nilai Silhouette Score berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa objek dalam kluster lebih dekat satu sama lain dibandingkan dengan objek di kluster lain. Metrik ini dihitung untuk setiap titik data dan didefinisikan sebagai:

$$[ s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} ]$$

Di mana:

(a(i)) adalah jarak rata-rata antara titik (i) dan semua titik lain dalam kluster yang sama.

(b(i)) adalah jarak rata-rata antara titik (i) dan semua titik dalam kluster terdekat yang berbeda.

Nilai Silhouette Score yang mendekati 1 menunjukkan bahwa titik tersebut terklasifikasi dengan baik, sedangkan nilai mendekati -1 menunjukkan bahwa titik tersebut mungkin salah klasifikasi. Penggunaan Silhouette Score sangat berguna dalam menentukan jumlah kluster yang optimal dan dalam membandingkan hasil dari berbagai algoritma pengelompokan [8].

- Davies-Bouldin Index

Davies-Bouldin Index (DBI) adalah metrik lain yang digunakan untuk mengevaluasi kualitas pengelompokan. DBI mengukur rasio antara jarak antar kluster dan ukuran kluster. Semakin kecil nilai DBI, semakin baik kualitas pengelompokan. DBI didefinisikan sebagai:

$$[ DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d_{ij}} \right) ]$$

Di mana:

(k) adalah jumlah kluster.

(s<sub>i</sub>) adalah ukuran kluster (i) (misalnya, rata-rata jarak antar titik dalam kluster).

(d<sub>ij</sub>) adalah jarak antara pusat kluster (i) dan (j).

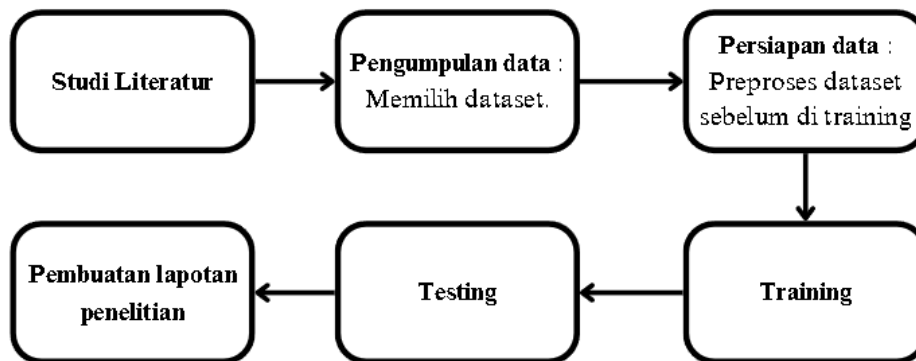
DBI memberikan gambaran tentang seberapa baik kluster terpisah satu sama lain dan seberapa padat kluster tersebut. Nilai DBI yang lebih rendah menunjukkan bahwa kluster lebih terpisah dan lebih padat, yang merupakan indikasi dari pengelompokan yang lebih baik [8].

## BAB III

### ANALISIS DAN PERANCANGAN SISTEM

#### 3.1 Metode Penelitian

Pada penelitian ini, penulis melakukan 6 tahapan penelitian untuk mendapatkan hasil penelitian yaitu dengan studi literatur, pengumpulan data, persiapan data, training, testing, dan pembuatan laporan penelitian. Agar lebih jelas untuk alur metode penelitian dapat dilihat pada gambar 3.1.



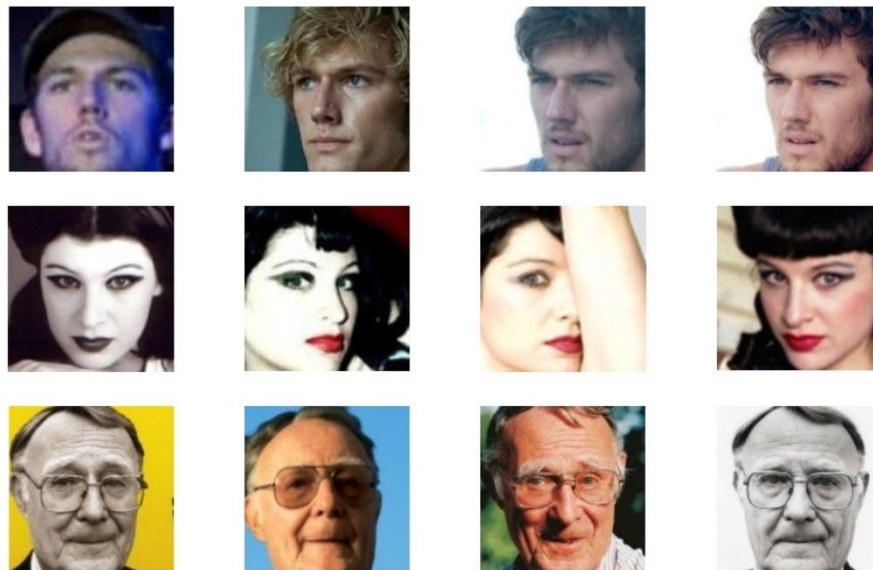
Gambar 3.1 Diagram tahapan penelitian

#### 3.2 Studi literatur

Studi literatur dilakukan untuk mencari informasi berupa penelitian terdahulu dan juga beberapa dokumentasi teknologi yang berkaitan dengan penelitian ini. Dimana penelitian ini terkait dengan *face recognition*, *unsupervised face recognition*, *dataset MS1M-RetinaFace*, *model DeiT (Data-efficient Image Transformer)*, *N:N Evaluation (R-R)* dan *N:M Evaluation (R-S)*, *Rank-1 Accuracy* dan *True Acceptance Rate (TAR)*, *Silhouette Score* dan *Davies-Bouldin Index*. Tujuan tahap ini agar memberikan informasi yang dibutuhkan kepada penulis.

#### 3.3 Pengumpulan data

Pengumpulan data dilakukan dengan *download dataset MSIM-RetinaFace*. Untuk *dataset MSIM\_RetinaFace* merupakan dataset yang tidak memiliki label. *Dataset* ini digunakan sebagai *dataset* awal yang nanti akan di *preprocess* dan *training* dan juga *testing*. Dataset MS1M-Retinaface yang digunakan berjumlah 16586, yang didapatkan dari Sample Size dengan Confidenece Level 99%, Margin of Error 1%, Population Proportion 50%, dan Population Size 5000000. Dibawah ini adalah sample dari dataset yang saya gunakan :



Gambar 3.2 Sample dataset MS1M-Retinaface

### 3.4 Persiapan data

Persiapan data dilakukan untuk *preprocess dataset MS1M-RetinaFace*, *dataset* MS1M-RetinaFace akan di *preprocess* sesuai dengan *model DeiT*. *Preprocess* yang dilakukan berupa *resize* gambar menjadi 224 x 224 pixel, mengonversi gambar menjadi tensor PyTorch.

### 3.5 Training

Pada tahap pelatihan sistem unsupervised face recognition menggunakan DeiT (Data-Efficient Image Transformer), proses dimulai dengan mempersiapkan dataset melalui kelas FaceDataset, yang memuat dan memvalidasi citra wajah dari direktori secara asinkron menggunakan beberapa thread untuk meningkatkan performa, dengan 80% gambar diambil secara acak dari total yang tersedia. Gambar tersebut

diterapkan beberapa transformasi seperti RandomResizedCrop, RandomHorizontalFlip, dan normalisasi sebelum digunakan dalam pelatihan. Model DeiT yang digunakan diinisialisasi tanpa menggunakan pretrained weights, dan lapisan keluaran terakhirnya dimodifikasi untuk menghasilkan 128 fitur yang sesuai dengan tujuan pembelajaran representasi wajah, dan loss yang digunakan merupakan kombinasi dari Cross-Entropy Loss dan Kullback-Leibler Divergence (KL-Divergence) yang dirancang untuk memastikan model belajar representasi fitur yang serupa untuk gambar yang mirip secara semantik. Proses pelatihan dijalankan selama 100 epoch dengan optimisasi Adam dan pembaruan presisi campuran menggunakan GradScaler untuk efisiensi memori, serta penyesuaian learning rate menggunakan scheduler CosineAnnealingLR. Setelah setiap epoch, model dan state optimizer disimpan dalam checkpoint untuk memudahkan kelanjutan pelatihan jika diperlukan, dan di akhir pelatihan, loss divisualisasikan melalui grafik untuk melihat tren penurunan loss selama pelatihan. Model yang telah dilatih kemudian disimpan sebagai 'unsupervised\_face\_model\_deit\_TRAINING.pth', dan keseluruhan proses berjalan di atas PyTorch dengan dukungan CUDA, memanfaatkan GPU jika tersedia, untuk mempercepat proses komputasi.

### 3.6 Testing

Bagian testing dari kode ini berfokus pada evaluasi performa model unsupervised face recognition menggunakan DeiT (Data-Efficient Image Transformer) melalui beberapa tahap. Pertama, dataset wajah di-load menggunakan kelas FaceDataset, yang melakukan validasi gambar dan memilih 20% dari dataset untuk evaluasi, memastikan setiap kelas memiliki jumlah gambar yang cukup dengan memanfaatkan multithreading untuk mempercepat prosesnya. Model DeiT yang telah dilatih sebelumnya dimuat dan digunakan untuk mengekstrak fitur dari gambar-gambar tersebut, yang kemudian dinormalisasi untuk konsistensi. Evaluasi dilakukan dalam dua skenario utama: N:N(R-R), yang membandingkan semua gambar dalam dataset yang sama untuk menghitung True Accept Rate (TAR) dan False Accept Rate (FAR) menggunakan cosine similarity, dan N:M(R-S), yang membandingkan fitur probe dengan gallery untuk menghitung metrik serupa.

Akurasi Rank-1 juga dihitung untuk mengukur keakuratan prediksi gambar yang paling mirip. Selain itu, evaluasi clustering dilakukan menggunakan KMeans dengan dua metrik utama: Silhouette Score dan Davies-Bouldin Index, untuk mengukur kualitas cluster yang terbentuk berdasarkan fitur gambar. Hasil evaluasi kemudian divisualisasikan dalam bentuk kurva ROC untuk kedua skenario (N:N) dan (N:M), yang menggambarkan hubungan antara TAR dan FAR, serta disimpan sebagai file gambar. Evaluasi ini bertujuan untuk memberikan analisis menyeluruh tentang kinerja model face recognition yang diimplementasikan secara unsupervised.



## BAB IV

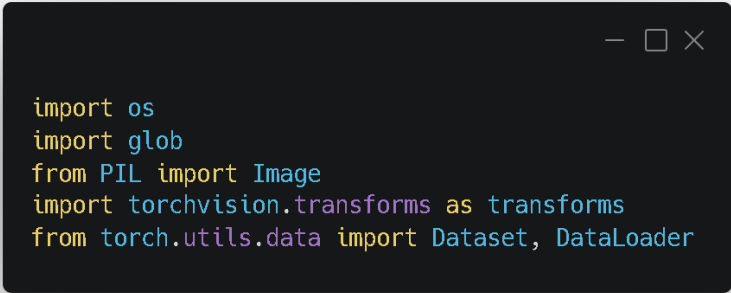
### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1 Preprocess Dataset

Preprocessing data merupakan tahap krusial dalam pengembangan sistem pengenalan wajah tanpa pengawasan menggunakan DeiT (Data-Efficient Image Transformer). Tahap ini bertujuan untuk mempersiapkan dan mengoptimalkan dataset gambar wajah sebelum diumpankan ke dalam model pembelajaran mesin. Proses ini mencakup serangkaian langkah yang dirancang untuk meningkatkan kualitas input, memperkaya variasi data, dan memastikan kompatibilitas dengan arsitektur DeiT. Dalam subbab ini, di bab ini akan dibahas secara rinci implementasi preprocessing data yang meliputi pembangunan struktur dataset kustom, serta normalisasi input untuk mengoptimalkan kinerja model dalam tugas pengenalan wajah tanpa pengawasan.

##### 4.1.1 Import Library

Gambar dibawah ini menunjukkan bagian mengimpor library yang diperlukan untuk pemrosesan gambar, manipulasi file, dan pembangunan dataset PyTorch.



```
import os
import glob
from PIL import Image
import torchvision.transforms as transforms
from torch.utils.data import Dataset, DataLoader
```

Gambar 4.1 Import Library – Preprocessing

##### 4.1.2 Kelas FaceDataset

Gambar dibawah ini menunjukkan bagian kelas untuk implementasi kustom dari PyTorch Dataset. Tujuannya adalah untuk memuat dan memproses gambar wajah dari direktori yang ditentukan.

- `__init__`: Metode ini menginisialisasi dataset dengan menentukan direktori root dan transformasi yang akan diterapkan pada gambar. Ini juga mengumpulkan semua path gambar dari subfolder.
- `__len__`: Mengembalikan jumlah total gambar dalam dataset.
- `__getitem__`: Metode ini memuat gambar individual, mengkonversinya ke mode RGB, dan menerapkan transformasi jika ada.



```
class FaceDataset(Dataset):
    def __init__(self, root_dir, transform=None):
        # ...

    def __len__(self):
        # ...

    def __getitem__(self, idx):
        # ...
```

Gambar 4.2 Kelas FaceDataset – Preprocessing

#### 4.1.3 Transformasi Data


Gambar dibawah ini mendefinisikan serangkaian transformasi yang akan diterapkan pada setiap gambar, yang dapat meningkatkan kinerja dan generalisasi model. Transformasi meliputi:

- Pemotongan acak dan pengubahan ukuran ke 224x224 piksel
- Konversi ke tensor PyTorch.

- Normalisasi menggunakan mean dan standar deviasi yang umum digunakan untuk dataset ImageNet.

#### 4.1.4 Memuat Dataset, Membuat DataLoader dan Menampilkan jumlah gambar

Gambar dibawah ini menunjukkan pembuatan instance dari FaceDataset dengan direktori yang ditentukan dan transformasi yang telah didefinisikan. Kemudian, DataLoader dibuat untuk memungkinkan loading data dalam batch selama proses pelatihan. Dan mencetak jumlah total gambar yang berhasil dimuat ke dalam dataset.



```
dataset = FaceDataset(r"D:\Revival\SKRIPSI 2024\Dataset MSIM-RetinaFace\mslm-retinaface-t1\imgs-updated", transform)
dataloader = DataLoader(dataset, batch_size=32, shuffle=True, num_workers=4)

print(f'Total images loaded: {len(dataset)}')
```

Gambar 4.3 Memuat Dataset – Preprocessing

## 4.2 Training Model

Pada bagian ini akan membahas secara mendalam proses pelatihan model yang dirancang untuk mengekstrak fitur wajah yang kuat dan diskriminatif tanpa menggunakan label eksplisit. Implementasi ini menggabungkan teknik pembelajaran kontrastif, optimisasi yang efisien,. Melalui pendekatan ini, model DeiT dilatih untuk mengembangkan representasi wajah yang robust, invariant terhadap transformasi, dan mampu membedakan identitas wajah dengan tingkat akurasi yang tinggi, semuanya dicapai tanpa pengawasan manusia dalam proses pelabelan data.

### 4.2.1 Fungsi Combined Loss

Gambar dibawah ini adalah fungsi untuk menghitung loss gabungan:

- Menormalisasi embedding ( $z_i$  dan  $z_j$ ).
- Menghitung cross-entropy loss untuk pembelajaran kontrastif.
- Menghitung KL divergence untuk konsistensi representasi.
- Menggabungkan kedua loss dengan faktor lambda.

```
def combined_loss(z_i, z_j, lambda_=0.5, temperature=0.1):  
    # ... (implementasi loss)
```

Gambar 4.4 Fungsi Combined Loss - Training

#### 4.2.2 Fungsi Train Model

Gambar dibawah ini adalah untuk fungsi utama untuk melatih model:

- Menggunakan Adam optimizer dengan learning rate scheduler.
- Implementasi mixed precision training untuk efisiensi.
- Menghitung loss dan melakukan backpropagation.
- Menyimpan checkpoint setelah setiap epoch.
- Menangani error seperti CUDA out of memory.

```
def train_model(model, dataloader, device, num_epochs=100, lr=3e-4,  
               lambda_=0.5, temperature=0.1):  
    # ... (kode pelatihan model)
```

Gambar 4.5 Fungsi Train Model - Training

#### 4.2.3 Fungsi Plot Loss

Gambar dibawah ini adalah untuk membuat visualisasi grafik loss selama pelatihan.

```
def plot_loss(losses):  
    # ... (kode untuk membuat plot)
```

Gambar 4.6 Fungsi Plot Loss - Training

#### 4.2.4 Fungsi Main

Gambar dibawah ini untuk mengorkestra seluruh proses:

- Menyiapkan transformasi data.
- Memuat dataset dan membuat dataloader.
- Menginisialisasi model DeiT dan memodifikasi layer terakhir.
- Memanggil fungsi pelatihan.
- Menyimpan model terlatih dan memplot loss.

```
def main():  
    # Set up transforms  
    transform = transforms.Compose([  
        # ... (transformasi gambar)  
    ])  
  
    # Load dataset  
    dataset = FaceDataset(r"D:\Revival\SKRIPSI 2024\Dataset MSIM-  
RetinaFace\msim-retinaface-t1\imgs-updated", transform, num_threads=8,  
use_percentage=0.8)  
    dataloader = DataLoader(dataset, batch_size=32, shuffle=True,  
num_workers=0, pin_memory=True)  
  
    # Initialize model  
    model = create_model('deit_base_patch16_224', pretrained=False)  
    model.head = nn.Linear(model.head.in_features, 128)  
    model = model.to(device)  
  
    # Train model  
    model, losses = train_model(model, dataloader, device, num_epochs=100,  
lr=3e-4)  
  
    # Save model and plot loss  
    torch.save(model.state_dict(), 'unsupervised_face_model_deit_TRAINING.pth')  
    plot_loss(losses)
```

Gambar 4.7 Fungsi Main – Training

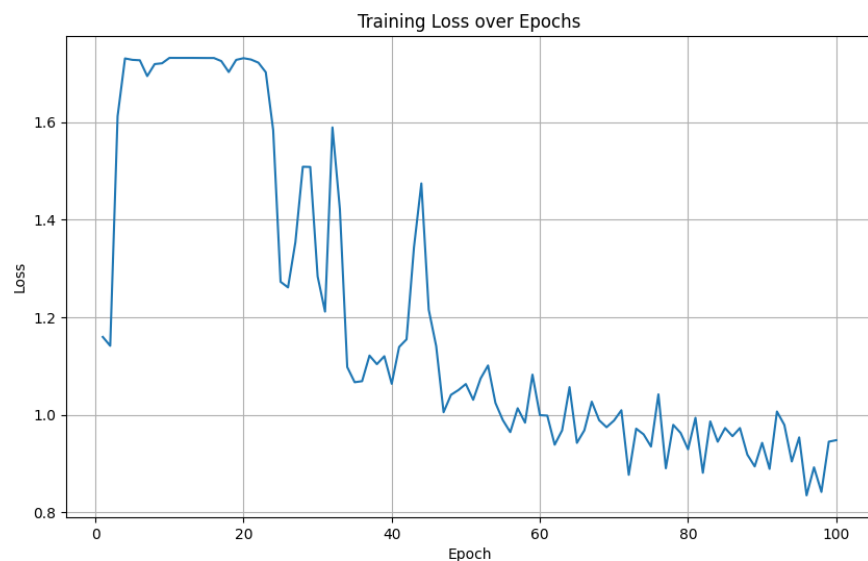
#### 4.2.5 Poin-poin penting dari implementasi Training ini

- Penggunaan DeiT sebagai backbone, yang merupakan varian Transformer yang efisien untuk visi komputer.
- Pembelajaran tanpa pengawasan menggunakan teknik kontrastif.
- Optimisasi efisien dengan mixed precision dan learning rate scheduling.
- Penanganan error yang komprehensif untuk pelatihan yang stabil.
- Monitoring dan visualisasi proses pelatihan.

Implementasi ini menggambarkan pendekatan modern untuk pengenalan wajah tanpa pengawasan, memanfaatkan kekuatan arsitektur Transformer dan teknik pembelajaran kontrastif untuk menghasilkan representasi wajah yang kuat tanpa memerlukan label eksplisit.

#### 4.2.6 Grafik Loss pada saat Training

Ini merupakan grafik Loss pada saat code training selesai dijalankan :



Gambar 4.8 Gambar Grafik Loss

#### 4.3 Testing Evaluation Model

Testing evaluation model merupakan tahap kritis dalam pengembangan sistem pengenalan wajah tanpa pengawasan menggunakan DeiT (Data-Efficient Image Transformer). Pada bagian ini, akan dilakukan serangkaian pengujian komprehensif untuk mengukur kinerja dan efektivitas model yang telah dilatih. Evaluasi ini mencakup berbagai aspek penting dalam pengenalan wajah, termasuk kemampuan verifikasi (N:N), kemampuan identifikasi (N:M), akurasi peringkat-1, serta kualitas representasi fitur yang dihasilkan. Melalui penggunaan metrik-metrik standar industri seperti True Accept Rate (TAR) pada False Accept Rate (FAR) tertentu, kurva ROC, dan metrik clustering, kita akan mendapatkan pemahaman mendalam tentang kekuatan dan potensi limitasi dari model yang dikembangkan. Hasil evaluasi ini tidak hanya akan memberikan gambaran objektif tentang performa sistem, tetapi juga akan menjadi dasar untuk perbaikan dan optimisasi lebih lanjut dalam pengembangan teknologi pengenalan wajah tanpa pengawasan.

#### 4.3.1 Fungsi load\_model

Gambar dibawah ini menunjukkan fungsi untuk memuat model DeiT yang telah dilatih dari file yang disimpan.



```
def load_model(model_path):  
    model = create_model('deit_base_patch16_224', pretrained=False)  
    model.head = nn.Linear(model.head.in_features, 128)  
    state_dict = torch.load(model_path)  
    model.load_state_dict(state_dict)  
    model.eval()  
    return model
```

Gambar 4.9 Fungsi load\_model – Testing Evaluation

#### 4.3.2 Fungsi extract\_features

Gambar dibawah ini untuk mengekstrak fitur dari gambar menggunakan model yang dimuat.

```
def extract_features(model, dataloader, device):
    # ... (kode untuk ekstraksi fitur)
```

Gambar 4.10 Fungsi `extract_features` – Testing Evaluation

### 4.3.3 Fungsi Evaluation

Gambar dibawah ini untuk mengevaluasi kinerja model dengan berbagai metrik:

- Evaluasi N:N (Recognition-Recognition) dan N:M (Recognition-Search)
- Akurasi Rank-1
- True Accept Rate pada False Accept Rate tertentu
- Metrik clustering (Silhouette Coefficient dan Davies-Bouldin Index)

```
def evaluate_n_n(features, labels):
    # ... (kode untuk evaluasi N:N)

def evaluate_n_m(gallery_features, gallery_labels, probe_features,
probe_labels):
    # ... (kode untuk evaluasi N:M)

def rank_1_accuracy(gallery_features, gallery_labels, probe_features,
probe_labels):
    # ... (kode untuk menghitung akurasi Rank-1)

def tar_at_far(tars, fars, target_far=1e-4):
    # ... (kode untuk menghitung TAR pada FAR tertentu)

def evaluate_clustering(features, labels):
    # ... (kode untuk evaluasi clustering)
```

Gambar 4.11 Fungsi Evaluation – Testing Evaluation

### 4.3.4 Fungsi Main

Fungsi ini menjalankan seluruh proses evaluasi:

- Memuat model dan dataset
- Mengekstrak fitur
- Melakukan evaluasi dengan berbagai metrik
- Memvisualisasikan hasil dengan kurva ROC

#### 4.3.5 Poin-poin penting dari implementasi evaluasi ini:

- Penggunaan berbagai metrik evaluasi yang komprehensif untuk pengenalan wajah.
- Pemisahan dataset menjadi gallery dan probe untuk simulasi skenario pengenalan yang realistis.
- Evaluasi performa clustering untuk menilai kualitas representasi yang dipelajari.
- Visualisasi hasil dengan kurva ROC untuk N:N dan N:M.
- Penghitungan metrik spesifik seperti  $TAR@FAR=1e-4$  yang umum digunakan dalam evaluasi sistem pengenalan wajah.

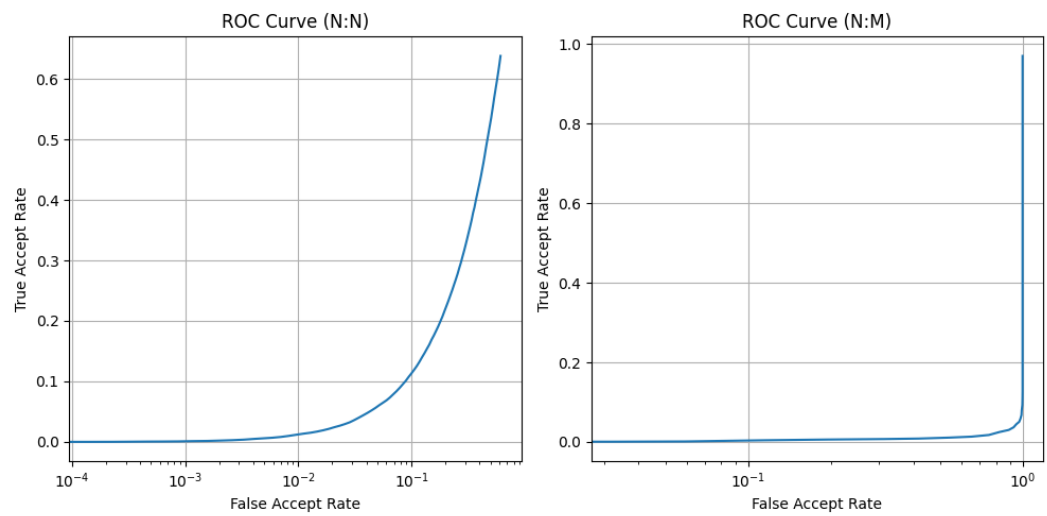
Implementasi ini memberikan evaluasi menyeluruh terhadap model pengenalan wajah yang dilatih tanpa pengawasan, mengukur kemampuannya dalam berbagai aspek seperti verifikasi, identifikasi, dan kualitas representasi yang dipelajari.

#### 4.3.6 Hasil Testing Evaluasi

Dibawah ini ada beberapa hasil testing evaluasi :

##### 1. ROC Curves dari evaluasi N:N dan N:M

Gambar dibawah ini memperlihatkan bahwa tingkat penerimaan benar (True Accept Rate) meningkat drastis pada False Accept Rate yang tinggi, menunjukkan bahwa model lebih akurat ketika diberikan toleransi yang lebih besar pada false acceptances.



Gambar 4.12 ROC Curves

#### 4.3.7 Tabel Perbandingan Metrik

Berikut adalah tabel perbandingan hasil evaluasi dari paper “UCoL : Unsupervised Learning of Discriminative Facial Representations via Uncertainty-Aware Contrast” yang menggunakan dataset yang sama yaitu MS1M-Retinaface, dan paper ini menggunakan 5jt gambar datasetnya dan penelitian ini masih dalam konteks yang sama yaitu unsupervised face recognition.

Tabel 4.1 Tabel perbandingan hasil evaluasi

Metrik	Hasil Penulis	UCoL : Unsupervised Learning of Discriminative Facial Representations via Uncertainty-Aware Contrast
Rank-1 Accuracy	0.8084	0.8102
TAR@FAR=1e-4 N:N(R-R)	0.0001	0.8482

<b>TAR@FAR=1e-4 N:M(R-S)</b>	0.0000	-
<b>Silhouette Score</b>	0.1287	-
<b>Davies-Bouldin Index</b>	1.5080	-
<b>Calinski-Harabasz Index</b>	57.7656	-

#### 4.4 Analisis

Hasil evaluasi menunjukkan beberapa aspek menarik yang perlu dianalisis lebih lanjut. Berikut adalah pembahasan terperinci setiap metrik:

- Rank-1 Accuracy : Rank-1 Accuracy sebesar 0.8084 menunjukkan bahwa model mampu mengidentifikasi wajah yang benar sebagai pilihan pertama dalam sekitar 80% kasus. Ini merupakan kinerja yang cukup baik untuk sebuah sistem pengenalan wajah tanpa pengawasan, menunjukkan bahwa model telah berhasil mempelajari fitur-fitur wajah yang relevan. Namun, masih ada ruang untuk peningkatan, terutama jika sistem ini akan digunakan dalam aplikasi yang memerlukan tingkat keamanan tinggi atau verifikasi identitas yang kritis.
- Hasil TAR@FAR=1e-4 untuk skenario N:N (Recognition vs Recognition) dan N:M (Recognition vs Search) : Hasilnya menunjukkan tantangan signifikan dalam sistem. Nilai yang sangat rendah ini mengindikasikan bahwa sistem sangat konservatif dalam melakukan pencocokan, yang mungkin menyebabkan banyak penolakan yang seharusnya diterima (false negatives). Ini bisa menjadi masalah serius dalam aplikasi praktis, di mana pengguna yang sah mungkin sering ditolak oleh sistem. Hasil ini menunjukkan perlunya penyesuaian pada ambang batas pencocokan atau peningkatan dalam ekstraksi dan representasi fitur wajah.
- Silhouette Score : Hasilnya sebesar 0.1287 mengindikasikan bahwa pemisahan antar kluster dalam ruang fitur wajah relatif lemah. Ini

menunjukkan bahwa banyak sampel mungkin berada di perbatasan antara kluster, yang bisa menyebabkan kesalahan klasifikasi. Dalam konteks pengenalan wajah, skor ini menunjukkan bahwa sistem mungkin kesulitan membedakan antara individu yang memiliki kemiripan fitur wajah.

- Davies-Bouldin Index : Hasilnya sebesar 1.5080 menunjukkan kualitas kluster yang moderat. Nilai ini mengindikasikan bahwa meskipun ada beberapa pemisahan antar kluster, masih ada tumpang tindih yang signifikan. Dalam sistem pengenalan wajah, ini bisa berarti bahwa representasi fitur untuk individu yang berbeda tidak cukup terdistribusi dengan baik dalam ruang fitur.
- Calinski-Harabasz Index : Hasilnya sebesar 57.7656 menunjukkan pemisahan kluster yang moderat. Meskipun ada beberapa struktur kluster yang terdeteksi, masih ada ruang untuk peningkatan dalam hal pemisahan fitur wajah. Nilai ini mengindikasikan bahwa sistem telah berhasil menangkap beberapa perbedaan antar individu, namun masih perlu ditingkatkan untuk mencapai pemisahan yang lebih jelas.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa sistem pengenalan wajah tanpa pengawasan memiliki dasar yang cukup baik, namun masih memerlukan perbaikan signifikan untuk mencapai kinerja yang optimal. Akurasi Rank-1 yang cukup tinggi menunjukkan bahwa model mampu menangkap fitur-fitur wajah yang relevan, namun nilai TAR yang sangat rendah pada FAR  $1e-4$  mengindikasikan masalah dalam pencocokan yang perlu segera diatasi.

Metrik clustering menunjukkan bahwa representasi fitur dari sistem belum optimal dalam memisahkan identitas yang berbeda. Ini mungkin menjadi akar dari masalah yang terlihat pada nilai TAR yang rendah.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Dari hasil penelitian ini, beberapa kesimpulan dapat diambil sebagai berikut: Unsupervised Face Recognition menggunakan model DeiT dengan pendekatan Unsupervised dapat digunakan untuk meningkatkan performa Unsupervised Face Recognition, dan mampu menghasilkan performa yang cukup baik pada beberapa metrik evaluasi namun pada beberapa metrik evaluasi lainnya menunjukkan nilai yang kurang baik dan memiliki peluang untuk bisa ditingkatkan, meskipun tanpa label data. Hal ini ditunjukkan dari hasil pengujian menggunakan beberapa metrik evaluasi seperti Rank-1 Accuracy mendapatkan hasil 0.8084, TAR pada FAR=1e-4 N:N sebesar 0.0000 dan N:M sebesar 0.0001, serta metrik clustering seperti Silhouette Score mendapatkan hasil 0.1287, Davies-Bouldin Index mendapatkan hasil 1.5080, dan Calinski-Harabasz Index sebesar 57.7656.

#### **5.2 Saran**

1. Perbaikan Arsitektur Model: Mengingat hasil dari beberapa metrik seperti Silhouette Coefficient dan Davies-Bouldin Index yang masih menunjukkan ruang perbaikan, penelitian lebih lanjut bisa difokuskan pada pengembangan arsitektur model transformer yang lebih efektif untuk tugas clustering dalam face recognition tanpa supervisi.
2. Penggunaan Dataset yang Lebih Beragam: Untuk meningkatkan performa model, penggunaan dataset yang lebih beragam dengan variasi pose, pencahayaan, dan ekspresi dapat membantu model dalam mempelajari representasi fitur wajah yang lebih robust.
3. Integrasi Model Pre-trained: Penelitian selanjutnya dapat mencoba menggabungkan pendekatan unsupervised dengan model pre-trained, yang mungkin dapat meningkatkan akurasi dan performa clustering dibandingkan model yang sepenuhnya dilatih dari awal (scratch).

## DAFTAR PUSTAKA

- [1] THALES, "Facial recognition: top 7 trends (tech, vendors, use cases)," [Online]. Available: <https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition>.
- [2] National Institute of Standards and Technology, "Face Recognition Vendor Test (FRVT)," [Online]. Available: <https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt>.
- [3] FasterCapital, "Labeling Face Recognition The Role of Labeling in Face Recognition Technology: A Business Perspective," 19 June 2024. [Online]. Available: <https://fastercapital.com/content/Labeling-Face-Recognition-The-Role-of-Labeling-in-Face-Recognition-Technology--A-Business-Perspective.html>.
- [4] F. Capital, "Labeling Face Recognition: Labeling Face Recognition Data: Key Considerations for Startups," 20 June 2024. [Online]. Available: <https://fastercapital.com/content/Labeling-Face-Recognition--Labeling-Face-Recognition-Data--Key-Considerations-for-Startups.html>.
- [5] I. Gordon and W. J. Tanaka, "The role of name labels in the formation of face representations in event-related potentials," [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21988390/>.
- [6] F. Boutros, M. Klemt, M. Fang, A. Kuijper and N. Damer, "Unsupervised Face Recognition using Unlabeled Synthetic Data," *Computer Vision and Pattern Recognition*, 2022 .
- [7] E. Solomon, A. Woubie and K. J. Cios, "UFace: An Unsupervised Deep Learning Face Verification System," 2022.
- [8] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles and H. Jegou, "Training data-efficient image transformers," 2021.
- [9] H. Wang, M. Li, Y. Song<sup>1</sup>, Y. Zhang<sup>2</sup> and L. Chi<sup>1</sup>, "UCoL: Unsupervised Learning of Discriminative Facial Representations via Uncertainty-Aware Contrast," *Uncertainty-aware Contrastive Learning (UCoL): a fully*

*unsupervised framework for discriminative facial representation learning*, vol. 37, 2023.

- [10] A. Jain, M. Sun and C. Zou, "Unsupervised Face Recognition in Television News Media".
- [11] C. Annubaha, A. P. Widodo and K. Adi, "Universitas Dipenogoro Institutional Repository," 16 November 2022. [Online]. Available: <https://eprints2.undip.ac.id/id/eprint/9721/>.
- [12] M. Munawir, L. Fitria and M. Hermansyah, "Jurnal Nasional Informatika dan Teknologi Jaringan," 2020. [Online]. Available: <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/2333>.
- [13] Y. Zuo, Y. Man and Z. Luo, "Unsupervised Visual Learning: An Empirical Study," 2019.
- [14] Liu, Y., Wang, F., & Zhang, K. (2023). "A comprehensive survey of Vision Transformers in face recognition: Recent advances and new frontiers." *Pattern Recognition*, 135, 109116.
- [15] Wang, X., Li, J., & Zhou, J. (2024). "Self-supervised Vision Transformers for Face Recognition with Limited Labels." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [16] Zhang, B., Li, Y., & Wang, X. (2023). "DeiT-Face: Enhanced Face Recognition with Data-Efficient Training and Attention Distillation." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.



# LAMPIRAN

## 1. Lampiran Turnitin



Page 2 of 57 - Integrity Overview

Submission ID trn:oid::1:3037058626

### 19% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

#### Filtered from the Report

- Bibliography
- Quoted Text

#### Top Sources

- 19% Internet sources
- 4% Publications
- 13% Submitted works (Student Papers)



Page 2 of 57 - Integrity Overview

Submission ID trn:oid::1:3037058626

M U L T I M E D I A  
N U S A N T A R A

### Top Sources

19%  Internet sources  
 4%  Publications  
 13%  Submitted works (Student Papers)

### Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	kc.umn.ac.id	15%
2	Student papers	Academic Library Consortium	1%
3	Internet	www.journaltoes.ac.uk	1%
4	Student papers	Ateneo de Manila University	0%
5	Internet	eprints.uad.ac.id	0%
6	Publication	Pranowo Narjosoeripto, Toni Harsan, Para Mitta Purbosari. "PENGARUH CLASSMA...	0%
7	Internet	dblp.uni-trier.de	0%
8	Internet	id.123dok.com	0%
9	Internet	repository.unisbablitar.ac.id	0%
10	Internet	www.x-mol.com	0%
11	Internet	123dok.com	0%

U N I V E R S I T A S  
 M U L T I M E D I A  
 N U S A N T A R A

12	Internet	
polen.itu.edu.tr		0%
13	Internet	
catalog.libraries.psu.edu		0%
14	Internet	
fastercapital.com		0%

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## 2. Lampiran Konsultasi Bimbingan

**Form Bimbingan Skripsi  
Program Studi Computer Engineering  
Semester Gasal 2024/2025**



Nama : CHRISTIAN REVIVAL HIZKIA KAUNANG  
NIM : 00000030714  
Angkatan : 2018  
Dosen Pembimbing : Monica Pratiwi (Pembimbing)

No	Tanggal	Jam	Keterangan	Tanggal Approval
1	26 Juli 2024	09:00	<ul style="list-style-type: none"> <li>- Merujuk ke statistik, menggunakan tetap dataset ms1m-retinaface</li> <li>- Mempelajari = Sample Size Calculator, jika menggunakan data sekian, maka margin errornya berapa, dilaporan ditulis margin error</li> <li>- Mencari penelitian unsupervised dengan menggunakan ms1m-retinaface</li> <li>- Dataset menggunakan model apa, harus ada argumennya.</li> </ul>	10 Oktober 2024 10:44
2	30 Juli 2024	09:00	<ul style="list-style-type: none"> <li>- Mempelajari ini : <a href="https://medium.com/machine-intelligence-and-deep-learning-lab/vit-vision-transformer-cc56c8071a20">https://medium.com/machine-intelligence-and-deep-learning-lab/vit-vision-transformer-cc56c8071a20</a></li> <li>- Akan menggunakan vit tipe apa, dipilih berdasarkan penelitian sebelumnya</li> <li>- Cari data statistik implementasi terbanyak, face recognition itu banyak diaplikasikan di mana, contoh jika banyak digunakan di mobile berarti vit ini menggunakan resource yang kecil</li> <li>- Jika sudah tau banyak dipakai dimana, contoh misalkan bnyk menggunakan di mobile maka cari vit yang model yang ringan</li> <li>- Cari juga model vit yang ringan dan bisa digunakan di mobile dan website</li> <li>- Berangkat dari implementasi terbanyak atau penelitian terdahulu</li> <li>- Jika sudah dapat penelitiannya dilihat akurasi</li> </ul>	10 Oktober 2024 10:44
3	02 Agustus 2024	09:00	<ul style="list-style-type: none"> <li>- Penggunaan jumlah dataset ms1m-retinaface itu berapa dalam penelitian</li> <li>- Membandingkan arcface sama Vision Transformer untuk unsupervised face recognition Vision Transformer</li> </ul>	10 Oktober 2024 10:44
4	06 Agustus 2024	09:00	<ul style="list-style-type: none"> <li>- Mencari dimana diaplikasikan face recognition</li> <li>- Presentasi Penggunaan terbesar adalah Unlocking Devices</li> <li>- Dan model ViT yang cocok untuk unlocking devices adalah DeiT (Data-Efficient)</li> </ul>	10 Oktober 2024 10:44

M U L T I M E D I A  
N U S A N T A R A

**Form Bimbingan Skripsi**  
**Program Studi Computer Engineering**  
**Semester Gasal 2024/2025**



No	Tanggal	Jam	Keterangan	Tanggal Approval
			Image Transformer) - Mencari tau tentang DeiT Data-Efficient Image Transformer)	
5	08 Agustus 2024	09:00	- Proses training untuk model DeiT apakah sama jika unsupervised atau berbeda - Pengukuran model untuk unsupervised - Mulai Mencil BAB 1	10 Oktober 2024 10:44
6	13 Agustus 2024	09:00	- Proses training untuk model DeiT (Data-efficient Image Transformers) dalam konteks unsupervised face recognition memiliki beberapa perbedaan - Confusion Matrix / pengukuran model untuk unsupervised	10 Oktober 2024 10:44
7	09 September 2024	09:00	- Mencoba running code untuk Preproses dan code untuk training di Google Colab Pro + yang dipinjamkan oleh ibu Monica	10 Oktober 2024 10:44
8	13 September 2024	13:00	- Update terkait proses running code di Google Colab - Melanjutkan proses running sampai pada bagian testing & evaluasi di laptop local menggunakan Visual Studio Code (Dengan mengurangi Dataset yang digunakan menjadi 16586 yang didapat dari sample size calculator)	10 Oktober 2024 10:44

M U L T I M E D I A  
N U S A N T A R A