

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 Transformer).

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.

UMMN

**UNIVERSITAS
MULTIMEDIA
NUSANTARA**