

**SISTEM IDENTIFIKASI WAJAH PARSIAL  
MENGUNAKAN DATASET WAJAH DENGAN  
JUMLAH DAN TIPE POSE TERBATAS**

**RINGKASAN DISERTASI**

**MUHAMMAD DJAMALUDDIN  
NIM: 33218027  
(Program Studi Doktor Teknik Elektro dan Informatika)**



**INSTITUT TEKNOLOGI BANDUNG  
Desember 2023**

# **SISTEM IDENTIFIKASI WAJAH PARSIAL MENGUNAKAN DATASET WAJAH DENGAN JUMLAH DAN TIPE POSE TERBATAS**

**Muhammad Djamaluddin**

NIM: 33218027

## **I. Latar Belakang**

Pengenalan Wajah (*Face Recognition*) secara otomatis adalah salah satu problem utama dalam visi komputer yang berhubungan dengan interaksi antara manusia dan komputer. Riset mengenai ini sudah dimulai sejak tahun 1970-an dengan menggunakan pendekatan secara geometri dari titik fitur gambar wajah, kemudian berkembang ke pendekatan holistik dan fitur lokal, dan saat ini masih secara intensif dengan metode *state of the art* berbasis pembelajaran mendalam (*Deep Learning*). L. Li dkk. (2020) mendefinisikan pengenalan wajah sebagai teknologi biometri yang didasarkan pada pengidentifikasian fitur wajah seseorang. Problem umum pengenalan wajah secara ideal adalah sebagai berikut: bagaimana mengenali wajah satu atau beberapa orang dalam berbagai kondisi, termasuk pada kondisi yang tidak ideal di mana pose wajah mengalami orientasi sehingga tidak tegak lurus di depan kamera, ekspresi wajah berubah-ubah, pencahayaan yang tidak merata dan kualitas resolusi gambar yang tidak tinggi. Biasanya ini dinamakan problem pengenalan wajah di alam nyata (*faces in the wild*) (Sengupta dkk., 2016; Dhifli dan Diallo, 2016).

Ada dua macam tugas umum dari pengenalan wajah yaitu untuk identifikasi (*identification*) dan verifikasi (*verification*). Identifikasi (*1:N matching*) artinya diberikan sebuah gambar berisi wajah, sistem akan mengidentifikasi identitas orang yang memiliki wajah tersebut dari galeri basis data yang ada. Contoh aplikasi proses identifikasi adalah penggunaan oleh aparat keamanan dalam memecahkan identitas kriminal ketika sebuah sistem kamera *Closed Circuit Television* (CCTV) menangkap video pelaku tindakan kriminal dan dianalisis oleh sistem pengenalan

wajah. Problem verifikasi (*1:1 matching*) adalah diberikan wajah dan basis data berupa identitas dan wajah orang yang ingin diverifikasi, sistem akan memverifikasi apakah benar wajah tersebut dimiliki oleh identitas yang ada dalam basis data. Sistem verifikasi wajah banyak digunakan untuk otentikasi perangkat *mobile* seperti *handphone*, *tablet* maupun *laptop* ketika pengguna ingin *login* ke dalam perangkat melalui pengenalan wajah *Apple FaceID*, *Samsung Face Recognition* atau *Microsoft Windows Hello*.

Penelitian pengenalan wajah *frontal* atau *near frontal* sudah sangat maju, berkembang dan ‘*mature*’, dengan akurasi yang sangat tinggi pada dataset wajah hampir frontal yang standar seperti *Labeled Face in the Wild (LFW)*. **Wajah frontal** adalah wajah yang menghadap langsung ke depan kamera, sudut simpangnya (*yaw-angle*) adalah 0°. Sudut simpang (*yaw-angle*) adalah sudut yang terbentuk sekitar sumbu simpang atau sumbu vertikal. Untuk problem verifikasi wajah pada dataset *LFW*, teknik *DeepFace* dari Facebook mampu memperoleh tingkat akurasi 97.25 % (Taigman dkk., 2014), sementara *FaceNet* (Schroff dkk., 2015) dari Google akurasinya adalah 99.63%, dan ResNet-10 dengan *loss function ArcFace Loss* (Deng dkk., 2018) mencapai akurasi 99.83%. Tetapi penggunaan dataset *LFW* ini dikritik karena total jumlah datanya relatif kecil untuk ukuran dataset wajah pada saat ini yaitu hanya sekitar 13.000 sampel gambar wajah. Adrienne Lafrance (2016) dan Steve Lohr (2018) melaporkan bahwa algoritma pengenalan wajah yang memiliki akurasi tinggi pada dataset terbatas dan bias seperti *LFW*, akurasinya menjadi turun tajam ketika digunakan dataset lain yang lebih variatif apalagi di dunia nyata. Saat ini terdapat beberapa dataset publik lain seperti *MegaFace*, *VGGFace2* dan *MS-Celeb-1M* untuk menggantikan *LFW* sebagai dataset pelatihan pengenalan wajah dengan jumlah individu yang sangat besar dan jumlah pose per-individu juga besar.

Topik penelitian pengenalan wajah yang masih intensif dilakukan saat ini salah satunya berkaitan dengan pengenalan wajah parsial atau bentuk wajah tidak utuh karena tertutup (oklusi). Wajah parsial terbentuk karena pose wajah yang tidak frontal, wajah terhalang benda lain atau faktor pencahayaan yang tidak merata,

sehingga hanya sebagian wajah yang terlihat. **Wajah profil** (*profile face*) adalah kasus khusus dari wajah parsial di mana bagian wajah tertutup akibat pose individu yang berputar mendekati  $90^\circ$  pada sudut simpang (*yaw-angle*) dari arah kamera. Pose ini menyebabkan terjadi oklusi-diri (*self-occlusion*) pada wajah. Dalam disertasi ini istilah 'wajah profil' khusus mengacu pada wajah yang sudut simpangnya sama atau sangat mendekati  $90^\circ$  itu. Pengambilan foto wajah profil seseorang ini sering dilakukan pada beberapa institusi misalnya kepolisian ketika memfoto tersangka suatu tindakan kriminal (*mug-shot*), selain juga menyertakan foto wajah frontalnya.

Riset pengenalan wajah yang tertutupi oleh obyek sudah banyak dilakukan baik menggunakan berbasis pada ekstraksi fitur lokal berbasis titik kunci (*keypoint*) seperti *IFPLD*, *HOG*, *LBP* atau yang modern berbasis *deep learning*. Tipe wajah parsial karena oklusi yang diinvestigasi bisa berupa wajah yang ditutupi kaca mata hitam, masker, selendang. Weng dkk. (2016) menggunakan sekelompok metode pencocokan fitur berbasis pada detektor *SIFT*, *SURF* dan *LBP* yang disebut *Robust Point Set Matching* (RPSM) untuk mencocokkan 2 gambar wajah, 1 wajah parsial dan 1 lagi wajah frontal penuh. Metode ini bisa digunakan untuk verifikasi wajah parsial dengan akurasi pada dataset *LFW* 71%. L He dkk. (2018) menggunakan *Fully Convolutional Network* (FCN) untuk melakukan ekstraksi fitur wajah parsial yang ingin dicocokkan dan wajah-wajah di galeri, kemudian mencocokkan dengan fitur tersebut menggunakan teknik *Dynamic Feature Matching* (DFM). Metode ini mencapai akurasi 94.96% pada dataset *CASIA-NIR*. Tetapi kedua metode tersebut fokus pada wajah parsial yang dirotasi pada sudut putar (*rolling-angle*) tidak pada wajah parsial yang dirotasi pada sudut simpang (*yaw-angle*) karena dataset yang digunakan adalah dataset wajah frontal atau hampir frontal yang sengaja dihilangkan sebagian wajahnya.

Pengenalan wajah profil masih merupakan problem terbuka (*open problem*) yang terus diteliti dengan berbagai pendekatan modern terutama berbasis *deep learning*. Cao dkk. (2018) menggunakan blok *Deep Residual Equivariant Mapping* (DREAM) yang terdiri atas CNN dan blok residu untuk menghasilkan vektor

*embedding* wajah frontal dari sebuah wajah profil. Vektor *embedding* ini kemudian digunakan untuk proses verifikasi atau identifikasi. Tercatat akurasi identifikasi wajah sebanyak 94.6 % pada dataset *benchmark IJB-A*. Huang dan Ding (2021) menggabungkan CNN dan vektor *Attention* untuk membuat wajah profil menjadi frontal (*frontalization*) di ruang fitur. Metode ini dilaporkan mencapai akurasi verifikasi wajah 98.4% pada dataset *Celebrity Frontal Profile (CFP)*. Kedua metode di atas, membutuhkan dataset pelatihan yang sangat besar yaitu *MS-Celeb-1M* di mana dataset tersebut itu terdiri atas 10 juta wajah selebriti berbagai pose untuk dapat mengenali pola dari berbagai jenis wajah. Dari penelitian tersebut ternyata metode DREAM juga tidak berhasil merekonstruksi fitur wajah profil untuk sudut simpang mendekati atau sama dengan  $90^\circ$ , bahkan untuk wajah dengan sudut simpang di atas  $45^\circ$ , metode tersebut mengalami kesulitan (K. Cao dkk., 2018)<sup>o</sup>.

Tantangan lain yang dihadapi banyak institusi yang ingin menerapkan pengenalan wajah adalah institusi tersebut hanya memiliki dataset wajah dengan tipe pose terbatas dengan jumlah gambar per-subyek yang rendah untuk pelatihan sistem. Penelitian pengenalan wajah yang *state of the art* (SOTA) mensyaratkan di mana subyek memiliki ratusan foto berbagai pose sebagai data sampel untuk pelatihan sistem. Pada dataset *VggFace2* (Q. Cao dkk., 2017) misalnya, setiap subyek rata-rata memiliki 362 sampel gambar. Hal tersebut jarang terjadi di dunia nyata di mana banyak institusi memiliki basis data foto subyek hanya terdiri dari 1 atau beberapa buah foto wajah frontal (untuk KTP, Kartu Mahasiswa, SIM) dan mungkin foto tambahan menghadap ke samping.

Metode berbasis pembelajaran mendalam (*deep learning*) mensyaratkan dataset berjumlah besar dalam berbagai pose untuk bisa melakukan generalisasi dengan baik. Problemnnya adalah jika dataset pelatihan yang dimiliki sangat terbatas dari sisi jumlah kelas dan jumlah data per-kelas maka terjadi penurunan kinerja model yang drastis. Sebagai perbandingan, manusia bisa mengenali pola sebuah obyek jika diberikan hanya dengan mengamati beberapa sampel saja. Bahkan diberikan 1 contoh gambar saja manusia bisa mempelajari pola dan menggambarkan obyek

yang memiliki pola yang sama (Tiedemann dkk., 2022). Kaitannya dengan pengenalan wajah, jika diberikan beberapa sampel wajah orang, manusia bisa mengenali wajah orang yang sama di dalam galeri dengan wajah yang ditanyakan.

Teknik umum yang biasa digunakan untuk mengatasi kurangnya data adalah menggunakan *transfer learning* dan data augmentasi. Transfer learning melakukan transfer pengetahuan dari domain asal yang memiliki banyak data ke domain target yang memiliki sedikit data (Wang et al., 2020). Model pengenalan wajah yang sudah dilatih oleh dataset wajah yang berjumlah besar dapat digunakan untuk melakukan tugas yang sama pada dataset wajah berjumlah kecil dengan beberapa modifikasi seperti fine-tuning.

Selain *transfer learning*, terdapat konsep lain dalam menggunakan model yang dilatih untuk tugas lain pada tugas baru yaitu *meta learning*. Dalam *meta learning*, sistem belajar untuk mempelajari pola untuk dapat diadopsi pada tugas berbeda dengan cepat. Tujuannya adalah untuk membuat algoritma pembelajaran dasar lebih mudah beradaptasi dengan tugas-tugas baru, dengan mengekstraksi pengetahuan yang dapat ditransfer dari banyak tugas lainnya (G. Chen dkk., 2019). *Meta learning* digunakan untuk menyelesaikan problem *few-shot learning* yaitu problem di mana jumlah data sampel untuk belajar berjumlah terbatas.

Saat ini penelitian tentang *meta learning* dan *few-shot learning* mulai banyak dilakukan (Snell dkk., 2017; Y. Wang dkk., 2020 ; Li dkk., 2021), tetapi penggunaannya untuk problem pengenalan wajah baru beberapa tahun terakhir mulai diteliti secara intensif. Peng dan Zheng (2021) menggunakan metode *meta learning* bernama Model Agnostic Meta-Learning (MAML) untuk melakukan pengenalan wajah berbasis VGG network dengan 40 sampel data per-kelas dengan akurasi maksimum 98.5%. Guo dkk (2020) menggunakan teknik *meta learning* untuk mendesain sistem yang dapat dengan baik mengenali wajah yang tidak terdapat dalam dataset pelatihan (*unseen domain*) sehingga mengurangi kesalahan identifikasi dimana model dilatih menggunakan dataset berbasis MS-Celeb-1M dan diuji dengan dataset RFW.

## I.1 Tujuan dan Sasaran Penelitian

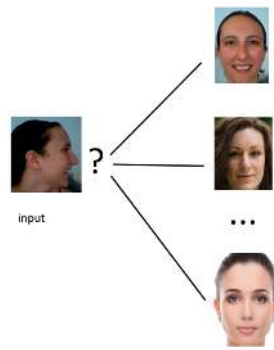
Maksud dari penelitian disertasi ini adalah untuk meneliti problem identifikasi wajah parsial dengan dataset pelatihan yang terbatas dari sisi jumlah foto per-individu. Secara spesifik wajah parsial yang akan diteliti adalah wajah profil, di mana sudut simpangnya hampir atau sama dengan  $90^\circ$ . Wajah profil menarik untuk diteliti karena merepresentasikan pose sudut simpang yang ekstrem, yang mana setengah bagian dari wajah tidak tampak sehingga merupakan sebuah tantangan penelitian. Tugas identifikasi wajah profil berarti mencari individu yang memiliki wajah profil yang ditanyakan (*data query*).

Banyak penelitian pengenalan wajah sekarang ini berfokus pada area ini, tetapi dataset pelatihan yang digunakan sangat besar seperti dataset *MS-Celeb-1M* yang berjumlah 10 juta gambar. Selain membutuhkan dataset besar, tentunya sumber daya komputasi untuk melakukan pelatihan sistem pengenalan wajah dengan dataset ini sangat besar. Maka untuk maksud penelitian ini, dibangun dataset tersendiri yang memang sesuai dengan kriteria tertentu yaitu setiap individu memiliki jumlah foto terbatas dan pose wajahnya terbatas jenisnya dengan karakteristik yang memang lebih sesuai dengan wajah Indonesia.

## II.2 Rumusan Masalah

Dari pemaparan latar belakang dan maksud penelitian di atas, terlihat bahwa masih terdapat tantangan dan kendala pada pengenalan wajah profil baik untuk problem verifikasi maupun identifikasi. Beranjak dari problem-problem tersebut maka rumusan masalah utama penelitian yang diangkat adalah ***jika diberikan wajah profil seseorang, apakah sistem yang dilatih dengan keterbatasan dataset bisa dengan baik melakukan identifikasi mana wajah frontal orang tersebut dari sekumpulan wajah frontal dalam galeri.*** Wajah frontal seseorang bisa menyatakan identitas dari seorang individu karena wajah frontal gampang dikenali. Problem ini ditunjukkan pada gambar I.1.

Adapun batasan penelitian disertasi ini adalah sebagai berikut:



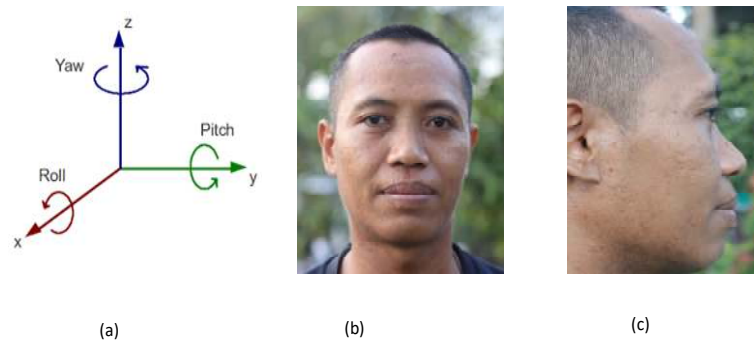
Gambar I.1 Mencari wajah frontal di dalam galeri yang cocok dengan wajah profil

1. Penelitian ini hanya fokus meneliti dan mengembangkan metode untuk tahap ekstraksi fitur dan tahap pengenalan/pencocokan wajah.
2. Penelitian tidak mengembangkan metode deteksi wajah walaupun menggunakan salah satu metode deteksi wajah untuk menyiapkan dataset. Selain itu penelitian juga tidak membahas *face alignment* atau penyelarasan wajah dengan mencari *facial landmark*.
3. Penelitian tidak membahas metode *face frontalization* yaitu mengubah wajah profil menjadi wajah frontal baik di ruang gambar maupun ruang fitur, yang walaupun sangat relevan dengan permasalahan, karena penelitian ini fokus pada aspek lain yaitu pembelajaran sistem identifikasi wajah dengan dataset terbatas.

## II. Metode Penelitian

Sesuai dengan *problem statement* yang telah diutarakan di bagian I, bahwa fokus utama penelitian ini adalah meneliti sistem identifikasi wajah seseorang dari wajah profilnya, di mana sistem dilatih dengan dataset yang sangat terbatas yaitu sebuah dataset wajah yang mana setiap subyek terdiri atas satu wajah depan (frontal) dan satu wajah samping (profil). Dataset ini dinamakan dengan *ITB Frontal Profile Limited Dataset (IFPLD)* yang terdiri atas 475 individu. Kriteria utama dataset *IFPLD* yaitu subyek memiliki hanya dua tipe pose yaitu depan (frontal) dan samping (profil), dan masing-masing pose hanya satu gambar. Jadi dalam dataset ini variasi pose wajah hanya rotasi pada sudut simpang (*yaw*) bukan sudut putar (*rolling*) atau sudut menukik (*pitch*) (gambar II.1).



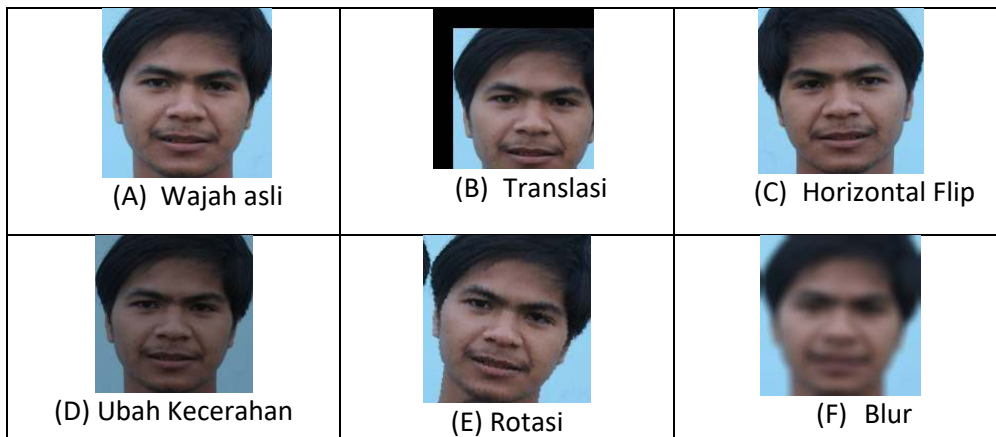


Gambar II.1 (a) Tipe Poros koordinat 3D (b) Wajah Frontal (depan)  
(c) Wajah Samping

Melihat keterbatasan jumlah data IFPLD yang memiliki kriteria ketat, yaitu 1 wajah frontal dan 1 wajah profil, maka metode sederhana mengatasi keterbatasan data ini adalah melakukan data augmentasi. Maka pada dataset *IFPLD* dilakukan proses data augmentasi sederhana yaitu melakukan 5 buah operasi pengolahan gambar sederhana untuk setiap wajah yaitu translasi, pencerminan, rotasi, perubahan kecerahan (*brightness*) dan blur (Gambar II.2).

Operasi translasi pada data augmentasi dilakukan dengan menggeser gambar 20 piksel ke kanan dan 20 piksel ke bawah. Operasi rotasi memutar gambar  $10^\circ$  arah jarum jam. Operasi perubahan kecerahan mengubah kecerahan dan kontras gambar dengan koefisien gain 0.7 dan bias 0.2. Operasi *blur* adalah melakukan pengaburan (*blurring*) pada gambar dengan *average/box filter* berukuran  $10 \times 10$ . Untuk memudahkan referensi dalam disertasi ini, dataset *IFPLD* teraugmentasi dengan ke-5 operasi pengolahan gambar tersebut di atas dinamakan *IFPLD+Aug*.

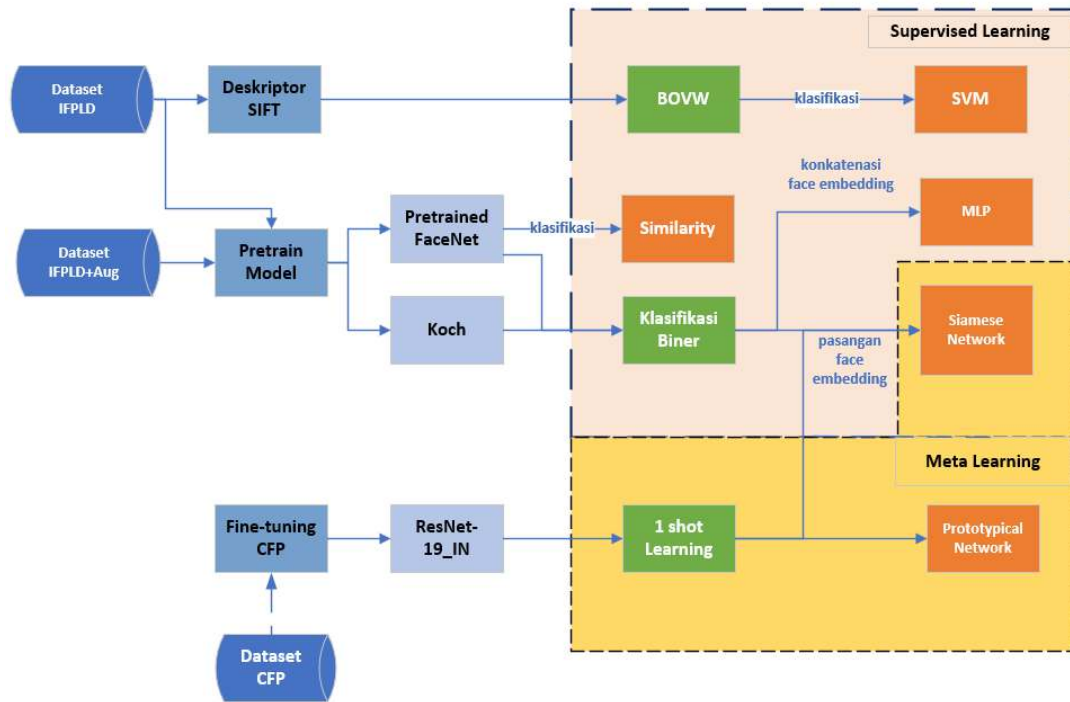
Secara ringkas metode-metode yang digunakan dalam penelitian ini digambarkan pada Gambar II.3. Dari sisi ekstraksi fitur yang digunakan ada dua yaitu detektor fitur lokal dan *deep learning*. Dan fitur representasi menggunakan *deep learning* didapatkan dengan 2 cara yaitu menggunakan vektor *embedding* sistem *FaceNet* dan melakukan ekstraksi fitur CNN yang sudah dilatih dan dilakukan *fine-tuning* dengan berbagai metode pembelajaran untuk mendapatkan *feature embedding* yang unik dan distingtif.



Gambar II.2 Augmentasi data gambar wajah seseorang (A) diubah menjadi 5 gambar berbeda sesuai operasi yang dilakukan

Dalam merumuskan permasalahan identifikasi wajah parsial pada dataset *IFPLD* dilakukan 2 pendekatan. Pendekatan pertama dengan menganggap problem pengenalan wajah tersebut sebagai problem *supervised learning* di mana setiap data ada labelnya, dan sistem dilatih berdasarkan data berlabel tersebut untuk bisa melakukan klasifikasi pada data lain yang berlabel pada data pengujian. Dalam pendekatan *supervised learning*, dilakukan perbandingan pendekatan pengenalan wajah dengan ekstraksi fitur tradisional dan modern untuk verifikasi dan identifikasi wajah. Pada kedua problem pengenalan wajah, *metrik similarity* dan metode klasifikasi digunakan untuk membandingkan pasangan wajah frontal dan profil atau menentukan individu (yang dinyatakan dengan wajah frontal dalam galeri) dari sebuah wajah profil.

Pendekatan kedua adalah menganggap permasalahan dataset *IFPLD* yang sangat terbatas sebagai permasalahan utama dan sangat krusial sehingga tidak cukup hanya menggunakan pendekatan *supervised learning* yang umumnya digunakan dalam pengenalan wajah. Pada studi literatur di bagian II.3 dijelaskan bahwa konsep pembelajaran *meta-learning* bertujuan melatih sistem 'mempelajari cara untuk belajar' (*learns to learn*) dan digunakan untuk menyelesaikan problem *few-shot learning*. Problem identifikasi wajah pada dataset *IFPLD* dianggap sebagai problem *N-way-1-shot learning*, yaitu mencari satu wajah profil pada N individu dan setiap individu memiliki satu wajah frontal.



Gambar II.3 Ringkasan Metode Penelitian

Di beberapa metode yang digunakan baik untuk pendekatan *supervised learning* maupun *meta learning*, model verifikasi wajah (1:1) diteliti sebagai cara awal untuk mencari kandidat model identifikasi wajah (1:N) yang kinerjanya baik. Contohnya pada *siamese network*. Siamese Network menghasilkan probabilitas yang bisa mengukur derajat kemiripan dua input data. Dari nilai probabilitas itu bisa digunakan untuk problem identifikasi wajah dengan konsep *one-shot validation*.

## II.1 Ekstraksi Fitur Klasik dengan SIFT

Dengan kriteria dataset wajah IFPLD yang jumlah dan tipe posenya terbatas untuk setiap orang, maka pendekatan pertama adalah identifikasi wajah sebagai problem klasifikasi obyek menggunakan fitur wajah lokal untuk ekstraksi fitur dan klasifikasi menggunakan metode *Machine Learning* klasik. Terdapat berbagai metode representasi fitur klasik dalam *computer vision* diantaranya adalah *Scale Invariant Feature Transform* (SIFT) yang mencari titik kunci (*keypoint*) yang unik pada sebuah gambar dengan kriteria tertentu

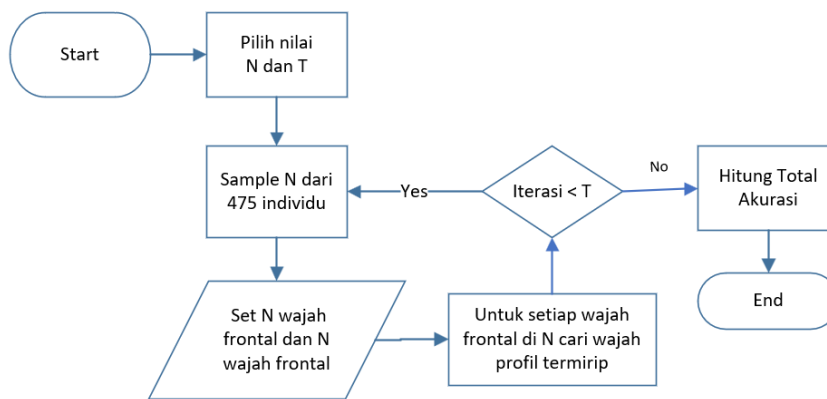
Setiap wajah pada IFPLD diubah ke ukuran 255x255 agar seragam dan wajah dari 475 individu di IFPLD diekstraksi dengan SIFT. Deskriptor SIFT dapat digunakan untuk melakukan pencocokan fitur obyek dari 2 buah gambar yang berbeda skala, rotasi atau translasi. Jika deskriptor SIFT digunakan untuk melakukan kategorisasi/klasifikasi salah satu metode yang dapat digunakan adalah metode *Bag of Visual Word* (BOVW). BoVW menggunakan deskriptor SIFT, SURF dll sebagai data pembentukan “kamus” visual. Jumlah *visual words* pada kamus ini ditentukan sebanyak nilai tertentu misalnya 50, 100 dan 500 *visual words* sebagai jumlah kluster *words* yang dibentuk. Setiap gambar wajah adalah kumpulan kata-kata tertentu dan diasumsikan setiap individu memiliki wajah frontal dan profil dengan kelompok visual kata yang mirip.

## II.2 Ekstraksi Fitur dengan Pretrained FaceNet

FaceNet menghasilkan vektor *embedding* yang unik setelah dilatih dengan sebuah dataset wajah yang berukuran besar dan kemudian bisa diuji atau dievaluasi dengan dataset wajah yang berbeda dengan dataset pelatihan. FaceNet yang dilatih dengan dataset VGG\_Face2 diuji dengan dataset LFW dan menghasilkan akurasi sangat tinggi yaitu sekitar 99.65% untuk verifikasi wajah dengan menggunakan jarak *euclidean* walaupun VGG\_Face2 dan LFW adalah dua dataset wajah yang sama sekali berbeda dari segi jumlah individu, foto per-individu dll.

Pada tahap ini, skenario yang sama juga dilakukan pada fitur FaceNet dari dataset IFPLD yaitu *pre-trained* FaceNet digunakan untuk ekstraksi fitur dataset IFPLD dan fitur ini digunakan untuk melakukan identifikasi wajah. Cara yang paling mudah untuk membandingkan kemiripan antara 2 pose wajah frontal dan profil adalah menggunakan jarak *euclidean* dan *cosine similarity* dari kedua fitur vektor.

Jika terdapat  $N$  identitas, maka pasangan vektor wajah frontal dan profil dari  $N$  identitas itu dibandingkan untuk dicari yang paling mirip.  $N$  di sini bisa diset maksimum 475 individu IFPLD, berapa individu yang akan dibandingkan. Dalam eksperimen dipilih  $N = 5, 20, 40, 95$  dan 475. Sementara  $T$  menunjukkan berapa kali  $N$  buah sample identitas dipilih dari 475 individu, dan  $T$  dipilih 400 dan 1000. Algoritma identifikasi wajah dengan similarity yang digunakan digambarkan pada flowchart di gambar II.3.



Gambar II.3 Algoritma perhitungan similarity vektor embedding FaceNet

Dalam problem verifikasi wajah, model membandingkan apakah sebuah data wajah *input* memiliki kecocokan (*matching*) dengan data wajah subyek yang tersimpan dalam basis data. Ini bisa dianggap sebagai problem klasifikasi biner (*Binary Classification*) dengan data *input* berupa 2 macam yaitu gabungan/konkatenasi dari 2 pasang vektor wajah frontal dan profil sebagai input *Multi Layer Perceptron* (MLP) atau masing-masing wajah frontal dan profil dimasukkan ke dalam *siamese network*. Kedua network kemudian dilatih menghasilkan model yang mampu mengklasifikasikan *input* itu ke salah satu dari 2 kelas yaitu kelas 1 jika kedua vektornya dari orang yang sama atau kelas 0 jika keduanya dari wajah berbeda.

Karena ada 475 subyek pada IFPLD maka diperoleh  $475 \times 475 = 225625$  kemungkinan pasangan *input* data, di mana hanya 475 *input* yang termasuk kategori kelas 1 (kedua wajah dari orang yang sama) dan sisanya adalah kelas 0. Sehingga kelas 1 hanya memiliki data 0.2 % atau sangat kecil sekali, alias dataset sangat

*imbalanced*. Pada dataset *IFPLD+Aug*, secara total data jumlah kelas 1 meningkat, tapi juga kelas 0 meningkat dengan semakin banyak kombinasi wajah perpasangan. Maka disusun skema data pelatihan, di mana kelas 0 hanya mencakup pasangan wajah frontal dan wajah profil yang asli sehingga jumlah kelas 0 sama dengan skema sebelumnya yaitu tanpa data augmentasi. Diperoleh total persentase kelas 1 adalah 7 % bukan lagi 0.2%.

Dalam penelitian ini dilakukan pelatihan yang menjaga keseimbangan data wajah sama dan wajah berbeda yang mengambil inspirasi dari konsep *K-fold cross-validation*. Dalam metode pelatihan yang dilakukan, dalam setiap *batch* pada pelatihan DNN dipastikan perbandingan data kelas 0 dan 1 seimbang. Sebuah *batch* dalam pelatihan menunjukkan jumlah data yang dilatih sebelum bobot model diupdate dan dalam setiap *batch*-nya terdapat 2K data. Untuk lebih memudahkan penyebutannya, pelatihan ini disebut dengan pelatihan *2K-balanced batch*, di mana K menunjukkan jumlah data kelas 1 yang digunakan.

### **II.3 Ekstraksi Fitur dengan *Feature Embedding Network* hasil fine-tuning**

Penggunaan *FaceNet* sebagai ekstraktor fitur untuk setiap data wajah *IFPLD* dalam problem pengenalan wajah menggunakan MLP atau Siamese pada pendekatan sebelumnya, sama artinya dengan menganggap selama *training* berlangsung model *FaceNet* dibekukan (*frozen*). Artinya bobot networknya tidak diupdate dalam *backpropagation*. Agar mengakomodir perbedaan dan karakteristik dataset *IFPLD* maka dilakukan fine-tuning pada sebuah model base CNN.

Model dasar (*baseline*) CNN yang digunakan untuk pembelajaran *meta-learning* pada problem *face recognition* adalah 3 model CNN, yaitu Model Koch (Koch dkk., 2016), *FaceNet* yang dilatih dengan VGG2 (dinamakan *FaceNet\_VGG2*) dan *ResNet-18* yang dilatih dengan ImageNet 1000 kelas (dinamakan *ResNet-18\_IN*).

Jika kedua model di atas, Model Koch dan *FaceNet\_VGG2* digunakan apa adanya sesuai struktur dan bobot yang ada (jika sudah dilakukan *pre-trained*), maka model

*baseline* ResNet-18\_IN merupakan model yang dasar yang akan digunakan membentuk *feature embedding network* dengan teknik *fine-tuning* dan pelatihan tertentu. Semua model network untuk *feature embedding* yang berasal atau berbasis pada ResNet-18\_IN akan disebut sebagai **ResNet-18\_IN\***.

Model **ResNet-18\_IN** dilakukan *fine-tuning* dengan dataset yang lebih spesifik pada problem pengenalan wajah khususnya wajah dengan karakteristik pose frontal dan profil. Dipilih dataset *Celebrity Frontal Profile* (CFP) (Sengupta et al., 2016), karena tipe data wajah pada dataset CFP memiliki kemiripan dengan dataset *IFPLD*. Perbedaannya adalah dataset *IFPLD* hanya memiliki 1 wajah frontal dan 1 profil untuk setiap individu sedangkan CFP terdiri atas 500 individu selebriti dan setiap individu memiliki 10 wajah frontal dan 4 wajah profil, sehingga total ada 7000 gambar wajah.

Tabel II.1 menunjukkan daftar model ResNet-18\_IN yang sudah dilakukan *fine-tuning* dengan dataset CFP menggunakan 3 skema pelatihan dan *loss function* yang berbeda. Untuk lebih memudahkan referensi, maka ketiga model ResNet-18 hasil *fine-tuning* dengan ketiga metode disebut **ResNet-18\_IN\_CFP\***.

*Siamese Network* sangat penting dalam *Metric Learning*, karena secara natural arsitektur ini membandingkan dua buah *input* atau lebih yang akan dicari perbedaan representasi fiturnya satu dengan yang lain berdasarkan kriteria *loss function* tertentu. Model ResNet-18\_IN\_CFP\* sebagai basis *Siamese Network* yang dilatih dengan *contrastive loss* disebut sebagai **ResNet-18\_IN\_CFP\* + Contrastive Loss**

Tabel II.1. Tipe Model ResNet-18\_IN\_CFP\* yang dilakukan fine-tuning dan Loss Function yang digunakan

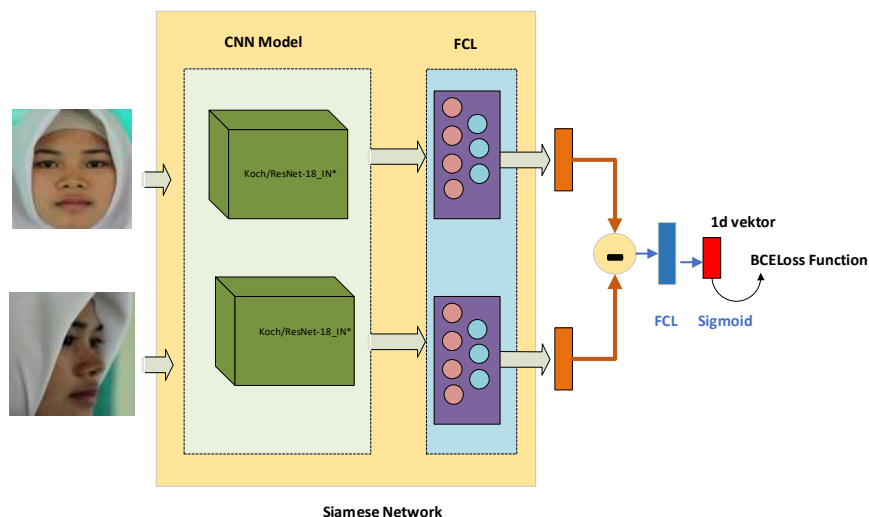
<b>Fine-tuned Model</b>	<b>Loss Function</b>
ResNet-18_IN_CFP1	Cross Entropy / Softmax Loss
ResNet-18_IN_CFP2	Binary Cross Entropy
ResNet-18_IN_CFP3	ArcFace Loss

sedangkan yang dilatih dengan *triplet loss* disebut dengan **ResNet-18\_IN\_CFP\* + Triplet Loss**.

Selain menggunakan konsep *metric learning* untuk mendapatkan *feature embedding network*, dilakukan juga pelatihan menggunakan *contrastive learning* dengan struktur network yang berbeda. Metode *contrastive learning* yang digunakan adalah SimCLR (T. Chen dkk., 2020) sehingga model ResNet-18\_IN\_CFP\* yang dihasilkan diberi nama **ResNet-18\_IN\_CFP\* + SimCLR**.

*Feature embedding network* yang di dapat dengan berbagai macam fine-tuning dan pelatihan tersebut, digunakan untuk verifikasi wajah dengan Siamese Network. Data pelatihan dan pengujian diperoleh dengan secara acak memilih 2 pasangan wajah orang yang sama atau berbeda dari data wajah *IFPLD+Aug* dengan kelas subyek sebanyak 380.

Proses verifikasi wajah dengan *siamese network* digambarkan pada Gambar II.4, di mana masing-masing gambar wajah frontal dan profil dimasukkan ke dua buah network seragam dari sisi arsitektur dan bobot, yang kemudian selisih vektor *embedding* yang dihasilkan diubah ke dalam bentuk probabilitas oleh fungsi



Gambar II.4 Verifikasi wajah menggunakan *Siamese Network* dengan output berupa nilai probabilitas yang menunjukkan kedua input dari wajah yang sama



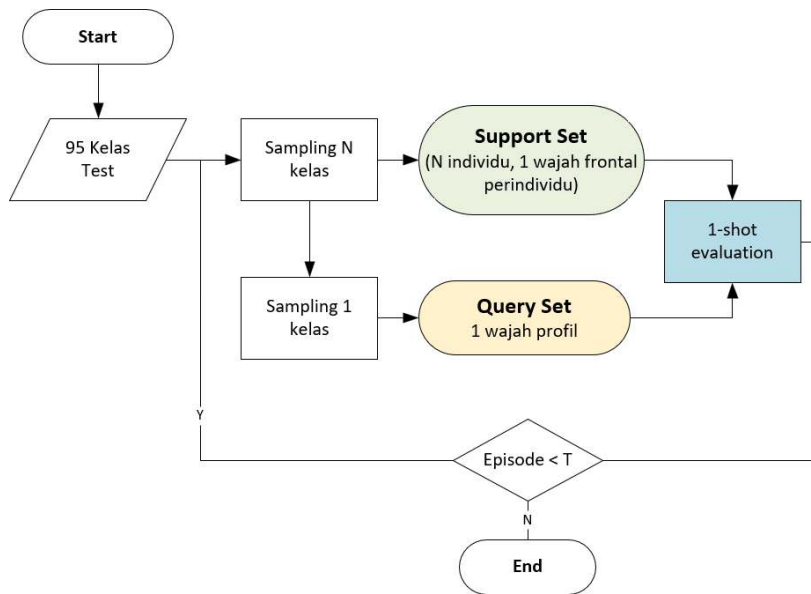
*sigmoid* (warna merah). Label untuk pelatihan adalah kelas 0 untuk sepasang wajah frontal dan profil berasal dari orang yang berbeda sementara kelas 1 sepasang wajah berasal dari orang yang sama. Output dari model Siamese berupa nilai fungsi *sigmoid* menunjukkan probabilitas kemiripan dari 2 gambar *input*, di mana nilai di atas 0.5 adalah kelas 1 dan nilai di bawahnya adalah kelas 0.

Keterbatasan jumlah dataset *IFPLD* membuat konsep *Few-Shot Learning* berupa klasifikasi *N-way-K-shot* sangat cocok karena diterapkan pada problem identifikasi wajah. Nilai *N* di sini menunjukkan jumlah dari kelas yang di-*sampel* dari sekelompok kelas untuk dijadikan *support class*, dan *K* menunjukkan jumlah data setiap kelasnya.

Seperti juga eksperimen sebelumnya, data pelatihan dan pengujian didapat dengan membagi 475 subyek pada *IFPLD* menjadi 380 subyek sebagai kelas data pelatihan dan validasi, dan 95 subyek untuk kelas data pengujian. Data pelatihan diperoleh dengan secara acak memilih 2 pasangan wajah orang yang sama atau berbeda dari data wajah *IFPLD+Aug* dengan kelas subyek sebanyak 380. Data ini kemudian diatur berbeda dalam setiap *batch*-nya sesuai metode yang digunakan.

Prosedur evaluasi *N-Way-1-shot learning* untuk Siamese Network digambarkan pada *flowchart* di gambar II.5, yang mana ada variabel *N* dan *T* yang dipilih. *T* menunjukkan jumlah maksimum episode pengujian yang dilakukan dan diset 400 atau 1000. Dari 95 kelas uji, dipilih *N* buah kelas secara acak, di mana *N* umumnya dipilih 5, 10, 20 atau 40. Kemudian dari masing-masing *N* kelas ini, diambil masing-masing wajah frontalnya sebagai *data support*. Setelah itu secara acak diambil sebuah kelas dari *N* berikut wajah profilnya sebagai *data query*. Wajah profil ini lalu dibandingkan dengan *N* buah wajah frontal untuk mencari kelas identitas dari individu yang memiliki wajah profil itu.

Metode lain dalam meta learning untuk problem *N-way-k-shot learning* adalah *Prototypical network*. *Prototypical network* bertujuan untuk membentuk sekumpulan *prototype* dari data di dalam *support set* berdasarkan kelas dari data



Gambar II.5 Skenario evaluasi *N-Way-1-Shot* pada Siamese Network

tersebut dan mengidentifikasi kelas dari data pada *query set* berdasarkan kedekatan data *query* ke *prototype* sebuah kelas. Pada saat pelatihan, untuk membentuk sebuah *support set*,  $N$  buah kelas latih, dipilih secara acak di mana setiap kelas hanya terdiri memiliki satu wajah frontal. Data *query* dibentuk dari satu atau beberapa kelas pada *support set*, yang diambil wajah profilnya. Dikarenakan hanya satu sampel data setiap  $N$  kelas maka problemnya ada  $N$ -way-1-shot learning.

Dalam sebuah episode pelatihan dan pengujian, terdapat rangkaian proses dari pembentukan *support set* dan *query set*, kemudian masing-masing set diambil *vektor embedding*-nya menggunakan model CNN, setelah itu titik *prototype* dihitung untuk setiap kelas dalam *support set*, sampai akhirnya penentuan kelas dari data di *query set* dari jarak terdekat ke *prototype* kelas yang mana. Titik *prototype* dihitung dengan rumus :

$$c_j = \frac{1}{|S_j|} \sum_{(x,y) \in S_j} f_{\phi}(x_i) \quad (\text{II.1})$$

Untuk analisa *prototypical network* digunakan 2 model. Yang pertama adalah model dasar CNN yang digunakan oleh beberapa metode *meta-learning* seperti Vinyals dkk. (2016), Snell dkk. (2017) dan Sung dkk. (2018) yaitu CNN dengan 4 blok khusus yang untuk mudahnya disebut *Model Standar FSL*. Model kedua

adalah model berbasis ResNet-18\_CFP\* yang sudah dilakukan fine-tuning dan memang dipersiapkan untuk mengekstraksi fitur gambar wajah.

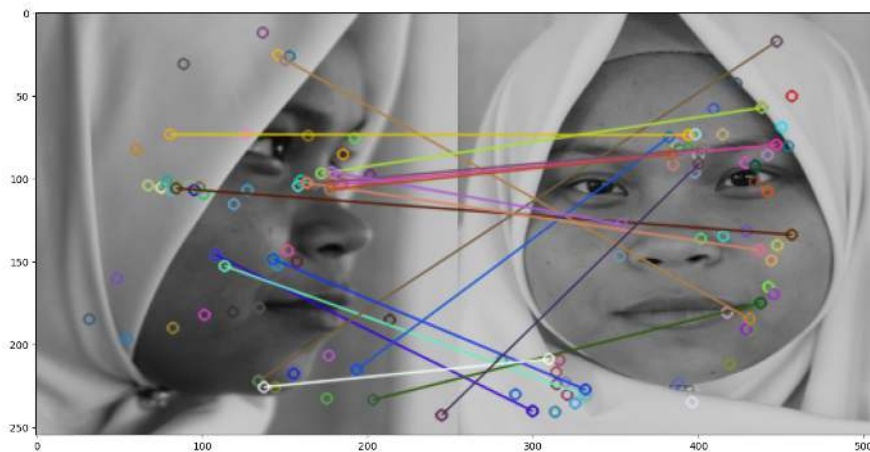
Ada 2 metode perhitungan jarak *prototypes* dan vektor *embedding query set* sehingga ada 2 tipe perhitungan *loss function* yang digunakan. Metode pertama menggunakan semua vektor *embedding* dari *query set* untuk dihitung jaraknya ke kelompok *prototype* pada *support set*. Sementara metode kedua, mencari titik pusat (*center point*) dari 'cluster' vektor *query set*, dan menghitung jarak titik pusat ini ke kelompok *Prototype*. Pusat kluster  $q$  dari sekumpulan vektor data *query*  $Q$  di definisikan sebagai:

$$q = \frac{1}{|Q|} \sum_{(x_i) \in Q} f_{\phi}(x_i) \quad (\text{II.2})$$

Terminologi  $N_P$  adalah jumlah individu/kelas dalam *support set* yang dipilih saat pelatihan dalam 1 episode,  $N_{PS}$  merupakan jumlah gambar wajah per-individu dalam *support set* dan  $N_{PQ}$  adalah jumlah gambar wajah per-individu dalam *query set*. Biasanya  $N_{PQ}=N_{PS}$ . Data pengujian berasal dari 95 kelas uji pada dataset IFPLD, yang individunya berbeda dengan kelas dan data pelatihan. Agar pengujian bersifat konsisten untuk semua model, maka dari 95 kelas dibentuk 5 buah dataset yang dinamakan  $D_1, \dots, D_5$  untuk setiap  $N_U$ , di mana  $N_U=5, 20, 40$ .

### III. Hasil dan Pembahasan

Untuk melakukan pencocokan fitur pada deskriptor titik kunci seperti SIFT, bisa digunakan *Metode Brute Force*, *Flann* atau *KNN Matcher*. Gambar. Terlihat hanya sedikit deskriptor yang dapat dicocokkan yaitu 17 buah, dari 75 deskriptor wajah profil dan 67 deskriptor wajah frontal pada Gambar III.1. Ini menunjukkan metode *brute force* pada SIFT tidak berhasil mengenali fitur wajah yang mirip dan bersesuaian pada kasus kedua gambar di atas



Gambar III.1 Penyocokan *keypoint* deskriptor SIFT pada gambar wajah profil dan frontal subyek yang sama

Salah satu cara untuk memproses fitur lokal dalam melakukan proses klasifikasi adalah metode *Bags of Visual World* (BOVW). Akurasi yang didapat dari klasifikasi SVM pada BOVW ini sangat kecil yaitu 5.2% untuk problem *closed-set face recognition* di mana pencocokan histogram wajah frontal dan wajah profil banyak mengalami kegagalan untuk kelas tes pada IFPLD.

Dalam perhitungan kemiripan antar 2 vektor embedding, *euclidean distance* dan *cosine similarity* digunakan langsung pada fitur *FaceNet* untuk IFPLD. *Cosine similarity* banyak digunakan dalam klasifikasi teks dan gambar termasuk dalam pengenalan wajah *closed-set* maupun *open-set*.

Untuk identifikasi wajah, dipilih N identitas dari maksimum 475 kelas IFPLD, dan pasangan vektor wajah frontal dan profil dari N identitas itu dibandingkan untuk dicari kedekatannya. N dipilih 5, 20, 40, 95 dan 475. Sementara berapa kali N individu dibandingkan ditentukan oleh T, yang diset 400 dan 1000. Tabel III.1 menunjukkan kinerja dari similarity pada FaceNet. Akurasi untuk N yang berbeda dari N=5 sampai N=475 hampir sekitar 47.0%. Jika dibanding kinerja fitur SIFT dengan BOVW yang hanya 5.2%, maka kinerja FaceNet menghitung *similarity* vektornya hampir 10 kali lipat. Artinya fitur FaceNet yang sudah di-pretrained dengan ribuan wajah menghasilkan fitur yang cukup unik untuk langsung

digunakan dalam problem identifikasi wajah IFPLD, dibandingkan *hand-crafted* fitur seperti SIFT.

Tabel III.1 Kinerja akurasi identifikasi wajah berbasis similarity pada fitur FaceNet (dalam %)

N	T	
	400	1000
5	47,5	47,7
20	47,5	46,9
40	47,1	46,9
95	47,0	47,2
475	47,0	47,0

Dalam verifikasi wajah dilakukan 2 pendekatan yaitu menggunakan MLP dan Siamese Network dengan format input yang berbeda seperti yang dijelaskan di bagian II.3. Akurasi dari MLP untuk IFPLD+Aug dengan metode pelatihan *2K-balanced* ini sama persis dengan akurasi MLP tanpa *2K-balanced* (tabel III.2), yaitu 91%. Tetapi dari *confusion matrix* yang ditampilkan pada Tabel III.3 terlihat nilai presisi dari kelas 1 meningkat 2% tetapi *recall* menurun dengan persentase yang sama sehingga *f1-score* tidak berubah yaitu 83%. Sehingga secara rata-rata untuk kedua kelas 0 dan 1, rata-rata presisi pada metode *2K-balanced* sedikit lebih tinggi yaitu 92% dibanding 91%. Sementara rata-rata *recall*/sensitivitas tidak berubah yaitu 87%.

Jika dibandingkan klasifikasi fitur vektor dari *FaceNet*, akurasi klasifikasi fitur SIFT sangat mengecewakan, tetapi bisa diprediksi dari hasil *feature matching* sebelumnya di mana sangat sedikit sekali dari 60 *keypoint* deskriptor SIFT pada fitur wajah frontal yang cocok (*match*) dengan 61 *keypoint* deskriptor fitur wajah profil menggunakan teknik *Brute Force Matcher*. Karena memang SIFT didesain

Tabel III.3 Kinerja MLP dengan pelatihan 2K-balanced pada data uji IFPLD+Aug untuk verifikasi wajah

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	Akurasi
0	0.91	0.98	0.94	0.91
1	0.93	0.75	0.83	

*invariant* pada rotasi pada sudut putar (*rolling-angle*) bukan *invariant* pada rotasi sudut simpang (*yaw-angle*). Keterbatasan SIFT untuk mendeteksi fitur deskriptor wajah profil yang sebagian wajah tertutup (*self-occlusion*), sementara deskriptor itu digunakan untuk proses *matching* dengan wajah frontal dan proses klasifikasi menggunakan BOVW, membuat akurasi model sangat kecil.

FaceNet sebagai fitur ekstraktor dataset IFPLD menghasilkan vektor *embedding* yang unik dan *distingtif*, sehingga bisa langsung digunakan dengan metrik *similarity* seperti *cosine similarity* atau *euclidean distance* untuk problem identifikasi wajah profil pada galeri N buah wajah frontal, dengan hasil yang jauh lebih baik dibandingkan dengan fitur lokal SIFT seperti yang ditunjukkan pada tabel III.1. Akurasi maksimum yang didapat untuk data tes IFPLD adalah 47.7%, dan hampir konstan untuk berbagai kombinasi N. Sementara Schroff dkk. (2015) melaporkan untuk dataset LFW sebagai data uji, FaceNet dengan dimensi vektor 128 menghasilkan akurasi  $99.63\% \pm 0.09$  dengan menggunakan metrik *similarity*. Fakta ini menunjukkan dataset IFPLD adalah dataset yang menantang dan sulit bagi FaceNet *embedding* vektor.

Pelatihan dengan metode *2K-balanced batch* pada model MLP walaupun tidak meningkatkan akurasi dan *f1-score* dibanding pelatihan biasa, tetapi meningkatkan sedikit presisi dari kelas 1. Dari plot *validation vs training loss*, kecenderungan bahwa network belajar dari pola dengan baik di mana *validation loss* mengikuti pola *training loss* dengan *gap* sangat kecil. Ini dikarenakan di setiap batch, terjadi keseimbangan data. Kekurangan metode ini adalah disebabkan jumlah data positif yang terbatas, K buah data positif pada setiap batch selalu sama, walaupun urutannya berbeda dalam setiap *epoch*, sehingga tidak secara signifikan meningkatkan performa dari model MLP.

Eksperimen selanjutnya adalah identifikasi wajah dengan feature *embedding* network. Pada tahap ini dilakukan investigasi untuk mencari model *siamese network* yang representatif untuk dapat melakukan verifikasi wajah dengan akurasi yang tinggi untuk setingan data pengujian berasal dari 95 kelas test IFPLD. Dari

penyelidikan model verifikasi wajah ini kemudian menjadi basis bagi identifikasi wajah pada Siamese Network.

Tabel III.4 Akurasi *Siamese Network* untuk Verifikasi Wajah dengan dataset uji IFPLD

Model Siamese Network	Akurasi
Koch	0.79
FaceNet VGG2	0.88
ResNet-18 IN	0.90
ResNet-18 IN CFP1	0.89
ResNet-18 IN CFP2	0.84
ResNet-18 IN CFP3	0.92
ResNet-18 IN CFP1 + Contrastive Loss	0.86
ResNet-18 IN CFP2 + Contrastive Loss	0.85
ResNet-18 IN CFP3 + Contrastive Loss	0.82
ResNet-18 IN CFP1 + Triplet Loss	0.90
ResNet-18 IN CFP2 + Triplet Loss	0.84
ResNet-18 IN CFP3 + Triplet Loss	0.92
ResNet-18 IN CFP1 + SimCLR	<b>0.93</b>
ResNet-18 IN CFP2 + SimCLR	0.68
ResNet-18 IN CFP3 + SimCLR	0.91

Jika membanding 6 model pertama dalam tabel III.4 yaitu ketiga model *baseline* (Koch, FaceNet\_VGG2, ResNet-18\_IN) dan ketiga model ResNet-18\_IN\_CFP\*, terlihat akurasi ResNet-18\_IN\_CFP3 (fine-tune dengan *arcFace loss*) tertinggi dengan 92% untuk dataset pengujian IFPLD sesuai dengan deskripsi di bab II. Ada 2% peningkatan akurasi dibanding model *baseline* ResNet-18\_IN yang menempati urutan kedua dan hampir 13% lebih dari model *baseline* lain yaitu model Koch. Ini berarti untuk ke-6 model tersebut ResNet-18 yang dilakukan *fine-tuning* dgn CFP menggunakan *loss function arcface* menghasilkan *feature embedding* terbaik untuk melakukan verifikasi wajah.

Penggunaan *loss function arcface* dibanding *cross-entropy* pada *fine-tuning* dataset wajah IFPLD memberikan peningkatan akurasi 3% dan *f1-score* 4%. ArcFace menambahkan margin minimum  $m$  pada ruang sudut untuk vektor *embedding* yang terbentuk sehingga batas antar vektor pada individu berbeda lebih jelas dibanding menggunakan *cross-entropy*. Pemisahan kluster kelas dengan *arcface loss* pada

ruang vektor menghasilkan model yang memiliki akurasi lebih baik untuk problem klasifikasi termasuk verifikasi wajah.

Tabel III.5 Akurasi dari evaluasi *One-shot learning* pada *Siamese Network* pada dataset uji IFPLD

Model Siamese Network	T = 400			T = 1000		
	N			N		
	5	20	40	5	20	40
Koch	0.51	0.27	0.17	0.51	0.26	0.18
FaceNet_VGG2	0.75	0.43	0.28	0.72	0.42	0.26
ResNet-18_IN	0.81	0.50	0.39	0.77	0.51	0.39
ResNet-18_IN_CFP1	0.71	0.45	0.33	0.68	0.47	0.29
ResNet-18_IN_CFP2	0.49	0.23	0.13	0.47	0.23	0.17
ResNet-18_IN_CFP3	0.63	0.43	0.3	0.67	0.42	0.31
ResNet-18_IN_CFP1 + Contrastive Loss	0.63	0.38	0.26	0.67	0.37	0.24
ResNet-18_IN_CFP2 + Contrastive Loss	0.68	0.4	0.23	0.73	0.39	0.31
ResNet-18_IN_CFP3 + Contrastive Loss	0.70	0.33	0.20	0.66	0.34	0.20
ResNet-18_IN_CFP1 + Triplet Loss	0.81	0.6	<b>0.54</b>	0.82	0.63	0.53
ResNet-18_IN_CFP2 + Triplet Loss	0.67	0.5	0.46	0.67	0.49	0.37
ResNet-18_IN_CFP3 + Triplet Loss	<b>0.89</b>	<b>0.72</b>	0.52	<b>0.89</b>	<b>0.69</b>	<b>0.58</b>
ResNet-18_IN_CFP1 + SimCLR	0.83	0.6	0.44	0.84	0.59	0.47
ResNet-18_IN_CFP2 + SimCLR	0.77	0.48	0.36	0.77	0.50	0.38
ResNet-18_IN_CFP3 + SimCLR	0.83	0.5	0.34	0.80	0.51	0.37

Kinerja *siamese network* berbasis FaceNet\_VGG2 walaupun lebih baik dari beberapa model lain tetapi bukan yang terbaik. Model berbasis ResNet-18\_IN\_CFP1 + SimCLR yang memberikan akurasi terbaik adalah 5% lebih tinggi dari FaceNet\_VGG2. Karakteristik unik dari dataset IFPLD yang terbatas dari sisi jumlah dan tipe pose serta sebagian individu memiliki asesori di wajah, menjadi tantangan bagi FaceNet\_VGG2 karena dilatih dengan dataset VGGFace2 yang lebih umum. Fine-tuning pada FaceNet\_VGG2 dengan membekukan lapisan depan tidak memberikan akurasi yang lebih baik dibanding ResNet-18\_IN\_CFP\* karena dataset VGG2 berbeda karakteristik dengan IFPLD. Tidak membekukan semua



lapisan FaceNet\_VGG2 memberikan hasil terbaik tetapi kinerja akurasi masih di bawah ResNet-18\_IN\_CFP1 + SimCLR.

Jika membandingkan semua kombinasi model yang diinvestigasi, *siamese network* dengan model dasar ResNet-18\_IN\_CFP3 tidak lebih baik dari ResNet-18\_IN\_CFP1 + SimCLR yang secara rata-rata sedikit lebih baik dari sisi akurasi maupun dari sisi *f1-score* kecuali ketika dilatih menggunakan *contrastive loss*.

Tabel III.6 Akurasi rata-rata *Prototypical Network* dengan Model Standar FSL dengan skenario training tertentu

SKENARIO PELATIHAN	PENGUJIAN 95 KELAS TES						
	N <sub>U</sub>	N <sub>US</sub>	N <sub>UQ</sub>	Akurasi Metode 1	Waktu (ms)	Akurasi Metode 2	Waktu (ms)
Episode = 400, N <sub>PS</sub> = 6, N <sub>PQ</sub> = 6, N <sub>P</sub> = 20	5	1	1	<b>0,833</b>	19,3	0,723	16,0
	5	2	2	0,818	24,8	0,757	16,5
	5	3	3	0,808	27,4	0,780	17,8
	5	4	4	0,818	30,7	0,834	19,8
	5	5	5	0,790	32,8	0,856	21,7
	5	6	6	0,792	34,6	<b>0,862</b>	23,4
	20	1	1	<b>0,602</b>	220,3	0,481	122,0
	20	2	2	0,549	181,8	0,512	147,5
	20	3	3	0,530	393,7	0,523	232,6
	20	4	4	0,553	381,7	0,601	271,0
	20	5	5	0,510	380,2	<b>0,663</b>	278,6
	20	6	6	0,517	403,2	<b>0,663</b>	281,7

*Siamese Network* umumnya digunakan selain untuk verifikasi wajah juga bisa digunakan untuk identifikasi wajah dengan menggunakan metode evaluasi N-way-1-shot learning, di mana sebuah wajah profil yang menjadi data *query* akan mencari wajah frontal yang secara probabilitas kemiripannya terbesar dari *N* identitas.

Dalam eksperimen, *N* buah kelas yang dipilih, di mana *N* bisa bernilai 5, 20 dan 40, masing-masing diambil wajah frontalnya sebagai *support* data. Kemudian dipilih 1 kelas dari *N*, katakan kelas *d*, diambil wajah profilnya sebagai data *query*. Wajah

profil ini kemudian dibandingkan dengan  $N$  buah wajah frontal dengan menggunakan *siamese network* di mana probabilitas terbesar menunjukkan kelas dari wajah profil tersebut.

Dari tabel terlihat bahwa nilai maksimum akurasi *prototypical network* dengan skenario pelatihan *20-way-6-shot learning* adalah untuk  $N_U=5$  adalah 97.4 % (saat  $N_{US}=6, N_{UQ}=3$ ), untuk  $N_U=20$  adalah 89% (Saat  $N_{US}=6, N_{UQ}=1$ ) dan untuk  $N_U=40$  adalah 82.2% (saat  $N_{US}=6, N_{UQ}=4$ ).

Semakin besar jumlah  $N_U$  atau jumlah subyek dengan wajah frontal yang akan dibandingkan dengan wajah profil, maka semakin kecil akurasinya. Secara teori, probabilitas cocoknya identitas dari seseorang yang memiliki 1 wajah profil jika

Tabel III.7 Akurasi rata-rata *Prototypical Network* dengan network dasar ResNet-18\_IN\_SP3 + SimCLR dengan skenario pelatihan tertentu

SKENARIO PELATIHAN	PENGUJIAN 95 KELAS TES						
	$N_U$	$N_{US}$	$N_{UQ}$	Akurasi Metode 1	Waktu (ms)	Akurasi Metode 2	Waktu (ms)
Episode = 400, $N_{PS} = 6,$ $N_{PQ} = 6,$ $N_P = 20$	5	1	1	0,972	15,7	0,955	16,0
	5	2	2	0,971	17,9	0,948	17,1
	5	3	3	<b>0,974</b>	18,9	0,953	19,2
	5	4	4	0,972	20,3	0,953	20,2
	5	5	5	0,970	23,2	0,952	22,0
	5	6	6	0,971	24,0	<b>0,956</b>	23,9
	20	1	1	<b>0,890</b>	79,0	0,833	73,7
	20	2	2	0,879	68,3	0,817	73,5
	20	3	3	0,887	79,6	0,823	79,2
	20	4	4	0,885	84,1	0,825	83,8
	20	5	5	0,880	87,9	0,832	87,8
	20	6	6	0,883	210,1	<b>0,838</b>	224,7
	40	1	1	<b>0,822</b>	547,3	0,772	168,4
	40	2	2	0,801	586,0	0,781	173,6
	40	3	3	0,816	585,1	0,794	175,7
	40	4	4	0,820	1116,7	0,802	199,2
	40	5	5	0,810	1943,3	<b>0,809</b>	1686,7
	40	6	6	0,815	2423,3	0,807	3010,0

dicocokkan dengan  $N$  buah wajah frontal adalah  $1/N$ . Artinya semakin besar  $N$  semakin kecil probabilitasnya. Tetapi oleh pelatihan *prototypical network* probabilitas ini menjadi jauh lebih besar dari yang dihitung secara teori.

Jack Snell (2017) menyebutkan melatih *prototypical network* dengan  $N_p$  yang lebih besar memberikan keuntungan dengan kinerja yang relatif lebih baik. Ini juga terkonfirmasi dari data eksperimen bahwa model  $N_p=20$  lebih baik dibanding  $N_p=5$ . Tetapi keuntungan tidak berdampak dengan meningkatkan nilai  $N_p$  lebih lanjut yaitu 40 yang ternyata tidak lebih baik dari  $N_p=20$ .

## 5. Kesimpulan dan Saran

Dalam eksperimen ditunjukkan bahwa dengan menyatakan problem identifikasi wajah sebagai problem *few-shot face identification* dan kemudian sistem dilatih dengan pendekatan *meta-learning*, telah menghasilkan model dengan kinerja terbaik. Model *prototypical network* yang dilatih untuk mengenali wajah dari  $N$  individu yang masing-masing memiliki 1 buah sampel wajah (*one-shot learning*), diubah menjadi problem *few-shot learning*, dengan masing-masing terdiri atas  $k$  buah sampel, di mana  $k-1$  berasal dari data augmentasi.

Dari perbandingan antara beberapa model CNN yang diteliti, model Koch, Model Standar FSL dan ResNet\_18\_IN\*, terlihat bahwa peran *Transfer Learning* sangat penting dalam meningkatkan kemampuan ekstraksi fitur CNN karena model CNN yang sudah diinisialisasi dengan bobot pengalaman sebelumnya, mampu menghasilkan fitur yang lebih baik untuk dataset problem dibandingkan dengan model yang belum ‘belajar’ sama sekali. Ini sesuai dengan salah satu premis yang digunakan dalam penelitian ini bahwa ”*Pembelajaran mendalam (deep learning) mampu belajar representasi fitur dengan baik pada sembarang pose wajah dengan menggunakan konsep transfer learning*”.

Dalam eksperimen, model *prototypical network* menghasilkan kinerja akurasi terbaik untuk identifikasi wajah pada *IFPLD* adalah dengan menggunakan metode 1 yang menghitung jarak *prototypes* ke *query set* secara langsung dengan majority

voting. Pelatihan model diubah dari pelatihan *N-way-1-shot* menjadi *N-way-6-shot* learning di mana 5 data diperoleh melalui proses augmentasi data secara *offline* saat pelatihan atau *online* saat pada pengujian.

Hasil dari eksperimen yang dilakukan menggunakan *siamese network* dan *prototypical network* dalam menyelesaikan identifikasi wajah pada IFPLD sebagai problem few-shot learning dengan memanfaatkan model CNN yang telah dilakukan fine-tuning membuktikan kebenaran hipotesis dari penelitian.

Ada beberapa saran untuk penelitian ke depan dari topik yang sudah diteliti dalam disertasi ini antara lain:

1. Keterbatasan dataset *IFPLD* selain dari jumlah dan tipe pose, juga ada bias dari dataset tersebut yaitu waktu pengambilan gambar wajah frontal dan wajah profil pada momen yang sama. Perlu pengembangan lebih lanjut dari dataset ini dengan misalnya menggabungkannya dengan sumber lain yang memenuhi kriteria yang sama yaitu 1 wajah frontal dan 1 wajah profil.
2. Penelitian ini tidak membahas skenario di mana subyek *query* data tidak terdapat pada database atau *support set* dalam skenario *N-way-1-shot* learning. Syarat penuh dari *open-set face recognition* adalah terdapat sampel data yang tidak termasuk dalam *N* kelas yang ada dan model mampu melakukan penolakan dari data tersebut ke dalam kelas tertentu misalnya kelas ‘subyek tidak dikenal’.
3. Penggunaan data augmentasi terbukti sangat penting dalam meningkatkan kinerja model dengan keterbatasan dataset *IFPLD* bahkan di kasus *prototypical network* sangat krusial. Penelitian kedepan adalah perlunya teknik data augmentasi secara *online* dengan metode generatif yang lebih kompleks seperti GAN, *Diffusion Model* yang saat ini sangat berkembang untuk menghasilkan gambar dan contoh yang ada.

## Riwayat Hidup

Promovendus lahir di Pancor, Lombok Timur 15 November 1973, sebagai anak ke 4 dari 6 bersaudara, dari pasangan alm. H.M Djalaluddin SH dan Hj Siti Rauhun Zainuddin SAg. Promovendus menyelesaikan studi S1 (BE) di Department Electrical and Computer System Monash University (1998), S2 (MKom) dari Prodi Teknik Informatika Institut Teknologi Sepuluh November Surabaya (2007). Promovendus merupakan suami dari Hj Dukha Yunitasari SH MPd, dan ayah dari tiga anak yaitu Fardan Ramadana Kalman (15), Zaky Akhtar Kalman (12) dan Nadia Khairatun Hisan (6).

Promovendus selama kurang lebih 18 tahun bekerja di bagian IT di beberapa perusahaan dan sejak tahun 2018 fokus sebagai dosen dan dekan Fakultas Teknik Universitas Hamzanwadi Selong.

Daftar Publikasi yang berkaitan dengan penelitian :

No	Tahun	Judul	Jenis	Penerbit
1	2023	Open-Set Profile-to-Frontal Face Recognition on a Very Limited Dataset	Jurnal International (Q1)	IEEE Access
2	2023	The Impact of Data Augmentation on Prototypical Networks for Open-Set Face Identification	Proceeding Konferensi IEEE	ICOIACT 2023

1. M. Djamaluddin, R. Munir, N. P. Utama and A. I. Kistijantoro, "Open-Set Profile-to-Frontal Face Recognition on a Very Limited Dataset," in IEEE Access, vol. 11, pp. 65787-65797, 2023, doi: 10.1109/ACCESS.2023.3289923.

## **Ucapan Terima kasih**

Dalam kesempatan ini penulis ingin menyampaikan terima kasih kepada alm. Prof. Dwi Hendratmo Widyantoro MSc (ketua pembimbing awal), Dr. Ir. Rinaldi Munir MT (ketua pembimbing), Nugraha Priya Utama ST. MA. PhD. dan Achmad Imam Kistijantoro ST. MSc. PhD. sebagai para pembimbing yang telah banyak memberikan bimbingan, saran, masukan dan nasehat selama berlangsungnya penelitian disertasi ini baik melalui pertemuan langsung tatap muka, zoom, chat maupun media komunikasi lain.

Ucapan terima kasih penulis sampaikan kepada keluarga besar di lombok yang selalu mendukung selama menempuh studi di Bandung. Terima kasih juga penulis ucapkan kepada rekan-rekan mahasiswa S3 di Lab Residensi lantai 3 STEI yang selama ini menjadi teman berdiskusi, berbagi cerita dan pengalaman. Tidak lupa, ucapan terima kasih dan penghargaan kepada para responden dan fotografer dari NWDI Media Center dalam membantu mempersiapkan dataset *IFPLD* yang sangat diperlukan dalam penelitian ini.

Penulis juga ingin menyampaikan terima kasih kepada Lembaga Pengelola Dana Pendidikan (LPDP) atas Beasiswa Unggulan Dosen Indonesia – Dalam Negeri (BUDI-DN) tahun 2018-2022 dan kepada Kementerian Riset Dikti atas hibah Penelitian Disertasi Doktor (PDD) tahun 2023. Beasiswa dan hibah penelitian ini sangat membantu penulis dalam menempuh pendidikan doktoral dan melakukan penelitian disertasi.

Akhirnya, banyak pengalaman dan pembelajaran sangat berharga yang penulis dapatkan dalam menempuh pendidikan doktoral ini dan semoga apa yang penulis lakukan mendapatkan keberkahan Allah SWT. Aamiin.