

## Pengenalan Wajah Resolusi Rendah Menggunakan Arsitektur Lightweight VarGFaceNet dengan Adaptive Margin Loss

Daffa Tama Ramadani <sup>1\*</sup>, Riza Ibnu Adam <sup>2\*</sup>, Jajam Haerul Jaman <sup>3\*</sup>, Chaerur Rozikin <sup>4\*</sup>, Garno <sup>5\*</sup>

\* Informatika, Universitas Singaperbangsa Karawang

[daffa.tama19013@student.unsika.ac.id](mailto:daffa.tama19013@student.unsika.ac.id) <sup>1</sup>, [riza.adam@staff.unsika.ac.id](mailto:riza.adam@staff.unsika.ac.id) <sup>2</sup>, [jajam.haeruljaman@staff.unsika.ac.id](mailto:jajam.haeruljaman@staff.unsika.ac.id) <sup>3</sup>,  
[chaerur.rozikin@staff.unsika.ac.id](mailto:chaerur.rozikin@staff.unsika.ac.id) <sup>4</sup>, [garno@staff.unsika.ac.id](mailto:garno@staff.unsika.ac.id) <sup>5</sup>

### Article Info

#### Article history:

Received 2023-07-07

Revised 2023-07-31

Accepted 2023-07-31

#### Keyword:

Face Recognition,  
Lightweight CNN,  
Low Resolution,  
Super Resolution.

### ABSTRACT

Face recognition is a modern security solution that is quickly and easily integrated into most existing devices, so this system is widely applied to several domains as one of the security authorizations. Developing face recognition models using mainstream architectures (AlexNet, VGGNet, GoogleNet, ResNet, and SENet) will make it difficult to implement the models on mobile devices and embedded systems. In addition, low resolution images, such as those from CCTV surveillance cameras or drones, pose challenges for the models to recognize faces, as the images lack sufficient details for identification. Therefore, this research aims to analyze the performance of a face recognition model developed using the lightweight VarGFaceNet architecture with the adaptive margin loss AdaFace on a low-resolution image dataset. Based on the evaluation results on the LFW dataset, an accuracy of 99.08% was achieved on high-resolution data (112x112 pixels), while on the lowest synthetic low-resolution data (14x14 pixels), an accuracy of 79.87% was obtained with the assistance of the Real-ESRGAN and GFP-GAN super-resolution models. On the TinyFace dataset, without fine-tuning, a Rank-1 accuracy of 46.08% was achieved without using super-resolution models and 45.03% when utilizing super-resolution models.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

### I. PENDAHULUAN

Pengenalan wajah merupakan solusi keamanan modern yang cepat dan mudah diintegrasikan pada kebanyakan *device* yang ada saat ini, sehingga sistem ini banyak diterapkan pada beberapa domain seperti militer, keuangan, keamanan publik dan kehidupan sehari-hari sebagai salah satu otorisasi keamanan [1]. Sistem pengenalan wajah juga dianggap *easy-to-use* karna dalam penerapannya, sistem ini menggunakan wajah sebagai biometrik untuk melakukan identifikasi sehingga pengguna tidak perlu mengingat kode atau sandi. Sistem pengenalan wajah juga sudah tersemat pada hampir setiap *smartphone* yang ada saat ini sebagai salah satu fitur keamanan untuk membuka kunci layar. Selain itu beberapa penerapan dari sistem pengenalan wajah lainnya yaitu sistem presensi kehadiran siswa [2], sistem penguncian rumah [3], serta untuk penerapan yang lebih *advance* yaitu sistem pengawasan buronan Interpol I-24/7 milik Divhubinter

Polri yang terdapat pada 37 bandara dan 90 pelabuhan internasional di Indonesia [4].

Pengembangan model pengenalan wajah menggunakan arsitektur CNN *mainstream* (AlexNet, VGGNet, GoogleNet, ResNet, dan SENet) dapat menyebabkan model pengenalan wajah sulit diimplementasikan pada perangkat *mobile* dan *embedded system*, hal ini dikarenakan arsitektur CNN *mainstream* memiliki jumlah parameter yang banyak serta komputasi yang kompleks. Disisi lain, arsitektur CNN *lightweight* seperti VarGFaceNet [5] menggunakan *variable group convolution* [6] untuk mengurangi kompleksitas serta mempertahankan jumlah parameter agar tetap sedikit meskipun memiliki layer yang cukup dalam, sehingga dapat berjalan dengan lebih ringan.

Selain dari masalah penggunaan arsitektur, gambar dengan resolusi yang rendah seperti pada *footage* kamera pengawas CCTV ataupun *drone* menyebabkan model kesulitan untuk mengenali wajah karena tidak punya cukup petunjuk untuk

dikenali. Beberapa metode telah dilakukan seperti menggunakan *super resolution* [7], melakukan *fine tune* pada data gambar resolusi rendah [8], serta penggunaan fungsi loss khusus seperti *adaptive margin loss* AdaFace [9].

Penelitian serupa [8] telah dilakukan sebelumnya. Penelitian ini membandingkan performa 2 model CNN *lightweight*, MobileFaceNet [10] dan ShuffleFaceNet [11] pada dataset resolusi rendah asli (SCFace [12], TinyFace [7], dan QMUL-SurvFace [13]) dan dataset resolusi rendah sintetis (LFW [14]). Pada penelitian ini juga dilakukan *fine tune* model pengenalan wajah pada dataset resolusi rendah sintetis yang dibuat menggunakan dataset CASIA WebFace [15] dengan metode *bicubic* dan *area interpolation*. Hasil model yang telah di-*fine tune* menunjukkan peningkatan akurasi yang signifikan, namun penggunaan metode ini membuat proses pengembangan model jauh lebih lama, karena model harus di latih 2 kali.

Penelitian lain [7] menggunakan model *super resolution* untuk mengatasi tantangan pengenalan wajah pada resolusi rendah. Namun penggunaan model *super resolution* menunjukkan hasil yang tidak stabil, karena terkadang penggunaan model *super resolution* malah menambahkan *noise* baru pada gambar yang membuat akurasi model pengenalan wajah jauh lebih buruk, tetapi di eksperimen lainnya, penggunaan model *super resolution* bisa meningkatkan akurasi dari model pengenalan wajah.

Disisi lain penggunaan *adaptive margin loss* AdaFace [9] menunjukkan performa paling baik pada dataset resolusi rendah di antara 2 penelitian sebelumnya. Dengan penggunaan AdaFace, pelatihan model akan difokuskan sesuai dengan kualitas dari gambar. Saat kualitas gambar rendah maka pelatihan akan difokuskan pada sampel data yang mudah dan juga sebaliknya, hal ini menyebabkan model bisa beradaptasi pada gambar dengan kualitas rendah dan menghindari *overfit* pada gambar dengan kualitas tinggi.

Oleh karena itu untuk menjawab tantangan pengenalan wajah pada resolusi rendah, penelitian ini akan menganalisis performa dari model pengenalan wajah yang dikembangkan dengan arsitektur *lightweight* VarGFaceNet [5] dengan fungsi *adaptive margin loss* AdaFace [9] pada data gambar resolusi rendah. Kriteria dari gambar resolusi rendah yang digunakan pada penelitian ini ialah gambar yang memiliki dimensi  $\leq 32 \times 32$  piksel sesuai dengan *tiny object criteria* [16]. Pada penelitian ini juga akan digunakan model *super resolution* Real-ESRGAN [17] dan model *face restoration* GFP-GAN [18] untuk menghilangkan degradasi pada gambar resolusi rendah serta merekonstruksi pose dan ekspresi dari wajah.

## II. METODE

Pada bagian ini akan diuraikan terkait arsitektur VarGFaceNet, *adaptive margin loss* AdaFace, dataset yang akan digunakan, penggunaan model *super resolution* pada dataset evaluasi, detail implementasi, serta protokol evaluasi yang akan digunakan untuk mengevaluasi model pengenalan wajah.

### A. VarGFaceNet

Untuk membuat model yang ringan, maka pada penelitian ini akan digunakan arsitektur VarGFaceNet [5]. VarGFaceNet atau *Variable Group Face Network* merupakan sebuah arsitektur *Convolutional Neural Network* (CNN) ringan yang dikembangkan untuk melatih model pengenalan wajah. Arsitektur ini dikembangkan berdasarkan arsitektur VarGNet [6] dengan beberapa perubahan untuk menyesuakannya dengan tugas pengenalan wajah. Dengan menggunakan VarGNet sebagai basis arsitekturnya, arsitektur ini dapat menyelesaikan masalah optimasi pada *embedded system* yang terjadi pada arsitektur ShuffleNet, dan ShuffleNetV2 yang juga merupakan arsitektur *light-weight*. Selain itu arsitektur VarGFaceNet memiliki performa akurasi yang lebih unggul dari arsitektur ShuffleFaceNet yang dikembangkan berdasarkan arsitektur ShuffleNetV2 serta dapat berjalan secara *realtime* pada *low power* GPU karena memiliki *memory footprint* yang lebih kecil dibandingkan dengan ShuffleFaceNet [19], [20]. VarGFaceNet menerima masukan dengan *shape*  $112 \times 112 \times 3$  yang berarti gambar wajah RGB dengan dimensi  $112 \times 112$  piksel dan menghasilkan *embedding* wajah dengan *shape* 512.

### B. AdaFace

Untuk membuat model yang dapat beradaptasi pada gambar wajah resolusi rendah, maka pada penelitian ini akan digunakan fungsi *adaptive margin loss* AdaFace [9] AdaFace merupakan fungsi *adaptive margin loss* yang dikembangkan untuk pembuatan model pengenalan wajah. Pada jurnal penelitian AdaFace, Kim dan rekan-rekannya mengelompokkan gambar wajah menjadi 3 jenis, yaitu: gambar yang mudah dikenali, gambar yang sulit dikenali, dan gambar yang tidak mungkin dikenali, dimana gambar dengan resolusi rendah masuk ke dalam kategori gambar wajah yang sulit dikenali dan bahkan tidak mungkin dikenali. Pelatihan model terhadap gambar wajah yang tidak mungkin dikenali dapat menyebabkan performa model menjadi buruk. Untuk mengatasi masalah tersebut, AdaFace akan mengidentifikasi kualitas gambar terlebih dahulu untuk menentukan fokus pelatihan. Saat kualitas gambar rendah maka pelatihan akan difokuskan pada sampel data yang mudah dan juga sebaliknya, hal ini menyebabkan model dapat mempelajari gambar dengan kualitas rendah tanpa terpengaruh oleh gambar wajah yang tidak mungkin dikenali dan menghindari *overfit* pada gambar dengan kualitas tinggi.

### C. Dataset

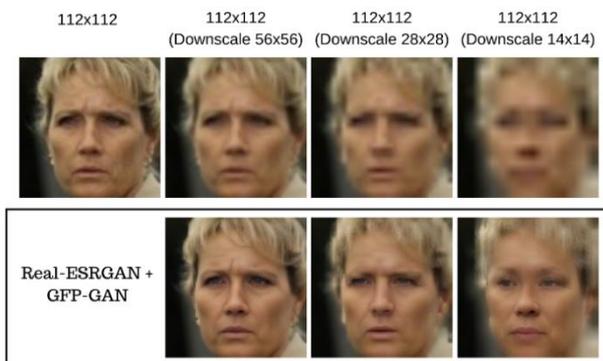
Terdapat dua dataset yakni dataset pelatihan dan evaluasi. Pada penelitian ini, dataset yang akan digunakan dalam pelatihan model adalah CASIA WebFace [15]. Dataset ini berisi kurang lebih 494.414 gambar dengan 10.575 subjek (label) dengan dimensi gambar yaitu  $112 \times 112$  piksel. Pada dataset CASIA, setiap wajah yang berasal dari identitas yang sama diletakkan dalam satu folder dan folder-folder ini dilabeli dengan nomor sehingga tidak ada identitas nama orang pada dataset ini. Dataset CASIA WebFace diambil dari repositori

InsightFace [21] dimana dataset ini merupakan dataset wajah yang telah diluruskan (*aligned*).

Terdapat 2 dataset evaluasi yang akan digunakan dalam penelitian ini. Dataset pertama yaitu LFW (Labeled Faces in the Wild) [14], Dataset ini memiliki dimensi 112x112 piksel dan terdiri dari 13.233 gambar wajah yang merupakan wajah dari 5.749 identitas berbeda yang dimana didalamnya terdapat 1.680 identitas yang memiliki gambar wajah 2 ataupun lebih. Dataset ini nantinya akan di *downscale* dengan metode *bicubic* ke dimensi 56x56, 28x28, dan 14x14 untuk membuat data resolusi rendah sintetis. Dataset LFW juga diambil dari repositori InsightFace [21] dimana dataset ini merupakan dataset wajah yang telah diluruskan (*aligned*). Dataset evaluasi kedua yang akan digunakan dalam penelitian ini yaitu TinyFace [7]. Dataset TinyFace diambil dari repositori AdaFace [9], dimana dataset ini berisi data wajah yang telah diluruskan (*aligned*) menggunakan model *face detection* MTCNN. Dataset ini berisi 5.139 identitas (label) dengan total 169.403 gambar wajah yang memiliki resolusi rendah dengan rata-rata dimensi yaitu 20x16 piksel.

#### D. Super Resolution (SR)

Sebagai bentuk eksperimen, akan digunakan juga model *super resolution* Real-ESRGAN [17] dan model *face restoration* GFP-GAN [18] untuk menghilangkan degradasi pada gambar resolusi rendah serta merestorasi pose dan ekspresi dari wajah. Kedua model ini akan digunakan untuk pra-proses data evaluasi lebih lanjut sehingga nantinya model akan diujikan dengan data evaluasi asli dan data evaluasi yang sudah di pra-proses menggunakan model Real-ESRGAN dan GFP-GAN. Sebelum di pra-proses menggunakan model Real-ESRGAN dan GFP-GAN, data asli akan di *upscale* ke dimensi 112x112 terlebih dahulu lalu selanjutnya digunakan model Real-ESRGAN ukuran S dan model GFP-GAN v1.3 dengan konfigurasi *outscale* 1 dan *denoise* 0.5.



Gambar 1. Gambar diatas merupakan hasil implementasi SR pada dataset LFW. Dataset asli berdimensi 112x112 diambil dari repositori InsightFace (<https://github.com/deepsight/insightface>)



Gambar 2. Gambar diatas merupakan hasil implementasi SR pada dataset TinyFace. Dataset berdimensi 112x112 yang sudah *aligned* diambil dari repositori AdaFace (<https://github.com/mk-minchul/AdaFace>)

Hasil implementasi SR pada dataset LFW bisa dilihat pada gambar 1, sedangkan gambar 2 menunjukkan hasil implementasi SR pada dataset TinyFace.

#### E. Detail Implementasi

Model akan dilatih menggunakan arsitektur *lightweight* VarGFaceNet dengan *adaptive margin loss* AdaFace dengan konfigurasi *batch size* 512, *save frequency* 1, *test frequency* 2, dan *total epoch* 26. Dataset CASIA WebFace berdimensi 112x112 digunakan sebagai data latih sedangkan LFW berdimensi 112x112 digunakan sebagai data evaluasi. Dalam pelatihan, digunakan *optimizer* Stochastic Gradient Descent (SGD) dengan *weight decay*  $5e-4$ , *learning rate* 0.1, dan *momentum* 0.9 dimana *learning rate* akan dibagi 10 pada *epochs* 10, 18, dan 22. Sedangkan untuk konfigurasi parameter AdaFace, digunakan *margin* 0.4, *hardness* 0.333, *scale* 64.0, dan *t\_alpha* 0.01. Selain itu juga dilakukan proses augmentasi pada *dataloader* untuk membuat data latih resolusi rendah sintetis yang meliputi:

- *Cropping*, yaitu dengan memotong gambar dengan bentuk persegi secara acak dan membiarkan area diluar persegi menjadi warna hitam layaknya hasil proses *align* pada TinyFace.
- *Photometric jittering*, yaitu mengubah nilai hue, kehatangan, dan kecerahan gambar secara acak.
- *Rescaling*, yaitu proses degradasi gambar dengan menurunkan dimensinya secara acak dan mengembalikannya ke dimensi semula.

Semua proses augmentasi diatas diatur dengan probabilitas 0.2. Adapun pelatihan model akan dilakukan menggunakan GPU Nvidia Tesla A100 dengan 40 GB RAM

#### F. Protokol Evaluasi

*Face verification* merupakan proses yang dilakukan untuk membandingkan kedua wajah dan menyimpulkan apakah kedua wajah tersebut adalah orang yang sama [22]. Protokol *face verification* akan digunakan untuk mengevaluasi model pada dataset LFW. Dalam protokol *face verification*, evaluasi model pengenalan wajah dilakukan dengan membagi dataset

uji menjadi 10 *folds* dengan jumlah yang sama, pada tiap *folds* akan terdapat pasangan wajah positif dan negatif dengan perbandingan 1:1 tanpa ada identitas yang berulang. Metrik yang digunakan untuk melakukan evaluasi adalah *mean accuracy* yang didapatkan dari hasil akurasi model pada tiap *folds* [23].

*Face identification* merupakan proses yang dilakukan untuk membandingkan wajah yang terdeteksi atau yang diujikan dengan wajah pada *gallery* untuk mengidentifikasi identitas dari wajah tersebut [22] [23]. *Face identification* atau yang biasa disebut *probe-gallery identification testing* memiliki 2 protokol pengujian yang berbeda yaitu *open-set* dan *closed-set*. Perbedaan antara kedua protokol tersebut ada pada wajah yang terdaftar pada *gallery* yang dimana pada *open-set* terdapat wajah yang tidak dikenali atau tidak terdaftar pada *gallery* [24]. Pada *probe-gallery identification testing* khususnya *closed-set*, performa model ditentukan berdasarkan akurasi identifikasi pada *Rank-1*, dimana pada *Rank-1* akan diambil wajah dari *gallery* yang memiliki jarak kemiripan terdekat dengan wajah yang diujikan [25]. Protokol *closed-set face identification* akan digunakan untuk mengevaluasi model pada dataset TinyFace.

### III. HASIL DAN PEMBAHASAN

Pada bagian ini akan diuraikan hasil dari pelatihan model, hasil *benchmark* efisiensi komputasi dari model yang telah dibuat, serta hasil evaluasi model pada dataset evaluasi resolusi rendah.

#### A. Hasil Pelatihan Model

TABEL I  
LOG HASIL PELATIHAN MODEL

Waktu	Epoch	Loss	Mean Accuracy
20230605-04:28:19	1	35.28%	-
20230605-04:34:16	2	29.55%	93.13%
20230605-04:40:54	3	21.92%	-
20230605-04:46:51	4	17.97%	97.00%
20230605-04:53:29	5	15.89%	-
20230605-04:59:26	6	14.57%	98.08%
20230605-05:06:03	7	13.65%	-
20230605-05:12:01	8	12.95%	97.81%
20230605-05:18:39	9	12.44%	-
20230605-05:24:36	10	8.58%	98.90%
20230605-05:31:14	11	7.32%	-
20230605-05:37:11	12	6.74%	99.01%
20230605-05:43:49	13	6.36%	-
20230605-05:49:46	14	6.10%	99.10%
20230605-05:56:24	15	5.90%	-
20230605-06:02:22	16	5.74%	98.81%
20230605-06:09:00	17	5.62%	-
20230605-06:14:58	18	4.62%	99.05%
20230605-06:21:36	19	4.38%	-
20230605-06:27:33	20	4.26%	99.06%
20230605-06:34:10	21	4.16%	-
20230605-06:40:08	22	3.99%	99.08%
20230605-06:46:46	23	3.97%	-

20230605-06:52:43	24	3.96%	99.03%
20230605-06:59:23	25	3.95%	-
20230605-07:05:20	26	3.94%	98.96%

Tabel I menunjukkan log pelatihan model yang telah ditransformasikan menjadi sebuah tabel. Bisa dilihat bahwa model dapat menyentuh *mean accuracy* 99% pada *epoch* 12 dan tidak ada perubahan yang signifikan di *epoch-epoch* selanjutnya, hal ini menandakan bahwa akurasi model sudah mencapai titik optimal atau biasa disebut dengan konvergen, sedangkan *loss* baru menyentuh titik konvergen pada *epoch* 22. *Mean accuracy* pada *epoch* terakhir atau *epoch* 26 juga mengalami penurunan sebesar 0.07% (99.03-98.96) dibanding dengan *epoch* 24, oleh karena itu dilakukan evaluasi model lebih lanjut untuk mendapatkan model terbaik. Evaluasi dilakukan dengan menguji ulang model yang sudah tersimpan dengan menggunakan data uji LFW 112x112 pada tugas *face verification* untuk mendapatkan *mean accuracy*. Pengujian dilakukan dari model terakhir dan didapat hasil sebagai berikut.

TABEL II  
HASIL EVALUASI UNTUK PEMILIHAN MODEL TERBAIK

Epoch	Loss	Mean Accuracy	Selisih Loss (epoch 26)	Selisih Mean Accuracy (epoch 26)
26	3.94%	98.96%	0%	0%
25	3.95%	98.98%	0.01%	0.03%
24	3.96%	99.03%	0.02%	0.07%
23	3.97%	99.02%	0.03%	0.06%
<b>22</b>	<b>3.99%</b>	<b>99.08%</b>	<b>0.05%</b>	<b>0.12%</b>
21	4.16%	99.13%	0.22%	0.17%

Berdasarkan hasil pengujian model dari *epoch* 26, 25, 24, 23, 22, dan 21 yang disajikan pada tabel II. Model 22 dipilih sebagai model terbaik karena memiliki selisih *mean accuracy* yang cukup terjauh terhadap *mean accuracy* dari model 26 dengan selisih *loss* yang tidak terlalu jauh. Oleh karena itu, maka proses evaluasi model terhadap dataset resolusi rendah akan dilakukan menggunakan model 22.

#### B. Efisiensi Komputasi

Meskipun sudah diketahui bahwa arsitektur VarGFaceNet merupakan arsitektur yang ringan (*lightweight*) namun pada penelitian ini akan dilakukan evaluasi kembali terkait efisiensi komputasi, penggunaan sumber daya, dan waktu inferensi dari arsitektur VarGFaceNet. Untuk melakukan *benchmark* terkait efisiensi komputasi digunakan *library* DeepSpeed [26] yang dikembangkan oleh Microsoft. Sebagai parameter pembanding, akan dihitung *size* model dalam MegaBytes (MB), Floating Point Operations (FLOPs), Multiply-Accumulate Operations (MACs), dan jumlah parameter. Berikut adalah hasil *benchmark* menggunakan

TABEL V  
HASIL EVALUASI MODEL PADA DATASET LFW

Model	Super Resolution	112x112	Downscale 56x56 (112x112)	Downscale 28x28 (112x112)	Downscale 14x14 (112x112)
		VarGFaceNet+AdaFace (CASIA WebFace 0.5M)	Tanpa SR	99.08%	<b>98.92%</b>
	Real-ESRGAN+GPF- GAN	-	98.62%	95.60%	<b>79.87%</b>

library DeepSpeed dan perbandingannya dengan arsitektur ShuffleFaceNet [11].

TABEL III  
HASIL BENCHMARK EFISIENSI KOMPUTASI MODEL

Model	Parameter	FLOPs	MACs	Size
Shuffle FaceNet [11]	<b>2.66 M</b>	<b>584.59 MFLOPs</b>	<b>282.88 MMACs</b>	<b>10.4 MB</b>
VarGFace Net *	6.46 M	1071 MFLOPs	512.59 MMACs	25.6 MB

Berdasarkan hasil *benchmark* pada tabel III, efisiensi komputasi pada arsitektur VarGFaceNet masih dibawah jika dibandingkan dengan arsitektur ShuffleFaceNet, hal ini dapat dilihat pada keseluruhan variabel yang menyatakan bahwa ShuffleFaceNet lebih ringan, namun perlu diingat juga bahwa fitur wajah yang dihasilkan ShuffleFaceNet adalah 128 atau 1/4 dari fitur VarGFaceNet yang berjumlah 512. Meskipun begitu, VarGFaceNet tetap dikategorikan sebagai arsitektur ringan (*lightweight*) jika dibandingkan dengan arsitektur CNN *mainstream*.

Selain itu pengukuran waktu inferensi, dilakukan pada CPU Intel(R) Xeon(R) @ 2.30GHz dan GPU Nvidia Tesla T4 12GB. Proses perhitungan dilakukan dengan menghitung waktu inferensi sebanyak 300 kali, kemudian dihitung rata-ratanya, namun sebelum itu dilakukan proses *warm-up* GPU dan CPU dengan menjalankan *testing* sebanyak 10 kali. Pengujian dilakukan dengan menggunakan GPU (T4) dan tanpa GPU (CPU). Berikut adalah hasil dari perhitungan waktu inferensi arsitektur VarGFaceNet dan perbandingannya dengan arsitektur ShuffleFaceNet [11].

TABEL IV  
HASIL PERHITUNGAN WAKTU INFERENSI

Model	Tanpa GPU (CPU)	GPU (T4)
ShuffleFaceNet [11]	<b>22.36ms</b>	<b>13.05ms</b>
VarGFaceNet *	45.99ms	17.68ms

Tabel IV Menunjukkan bahwa pada CPU, waktu inferensi VarGFaceNet lebih lambat 23,32ms (45.99-22.36) dibandingkan dengan ShuffleFaceNet, begitu juga pada GPU, ShuffleFaceNet lebih unggul 4.63ms (17.68-13.05). Maka berdasarkan semua variabel *benchmark* yang telah disajikan pada tabel III dan tabel IV, dapat dikatakan bahwa arsitektur ShuffleFaceNet memiliki efisiensi komputasi lebih baik

dibandingkan dengan VarGFaceNet, namun berdasarkan [20] ShuffleFaceNet tidak bisa berjalan secara *realtime* pada *low power* GPU seperti pada GPU Laptop GTX 1050Ti karena memiliki *memory footprint* yang lebih besar.

### C. Evaluasi LFW

Untuk mengevaluasi performa model dalam tugas *face verification* maka model pengenalan wajah akan diujikan pada dataset LFW, khususnya pada data resolusi rendah sintesis yang telah dibuat sebelumnya. Evaluasi akan dilakukan menggunakan protokol *face verification* dengan metrik yaitu *mean accuracy*. Sebelum masuk ke tahap evaluasi, data gambar akan dibagi menjadi pasangan positif dan pasangan negatif. Dimana pasangan positif merupakan pasangan wajah untuk identitas yang sama sedangkan pasangan negatif merupakan pasangan wajah untuk identitas yang berbeda. Dari hasil evaluasi model pada tabel V terlihat bahwa penggunaan model *super resolution* Real-ESRGAN dan GFP-GAN dapat mengurangi performa dari model terkecuali untuk data gambar yang sebelumnya telah di *downscale* ke 14x14 piksel mendapatkan peningkatan akurasi sebesar 8.59% (79.87-71.28).

### D. Evaluasi TinyFace

Untuk mengevaluasi performa model dalam tugas *face identification* maka model pengenalan wajah akan diujikan pada dataset TinyFace. Evaluasi akan dilakukan menggunakan protokol *face identification* dengan *Rank-1* sebagai matriks pengujian. Seperti pada evaluasi LFW, pada TinyFace juga dilakukan 2 skema evaluasi. Untuk skema pertama dataset TinyFace yang digunakan adalah dataset yang tidak melewati proses *super resolution* dengan Real-ESRGAN dan GFP-GAN dan pada skema 2 digunakan data yang telah melewati proses *super resolution*.

TABEL VI  
HASIL EVALUASI MODEL PADA DATASET TINYFACE

Model	Super Resolution	Rank-1	Rank-5
VarGFaceNet+ AdaFace(Casia WebFace 0.5M)	Tanpa SR	46.08%	52.57%
	Real- ESRGAN+GPF- GAN	45.03%	51.63%

Sama seperti pada dataset LFW, hasil evaluasi model pada tabel VI juga menunjukkan bahwa penggunaan model *super*

TABEL VII  
PERBANDINGAN PERFORMA DENGAN MODEL SOTA PADA DATASET TINYFACE

No.	Model	Dataset	Super Resolution	Fine Tune	Rank-1
1	ShuffleFaceNet+ArcFace [8]	MS1M 5.1M	Tanpa SR	Ya	43.10%
2	VarGFaceNet+AdaFace *	CASIA WebFace 0.5M	Tanpa SR	Tidak	<b>46.08%</b>
3	VarGFaceNet+AdaFace *	CASIA WebFace 0.5M	Real-ESRGAN+GPF-GAN	Tidak	45.03%

*resolution* Real-ESRGAN dan GFP-GAN menurunkan performa dari model pengenalan wajah.

Pada tabel VII ditunjukkan performa dari model (1) yang dikembangkan pada penelitian [8] dan model (2, 3) yang bertanda \*, merupakan model yang dikembangkan pada penelitian ini. Berdasarkan akurasi pada *Rank-1*, model yang dikembangkan menggunakan VarGFaceNet dengan *adaptive margin loss* AdaFace dapat mengungguli model yang dikembangkan menggunakan ShuffleFaceNet dengan *margin loss* ArcFace dengan selisih 2.98% (46.08-43.10) tanpa melakukan *fine tune* pada dataset TinyFace serta menggunakan data latihan dengan jumlah yang lebih sedikit.

#### E. Analisis Hasil dan Pembahasan

Dalam penelitian ini telah dilakukan pengembangan model pengenalan wajah ringan menggunakan arsitektur *lightweight* VarGFaceNet dengan *adaptive margin loss* AdaFace untuk pengenalan wajah pada resolusi rendah dengan data latihan CASIA WebFace dengan dimensi 112x112 piksel dan data evaluasi LFW berdimensi 112x112, 56x56, 28x28, 14x14, dan TinyFace.

Sebagai arsitektur *lightweight*, VarGFaceNet memiliki parameter 6.5M dan menggunakan memori sebesar 25MB dengan Float Point Operations 1G untuk mengekstrak 512 fitur wajah dari gambar 112x112 piksel dalam 45.99ms pada CPU dan 17.68ms pada GPU Nvidia Tesla T4. Namun berdasarkan tabel III dan IV, arsitektur ShuffleFaceNet [11] memiliki efisiensi komputasi serta waktu inferensi yang lebih unggul dari VarGFaceNet.

Untuk tugas *face verification* pada dataset LFW, di dapatkan akurasi sebesar 99.08% pada data resolusi tinggi dengan dimensi 112x112 piksel, sedangkan untuk data resolusi rendah sintetis dengan resolusi terendah (14x14) didapatkan akurasi 79.87% dengan batuan model *super resolution* Real-ESRGAN dan GFP-GAN. Semua hasil diatas telah disajikan pada tabel V, dan berdasarkan tabel tersebut juga dapat dilihat bahwa pada data resolusi rendah sintetis beresolusi 56x56 dan 28x28 penggunaan model *super resolution* justru mengurangi akurasi dari model pengenalan wajah yang telah dikembangkan. Ini juga terjadi pada hasil evaluasi model untuk tugas *face identification* pada dataset TinyFace, dimana di dapatkan perbedaan akurasi *Rank-1* sebesar 0.64% (52.27-51.63) antara data yang di proses dengan model *super resolution* dan tidak. Hal ini diasumsikan karena:

- Penggunaan model *super resolution* Real-ESRGAN dan GFP-GAN pada gambar yang telah di *downscale*

ke dimensi 56x56 dan 28x28 menghasilkan gambar yang lebih detail dibandingkan gambar sebelum *super resolution* dan bahkan lebih detail dari gambar asli (112x112) sehingga meningkatkan kompleksitas dari gambar tersebut, serta arah pelatihan model adalah gambar dengan resolusi rendah bukan resolusi tinggi (perbandingan gambar bisa dilihat pada gambar 1).

- Penggunaan model *super resolution* Real-ESRGAN dan GFP-GAN pada gambar yang telah di *downscale* ke dimensi 14x14 menghasilkan gambar wajah dengan fitur yang lebih baik dibandingkan dengan gambar sebelum *super resolution* meskipun terkadang dapat mengubah fitur dari wajah asli, namun karna kompleksitas dari dataset LFW yang cukup rendah maka hal tersebut tidak mempengaruhi akurasi yang didapatkan (perbandingan gambar bisa dilihat pada gambar 1). Sedangkan pada TinyFace yang memiliki kompleksitas lebih tinggi, perubahan fitur pada wajah tersebut malah merugikan dikarenakan hasil yang tidak konsisten pada identitas wajah yang sama seperti contoh pada gambar 3 berikut.



Gambar 3. Contoh di atas adalah hasil dari GFP-GAN yang tidak konsisten pada identitas yang sama sehingga menghasilkan representasi fitur wajah yang berbeda.

Gambar 3 merupakan salah satu contoh dari hasil GFP-GAN pada dataset TinyFace yang menunjukkan hasil yang tidak konsisten pada identitas yang sama sehingga menghasilkan representasi fitur wajah yang berbeda. Hal tersebut dapat memperburuk akurasi dari model pengenalan wajah.

Namun asumsi terkait hasil dari penggunaan model *super resolution* Real-ESRGAN dan GFP-GAN pada evaluasi model dengan dataset LFW belum bisa dikonfirmasi karena tidak adanya pembandingan yang setara, misalnya model yang dilatih dengan fungsi *loss* lain yang tidak berfokus pada gambar resolusi rendah seperti ArcFace [27] dan CosFace [28].

Di luar dari pada itu, berdasarkan hasil evaluasi model yang telah dikembangkan pada penelitian ini. Model pengenalan wajah yang dikembangkan dengan arsitektur *lightweight* VarGFaceNet dengan *adaptive margin loss* AdaFace pada dataset CASIA WebFace belum dapat mengenali gambar wajah resolusi rendah dengan baik. Namun berdasarkan perbandingan dengan model *SoTA* pada tabel VII, model yang dihasilkan dari penggunaan arsitektur *lightweight* VarGFaceNet dengan *adaptive margin loss* AdaFace dapat lebih unggul dibandingkan model pembanding yang juga dikembangkan menggunakan arsitektur *lightweight* tanpa melakukan *fine tune* pada data TinyFace serta menggunakan jumlah data latihan yang lebih sedikit.

#### IV. KESIMPULAN

Dari keseluruhan tahap penelitian yang telah dilakukan, model pengenalan wajah yang telah dikembangkan menggunakan arsitektur *lightweight* VarGFaceNet dengan fungsi *adaptive margin loss* AdaFace dapat mengekstrak 512 fitur wajah dari gambar 112x112 piksel dalam 45.99ms pada CPU Intel(R) Xeon(R) @ 2.30GHz dan 17.68ms pada GPU Nvidia Tesla T4. Dari hasil evaluasi model pada *dataset* LFW, didapatkan akurasi sebesar 99.08% pada data resolusi tinggi (112x112 piksel), sedangkan pada data resolusi rendah sintesis dengan resolusi terendah (14x14 piksel) didapatkan akurasi sebesar 79.87% dengan bantuan model *super resolution* Real-ESRGAN dan GFP-GAN. Pada *dataset* TinyFace, tanpa melakukan *fine tune*, didapatkan akurasi *Rank-1* sebesar 46.08% tanpa menggunakan model *super resolution* dan 45.03% dengan menggunakan model *super resolution*.

#### DAFTAR PUSTAKA

- [1] M. Heidari and K. Fouladi-Ghaleh, "Using Siamese Networks with Transfer Learning for Face Recognition on Small-Samples Datasets," in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, IEEE Computer Society, Feb. 2020. doi: 10.1109/MVIP49855.2020.9116915.
- [2] R. Prathivi and Y. Kurniawati, "Sistem Presensi Kelas Menggunakan Pengenalan Wajah Dengan Metode Haar Cascade Classifier," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 1, pp. 135–142, Apr. 2020, doi: 10.24176/SIMET.V11I1.3754.
- [3] H. Muchtar and R. Apriadi, "Implementasi Pengenalan Wajah Pada Sistem Penguncian Rumah Dengan Metode Template Matching Menggunakan Open Source Computer Vision Library (OpenCV)," *RESISTOR (Elektronika Kendali Telekomunikasi Tenaga Listrik Komputer)*, vol. 2, no. 1, pp. 39–42, May 2019, doi: 10.24853/RESISTOR.2.1.39-42.
- [4] A. M. R., "Polri Kenalkan Sistem Pengawasan Buronan I-24/7 di Markas Interpol di Lyon," *detikNews*, Dec. 04, 2022. <https://news.detik.com/berita/d-6436076/polri-kenalkan-sistem-pengawasan-buronan-i-247-di-markas-interpol-di-lyon> (accessed Apr. 12, 2023).
- [5] M. Yan, M. Zhao, Z. Xu, Q. Zhang, G. Wang, and Z. Su, "VarGFaceNet: An Efficient Variable Group Convolutional Neural Network for Lightweight Face Recognition," in *2019 International Conference on Computer Vision Workshop, ICCVW 2019*, Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 2647–2654. doi: 10.1109/ICCVW.2019.00323.
- [6] Q. Zhang *et al.*, "VarGNet: Variable Group Convolutional Neural Network for Efficient Embedded Computing," *arXiv:1907.05653*, Jul. 2019. [Online]. Available: <http://arxiv.org/abs/1907.05653>
- [7] Z. Cheng, X. Zhu, and S. Gong, "Low-Resolution Face Recognition," *arXiv:1811.08965v2*, Nov. 2018, [Online]. Available: <http://arxiv.org/abs/1811.08965>
- [8] Y. Martínez-Díaz, H. Méndez-Vázquez, L. S. Luevano, L. Chang, and M. Gonzalez-Mendoza, "Lightweight low-resolution face recognition for surveillance applications," in *International Conference on Pattern Recognition*, Milan: Institute of Electrical and Electronics Engineers Inc., Jan. 2020, pp. 5421–5428. doi: 10.1109/ICPR48806.2021.9412280.
- [9] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality Adaptive Margin for Face Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2022–June, pp. 18729–18738, Apr. 2022, doi: 10.1109/CVPR52688.2022.01819.
- [10] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices," in *Biometric Recognition*, J. Zhou, Y. Wang, Z. Sun, Z. Jia, J. Feng, S. Shan, K. Ubul, and Z. Guo, Eds., Cham: Springer International Publishing, 2018, pp. 428–438.
- [11] Y. Martindez-Diaz, L. S. Luevano, H. Mendez-Vazquez, M. Nicolas-Diaz, L. Chang, and M. Gonzalez-Mendoza, "ShuffleFaceNet: A lightweight face architecture for efficient and highly-accurate face recognition," in *International Conference on Computer Vision Workshop, ICCVW 2019*, Seoul: Institute of Electrical and Electronics Engineers Inc., Oct. 2019, pp. 2721–2728. doi: 10.1109/ICCVW.2019.00333.
- [12] M. Grgic, K. Delac, and S. Grgic, "SCface - Surveillance cameras face database," *Multimed Tools Appl*, vol. 51, no. 3, pp. 863–879, Feb. 2011, doi: 10.1007/s11042-009-0417-2.
- [13] Z. Cheng, X. Zhu, and S. Gong, "Surveillance Face Recognition Challenge," *arXiv: 1804.09691*, Apr. 2018, [Online]. Available: <http://arxiv.org/abs/1804.09691>
- [14] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Amherst, Oct. 2007. [Online]. Available: <http://vis-www.cs.umass.edu/lfw/>.
- [15] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," *arXiv:1411.7923v1*, Nov. 2014, [Online]. Available: <http://arxiv.org/abs/1411.7923>
- [16] A. Torralba, R. Fergus, and W. T. Freeman, "80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 30, no. 11, pp. 1958–1970, 2008, doi: 10.1109/TPAMI.2008.128.
- [17] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," *arXiv:2107.10833*, pp. 1905–1914, Jul. 2021, doi: 10.1109/ICCVW54120.2021.00217.
- [18] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards Real-World Blind Face Restoration with Generative Facial Prior," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 9164–9174, Jan. 2021, doi: 10.1109/CVPR46437.2021.00905.
- [19] Y. Martínez-Díaz *et al.*, "Benchmarking lightweight face architectures on specific face recognition scenarios," *Artif Intell Rev*, vol. 54, no. 8, pp. 6201–6244, 2021, doi: 10.1007/s10462-021-09974-2.
- [20] L. S. Luevano, L. Chang, H. Heydi Mendez-Vazquez, Y. Martinez-Diaz, and M. Gonzalez-Mendoza, "A Study on the Performance of Unconstrained Very Low Resolution Face Recognition: Analyzing Current Trends and New Research Directions," *IEEE Access*, vol. 9, no. 9, pp. 75470–75493, May 2021, doi: 10.1109/ACCESS.2021.3080712.
- [21] InsightFace, "InsightFace: an open source 2D&3D deep face analysis library," *InsightFace*. <https://insightface.ai/> (accessed Apr. 27, 2023).
- [22] S. Z. Li and A. K. Jain, "Introduction," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., 2nd ed. London: Springer London, 2011, pp. 1–15. doi: 10.1007/978-0-85729-932-1.
- [23] D. E. Kurniawan, K. Adi, and A. F. Rohim, "Sistem Identifikasi Biometrika Wajah Menggunakan Metode Gabor KPCA dan

- Mahalanobis Distance,” JSINBIS J. Sist. Inf. Bisnis, vol. 2, no. 1, pp. 006–010, Jan. 2014, doi: 10.21456/vol2iss1pp006-010.
- [24] X. Wu, R. He, Z. Sun, and T. Tan, “A Light CNN for Deep Face Representation with Noisy Labels,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2015, doi: 10.1109/TIFS.2018.2833032.
- [25] M. Wang and W. Deng, “Deep face recognition: A survey,” *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021, doi: 10.1016/J.NEUCOM.2020.10.081.
- [26] Microsoft Research, “DeepSpeed.” <https://www.microsoft.com/en-us/research/project/deepspeed/> (accessed Jul. 31, 2023).
- [27] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4690–4699.
- [28] H. Wang *et al.*, “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Jan. 2018, pp. 5265–5274. doi: 10.1109/CVPR.2018.00552.