

PENINGKATAN PERFORMA PENGENALAN WAJAH PADA GAMBAR *LOW-RESOLUTION* MENGGUNAKAN METODE *SUPER-RESOLUTION*

Muhammad Imaduddin Abdur Rohim^{*1}, Auliati Nisa², Muhammad Nurkhoiri Hindratno³, Radhiyatul Fajri⁴, Gembong Satrio Wibowanto⁵, Nova Hadi Lestriandoko⁶, Pesigrihastamadya Normakristagaluh⁷

^{1,2,3,4,5,6,7}Badan Riset dan Inovasi Nasional

Email: ¹muha228@brin.go.id, ²auliati.nisa@brin.go.id, ³muha217@brin.go.id, ⁴radh001@brin.go.id, ⁵gemb001@brin.go.id, ⁶nova004@brin.go.id, ⁷pesi001@brin.go.id

*Penulis Korespondensi

(Naskah masuk: 31 Oktober 2023, diterima untuk diterbitkan: 29 Desember 2023)

Abstrak

Kartu Tanda Penduduk Elektronik (KTP-el) merupakan identitas wajib bagi penduduk Indonesia. Penyimpanan pada cip KTP-el yang mana selain digunakan untuk menyimpan gambar potret wajah individu, juga harus dapat menyimpan identitas lain seperti biodata, tanda tangan, dan sidik jari kiri dan kanan. Keterbatasan tersebut mengharuskan gambar potret wajah disimpan pada ukuran *low-resolution* (LR) sehingga sistem pengenalan wajah tidak optimal. Dalam penelitian ini, kami menggunakan *Poznan University of Technology* (PUT) *Face database* yang terdiri atas 200 gambar dari 100 individu. Data tersebut dilakukan proses *downsampling* menggunakan *bicubic interpolation* untuk menghasilkan data LR. Kami menginvestigasi penggunaan metode *super-resolution* (SR) berbasis *deep learning*, termasuk DFDNet, LapSRN, GFPGAN, Real-ESRGAN, Real-ESRGAN+GFPGAN, dan FaceSPARNet. Hal ini bertujuan untuk meningkatkan kualitas gambar LR. Evaluasi performa dilakukan dengan menggunakan matriks *False Rejection Rate* (FRR) pada beberapa tingkatan *False Acceptance Rate* (FAR). Hasil penelitian menunjukkan bahwa beberapa metode SR terutama FaceSPARNet menunjukkan peningkatan performa *face recognition* hingga 2%. Sedangkan, metode SR yang berbasis GAN (GFPGAN, Real-ESRGAN, Real-ESRGAN+GFPGAN) cenderung meningkatkan false reject rate. Penelitian ini menunjukkan bahwa metode SR dari kategori *General Basic CNN-based FSR* dapat digunakan untuk meningkatkan kinerja *face recognition* pada gambar LR, seperti pada KTP-el.

Kata kunci: *downsampling, low-resolution, super-resolution, pengenalan wajah, gambar potret wajah*

IMPROVING FACE RECOGNITION PERFORMANCE ON LOW-RESOLUTION IMAGES USING SUPER-RESOLUTION METHODS

Abstract

The Indonesian Electronic Identity Card (KTP-el) is a mandatory identification for the citizens of Indonesia. Information stored on the KTP-el chip is used not only to preserve an individual's facial portrait but also to store other identification details such as biodata, signature, and left and right fingerprints. These limitations necessitate storing the facial portrait image at a low-resolution (LR), resulting in the suboptimal performance of the facial recognition system. In this research, we utilized the Poznan University of Technology (PUT) Face database, comprising 200 images from 100 individuals. The data underwent downsampling using bicubic interpolation to generate low-resolution (LR) data. We explored the application of deep learning-based super-resolution (SR) methods, including DFDNet, LapSRN, GFPGAN, Real-ESRGAN, Real-ESRGAN+GFPGAN, and FaceSPARNet, to enhance the quality of LR images. Performance evaluation was conducted using the False Rejection Rate (FRR) matrix at various False Acceptance Rate (FAR) levels. The research results indicate that SR methods, particularly FaceSPARNet, improved face recognition performance by up to 2%. In contrast, SR methods based on Generative Adversarial Networks (GANs) (GFPGAN, Real-ESRGAN, Real-ESRGAN+GFPGAN) tended to yield inferior performance. This study demonstrates that SR approaches of General Basic CNN-based FSR improved face recognition performance on LR images, such as those found in KTP-el.

Keywords: *downsampling, low-resolution, super-resolution, face recognition, face portrait image*

1. PENDAHULUAN

Kartu Tanda Penduduk Elektronik (KTP-el) merupakan identitas wajib bagi penduduk Indonesia yang telah berusia 17 tahun atau lebih, atau yang telah menikah (UU Republik Indonesia No. 23 Tahun 2006 Tentang Administrasi Kependudukan, 2006). Dalam upaya untuk memaksimalkan pemanfaatan data pada cip KTP-el, salah satu data penting yang perlu diperhatikan adalah gambar potret wajah dari setiap individu. Gambar potret wajah ini mencakup gambar wajah yang menyimpan banyak informasi identitas suatu individu (Cheng & Meng, 2021).

Pengelolaan gambar potret wajah dalam KTP-el tidaklah tanpa kendala. Saat ini, penyimpanan minimal yang tersedia pada cip KTP-el adalah 32 kB, yang juga harus mencakup data-data identitas lainnya seperti biodata, tanda tangan, dan sidik jari kiri dan kanan. Meskipun ukuran minimum gambar potret wajah diatur sebesar 3 kB (Standar dan Spesifikasi Perangkat Keras, Perangkat Lunak, dan Cip Kartu Tanda Penduduk Elektronik serta Penyelenggaraan Identitas Kependudukan Digital), tetapi ukuran gambar potret wajah pada cip KTP-el yang banyak dipakai saat ini masih beresolusi rendah. Penggunaan gambar dengan resolusi rendah memiliki dampak signifikan terhadap akurasi sistem pengenalan wajah. Penurunan kualitas gambar sering kali mengakibatkan hilangnya banyak detail penting selama proses penurunan resolusi gambar (Bouderbal et al., 2021). Hal ini mengakibatkan kesulitan sistem pengenalan wajah dalam mengidentifikasi individu dengan tepat (Z. Cheng et al., 2018).

Salah satu cara untuk menilai apakah suatu dataset memiliki kualitas yang baik dan kemungkinan akan dibaca dengan akurat oleh algoritma pengenalan wajah adalah *Face Image Quality Assessment* (FIQA) (Ou et al., 2021). Dalam algoritma tersebut, salah satu aspek penentu dalam menilai kualitas gambar wajah adalah jarak antara kedua mata (*inter-eye distance*) dan resolusi gambar (Schlett et al., 2020). Saat ini, metode pengenalan wajah yang menggunakan model state-of-the-art (Deng et al., 2018) telah mencapai akurasi yang sangat baik. Namun, metode ini masih menghadapi tantangan yang signifikan dalam pemadanan wajah pada gambar beresolusi rendah (Li et al., 2018; 2019).

Beberapa tantangan yang dihadapi dalam pemadanan wajah pada gambar resolusi rendah termasuk proses deteksi wajah (*face detection*), penyalarsan wajah (*face alignment*), dan pemadanan (*matching*). Sebagai contoh, algoritma *landmark detection* yang dilatih pada gambar berkualitas tinggi sering kali kurang efektif pada dataset yang berisi gambar resolusi rendah (Kumar & Chellappa, 2019; Peng et al., 2019).

Selain kendala dalam pengelolaan gambar wajah, masalah keamanan juga menjadi fokus utama dalam penggunaan gambar potret wajah pada cip KTP-el sebagai alat verifikasi. Gambar wajah yang

memiliki kualitas rendah pada cip KTP-el dapat mengarah pada kesalahan identifikasi identitas. Kesalahan ini dapat membuka peluang kerentanan pada sistem pengenalan wajah, yang dapat mengakibatkan akses yang tidak sah atau tindakan yang tidak diinginkan (Y. Cheng & Meng, 2021).

Dalam rangka menghasilkan gambar dengan ukuran penyimpanan yang sesuai dengan kapasitas cip KTP-el yang terbatas, seperti diatur dalam Standar Dan Spesifikasi Perangkat Keras, Perangkat Lunak, Dan Cip Kartu Tanda Penduduk Elektronik Serta Penyelenggaraan Identitas Kependudukan Digital tahun 2022 (Standar dan Spesifikasi Perangkat Keras, Perangkat Lunak, dan Cip Kartu Tanda Penduduk Elektronik serta Penyelenggaraan Identitas Kependudukan Digital), kami melakukan proses *downsampling* dari gambar asli menggunakan metode bicubic interpolation. Penelitian terkait sebelumnya menunjukkan bahwa metode bicubic interpolation dapat meningkatkan performa hingga 8,41% pada spesifik FAR dibandingkan dengan metode *downsampling* lain (Hindratno et al., 2023).

Dalam menghadapi tantangan ini, sebuah pendekatan yang menjanjikan adalah dengan memanfaatkan metode *super-resolution*. Metode *super-resolution* digunakan untuk meningkatkan kualitas gambar beresolusi rendah atau *low-resolution* (LR), sehingga gambar tersebut dapat mengandung lebih banyak detail dan informasi identitas atau *super-resolution* (SR). Beberapa metode *super-resolution* saat ini telah mengklaim kemampuan untuk mengurangi dampak gambar berkualitas rendah yang disebabkan oleh keterbatasan perangkat keras, seperti penyimpanan dan komputasi (Cao et al., 2021).

Metode *super-resolution* berbasis *deep learning* telah menunjukkan potensi tinggi dalam bidang interpolasi dan restorasi gambar dibandingkan dengan algoritma interpolasi berbasis piksel konvensional (Kim et al., 2021). Beberapa metode *super-resolution* terkini yang berbasis CNN (Convolutional Neural Network) (Nasrollahi et al., 2020; J. Wang et al., 2022) dan GAN (Generative Adversarial Network) (Ren et al., 2020; Tian et al., 2022) telah membuktikan kemampuannya dalam meningkatkan resolusi gambar yang rendah. Namun, untuk metode berbasis GAN, training GAN sering kali tidak stabil dan sering menghasilkan artefak yang tidak diinginkan bersama dengan detail yang direkonstruksi (Liang et al., 2022). Hal serupa terjadi pada metode berbasis CNN yang dapat menghasilkan artefak berupa blok dan ringing (Kim et al., 2021).

Algoritma pengenalan wajah saat ini mencapai tingkat akurasi sebesar 99,23% dalam mencocokkan gambar-gambar dengan kualitas yang baik (Z. Cheng et al., 2018). Namun, angka akurasi tersebut menurun drastis menjadi hanya 55% ketika gambar yang digunakan memiliki resolusi rendah (Z. Cheng et al., 2018). Salah satu solusi untuk meningkatkan kualitas

gambar wajah adalah menggunakan metode super resolusi. Namun, perlu diingat bahwa hasil dari metode super resolusi ini sering kali menciptakan artefak atau menambahkan fitur tambahan pada gambar wajah, yang pada akhirnya dapat memengaruhi akurasi sistem pengenalan wajah. Oleh karena itu, sangat penting untuk memilih metode super resolusi yang tepat, yang tidak hanya meningkatkan kualitas visual dari gambar wajah, tetapi juga dapat memulihkan identitas wajah pada gambar beresolusi rendah.

Penelitian ini akan memberikan beberapa kontribusi, antara lain: 1) komparasi dari keenam metode *super-resolution* terhadap akurasi dalam konteks sistem pemadanan (verifikasi wajah), 2) wawasan tentang performa pengenalan wajah antara gambar resolusi rendah dengan metode *super-resolution* terhadap gambar wajah asli, dan 3) rekomendasi metode *super-resolution* yang paling efektif dalam meningkatkan akurasi pengenalan wajah untuk KTP-el.

Pendahuluan, latar belakang dan tujuan dari penelitian dijelaskan pada bagian ini. Selanjutnya adalah metode penelitian yang memaparkan metode-metode yang dipakai dalam melaksanakan penelitian seperti dataset, *downsampling*, sistem pemadanan wajah, dan setiap metode *super-resolution* yang digunakan. Bagian berikutnya adalah pemaparan hasil dan penjelasan data yang didapatkan dari penelitian yang dilakukan, serta diakhiri kesimpulan.

2. METODE PENELITIAN

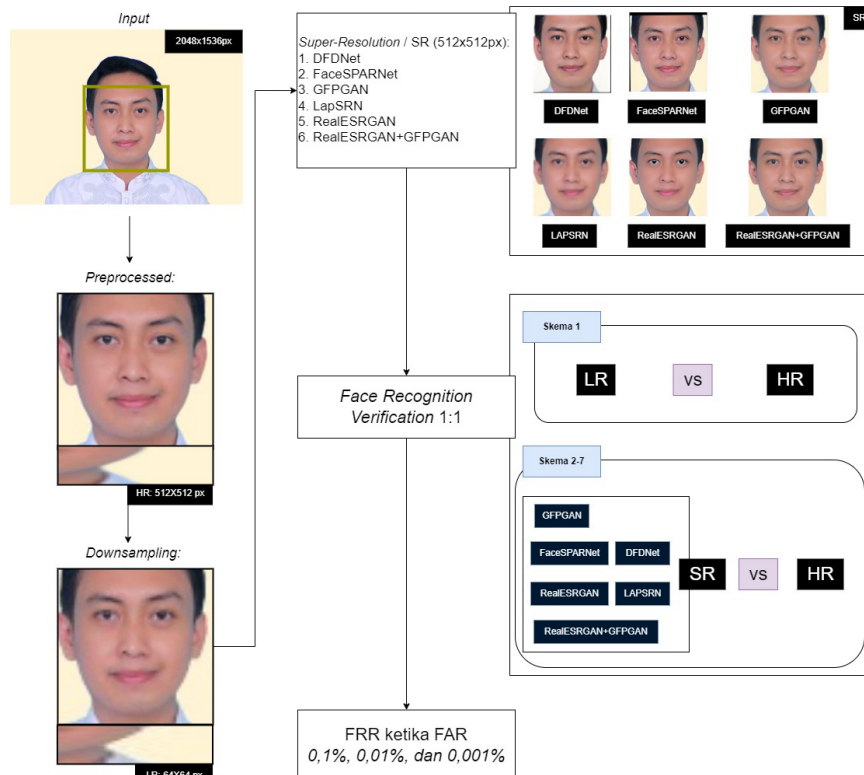
Kami menggunakan 3 kategori metode *super-resolution*: 1) kategori *General Face Super*

Resolution (FSR) berbasis *Generative Adversarial Network* (GAN) seperti GFPGAN, Real-ESRGAN, GFPGAN, Real-ESRGAN+GFPGAN, 2) kategori *General FSR* berbasis *basic Convolutional Neural Network* (CNN) seperti FaceSPARNet dan LapSRN, dan 3) kategori *Reference FSR* seperti DFDNet. Kemudian performa sistem pengenalan wajah dihitung dengan menggunakan matriks *False Rejection Rate* (FRR) pada beberapa tingkat *False Acceptance Rate* (FAR).

Secara garis besar, alur dimulai dengan Poznan University of Technology (PUT) Face Database (Kasiński et al., 2008) diolah menjadi tiga jenis dataset yaitu: *high-resolution* (HR), LR, dan *super-resolution* (SR). Kemudian dilakukan proses pemadanan dengan metode verifikasi wajah 1:1 dengan skema pemadanan antara LR dengan HR (Skema 1) dan skema *super-resolution* antara SR (6 Metode) dengan HR (Skema 2,3,4,5,6,7). Selanjutnya, dilakukan evaluasi menggunakan FRR pada nilai FAR tertentu. Keseluruhan proses ini dapat dilihat pada Gambar 1.

2.1. Dataset

Dalam paper ini, kami menggunakan Poznan University of Technology (PUT) Face Database (Kasiński et al., 2008) yang tersedia secara publik. Dataset ini merupakan dataset dengan skema warna RGB beresolusi 2048 x 1536 piksel dengan format jpeg dan memiliki ukuran berkisar antara 1,1 megabyte hingga 1,4 megabyte. Dataset ini terdiri dari 100 individu dengan komposisi 89 laki-laki dan 11 perempuan.

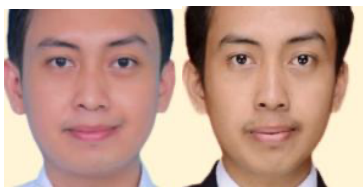


Gambar 1. Flowchart Penelitian

Setiap individu memiliki beberapa gambar dengan pose wajah yang berbeda. Pada eksperimen ini, untuk setiap individu kami memilih dua varian gambar dengan pose wajah menghadap kedepan (*frontal face*) sehingga total gambar wajah yang digunakan sebanyak 200 gambar. Pemilihan varian gambar disesuaikan dengan kondisi gambar pada KTP-el. Dua varian gambar tersebut dapat dilihat pada Gambar 2. Setelah memilih varian gambar, selanjutnya kami melakukan *pre-processing* dengan cara mendeteksi bagian wajah menggunakan *pre-trained model* dari dlib dengan menentukan 68 titik pada wajah, lalu dilakukan proses *aligning* wajah. Wajah yang sudah dilakukan *alignment* tersebut kemudian dipotong (*crop*) dan disimpan dengan resolusi 512 x 512 piksel dengan format jpeg dan memiliki ukuran berkisar 53 - 83 *kilobyte*. Data ini kami sebut sebagai dataset HR yang akan digunakan sebagai *ground-truth* pada eksperimen ini. Sampel data setelah dilakukan *pre-processing* dapat dilihat pada Gambar 3.



Gambar 2. Varian wajah



Gambar 3. Data wajah yang telah dilakukan *pre-processing*

2.2. Metode *Downsampling*

Setelah melakukan *pre-processing* dan mendapatkan dataset HR, selanjutnya dilakukan pembentukan dataset LR dengan cara melakukan proses *downsampling* pada dataset HR tersebut menggunakan metode *bicubic interpolation*. Metode ini merupakan teknik interpolasi dengan menggunakan 4 x 4 piksel di sekitar titik interpolasi sebagai referensi untuk menghitung nilai piksel pada titik interpolasi (Parsania & V. Virparia, 2015). Proses *downsampling* ini dilakukan menggunakan OpenCV versi 4.7.0. Data yang sebelumnya memiliki resolusi 512 x 512 piksel diturunkan resolusinya menjadi 64 x 64 piksel dengan format jpeg dan memiliki ukuran 3 *kilobyte*. Data berukuran 64 x 64 piksel ini kami sebut sebagai dataset LR.

2.3. *Super-Resolution*

Teknik-teknik *super-resolution* memainkan peran penting dalam meningkatkan kualitas gambar

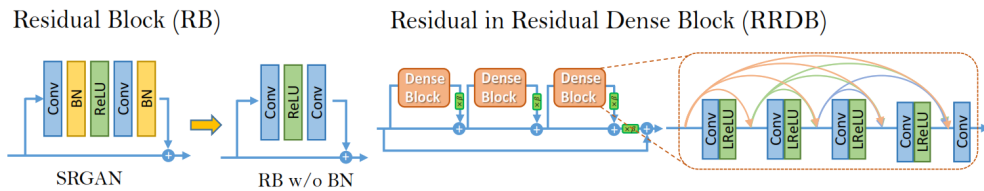
(Rakotonirina & Rasoanaivo, 2020). Saat ini terdapat berbagai metode *super-resolution* yang telah dikembangkan. Berdasarkan *survey deep learning-based Face Super-Resolution* yang dilakukan oleh Junjun Jiang dalam paper (Jiang et al., 2021), metode *super-resolution* dipetakan ke dalam taksonomi berdasarkan kategorinya. Dalam eksperimen ini, kami menggunakan metode-metode *super-resolution* yang berasal dari kategori *General FSR* dan *Reference FSR*. *General FSR* berisi metode-metode yang berfokus dalam mendesain jaringan yang efisien dan mengeksplorasi struktur jaringan potensial untuk FSR tanpa menggunakan karakteristik wajah. Sementara itu, *Reference FSR* adalah kategori di mana jaringan FSR menggunakan gambar referensi untuk dapat memproses gambar LR menjadi SR. Jumlah gambar wajah referensinya bisa hanya satu atau beberapa gambar.

Pada eksperimen ini, kami menggunakan enam metode *super-resolution* yaitu DFDNet, FaceSPARNet, GFPGAN, LAPSRN, Real-ESRGAN, dan Real-ESRGAN yang telah dikombinasikan dengan GFPGAN. Proses *super-resolution* dilakukan dengan menggunakan metode tersebut pada dataset LR untuk menghasilkan dataset berukuran 512 x 512 piksel dengan format jpeg yang kami sebut sebagai dataset SR.

2.3.1. ESRGAN

ESRGAN (*Enhanced Generative Adversarial Networks*) merupakan metode *super-resolution* yang menggunakan basis *Generative Adversarial Network* (GAN). Terdapat beberapa bagian utama dalam ESRGAN (X. Wang et al., 2018):

- Generator* yang mengubah input gambar beresolusi rendah menjadi gambar beresolusi tinggi. Bagian ini meningkatkan kemampuan model untuk menghasilkan image berkualitas tinggi menggunakan koneksi dan metode khusus.
- Diskriminator yang menilai seberapa berhasil generator dalam menghasilkan gambar beresolusi tinggi. Discriminator tersebut memberikan umpan balik kepada generator tentang cara untuk memperbaiki dengan membandingkan gambar yang dihasilkan dengan gambar resolusi tinggi asli.
- Jaringan ekstraksi fitur yang membantu dalam membuat gambar dari *Generator* yang lebih menyerupai foto resolusi tinggi asli. Hal ini dicapai dengan memeriksa apakah aspek-aspek tertentu dari foto cocok atau tidak. Ketiga komponen ini digabungkan oleh ESRGAN untuk menghasilkan gambar super-resolusi berkualitas tinggi yang sangat menyerupai gambar resolusi tinggi asli.
- Metode ini menggunakan model ESRGAN yang sudah dilatih sebelumnya untuk menghasilkan gambar SR dari gambar resolusi rendah yang berukuran 480 x 600 piksel.



Gambar 4. Arsitektur ESRGAN (X. Wang, Xie, et al., 2021).

Adversarial loss dan total loss generator pada ESRGAN dituliskan dalam formula sebagai berikut:

$$L_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_f, x_r))] \quad (1)$$

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (2)$$

L_1 adalah content loss yang mengevaluasi jarak 1-norm antara image yang dipulihkan dan ground-truth y . Sementara, λ dan η adalah koefisien untuk menyeimbangkan perbedaan berbagai jenis loss.

2.3.2. Real-ESRGAN

Real-ESRGAN merupakan sebuah model super-resolution yang dapat menghasilkan gambar berkualitas tinggi dari input gambar beresolusi rendah. Model ini merupakan pengembangan dari ESRGAN. Berikut beberapa deskripsi singkat terkait arsitekturnya (X. Wang, Xie, et al., 2021):

- a. *Generator Network*: Jaringan ini menggunakan jaringan saraf dalam dengan koneksi residual dan blok padat untuk menghasilkan gambar beresolusi tinggi dari input resolusi rendah.
- b. *Diskriminator Network*: Dengan menggunakan Deep Neural Network dengan struktur piramida gambar. Sehingga jaringan ini dapat membedakan gambar yang diciptakan dari gambar beresolusi tinggi sebenarnya.
- c. *Synthetic Data Generation*: Dataset DIV2K digunakan untuk membuat data pelatihan sintetis. Model ESRGAN digunakan untuk membuat gambar beresolusi tinggi yang sesuai dari gambar resolusi tinggi datataset. Setelah itu, gambar beresolusi rendah dihasilkan untuk melatih model Real-ESRGAN

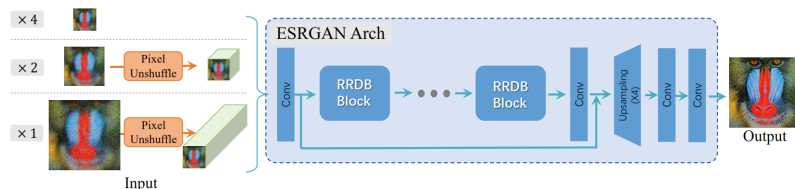
- d. *Progress Training*: Untuk meningkatkan kualitas gambar super-resolved, ukuran gambar input terus ditingkatkan selama proses training.

Pada Real-ESRGAN, terdapat pilihan untuk meningkatkan kualitas gambar, khusus untuk gambar wajah dengan menggabungkan model Real-ESRGAN dan GFGAN. Penggabungan model tersebut digunakan dalam eksperimen kami untuk melihat perbedaan gambar super-resolution yang dihasilkan.

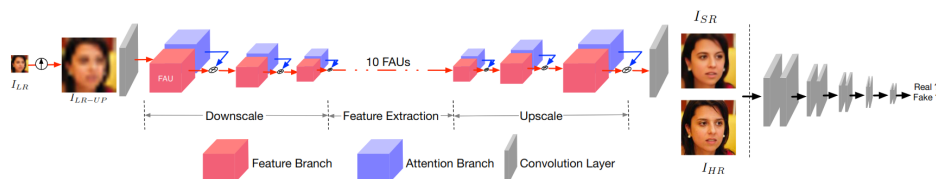
Model Real-ESRGAN dilatih dengan gabungan dari $L1$ loss, perceptual loss, dan GAN loss, dengan bobot berturut-turut $\{1,1,0.1\}$. Real-ESRGAN menggunakan feature maps sebelum diaktivasi dengan pre-trained VGG19 network sebagai perceptual loss.

2.3.3. Face-SPARNet

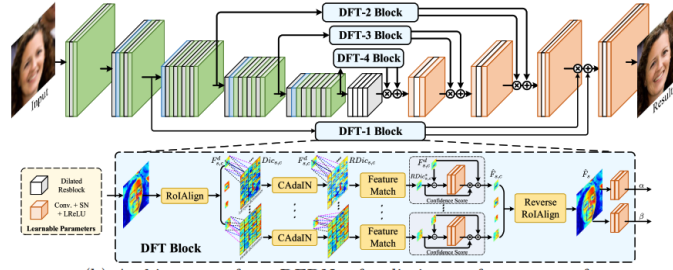
SPARNet adalah teknik super-resolusi yang menekankan peningkatan kualitas gambar wajah. Selama proses super-resolusi, arsitektur ini menyoroti fitur wajah penting melalui mekanisme spatial attention. Arsitektur ini mencakup feature extraction network, spatial attention module, many upsampling layers, dan kombinasi dari content loss dan adversarial loss functions. Model SPARNet menggunakan kerangka pembelajaran PyTorch. Secara umum, arsitektur SPARNet menghasilkan gambar wajah super-resolusi berkualitas tinggi dengan fokus pada fitur wajah yang signifikan dan memanfaatkan metode pelatihan yang efisien (Chen et al., 2020). Dalam eksperimen ini, kami menggunakan model SPARNetHD-Attn2D untuk menghasilkan gambar super-resolution dengan ukuran 512x512 piksel.



Gambar 5. Arsitektur Real-ESRGAN (X. Wang, Xie, et al., 2021).



Gambar 6. Arsitektur Face-SPARNet (Chen et al., 2020).



Gambar 7. Arsitektur DFDNet untuk *dictionary feature transfer* (Li et al., 2020).

SPARNet dioptimasi dengan meminimalisir *pixel-level* L2 *loss* sesuai persamaan (3). Simbol F_{SPAR} dan Θ menunjukkan SPARNET dan parameter.

$$C_{\text{pixel}}(\Theta) = \frac{1}{N} \sum |F_{SPAR}(I_R - U_P, \Theta) - T_{HR}|^2 \quad (3)$$

2.3.4. DFDNet

DFDNet bertujuan menghasilkan gambar berkualitas tinggi (I^h) dengan *component dictionaries* yang telah dibuat sebelumnya. Pada tahap pertama, *component dictionaries* dibentuk dari gambar berkualitas tinggi. *component dictionaries* ini dapat dipilih sebagai komponen referensi. Pada tahap kedua, DFDNet memilih fitur *dictionary* yang sesuai dengan gambar yang mengalami degradasi. DFDNet juga menggunakan skor kepercayaan untuk mengatasi tingkat degradasi yang berbeda. Pendekatan progresif dari kasar ke halus juga membantu dalam proses pemulihan gambar (X. Li et al., 2020).

Objektif *learning model* dari DFDNet terdiri menjadi dua bagian, 1) *reconstruction loss* yang membatasi hasil \hat{I} dekat dengan *ground-truth* I^h , 2) *adversarial loss* untuk memulihkan detail yang realistis. Adapun formulanya adalah sebagai berikut.

Reconstruction Loss.

$$\mathcal{L}_{rec} = \lambda_{12} \|\hat{I} - I^h\|^2 + \sum_{m=1}^M \frac{\lambda_{p,m}}{C_m H_m W_m} \|\psi_m(\hat{I}) - \psi_m(I^h)\|^2 \quad (4)$$

Pada Persamaan (4), ψ_m mewakili *convolution layer* ke- m . Sedangkan, ψ , C , H , dan W adalah *channel*, tinggi dan lebar.

Adversarial Loss.

$$\mathcal{L}_{adv, D_r} \quad (5)$$

$$= \sum_r E_{I_{I_r}^h \sim P(I_{I_r}^h)} [\min(0, D_r(I_{I_r}^h) - 1)] + E_{\hat{I}_{I_r} \sim P(\hat{I}_{I_r})} [\min(0, -1 - D_r(\hat{I}_{I_r}))]$$

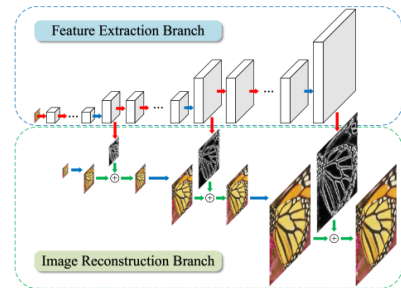
Pada persamaan (5), I_r mewakili operasi *downsampling* dengan skala faktor r dengan nilai $\{1, 2, 4, 8\}$. Kemudian persamaan ke (5) diturunkan menjadi fungsi *loss* pada *training generator* pada persamaan (6).

$$\mathcal{L}_{adv, G} \quad (6)$$

$$= -\lambda_{a,r} \sum_r E_{I^d \sim P(I^d)} [D_r(F(I^d | L^d, Disc))]$$

Pada persamaan (6), $\lambda_{a,r}$ adalah parameter pertukaran untuk setiap diskriminator skala.

2.3.5. LapSRN



Gambar 8. Arsitektur LapSRN (Lai et al., 2017).

Metode LapSRN adalah jaringan yang menggunakan kerangka piramida Laplacian untuk meningkatkan resolusi gambar LR dengan faktor skala. Jaringan ini terdiri dari dua cabang: ekstraksi fitur dan rekonstruksi gambar. Pada setiap level, cabang ekstraksi fitur menggunakan lapisan konvolusi dan *upsampling* untuk memperbesar fitur. Hasilnya digunakan untuk merekonstruksi gambar residu pada level tersebut dan mengekstrak fitur pada level yang lebih halus. Cabang rekonstruksi gambar memperbesar gambar *input*, menggabungkannya dengan gambar residu, dan menghasilkan gambar output resolusi tinggi. Jaringan ini membagikan representasi fitur antar-

level untuk meningkatkan kompleksitas pemetaan pada level yang lebih halus. (Lai et al., 2017).

Objektif *learning model* dari LapSRN menggunakan *robust loss function* untuk mengatasi hasil yang *outlier*. Keseluruhan model *loss function* adalah sebagai berikut:

$$\mathcal{L}(\hat{y}, y; \theta) = \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^L \rho(\hat{y}_s^{(i)} - y_s^{(i)}) \quad (7)$$

Pada persamaan tersebut, $\rho(x) = \sqrt{x^2 + \varepsilon^2}$ adalah fungsi *charbonnier penalty*, N adalah jumlah sampel pelatihan setiap batch, ε merupakan penetapan dengan nilai tertentu dan L adalah jumlah level dari piramida.

2.3.6. GFPGAN

Pada Gambar 9 dapat dilihat bahwa, GFP-GAN adalah model yang terdiri dari modul penghilangan degradasi (*U-Net*) dan *face GAN* yang telah dilatih sebelumnya. Kedua komponen ini terhubung melalui pemetaan kode laten dan beberapa lapisan *Channel-Split Spatial Feature Transform* (CS-SFT). Selama pelatihan, GFP-GAN menggunakan tiga tingkat kerugian: pemulihan untuk menghilangkan degradasi, komponen wajah dengan diskriminator untuk meningkatkan fitur wajah, dan mempertahankan identitas untuk mempertahankan identitas wajah. Fitur-fitur yang dihasilkan oleh model ini mencakup fitur laten, fitur spasial multi-resolusi, dan fitur konvolusi perantara untuk menangkap detail wajah yang rumit dan menghasilkan output realistis. (X. Wang, Li, et al., 2021).

Objektif *learning model* dari DFDNet terdiri menjadi: 1) *reconstruction loss* yang membatasi hasil \hat{y} dekat dengan *ground-truth* y , 2) *adversarial loss* untuk memulihkan detail yang realistis, 3) proposisi *facial component loss* untuk lebih meningkatkan rincian wajah, dan 4) *identity preserving loss*

Reconstruction Loss.

$$\mathcal{L}_{rec} = \lambda_{l1} \|\hat{y} - y\|_1 + \lambda_{per} \|\Phi(\hat{y}) - \Phi(y)\|_1 \quad (8)$$

Pada persamaan (8), Φ adalah pretrained VGG-19 Network. Sedangkan, λ_{l1} dan λ_{per} mewakili loss weights dari L1 dan *preceptual loss*.

Adversarial Loss.

$$\mathcal{L}_{adv} = -\lambda_{adv} \mathbb{E}_{\hat{y}} \text{softplus}(D(\hat{y})) \quad (9)$$

Pada persamaan (9), D mewakili diskriminator dan $-\lambda_{adv}$ adalah *adversarial loss weight*.

Facial Component Loss

$$\mathcal{L}_{comp} = \sum_{ROI} \lambda_{local} \mathbb{E}_{\hat{y}_{ROI}} [\log(1 - D_{ROI}(\hat{y}_{ROI})) + \lambda_{fs} \|\text{Gram}(\psi(\hat{y}_{ROI})) - \text{Gram}(\psi(y_{ROI}))\|_1] \quad (10)$$

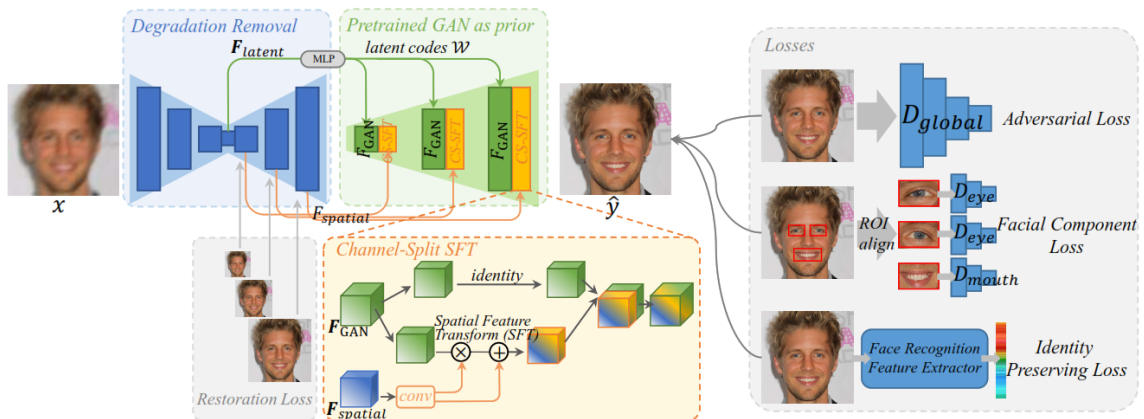
Pada persamaan (10), ROI adalah *region of interest* dari koleksi komponen {mata kiri, mata kanan, mulut}. D_{ROI} adalah diskriminator lokal untuk masing masing *region*. ψ mewakili fitur *multi-resolution* dari *learned discriminator*. λ_{local} dan λ_{fs} merupakan *loss weight* dari *local discriminator loss* dan *feature style loss*.

Identity Preserving Loss

$$\mathcal{L}_{id} = -\lambda_{id} \|\eta(\hat{y}) - \eta(y)\|_1 \quad (11)$$

2.4. Pada persamaan (11), η mewakili ekstraktor fitur wajah. Sedangkan, λ_{id} adalah bobot dari *identity preserving loss*. **Face Recognition**

Proses *face recognition* dilakukan dengan menggunakan *pre-trained* resnet34 model yang telah dilakukan proses *fine-tuning* menggunakan dataset pribadi milik kami untuk melakukan proses *feature extraction* dan mendapatkan nilai *embedded* dari gambar. Proses verifikasi wajah kemudian dilakukan dengan menghitung jarak *euclidean* berdasarkan nilai *embedding* dari masing-masing gambar yang dipadankan.



Gambar 9. Kerangka GFP-GAN (X. Wang, Li, et al., 2021).

Pada skema pertama, kami melakukan pemadanan wajah pada gambar LR berukuran 64 x 64 piksel dengan dengan gambar HR berukuran 512 x 512 piksel, kemudian pada skema selanjutnya kami melakukan pemadanan wajah pada gambar SR berukuran 512 x 512 piksel yang dihasilkan dari beberapa metode super-resolution dengan gambar HR berukuran 512 x 512 piksel. Evaluasi performa dilakukan dengan menggunakan matriks FRR pada FAR 0,1%, 0,01%, dan 0,001%.

3. HASIL DAN PEMBAHASAN

Pada Gambar 10 dapat dilihat bahwa pemadanan LR dan SR cenderung memberikan nilai FRR yang berbeda tergantung dengan metode dan tingkat FAR yang digunakan. Tingkat FAR tersebut adalah 0,1%, 0,01%, dan 0,001% yang mana menghasilkan nilai FRR dengan rentang 1,0% sampai dengan 9,8%. Pemadanan LR pada gambar tersebut digunakan sebagai nilai acuan dasar. Sementara itu, proses pemadanan *super-resolution* dengan skema 2 – 7 ditunjukkan di gambar dengan warna yang berbeda beda.

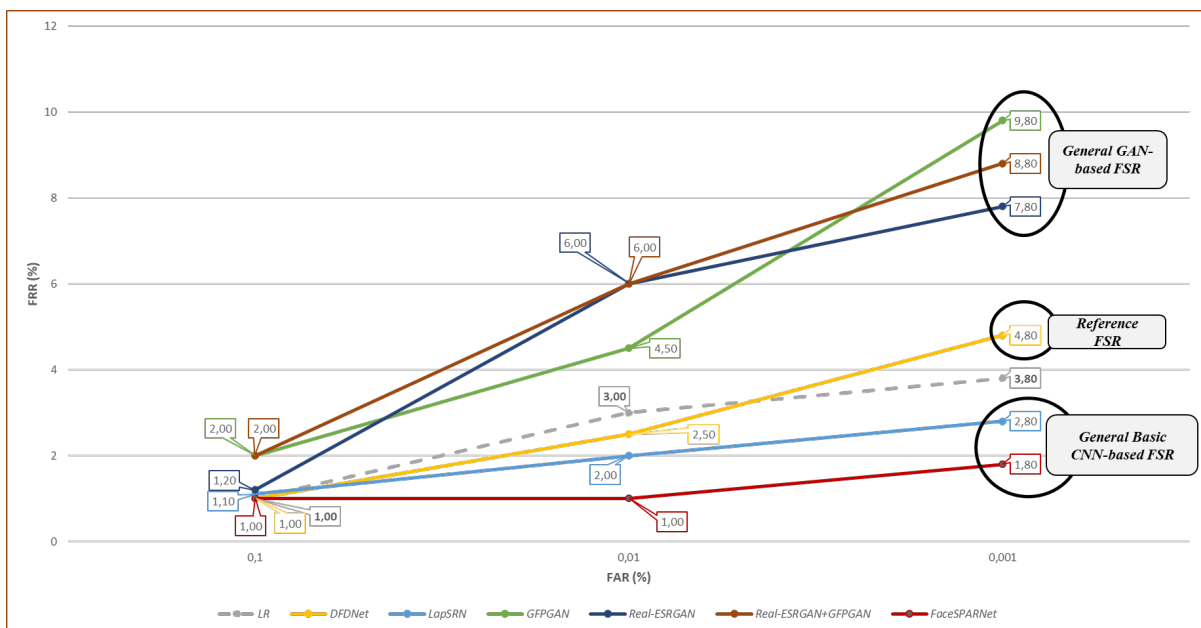
Secara umum, metode *super-resolution* tertentu dapat memberikan nilai FRR yang lebih baik dibandingkan dengan gambar LR pada level FAR tertentu. Metode FaceSPARNet merupakan pilihan metode *super-resolution* terbaik untuk peningkatan kinerja *Face Recognition* pada setiap ambang level FAR (0,1%, 0,01%, dan 0,001%) dengan nilai FRR sebesar (1,0%, 1,0%, 1,8%) secara berurutan dan juga lebih baik dari gambar di LR yang FRRnya bernilai (1,0%, 3,0%, 3,8%).

Jika dibandingkan dengan LR, LapSRN memberikan hasil lebih baik di tingkat FAR yang lebih rendah yaitu pada tingkat 0,01% dan 0,001%. Hal ini berbanding terbalik dengan DFDNet, yang menunjukkan performa FRR lebih baik pada tingkat yang lebih tinggi ketika FAR 0,1% dan 0,01%. Sementara itu, 3 metode dalam kategori *General GAN-based FSR* (Real-ESRGAN, Real-ESRGAN+GFPGAN, dan GFPGAN) secara konsisten memberikan nilai FRR yang lebih besar dibandingkan LR dan metode yang lain untuk semua tingkat FAR.

Jika dilihat dari jenis FSRnya, dapat dilihat bahwa FRR *General basic CNN-based FSR* cenderung lebih baik dari pada LR dan metode lain, sedangkan *reference FSR* cenderung meningkat *false reject*-nya pada tingkat FAR yang lebih rendah. Setelah itu, *General GAN-based FSR* selalu menduduki peringkat terendah dalam performa *face recognition*.

Berdasarkan hasil FRR untuk setiap metode FSR tersebut, kemudian kami melakukan rerata metode untuk masing-masing jenisnya. Adapun data pengelompokannya bisa dilihat pada Tabel 1. Pengelompokan Metode FSR dapat dilihat bahwa *General Basic CNN-based FSR* yang terdiri dari metode LapSRN dan FaceSPARNet mengungguli jenis metode FSR yang lain dengan nilai rerata FRR sebesar 1,34%.

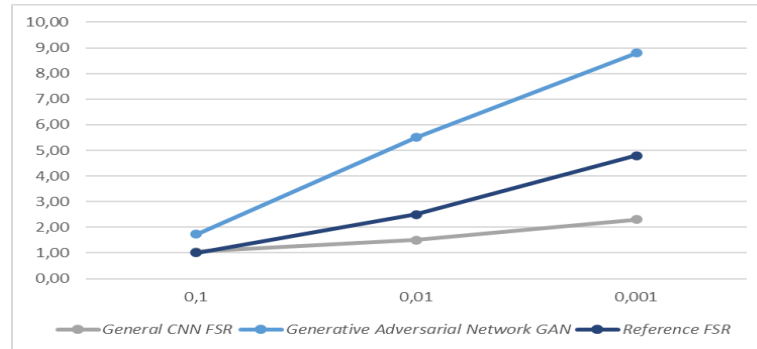
Dalam pengkajian yang lebih mendalam seperti pada Gambar 11, dapat dilihat bahwa metode *super-resolution* pada kategori *General Basic CNN-based FSR* cenderung konsisten memiliki performa verifikasi wajah lebih baik dari pada metode *super-resolution* yang lain untuk setiap FAR.



Gambar 10. Performa pengenalan wajah untuk LR dan seluruh metode SR pada setiap tingkat FAR

Tabel 1. Pengelompokan Metode *Super-resolution* Wajah

Jenis	Metode	Rerata <i>FRR</i> (%)
<i>General Basic CNN-based FSR</i>	LapSRN dan FaceSPARNet	1,34
<i>Reference FSR</i>	DFDNet	2,08
<i>General GAN-based FSR</i>	GFPGAN, Real-ESRGAN, dan Real-ESRGAN+GFPGAN	4,01



Gambar 11. Rata-rata performa FSR pada tiap FAR

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah metode SR tertentu terbukti mampu meningkatkan performa *face recognition* dibandingkan dengan pemadanan gambar LR. Secara spesifik, FaceSPARNet sebagai salah satu metode SR merupakan pilihan terbaik untuk meningkatkan performa *face recognition* karena telah menunjukkan kinerja unggul dibandingkan dengan metode SR lainnya pada setiap tingkat FAR hingga 2%. Jika dibandingkan dengan LR, metode LapSRN menunjukkan keunggulan pada tingkat FAR yang lebih rendah. Sebaliknya, DFDNet memberikan hasil yang lebih baik pada tingkat FAR yang lebih tinggi. Sementara itu, metode berbasis *General GAN-based FSR* yaitu GFPGAN, Real-ESRGAN, dan Real-ESRGAN+GFPGAN tidak direkomendasikan sebagai metode *super-resolution* dalam hal peningkatan performa *face recognition* karena memiliki *false reject rate* yang tinggi daripada LR untuk setiap tingkatan FAR. Jika dilihat dari jenis metode *face super-resolution*, metode yang termasuk dalam kategori *General Basic CNN-based FSR* memiliki performa lebih baik sehingga lebih disarankan dari pada *reference FSR* dan *General GAN-based FSR*.

Adapun saran untuk penelitian selanjutnya adalah penambahan jumlah data dengan mencakup lebih banyak variasi karakter wajah Indonesia, dengan variasi penambahan dataset yang mencakup berbagai etnis, usia, jenis kelamin, dan ekspresi wajah, sehingga memberikan hasil yang lebih akurat, relevan, dan berkesinambungan dalam mendukung perkembangan dan pemanfaatan teknologi pengenalan wajah di Indonesia.

DAFTAR PUSTAKA

Bouderbal, I., Amamra, A., & Benatia, M. A. (2021). *How Would Image Down-Sampling and*

Compression Impact Object Detection in the Context of Self-driving Vehicles? (pp. 25–37). https://doi.org/10.1007/978-3-030-69418-0_3

Cao, M., Liu, Z., Huang, X., & Shen, Z. (2021). Research for Face Image Super-Resolution Reconstruction Based on Wavelet Transform and SRGAN. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 448–451. <https://doi.org/10.1109/IAEAC50856.2021.9390748>

Chen, C., Gong, D., Wang, H., Li, Z., & Wong, K.-Y. K. (2020). *Learning Spatial Attention for Face Super-Resolution*. <https://doi.org/10.1109/TIP.2020.3043093>

Cheng, Y., & Meng, H. (2021). Research and implementation of network information security management system based on face recognition. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 294–302. <https://doi.org/10.1109/ICBAIE52039.2021.9389893>

Cheng, Z., Zhu, X., & Gong, S. (2018). *Low-Resolution Face Recognition*.

Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., & Zafeiriou, S. (2018). *ArcFace: Additive Angular Margin Loss for Deep Face Recognition*. <https://doi.org/10.1109/TPAMI.2021.3087709>

Hindratno, M. N., Nisa, A., Rohim, M. I. A., Fajri, R., Hamdani, M., Wibowanto, G. S., Lestriandoko, N. H., & Normakristagaluh, P. (2023). The Impact of Downsampling Methods on Face Recognition in Electronic Identity Card. *2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 412–417.

- <https://doi.org/10.1109/IC3INA60834.2023.10285777>
- Jiang, J., Wang, C., Liu, X., & Ma, J. (2021). *Deep Learning-based Face Super-Resolution: A Survey*.
- Kasinski, A. J., Florek, A., & Schmidt, A. (2008). The put face database. *Image Processing and Communications*, 13, 59–64. <https://api.semanticscholar.org/CorpusID:54105225>
- Kasiński, A., Schmidt, A., Kasinski, A., & Florek, A. (2008). *Article in Image Processing & Communications*. <https://www.researchgate.net/publication/232085001>
- Standar dan Spesifikasi Perangkat Keras, Perangkat Lunak, dan Blangko Kartu Tanda Penduduk Elektronik serta Penyelenggaraan Identitas Kependudukan Digital, Pub. L. No. 72 Year of 2022, Direktur Jenderal Peraturan Perundang-undangan Kementerian Hukum dan HAM Republik Indonesia (2022).
- Kim, S., Jun, D., Kim, B.-G., Lee, H., & Rhee, E. (2021). Single Image Super-Resolution Method Using CNN-Based Lightweight Neural Networks. *Applied Sciences*, 11(3), 1092. <https://doi.org/10.3390/app11031092>
- Kumar, A., & Chellappa, R. (2019). *Landmark Detection in Low Resolution Faces with Semi-Supervised Learning*.
- Lai, W.-S., Huang, J.-B., Ahuja, N., & Yang, M.-H. (2017). *Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution*.
- Li, P., Prieto, L., Mery, D., & Flynn, P. (2018). *Face Recognition in Low Quality Images: A Survey*.
- Li, P., Prieto, L., Mery, D., & Flynn, P. J. (2019). On Low-Resolution Face Recognition in the Wild: Comparisons and New Techniques. *IEEE Transactions on Information Forensics and Security*, 14(8), 2000–2012. <https://doi.org/10.1109/TIFS.2018.2890812>
- Li, X., Chen, C., Zhou, S., Lin, X., Zuo, W., & Zhang, L. (2020). *Blind Face Restoration via Deep Multi-scale Component Dictionaries*.
- Liang, J., Zeng, H., & Zhang, L. (2022). *Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution*.
- Nasrollahi, H., Farajzadeh, K., Hosseini, V., Zarezadeh, E., & Abdollahzadeh, M. (2020). *Deep Artifact-Free Residual Network for Single Image Super-Resolution*. <https://doi.org/10.1007/s11760-019-01569-3>
- Ou, F.-Z., Chen, X., Zhang, R., Huang, Y., Li, S., Li, J., Li, Y., Cao, L., & Wang, Y.-G. (2021). *SDD-FIQ: Unsupervised Face Image Quality Assessment with Similarity Distribution Distance*.
- Parsania, Prof. P. S., & V. Virparia, Dr. P. (2015). A Review: Image Interpolation Techniques for Image Scaling. *International Journal of Innovative Research in Computer and Communication Engineering*, 02(12), 7409–7414. <https://doi.org/10.15680/IJRCCE.2014.0212024>
- UU Republik Indonesia No. 23 Tahun 2006 tentang Administrasi Kependudukan, Pub. L. No. 23, Kementerian Hukum dan HAM RI (2006).
- Peng, Y., Spreewers, L. J., & Veldhuis, R. N. J. (2019). Low-resolution face recognition and the importance of proper alignment. *IET Biometrics*, 8(4), 267–276. <https://doi.org/10.1049/iet-bmt.2018.5008>
- Rakotonirina, N. C., & Rasoanaivo, A. (2020). *ESRGAN+: Further Improving Enhanced Super-Resolution Generative Adversarial Network*. <https://doi.org/10.1109/ICASSP40776.2020.9054071>
- Ren, H., Kheradmand, A., El-Khomy, M., Wang, S., Bai, D., & Lee, J. (2020). Real-World Super-Resolution using Generative Adversarial Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1760–1768. <https://doi.org/10.1109/CVPRW50498.2020.00226>
- Schlett, T., Rathgeb, C., Henniger, O., Galbally, J., Fierrez, J., & Busch, C. (2020). *Face Image Quality Assessment: A Literature Survey*. <https://doi.org/10.1145/3507901>
- Tian, C., Zhang, X., Lin, J. C.-W., Zuo, W., Zhang, Y., & Lin, C.-W. (2022). *Generative Adversarial Networks for Image Super-Resolution: A Survey*.
- Wang, J., Shao, Z., Huang, X., Lu, T., Zhang, R., & Li, Y. (2022). From Artifact Removal to Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–15. <https://doi.org/10.1109/TGRS.2022.3196709>
- Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards Real-World Blind Face Restoration with Generative Facial Prior. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9164–9174. <https://doi.org/10.1109/CVPR46437.2021.00905>
- Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*.
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*.