

Enhancing Hazy Image Quality with a Modular CNN Encoder–Decoder

Article History:

Received

24 November 2025

Revised

6 January 2026

Accepted

13 January 2026

ANDIKA MUHAMMAD NUR KHOLIQ¹, ARIEF SURYADI SATYAWAN², MOKH MIRZA ETNISA HAQIQI³, FAJAR RAHMAT AKBAR⁴, IASYA FAIQOH NURROHMAH⁵, AULIA ADAWIYAH⁶, ESTI FITRIA WULANDARI⁷, RENDI TRI SUGIAN⁸

^{1,3,4,5,6,7,8}Fakultas Teknik, Universitas Garut, Indonesia

²Pusat Penelitian Telekomunikasi, Badan Riset dan Inovasi Nasional (BRIN), Indonesia

Email : andikamuhammad643@gmail.com

ABSTRAK

Penelitian ini mengembangkan kerangka encoder–decoder berbasis CNN modular untuk tugas image dehazing, mengganti bottleneck konvensional dengan mekanisme token-mixing seperti FNet, Spatial-FNet, MLP-Mixer, dan gMLP. Pipeline mencakup pra-pemrosesan adaptif (CLAHE, histogram matching), augmentasi sintetik, serta pelatihan pada subset SOTS. Evaluasi numerik dan visual menunjukkan peningkatan signifikan dibanding baseline: rata-rata PSNR naik dari ≈ 18.4 dB menjadi ≈ 23.0 – 24.0 dB dan SSIM meningkat dari ≈ 0.75 menjadi ≈ 0.89 – 0.91 . Temuan ini memberikan pedoman arsitektural dan strategi pra-pemrosesan bagi sistem visi dunia nyata seperti kendaraan otonom. Rekomendasi meliputi evaluasi pada dataset nyata lebih luas, tuning hyperparameter, analisis efisiensi komputasi, dan latensi sistem.

Kata kunci: Image Dehazing, CNN, Token-mixing, Kendaraan otonom

ABSTRACT

This study develops a modular CNN encoder–decoder framework for single-image dehazing by replacing the conventional bottleneck with interchangeable token-mixing modules such as FNet, Spatial-FNet, MLP-Mixer, and gMLP-style designs. The pipeline integrates adaptive preprocessing (CLAHE and histogram matching), photometric augmentations, and training on a controlled subset of the SOTS dataset. Comprehensive quantitative and qualitative evaluations demonstrate substantial improvements over a baseline CNN, with mean PSNR increasing from approximately 18.4 dB to the 23.0–24.0 dB range and SSIM rising from about 0.75 to roughly 0.89–0.91. However, several variants require careful hyperparameter selection and loss-weight tuning to achieve stable performance. The results offer practical guidance for deployment in real-world vision systems.

Keywords: Image dehazing, CNN, Modular Token-mixing, Autonomous driving

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



1. PENDAHULUAN

Citra berkabut (*hazy image*) menjadi tantangan serius dalam bidang visi komputer karena penurunan kontras, pencahayaan yang tidak merata, dan kehilangan detail akibat hamburan cahaya di atmosfer (**Zheng, dkk, 2025**). Kondisi tersebut menyebabkan sistem berbasis visi komputer seperti kendaraan otonom, pemantauan lalu lintas, dan pengawasan keamanan gagal mengenali objek secara akurat (**Deshmukh, dkk, 2024**). Sebagaimana dijelaskan oleh Zheng, dkk partikel kabut menyebabkan cahaya menyebar dan terpantul, menghasilkan efek buram serta menurunkan kejernihan citra. Hal ini menurunkan keandalan sistem seperti deteksi objek, segmentasi semantik, dan pelacakan visual yang bergantung pada citra berkualitas tinggi (**Ayoub, dkk, 2025**). Oleh karena itu, *image dehazing* (penghilangan kabut) menjadi langkah praproses penting untuk meningkatkan kualitas visual sebelum proses analisis lanjutan (**Gui, dkk, 2021**).

Namun, pemrosesan citra berkabut merupakan masalah yang bersifat *ill-posed* karena banyak variabel atmosferik yang tidak diketahui, seperti peta transmisi dan intensitas cahaya latar belakang (**Gui, dkk, 2022**). Model fisika klasik seperti *Dark Channel Prior (DCP)* telah banyak digunakan, tetapi performanya menurun pada citra dengan kabut tebal atau area terang seperti langit (**Golts, dkk, 2020**). Selain itu, algoritma berbasis *prior* sering gagal mempertahankan detail lokal, menghasilkan citra yang terlalu gelap atau berkontras tidak alami (**Filin, dkk, 2023**). Penelitian terkini menunjukkan bahwa metode berbasis pembelajaran mendalam (*deep learning*) mampu mengatasi keterbatasan ini karena dapat mempelajari representasi fitur kompleks secara otomatis dari data (**Dakshinamurthi, dkk, 2023**). CNN (*Convolutional Neural Network*) secara khusus menjadi pilihan utama untuk tugas dehazing berkat kemampuannya mengekstraksi fitur spasial secara hierarkis dan efisien (**Pavethra, dkk 2024**) (**Jing, dkk, 2024**). Kondisi kabut/haze juga menjadi isu penting pada skenario kendaraan otonom, karena penurunan kontras dan kejernihan citra dapat mengganggu persepsi lingkungan dan menurunkan akurasi deteksi objek, sehingga berimplikasi pada keselamatan sistem berbasis visi komputer. Sejalan dengan itu, pendekatan berbasis *deep learning* tidak hanya berkembang pada CNN, tetapi juga pada model generatif seperti CycleGAN yang memungkinkan translasi citra berkabut ke citra bersih tanpa pasangan data (*unpaired*) relevan ketika *ground-truth* citra bersih sulit diperoleh pada kondisi jalan nyata. (**Sopian, dkk, 2025**)

Arsitektur CNN encoder-decoder modern untuk dehazing mengadopsi desain modular yang terdiri dari *encoder* untuk ekstraksi fitur, *bottleneck* untuk transformasi representasi, dan *decoder* untuk rekonstruksi spasial (**Sharma, dkk 2024**). Pada penelitian ini, digunakan mekanisme *token-mixing* pada lapisan bottleneck untuk menangkap konteks global, meliputi *self-attention* (**Vaswani, dkk, 2023**), *FNet* berbasis Fourier Transform (**Thorp, dkk, 2022**), *MLP-Mixer* (**Tolstikhin, dkk, 2021**), dan *gMLP* (**Liu, dkk, 2021**). Pendekatan ini memungkinkan jaringan memahami hubungan jangka panjang antar piksel tanpa kehilangan detail lokal. Selain itu, proses *preprocessing* seperti Contrast Limited Adaptive Histogram Equalization (CLAHE) (**Stimper, dkk, 2019**) dan *histogram matching* diterapkan untuk menormalkan distribusi intensitas citra dan meningkatkan kontras sebelum pemrosesan oleh CNN. Strategi ini memperkaya variasi data dan meningkatkan kemampuan generalisasi model terhadap kondisi kabut nyata. (**Gonzalez, dkk, 2018**)

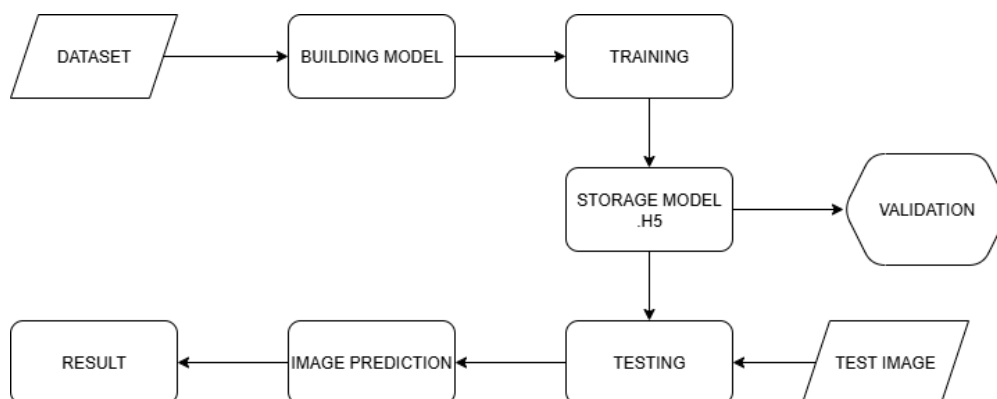
Evaluasi kuantitatif dan kualitatif menunjukkan bahwa pendekatan CNN modular dengan *token-mixing* dan *preprocessing* adaptif mampu menghasilkan peningkatan signifikan dalam metrik PSNR dan SSIM dibandingkan metode konvensional (**Heaton, dkk, 2018**). Misalnya, modifikasi berbasis *token-reduction* dengan CLAHE, melampaui baseline CNN. Peningkatan ini

juga tercermin pada hasil visual yang menunjukkan detail tekstur lebih tajam dan warna yang lebih alami (**Szeliski, 2021**). Penelitian ini memiliki kesamaan dengan beberapa studi pada berkas referensi dalam penggunaan pendekatan pembelajaran mendalam dan praktik *preprocessing* untuk meningkatkan performa model pada dataset citra yang relatif kecil, misalnya aplikasi CNN (**Supandi, dkk., 2025**) dan varian pra-latih (MobileNetV2)(**Lestari, dkk., 2024**) untuk klasifikasi dan Mask-RCNN (**Tyas, dkk., 2023**) untuk segmentasi, yang menekankan ekstraksi fitur visual dan optimisasi arsitektur untuk tugas spesifik. Namun, berbeda dari penelitian-penelitian tersebut yang berfokus pada klasifikasi atau segmentasi objek berbasis fitur warna/struktur, penelitian ini berorientasi pada restorasi citra (dehazing) dengan tujuan memperbaiki kualitas input bagi sistem visi kendaraan otonom.

2. METODE PENELITIAN

2.1 Alur Penelitian

Penelitian dimulai dari tahap pengumpulan dan pemilihan dataset, dilanjutkan dengan tahapan pra-pemrosesan citra yang meliputi normalisasi ukuran dan intensitas, kemudian pembentukan partisi data untuk pelatihan dan validasi. Selanjutnya dilakukan perancangan dan implementasi beberapa varian arsitektur model berdasar arsitektur encoder–decoder, pelatihan model dengan strategi augmentasi dan loss yang sesuai tiap varian, evaluasi pada data validasi, serta pengujian akhir pada citra uji terpisah. Alur ini digambarkan secara sistematis pada Gambar 1 sehingga memudahkan reproduksi eksperimen dan penelusuran setiap tahap eksperimen.



Gambar 1. Diagram Proses Metodologi Penelitian

2.2 Dataset dan Pembagian Data

Dataset yang digunakan merupakan bagian dari *Synthetic Object Testing Sets (SOTS)* pada platform Kaggle dan berjumlah 200 citra. Dari keseluruhan 200 citra tersebut, pembagian data untuk pelatihan dan validasi dilakukan dengan komposisi 80% untuk *training* dan 20% untuk *validation*, yang secara numerik menghasilkan 160 citra untuk pelatihan dan 40 citra untuk validasi. Selain itu, disediakan satu himpunan uji terpisah berukuran 9 citra yang digunakan sebagai *held-out test set* untuk pengukuran akhir performa model.

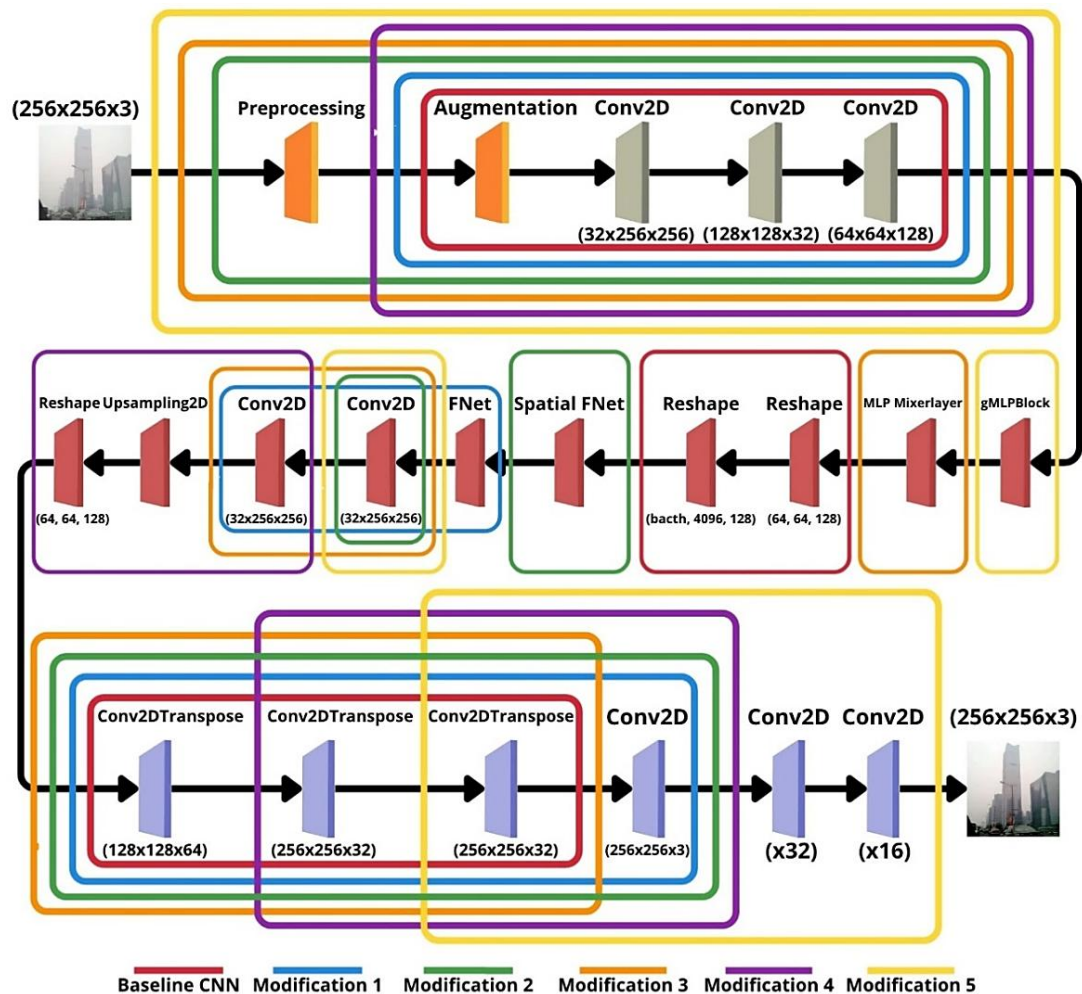
2.3 Pra-pemrosesan dan Augmentasi

Tahapan pra-pemrosesan dasar yang diterapkan pada semua eksperimen meliputi perubahan ukuran (*resize*) citra ke dimensi masukan model dan normalisasi intensitas pixel. Beberapa modifikasi menambahkan langkah-langkah pra-pemrosesan khusus seperti *histogram-matching* dan *CLAHE* (Contrast Limited Adaptive Histogram Equalization) untuk menormalkan

distribusi intensitas dan meningkatkan kontras lokal, sebagaimana dirinci pada perbandingan karakteristik model.

2.4 Arsitektur Model dan Modifikasi

Secara garis besar arsitektur yang menjadi dasar eksperimen adalah sebuah arsitektur CNN encoder-decoder standar yang mengandung mekanisme seperti Batch Normalization, Dropout, skip-add, dan bottleneck dengan self-attention pada titik sempitnya (bottleneck). Arsitektur dasar ini dan lima modifikasi utamanya dijabarkan di bawah sesuai ringkasan pada Gambar 2.



Gambar 2. Model Arsitektur Penghapusan Kabut

2.4.1 Baseline (Arsitektur CNN standar)

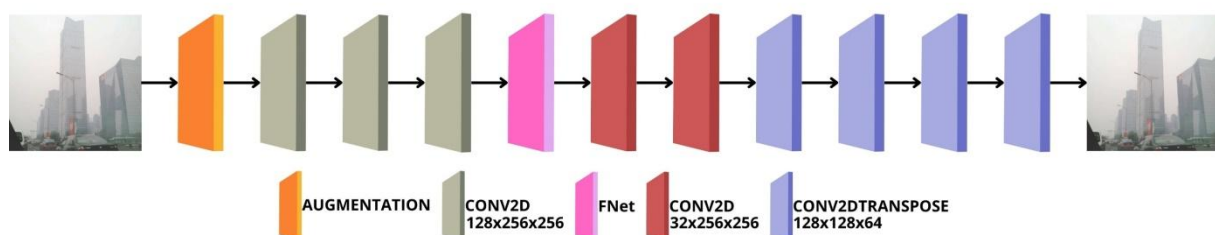
Model *baseline* adalah encoder-decoder berbasis CNN yang menggunakan operasi konvolusi berulang untuk ekstraksi fitur pada encoder dan penggabungan informasi spasial pada decoder. Mekanisme regulasi berupa BatchNorm dan Dropout dipakai untuk stabilitas pelatihan dan mencegah *overfitting*, sedangkan koneksi *skip-add* mempertahankan informasi resolusi tinggi yang hilang selama down-sampling. Fungsi loss utama yang digunakan pada baseline adalah Mean Absolute Error (MAE), yang mengukur perbedaan intensitas pixel antara keluaran model dan citra bersih target. Spesifikasi ini tercantum pada Tabel 1.

Tabel 1. Perbandingan Karakteristik Model

Model	Modifikasi dari Program Baseline	Penjelasan Modifikasi
CNN		Arsitektur CNN encoder–decoder standar dengan BatchNorm, Dropout, skip-add, bottleneck menggunakan self-attention; preprocessing hanya resize/normalisasi; loss utama MAE.
Modifikasi 1	Mengganti/menambah mekanisme bottleneck attention dengan FNet (FFT-based) + augmentasi fotometrik/geometrik dan simulasi kabut.	FNet melakukan token-mixing di domain frekuensi untuk menangkap konteks global dengan biaya komputasi lebih rendah; augmentasi sintesis meningkatkan generalisasi pada dataset kecil.
Modifikasi 2	Menambahkan preprocessing histogram-matching dan CLAHE, serta mengaplikasikan FNet pada feature-map (spatial FNet).	Preprocessing menormalkan distribusi intensitas antara domain, sementara Spatial-FNet memadukan informasi frekuensi pada level fitur untuk memperbaiki rekonstruksi struktur visual.
Modifikasi 3	Menempatkan MLP-Mixer di bottleneck dengan patching pada feature-map menggantikan attention.	MLP-Mixer melakukan token-mixing lewat dense layer sehingga menggabungkan informasi spasial-token secara efektif dan mengoptimalkan pemulihan detail struktural dengan kompleksitas parameter yang terkontrol.
Modifikasi 4	MLP-Mixer terstruktur (configurable patch/size, default input 128×128) + loss gabungan (MAE + (1–SSIM) ± perceptual) dan augmentasi lanjutan.	Bertujuan meningkatkan kualitas perseptual keluaran; membutuhkan penyetelan hyperparameter (ukuran input, bobot loss) agar konvergensi dan kinerja numerik optimal.
Modifikasi 5	Mengaplikasikan gMLP-style: pooling atau token-reduction sebelum token-mixing lalu upsample kembali; ditambah CLAHE/histmatch.	Pooling mengurangi panjang urutan token sehingga menekan biaya komputasi saat mixing, mempertahankan konteks global namun berpotensi kehilangan detail halus yang memerlukan tuning pool size.

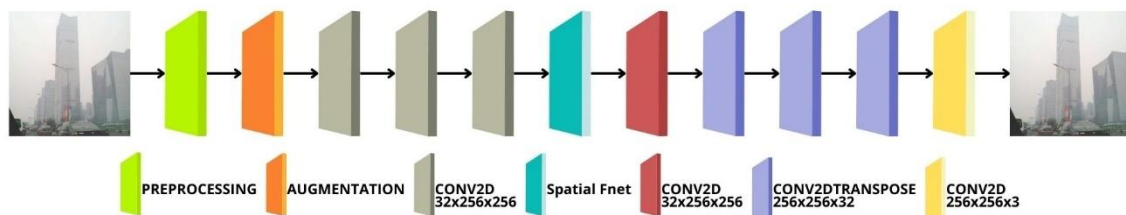
2.4.2 Modifikasi 1 — FNet (FFT-based) + Augmentasi Sintetis

Modifikasi 1 menggantikan mekanisme *bottleneck attention* tradisional dengan FNet (token-mixing berbasis FFT) untuk menangkap konteks global pada biaya komputasi rendah. Pada proses *forward*, encoder CNN mengekstrak fitur, lalu FNet melakukan pencampuran token di bottleneck untuk menyebarkan konteks global, dan decoder merekonstruksi citra bersih. Gambar 3 memperlihatkan arsitektur model Modifikasi 1.

**Gambar 3. Arsitektur Model pada Modifikasi 1**

2.4.3 Modifikasi 2 — Spatial-FNet + Histogram-matching/CLAHE

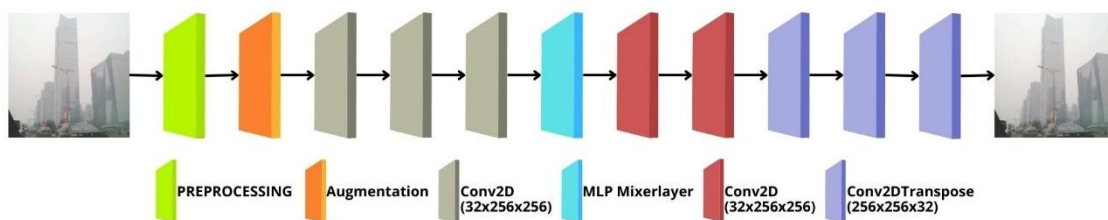
Modifikasi 2 menambahkan pra-pemrosesan berbasis histogram-matching dan CLAHE untuk menormalkan distribusi intensitas antardomain, serta menerapkan FNet pada peta fitur internal (*spatial FNet*). Dengan pendekatan ini, pencampuran di domain frekuensi tidak hanya terjadi pada token input awal, tetapi juga pada representasi fitur selama pemrosesan, sehingga model menggabungkan informasi frekuensi dan spasial pada tingkat fitur. Proses kerjanya dimulai dari citra input yang menjalani pra-pemrosesan (histogram-matching/CLAHE), kemudian melewati encoder; pada bottleneck, *spatial-FNet* melakukan pencampuran token pada peta fitur, lalu decoder merekonstruksi citra akhir. Kombinasi tersebut ditujukan meningkatkan rekonstruksi struktur visual yang sangat peka terhadap perbedaan sebaran intensitas. Gambar 4 menunjukkan arsitektur Modifikasi 2.



Gambar 4. Arsitektur Model pada Modifikasi 2

2.4.4 Modifikasi 3 — MLP-Mixer di Bottleneck

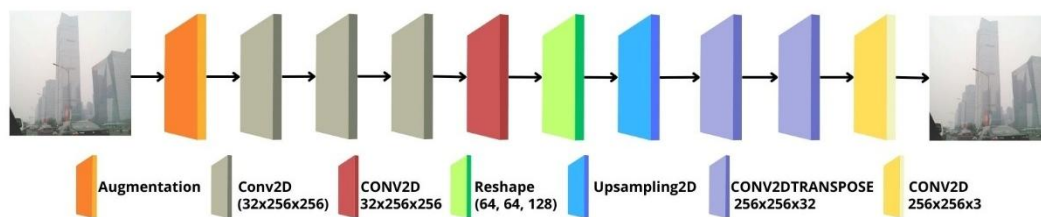
Pada Modifikasi 3, mekanisme *attention* di bottleneck digantikan oleh MLP-Mixer yang bekerja dengan membagi feature-map menjadi *patches*. Setiap patch diproses melalui lapisan-lapisan fully-connected (dense) sehingga terjadi pertukaran konteks antar-token secara efektif tanpa mengandalkan perhatian tradisional. Proses kerja mencakup pemisahan peta fitur menjadi sejumlah *patch*, pemrosesan antar-token melalui lapisan MLP untuk pertukaran konteks, dan kemudian penggabungan kembali patch sebelum diteruskan ke decoder. Pendekatan ini bertujuan untuk memulihkan detail struktural citra sambil menjaga kompleksitas parameter tetap terkontrol. Gambar 5 mengilustrasikan arsitektur Modifikasi 3.



Gambar 5. Arsitektur Model pada Modifikasi 3

2.4.5 Modifikasi 4 — MLP-Mixer Terstruktur + Loss Gabungan

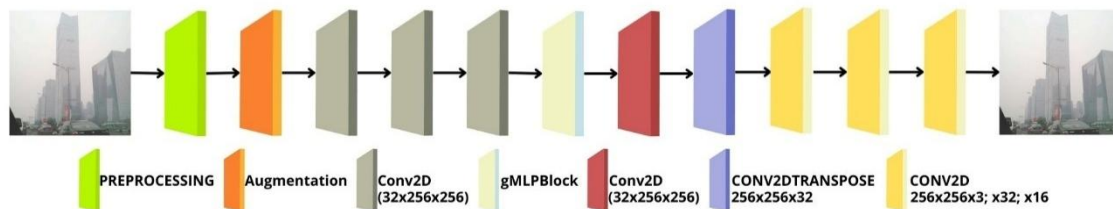
Modifikasi 4 mengembangkan Modifikasi 3 dengan MLP-Mixer yang lebih terstruktur (ukuran patch dapat dikonfigurasi; input default 128×128) serta penerapan fungsi loss gabungan. Strategi ini bertujuan meningkatkan kualitas perseptual keluaran agar hasil rekonstruksi tidak hanya minim kesalahan numerik tetapi juga tampil lebih tajam dan alami secara visual. Gambar 6 memperlihatkan arsitektur Modifikasi 4.



Gambar 6. Arsitektur Model pada Modifikasi 4

2.4.6 Modifikasi 5 — gMLP-style dengan Token-reduction + CLAHE/Histmatch

Modifikasi 5 mengadopsi gaya gMLP dengan mekanisme pooling/token-reduction sebelum tahap token-mixing, dilanjutkan upsampling sebelum decoding. Dengan mengurangi panjang urutan token, pendekatan ini menekan biaya komputasi pada tahap pencampuran, tetap mempertahankan konteks global sekaligus berisiko menghilangkan beberapa detail halus yang memerlukan penyetelan ukuran pooling yang tepat. Gambar 7 menampilkan arsitektur Modifikasi 5.



Gambar 7. Arsitektur Model pada Modifikasi 5

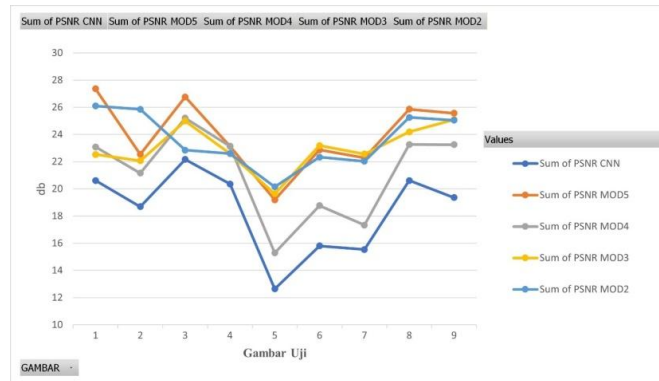
2.5 Prosedur Pelatihan, Validasi, dan Pengujian

Pelatihan setiap varian model mengikuti prosedur yang sama: data pelatihan (160 citra) diberikan augmentasi dan pra-pemrosesan sesuai desain eksperimen, lalu model dilatih untuk meminimalkan fungsi loss tertentu. Kinerja model dipantau pada set validasi (40 citra) untuk memilih *checkpoint* terbaik dan menentukan *early stopping*. Setelah model terbaik terpilih berdasarkan metrik pada data validasi, evaluasi akhir dilakukan pada himpunan uji terpisah (9 citra) untuk mengukur kemampuan generalisasi model pada kondisi yang belum pernah dilihat sebelumnya.

3. HASIL DAN PEMBAHASAN

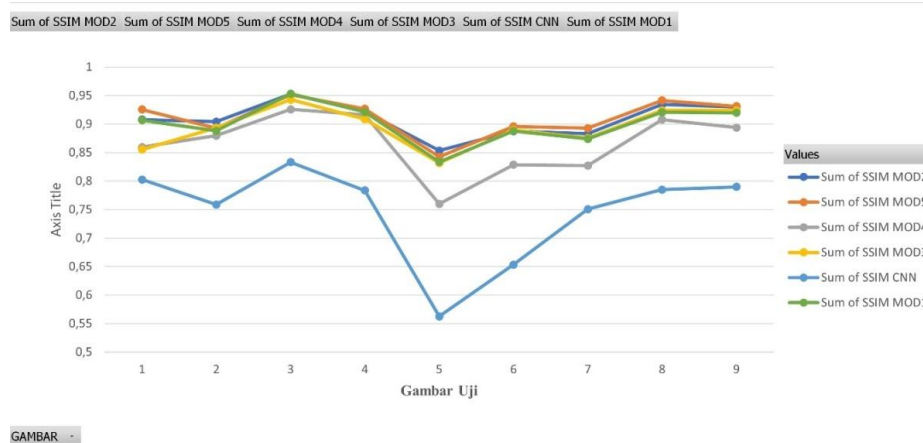
Pada bagian hasil dan pembahasan disajikan evaluasi kinerja model dehazing secara kuantitatif dan kualitatif, dimulai dari perbandingan metrik PSNR dan SSIM pada baseline CNN serta lima varian modifikasi untuk menilai peningkatan kualitas rekonstruksi dan konsistensi hasil.

Selanjutnya, analisis visual pada kondisi outdoor, indoor, dan citra nyata digunakan untuk meninjau pemulihan kontras, detail, dan kestabilan warna, sekaligus mengidentifikasi artefak seperti residu kabut, pergeseran tonal, dan potensi over-enhancement.



Gambar 8. Perbandingan Nilai PSNR untuk Setiap Gambar Uji

Gambar 8 dan Gambar 9 memperlihatkan perbandingan PSNR dan SSIM untuk setiap gambar uji. Secara numerik, model baseline CNN mencapai PSNR rata-rata ≈ 18.42 dB dan SSIM rata-rata ≈ 0.747 , sedangkan semua modifikasi arsitektur memberikan peningkatan signifikan dengan PSNR rata-rata berkisar ≈ 22.98 – 23.95 dB dan SSIM rata-rata ≈ 0.894 – 0.911 .



Gambar 9. Perbandingan Nilai SSIM untuk Setiap Gambar Uji

Data pada Tabel 2 menunjukkan bahwa hasil baseline mengalami variasi yang lebih besar dan beberapa sampel memiliki PSNR rendah, mencerminkan kegagalan relatif dalam memulihkan citra yang sangat terdegradasi. Secara keseluruhan, peningkatan angka tersebut menandakan perbaikan baik dalam kesamaan piksel murni (PSNR) maupun kesamaan struktural/perseptual (SSIM), yang lebih relevan untuk kualitas visual.

Tabel 2. Hasil Pengujian Model CNN

Model	PSNR (dB)	SSIM
CNN	20,62	0,8027
	18,69	0,7587
	22,17	0,8332
	20,36	0,7835
	12,64	0,5626
	15,8	0,6533
	15,54	0,7507
	20,61	0,7851
	19,36	0,7898

Modifikasi 1 menunjukkan peningkatan substansial dibandingkan baseline: PSNR rata-rata ≈ 23.74 dB dan SSIM rata-rata ≈ 0.901 . Nilai pada Tabel 3 memperlihatkan PSNR sangat tinggi pada beberapa sampel dan SSIM mendekati atau melebihi 0.90 pada mayoritas sampel. Peningkatan ini sesuai dengan desain Modifikasi 1 yang menggantikan bottleneck attention dengan FNet (token-mixing berbasis FFT) dan menggunakan augmentasi sintetis, yang bersama-sama meningkatkan kemampuan model menangkap konteks global dan memperbaiki generalisasi pada dataset terbatas. Hasil pada Tabel 3 mendukung argumen bahwa pemrosesan di domain frekuensi (FNet) beserta augmentasi sintetis efektif meningkatkan rekonstruksi citra dehazing.

Tabel 3. Hasil Pengujian Modifikasi 1

Model	PSNR (dB)	SSIM
Modifikasi 1	26,6941	0,9068
	22,3736	0,8881
	26,6358	0,9531
	23,2667	0,9213
	19,487	0,8338
	22,5615	0,8879
	22,0321	0,873968
	24,2468	0,9208
	26,3308	0,9199

Modifikasi 2 mempertahankan kinerja tinggi dengan PSNR rata-rata ≈ 23.58 dB dan SSIM rata-rata ≈ 0.908 . Data pada Tabel 4 menunjukkan Modifikasi 2 cenderung menghasilkan nilai SSIM sangat baik, mengindikasikan rekonstruksi struktur visual yang sangat baik. Hal ini konsisten dengan strategi pra-pemrosesan (histogram-matching dan CLAHE) yang menormalkan distribusi intensitas sehingga spatial-FNet dapat bekerja lebih stabil pada level fitur dan mempertahankan struktur lokal.

Table 4. Hasil Pengujian Modifikasi 2

Model	PSNR (dB)	SSIM
Modifikasi 2	26,1	0,9078
	25,85	0,9041
	22,85	0,9528
	22,6	0,9219
	20,15	0,8533
	22,33	0,8877
	22,03	0,8829
	25,26	0,9342
	25,05	0,93

Modifikasi 3 memperlihatkan peningkatan terhadap baseline dengan rata-rata PSNR ≈ 22.98 dB dan SSIM ≈ 0.894 . Nilai-nilai pada Tabel 5 menunjukkan konsistensi performa pada sebagian besar sampel, menandakan bahwa MLP-Mixer mampu menggabungkan konteks spasial-token secara efektif dan memulihkan detail struktural tanpa kompleksitas attention penuh.

Table 5. Hasil Pengujian Modifikasi 3

Model	PSNR (dB)	SSIM
Modifikasi 3	22,53	0,8555
	22,07	0,8936
	24,99	0,9428
	22,59	0,9088
	19,64	0,8318
	23,18	0,8895
	22,56	0,8755
	24,19	0,9242
	25,09	0,924

Modifikasi 4 menampilkan variasi hasil yang lebih besar; PSNR rata-rata ≈ 21.16 dB dan SSIM rata-rata ≈ 0.867 , secara umum lebih rendah dibanding modifikasi lain kecuali baseline. Tabel 6 menunjukkan beberapa sampel dengan PSNR dan SSIM tinggi serta beberapa sampel dengan penurunan performa, pola yang konsisten dengan sifat Modifikasi 4 yang memperkenalkan loss gabungan dan konfigurasi patch yang lebih sensitif. Artinya, Modifikasi 4 dapat menghasilkan kualitas perseptual yang sangat baik jika bobot loss dan *hyperparameter* disetel dengan tepat, namun juga rentan terhadap ketidaktepatan penyetelan sehingga menimbulkan variasi performa antarsampel.

Table 6. Hasil Pengujian Modifikasi 4`

Model	PSNR (dB)	SSIM
Modifikasi 4	23,08	0,8597
	21,16	0,8796
	25,21	0,9259
	23,12	0,9163
	15,29	0,7599
	18,77	0,8286
	17,33	0,8273
	23,26	0,9078
	23,25	0,8938

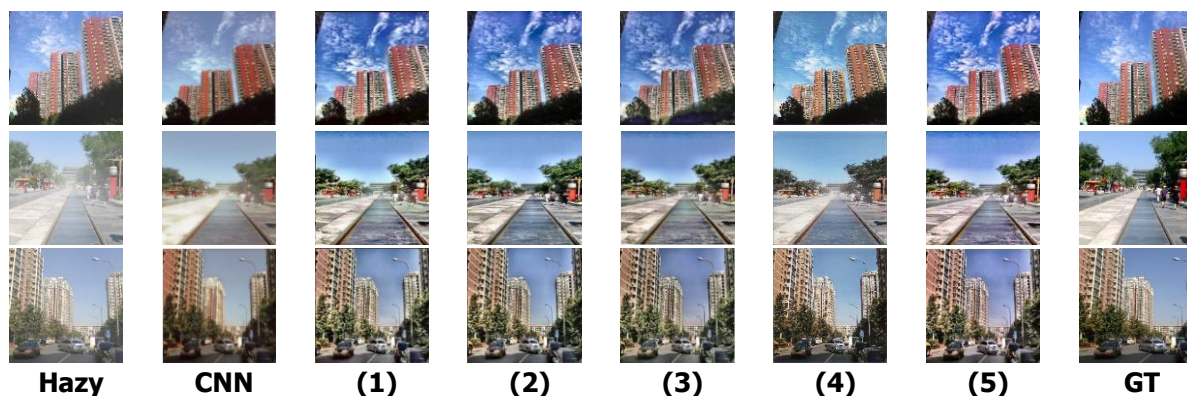
Modifikasi 5 mencatat performa terbaik secara agregat: PSNR rata-rata ≈ 23.95 dB dan SSIM rata-rata ≈ 0.911 . Tabel 7 menunjukkan beberapa nilai PSNR/SSIM tertinggi pada kumpulan sampel (misalnya PSNR ≈ 27.36 dB dengan SSIM ≈ 0.9255), menandakan kemampuan Modifikasi 5 menjaga konteks global melalui token-reduction sekaligus memulihkan detail struktural relevan. Hasil ini mengindikasikan bahwa strategi pooling/token-reduction yang hati-hati dapat menekan biaya komputasi selama pencampuran token tanpa mengorbankan kualitas rekonstruksi jika ukuran pooling dioptimalkan; namun terdapat risiko kehilangan detail halus jika ukuran pooling tidak disesuaikan dengan tepat. Selain unggul secara metrik, keluaran Modifikasi 5 cenderung lebih stabil pada variasi tingkat kabut karena kombinasi token-reduction dan pra-pemrosesan adaptif membantu menyeimbangkan pemulihan kontras dengan konservasi struktur utama.

Table 7. Hasil Pengujian Modifikasi 5

Model	PSNR (dB)	SSIM
Modifikasi 5	27,36	0,9255
	22,54	0,8926
	26,76	0,9515
	23,13	0,9267
	19,19	0,8425
	22,87	0,8959
	22,27	0,8927
	25,86	0,9415
	25,56	0,9313

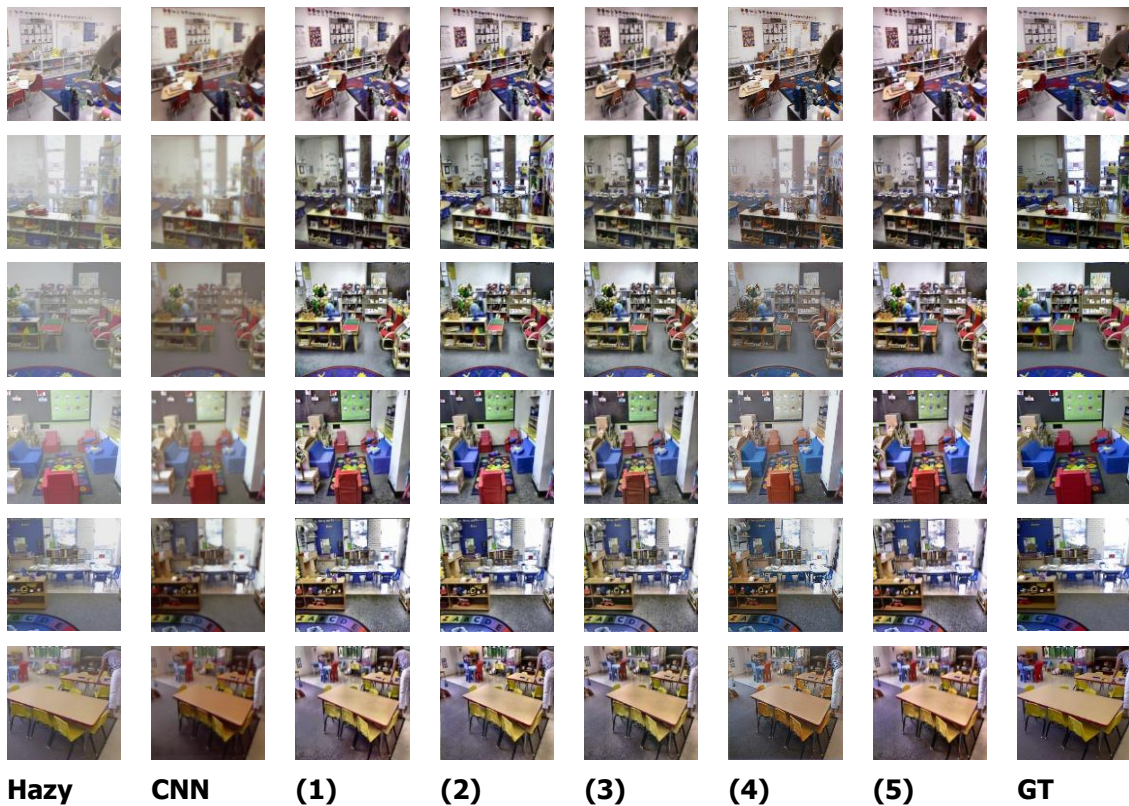
Pada Gambar 10 disajikan perbandingan hasil *dehazing* pada kondisi outdoor dengan urutan kolom Hazy (citra input berkabut), CNN (prediksi model *baseline*), (1)–(5) (prediksi dari lima varian model hasil modifikasi), serta GT (citra bersih/*ground truth*). Secara kualitatif, citra Hazy menunjukkan penurunan kontras global dan hilangnya detail pada area berjarak (misalnya tekstur fasad bangunan, batas vegetasi, serta permukaan jalan/rel) akibat dominasi kabut yang meningkatkan *veiling light*. Hasil CNN telah mampu mengurangi kabut dan memulihkan sebagian struktur, namun masih tampak adanya residu kabut pada area tertentu serta ketidakstabilan reproduksi warna/kontras pada beberapa sampel (misalnya langit cenderung kurang natural atau bayangan menjadi lebih pekat).

Dibandingkan *baseline*, keluaran model (1)–(5) umumnya memperlihatkan pemulihan visibilitas yang lebih baik melalui peningkatan kontras dan ketegasan tepi objek, sehingga detail visual tampak lebih mendekati GT.

**Gambar 10. Visualisasi Hasil Pengujian pada Kondisi Outdoor**

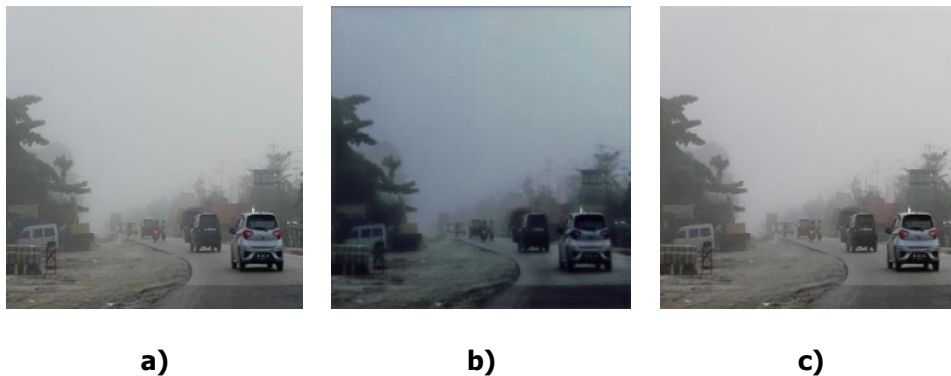
Gambar 11 menampilkan hasil dehazing pada kondisi indoor dengan format kolom yang sama (Hazy, CNN, (1)–(5), dan GT). Pada skenario indoor, dehazing lebih menantang karena pencahayaan tidak merata, pantulan, serta kontras terang–gelap, sehingga citra Hazy tampak washed-out dan detail kecil mudah hilang. Prediksi CNN mampu mengurangi kabut, tetapi pada beberapa contoh masih muncul ketidakakonsistenan tonal seperti citra lebih gelap atau color cast sehingga kemiripannya terhadap GT belum stabil. Secara umum, modifikasi (1)–(5) meningkatkan keterbacaan struktur dan kontras lokal sehingga detail spasial lebih jelas dan hasil lebih mendekati GT, namun peningkatan kontras juga berpotensi memicu amplifikasi

noise pada area gelap, pergeseran warna, atau residu kabut tipis pada area backlight; karena itu diperlukan keseimbangan antara penguatan detail dan kestabilan warna agar rekonstruksi tetap realistis lintas adegan.



Gambar 11. Visualisasi Hasil Pengujian pada Kondisi Indoor

Gambar 12 menyajikan perbandingan hasil prediksi dehazing pada citra nyata, yang terdiri atas (a) citra hazy sebagai masukan, (b) keluaran model baseline CNN, dan (c) keluaran model Modifikasi 5 sebagai model dengan nilai agregat tertinggi. Berdasarkan Gambar 12, model baseline CNN (b) cenderung menghasilkan perubahan tonal dan pergeseran warna yang lebih kuat, ditunjukkan oleh tampilan yang lebih gelap serta munculnya bias warna kebiruan/keunguan, sehingga naturalitas citra menurun meskipun terjadi peningkatan kontras pada beberapa area. Sebaliknya, model Modifikasi 5 (c) memberikan hasil yang lebih stabil secara visual



Gambar 12. Hasil Prediksi pada Citra Nyata; a) Hazy; b) Baseline CNN; c) Modifikasi 5

4. KESIMPULAN

Berdasarkan serangkaian eksperimen dan analisis yang dijelaskan mulai dari pendahuluan hingga hasil penelitian, dapat disimpulkan bahwa integrasi arsitektur CNN encoder–decoder dengan mekanisme token-mixing modern (FNet/FFT-based, MLP-Mixer, gMLP) serta strategi pra-pemrosesan adaptif (CLAHE dan histogram-matching) secara konsisten meningkatkan kualitas pemulihan citra berkabut dibandingkan baseline CNN konvensional; hal ini tercermin pada kenaikan metrik numerik dan perbaikan perseptual yang nyata, di mana baseline CNN memiliki kinerja rata-rata PSNR ≈ 18.42 dB dan SSIM ≈ 0.747 , sedangkan variasi modifikasi menghasilkan rentang rata-rata PSNR ≈ 22.98 – 23.95 dB dan SSIM ≈ 0.894 – 0.911 , dengan Modifikasi 1 (FNet + augmentasi) dan Modifikasi 5 (gMLP-style + CLAHE/histmatch) menempati posisi terbaik secara agregat, namun demikian hasil eksperimen juga menunjukkan bahwa meskipun beberapa varian berpotensi menghasilkan keluaran perseptual yang sangat baik, pendekatan tersebut rentan terhadap varians performa apabila hyperparameter dan bobot loss tidak disetel secara cermat, sehingga temuan ini menegaskan pentingnya penyetelan hyperparameter dan pemilihan bobot loss yang teliti untuk merealisasikan keuntungan teoretis metode-metode tersebut.

UCAPAN TERIMA KASIH

Dalam penelitian ini dari penulis ingin mengucapkan terima kasih kepada Fakultas Teknik Universitas Garut dan Badan Riset dan Inovasi Nasional (BRIN) atas penyediaan sumber daya dan bentuk fasilitas yang mendukung proses pelaksanaan penelitian.

DAFTAR RUJUKAN

- Abeer Ayoub, Walid El-Shafai, Fathi E. Abd El-Samie, Ehab K. I. Hamad, & El-Sayed M. EL-Rabaie. (2025). Video and image quality improvement using an enhanced optimized dehazing technique. *Multimedia Tools and Applications*, 84, 22681–22699.
- Alona Golts, Daniel Freedman, & Michael Elad. (2020). Unsupervised Single Image Dehazing Using Dark Channel Prior Loss. *IEEE Transactions on Image Processing*, 29, 2692–2701.
- Dakshinamurthi, V., G.P., S., P, M., & Hussain, Sk. R. (2023). DEEP LEARNING-BASED IMAGE DEHAZING AND VISIBILITY ENHANCEMENT FOR IMPROVED VISUAL PERCEPTION. *ICTACT Journal on Image and Video Processing*, 14(2), 3122–3128. <https://doi.org/10.21917/ijivp.2023.0444>
- Deshmukh, P. V., Kumar, A., & Chakrabarti, P. (2024). Review of Deep Learning Based Image Dehazing for Autonomous Vehicle. In *Tuijin Jishu/Journal of Propulsion Technology* (Vol. 45, Issue 4).
- Filin, A., Kopylov, A., & Gracheva, I. (2023). A SINGLE IMAGE DEHAZING DATASET with LOW-LIGHT REAL-WORLD INDOOR IMAGES, DEPTH MAPS and INFRARED IMAGES. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information*

- Sciences - ISPRS Archives*, 48(2/W3-2023), 53–57. <https://doi.org/10.5194/isprs-archives-XLVIII-2-W3-2023-53-2023>
- Gonzalez, R. C. ., & Woods, R. E. . (2018). *Digital image processing*. Pearson.
- Gui, J., Cong, X., Cao, Y., Ren, W., Zhang, J., Zhang, J., Cao, J., & Tao, D. (2022). *A Comprehensive Survey and Taxonomy on Single Image Dehazing Based on Deep Learning*. <http://arxiv.org/abs/2106.03323>
- Gui, J., Cong, X., Cao, Y., Ren, W., Zhang, J., Zhang, J., & Tao, D. (2021). *A Comprehensive Survey on Image Dehazing Based on Deep Learning*.
- Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines*, 19(1–2), 305–307. <https://doi.org/10.1007/s10710-017-9314-z>
- Jing, H., Chen, J., Zhang, C., Wei, S., Chen, A., & Zhang, M. (2024). Adaptive Multi-Feature Attention Network for Image Dehazing. *Electronics (Switzerland)*, 13(18). <https://doi.org/10.3390/electronics13183706>
- Kaiming He, Xiaoou Tang, & Jian Sun. (2011). *Single image haze removal using dark channel prior* (Vol. 33). IEEE.
- Lee-Thorp, J., Ainslie, J., Eckstein, I., & Ontanon, S. (2022). *FNet: Mixing Tokens with Fourier Transforms*. <http://arxiv.org/abs/2105.03824>
- Lestari, I. S., Jumadi, J., & Lukman, N. (2024). IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK DENGAN PRE-TRAINED MODEL MOBILENETV2 UNTUK DETEKSI KOLESTEROL. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 9(2), 173–183. <https://doi.org/10.36341/rabit.v9i2.4732>
- Liu, H., Dai, Z., So, D. R., & Le, Q. V. (2021). *Pay Attention to MLPs*. <http://arxiv.org/abs/2105.08050>
- Pavethra, M., & Uma Devi, M. (2024). A cross layer graphical neural network based convolutional neural network framework for image dehazing. *Automatika*, 65(3), 1139–1153. <https://doi.org/10.1080/00051144.2024.2346964>
- Sharma, T., & Verma, N. K. (2024). *Artificial Intelligent Algorithms for Image Dehazing and Non-Uniform Illumination Enhancement*. Springer Nature Singapore. <https://books.google.co.id/books?id=3wIPEQAAQBAJ>
- Sopian, S. M., Satyawan, A. S., Haqiqi, M. M. E., Susilawati, H., Wijaya, B., Artemysia, K. A., Firman, & Samie, M. I. (2025). Development of a Modified CycleGAN Model with Residual Blocks and Perceptual Loss for Image Dehazing. *Teknika*, 14(2), 232–238. <https://doi.org/10.34148/teknika.v14i2.1235>

- Stimper, V., Bauer, S., Ernstorfer, R., Schölkopf, B., & Xian, R. P. (2019). Multidimensional Contrast Limited Adaptive Histogram Equalization. *IEEE Access*, 7, 165437–165447. <https://doi.org/10.1109/ACCESS.2019.2952899>
- Supandi S, Ghofur S, & Santoso F. (2025). Deteksi Kualitas Rumput Laut Menggunakan Metode Convelution Neural Network (Cnn) Berdasarkan Citra Digital (Studi Kasus : Desa Alasmalang Kecamatan Raas Sumenep). *Rabit: J. Teknol. Dan Sistem Informasi Univrab*, 10, 604–614. <https://doi.org/10.36341/rabit.v10i2.7273>
- Szeliski, R. (2021). *Computer Vision: Algorithms and Applications 2nd Edition*. <https://szeliski.org/Book>,
- Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., & Dosovitskiy, A. (2021). *MLP-Mixer: An all-MLP Architecture for Vision*. <http://arxiv.org/abs/2105.01601>
- Tyas, D. A., & Ratnaningsih, T. (2023). Implementasi Mask-Rcnn Pada Dataset Kecil Citra Sel Darah Merah Berdasarkan Kriteria Warna Sel. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 8(1), 100–104. <https://doi.org/10.36341/rabit.v8i1.3026>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. <http://arxiv.org/abs/1706.03762>
- Zheng, C., Ying, W., & Hu, Q. (2025). Comparative analysis of dehazing algorithms on real-world hazy images. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-95510-z>