ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Pengembangan Arsitektur Modular Berbasis CNN pada Peningkatan Kualitas Citra Berkabut

Andika Muhammad Nur Kholiq[1], Arief Suryadi Satyawan[2], Mokh Mirza Etnisa Haqiqi[3], Fajar Rahmat Akbar[4], Iasya Faiqoh Nurrohmah[5], Aulia Adawiyah[6], Esti Fitria Wulandari[7], Rendi Tri Sugian[8], Nurul Fazri[9]

[1,3,4,5,6,7,8,9] Teknik Elektro, Universitas Garut
[2] Badan Riset dan Inovasi Nasional
[1]24052122040@ftenik.uniga.ac.id
[2]arie021@brin.go.id
[3]mirza@uniga.ac.id
[4]24052122044@fteknik.uniga.ac.id
[5]24052122011@fteknik.uniga.ac.id
[6]24052122002@fteknik.uniga.ac.id
[7]24052122003@fteknik.uniga.ac.id
[8]24052122025@fteknik.uniga.ac.id
[9]24052320032@fteknik.uniga.ac.id

*Corresponding author email: 24052122040@fteknik.uniga.ac.id

***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 hiperparameter, analisis efisiensi komputasi, dan latensi sistem.**

Kata Kunci— **Image Dehazing, CNN, Token-mixing, Kendaraan Otonom.**

I. PENDAHULUAN

Citra berkabut (*haze*) mengakibatkan penurunan kualitas visual yang signifikan akibat hamburan partikel di atmosfer, yang berdampak pada berkurangnya visibilitas, kontras, serta terjadinya distorsi warna pada citra yang tertangkap[1]. Kondisi ini menyebabkan citra hasil tangkapan tidak memadai untuk tugas-tugas visi komputer kritis, seperti deteksi objek dan pelacakan, terutama pada sistem nyata seperti kendaraan otonom dan pengawasan keamanan[2]. Degradasi kualitas tersebut secara langsung menurunkan keandalan sistem pengenalan otomatis, sehingga *image dehazing* menjadi tahap prapemrosesan yang krusial sebelum analisis lebih lanjut dilakukan[3].

Persoalan *dehazing* citra satu *frame* bersifat *ill-posed* karena model hamburan atmosfer (*Atmospheric Scattering Model*/ASM) membutuhkan estimasi parameter yang kompleks dan tidak diketahui secara langsung[4]. Metode berbasis *prior* seperti *Dark Channel Prior* (DCP) rentan menghasilkan estimasi transmisi yang tidak akurat serta artefak visual, khususnya pada area langit cerah yang tidak sesuai dengan asumsi DCP[5].

Convolutional Neural Network (CNN) mampu memetakan secara langsung citra berkabut ke citra bersih tanpa bergantung pada model *priors[6]*. Pengembangan lebih lanjut mengintegrasikan modul *token-mixing* seperti FNet, MLP-Mixer, dan gMLP ke dalam kerangka *encoder–decoder* CNN untuk menangkap konteks spasial global yang tidak terjangkau oleh CNN konvensional[7]. Pendekatan ini terbukti meningkatkan metrik kualitas visual secara signifikan, dengan rata-rata PSNR meningkat dari sekitar 18,4 dB menjadi 23,0–24,0 dB dan SSIM dari sekitar 0,75 menjadi 0,89–0,91 dibandingkan *baseline* CNN tanpa modul *token-mixing[8]*. Prapemrosesan adaptif berbasis CLAHE dan pencocokan histogram juga terbukti memperkuat generalisasi model dengan menormalkan distribusi intensitas citra sebelum masuk ke jaringan[9].

Meskipun demikian, terdapat *research gap* yang masih belum terselesaikan. Sebagian besar metode dilatih menggunakan pasangan citra sintetis seperti dataset RESIDE/SOTS yang karakteristik kabutnya berbeda secara fundamental dengan kondisi nyata, sementara dataset berpasangan dari dunia nyata masih sangat terbatas skalanya[10]. Metode *semi-supervised* maupun *unsupervised* yang dikembangkan untuk mengatasi kelangkaan data nyata, seperti varian CycleGAN, masih sangat bergantung pada data sintetik dan *prior-based constraint* yang tidak selalu berlaku[11]. Kondisi ini mengakibatkan kesenjangan performa yang nyata antara pengujian pada *benchmark* sintetis dan generalisasi ke kondisi lapangan, yang diperparah oleh minimnya evaluasi terintegrasi pada skenario dunia nyata[12]. Dalam konteks kendaraan otonom, kesenjangan ini sangat kritis karena tidak ada satu pun

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

algoritma *dehazing* yang secara konsisten menghasilkan kualitas tinggi pada seluruh kondisi pengujian nyata[13].
Berdasarkan identifikasi *gap* tersebut, penelitian ini mengembangkan arsitektur CNN modular berbasis *token-mixing* yang dipadukan dengan prapemrosesan adaptif dan dievaluasi secara sistematis pada kondisi kabut nyata. Berbeda dari kajian sebelumnya yang umumnya terbatas pada pengujian *benchmark* sintetis, penelitian ini secara spesifik menargetkan restorasi citra berkabut sebagai tahap prapemrosesan kritis guna meningkatkan kualitas masukan sistem visi kendaraan otonom pada kondisi dunia nyata. Oleh karena itu, penelitian ini memilki gap dengan beberapa referensi lain yang dijelaskan pada tabel 1.

Tabel 1.
State of The Art

Referensi	Metode Inti	Domain Tugas	Kontribusi Kunci	Gap
[5]	Prior-based dehazing menggunakan *Dark Channel Prior* dan model hamburan atmosfer	Single-image dehazing pada citra outdoor	Menunjukkan bahwa prior statistik sederhana dapat dipakai untuk memperkirakan *transmission* dan memulihkan citra berkabut tanpa proses pelatihan data	Metode ini bergantung pada asumsi statistik citra, sehingga kurang efektif pada kabut tebal dan area terang. Penelitian ini mengatasi keterbatasan tersebut melalui pendekatan deep learning yang lebih adaptif.
[12]	Encoder–decoder CNN untuk estimasi *transmission map*, lalu restorasi citra menggunakan model fisik	Single-image dehazing untuk aplikasi visi, termasuk kendaraan otonom	Mengintegrasikan pembelajaran mendalam dengan parameter fisik dehazing sehingga proses restorasi lebih terarah	Pendekatan masih bergantung pada estimasi parameter fisik sehingga belum sepenuhnya end-to-end. Penelitian ini menggunakan pendekatan end-to-end berbasis CNN modular.
[14]	GAN-based dehazing dengan CycleGAN yang dimodifikasi pada generator dan fungsi loss	Single-image dehazing pada data *unpaired*	Memungkinkan pelatihan tanpa pasangan hazy–clear, sehingga lebih fleksibel saat data berpasangan terbatas	Pelatihan GAN cenderung tidak stabil dan hasil bergantung pada desain loss. Penelitian ini menggunakan pendekatan supervised CNN yang lebih stabil.
[15]	Transformer end-to-end dengan satu encoder–decoder, *intra-patch transformer blocks*, dan *learnable weather type embeddings*	Restorasi citra cuaca buruk: hujan, kabut, dan salju	Menawarkan satu model efisien untuk beberapa jenis degradasi cuaca dan menunjukkan peningkatan pada berbagai dataset uji	Fokus pada multi-weather restoration dan memiliki kompleksitas tinggi. Penelitian ini berfokus khusus pada dehazing dengan metode yang lebih ringan.
[16]	Vision Transformer yang dimodifikasi melalui normalisasi, aktivasi, dan skema agregasi informasi spasial	Single-image dehazing pada berbagai dataset	Membuktikan bahwa transformer dapat diadaptasi untuk dehazing dan mencapai performa tinggi pada SOTS indoor, serta memperkenalkan dataset realistis untuk haze non-homogen	Memerlukan arsitektur transformer yang kompleks dan komputasi tinggi. Penelitian ini menggunakan token-mixing alternatif yang lebih efisien dalam kerangka CNN.

II. TINJAUAN PUSTAKA

A. Pendekatan Klasik Berbasis Prior

Pendekatan klasik *dehazing* umumnya memanfaatkan prior statistik pada citra alami, seperti *Dark Channel Prior (DCP),* untuk membantu estimasi komponen transmisi dan *airlight*. Walaupun efektif pada kondisi tertentu, metode berbasis prior cenderung menurun kinerjanya pada kabut yang sangat tebal atau area terang seperti langit, serta berisiko menghasilkan kontras yang tidak natural atau kehilangan detail lokal. Ilustrasi konsep saluran gelap dan perannya dalam pemulihan citra pada pendekatan *DCP* ditunjukkan pada Gambar 1.

Gambar 1. Ilustrasi Dark Channel : Citra bebas kabut (atas), saluran gelap yang bersesuaian (bawah), serta citra berkabut beserta saluran gelapanya (kanan)

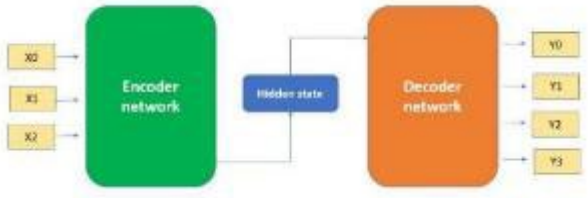
ISSN (Online) 2747-0563
Volume 6 Tahun 2026

B. Pendekatan Berbasis Pembelajaran Mendalam

Metode *deep learning* mengatasi keterbatasan pendekatan prior dengan mempelajari pemetaan kompleks dari citra berkabut menuju citra bersih secara *data-driven*. *CNN* banyak digunakan karena efisien dalam mengekstraksi fitur spasial secara hierarkis dan cocok untuk tugas restorasi berbasis piksel. Dalam praktiknya, strategi *dehazing* berbasis *deep learning* berkembang dari estimasi peta transmisi hingga restorasi *end-to-end*, termasuk penggunaan desain *loss* yang lebih kaya untuk meningkatkan kualitas visual dan mengurangi artefak [15], [17]. Sejalan dengan itu, arsitektur berbasis *transformer* mulai diadopsi untuk memperkuat pemodelan konteks global yang sulit ditangkap oleh konvolusi lokal murni, meskipun biasanya disertai perhatian pada biaya komputasi dan risiko *overfitting* pada data terbatas [17].

C. Arsitektur CNN Encoder-Decoder untuk Restorasi Citra

Arsitektur *encoder–decoder* merupakan pola umum pada restorasi citra karena menyediakan alur ekstraksi fitur melalui downsampling pada encoder, pemadatan representasi pada *bottleneck*, dan rekonstruksi spasial melalui upsampling pada *decoder*. Untuk menjaga informasi resolusi tinggi yang hilang saat *downsampling*, umumnya ditambahkan koneksi *skip* (*skip-add/skip-connection*) sehingga detail lokal tetap terpelihara saat proses rekonstruksi [18]. Skema umum arsitektur *encoder–decoder CNN* ditunjukkan pada Gambar 2.



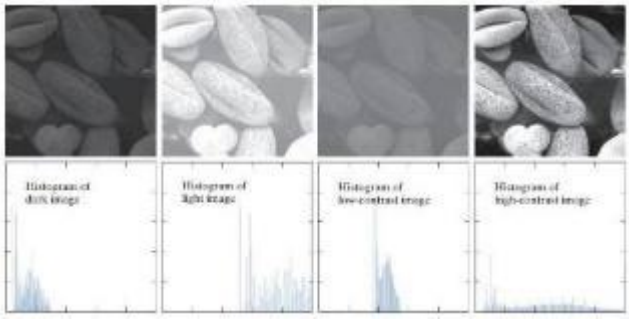
Gambar 2. Skema urutan arsitektur CNN Encoder-Decoder

D. Mekanisme Token-Mixing untuk Menangkap Konteks Global

Token-mixing bertujuan memperkaya interaksi antarbagian citra dalam cakupan luas untuk membantu pemulihan yang lebih konsisten. *Self-attention* pada *transformer* merupakan mekanisme *token-mixing* yang populer karena mampu mengaitkan informasi lintas token secara adaptif, namun kompleksitasnya dapat meningkat seiring panjang urutan token, sehingga kurang efisien pada resolusi tinggi[19].

E. Pra-pemrosesan dan Augmentasi sebagai Penguat Generalisasi

Performa *dehazing* dipengaruhi oleh kesesuaian distribusi data latih dan data uji. Pra-pemrosesan adaptif dapat digunakan untuk menstabilkan karakteristik masukan sebelum pemrosesan jaringan, terutama pada kondisi kontras rendah akibat kabut. *CLAHE* sering digunakan untuk meningkatkan kontras lokal secara adaptif dengan pembatasan penguatan agar *noise* tidak teramplifikasi berlebihan, sehingga detail tekstur yang semula tertutup kabut lebih mudah ditangkap model [20].



Gambar 3. Empat tipe citra beserta histogram yang bersesuaian; (a) gelap, (b) terang, (c) kontras rendah, (d) kontras tinggi

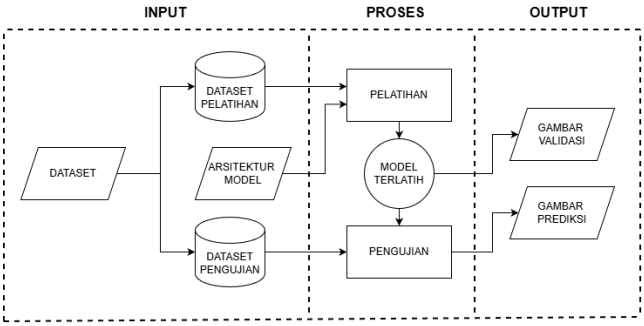
F. Dataset dan Metrik Evaluasi Kualitas Dehazing

Evaluasi *dehazing* umumnya menggunakan kombinasi *benchmark* sintetis dan data dunia nyata. Dataset sintetis lazim dipakai karena menyediakan pasangan *hazy-clean* untuk pelatihan dan pengukuran kuantitatif, sedangkan data dunia nyata penting untuk menguji robustnes karena pola dan densitas kabut di lapangan sering berbeda dari sintetis [11], [17].

III. METODE PENELITIAN

A. 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 4 sehingga memudahkan reproduksi eksperimen dan penelusuran setiap tahap eksperimen.



Gambar 4. Diagram Blok Model

B. 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,

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

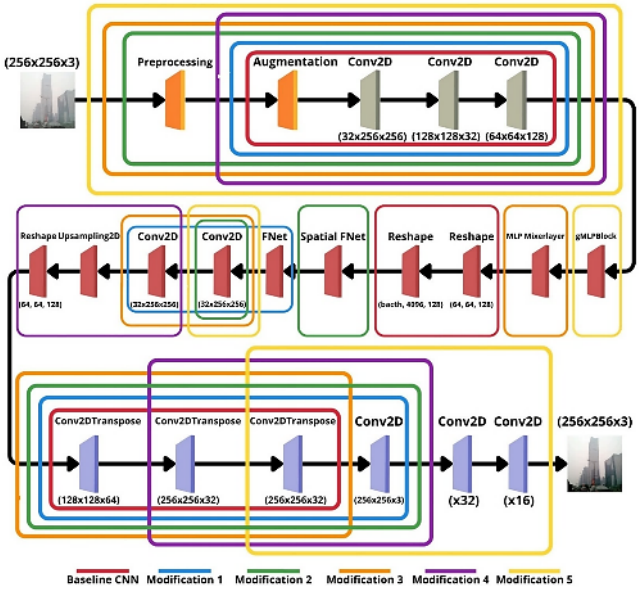
pembagian data untuk pelatihan dan validasi dilakukan dengan komposisi 80% untuk training dan 20% untuk validation, yang secara numerik menghasilkan 160 citra untuk pelatihan dan 40 citra untuk validasi. Selain itu, disediakan satu himpunan uji terpisah berukuran 9 citra yang digunakan sebagai held-out test set untuk pengukuran akhir performa model.

C. 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.

D. 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 5.



Gambar 5. Arsitektur Pengembangan Model

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. Spesifikasi ini tercantum pada tabel 2.

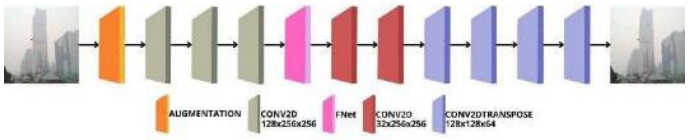
TABELl 2.
Perbandingan Karakteristik Model

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.
Mod 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 sintetis meningkatkan generalisasi pada dataset kecil.
Mod 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.
Mod 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.
Mod 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.
Mod 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.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

2) Modifikasi 1 : Fnet (FFT Based) + Augmentasi Sintesis

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 6 memperlihatkan arsitektur model Modifikasi 1.

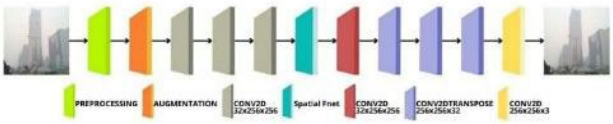


Gambar 6. Arsitektur Model pada Modifikasi 1

Konfigurasi parameter utama mencakup jumlah filter untuk kedalaman saluran dan *dropout rate* guna mencegah *overfitting* saat *feature mixing*. Model ini diperkuat dengan augmentasi data (probabilitas 0,7), yang meliputi ekstraksi *dark channel prior* (*patch* 15), parameter hamburan kabut acak (*beta* 0,5–2,0, intensitas cahaya 0,7–1,0), serta transformasi fotometrik dan geometris (rotasi maksimal 15 derajat). Pelatihan dioptimasi menggunakan Adam dengan *learning rate* awal 0,001, ukuran *batch* kecil untuk efisiensi CPU, mekanisme *early stopping*, serta reduksi *learning rate* dinamis untuk stabilitas konvergensi.

3) Modifikasi 2 : Spatial-Fnet + 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 7 menunjukkan arsitektur Modifikasi 2.

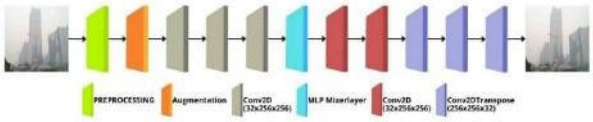


Gambar 7. Arsitektur Model pada Modifikasi 2

Modifikasi kedua mengeksplorasi *LambdaNetworks* melalui modul *LambdaLayer* untuk menangkap konteks spasial global secara efisien. Layer ini dikonfigurasi menggunakan empat *attention heads* dengan dimensi 16 per *head* (total kapasitas 64), serta matriks kernel konvolusi 3x3 untuk penyandian posisi spasial. Kualitas data latih ditingkatkan melalui prapemrosesan *Histogram Matching* dan CLAHE, yang diatur dengan *clip limit* 2,0 dan *tile grid* 8x8 piksel guna menjaga distribusi kontras yang natural. Performa representasi fitur ini kemudian divalidasi menggunakan proporsi pemisahan data uji sebesar 10%.

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 8 mengilustrasikan arsitektur Modifikasi 3.

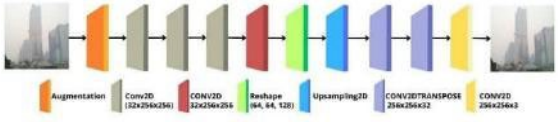


Gambar 8. Arsitektur Model pada Modifikasi 3

Ekstraksi token dari peta fitur diatur oleh ukuran *patch* sebesar 2, yang diproses secara hierarkis melalui dua blok perulangan (*depth*). Kapasitas komputasi dikendalikan oleh dimensi *hidden layer* sebesar 128 untuk tahapan pembauran antar-token (spasial) dan dimensi 512 untuk pembauran antar-saluran (*channel mixing*). Selain itu, regularisasi stokhastik berupa *dropout* pada rentang 0,0 hingga 0,1 diterapkan pada sub-blok Dense untuk menjaga kemampuan generalisasi jaringan.

5) Modifikasi 4 : MLP Mixer + 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 9 memperlihatkan arsitektur Modifikasi 4.



Gambar 9. Arsitektur Model pada Modifikasi 4

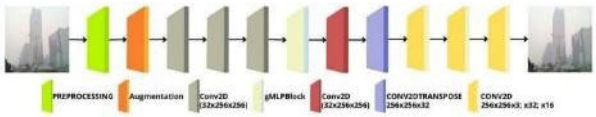
Ekstraksi token difokuskan pada peta fitur tingkat lanjut beresolusi 32x32 menggunakan ukuran *patch* 8, yang menghasilkan 16 *patch* diskrit untuk diproses oleh tiga tumpukan blok *Mixer* berdimensi *embedding* 64. Optimasi fungsi kerugian (*loss*) dieksekusi secara kombinasi linier

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

dengan parameter bobot 0,7 untuk *Mean Absolute Error* (MAE) dan 0,3 untuk *Structural Similarity Index* (SSIM), ditambah parameter penalti opsional sebesar 0,1 untuk kerugian perseptual berbasis model VGG19.

6) Modifikasi 5 : gMLP Style dengan Token-reduction + CLAHE

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 10 menampilkan arsitektur Modifikasi 5.



Gambar 10. Arsitektur Model pada Modifikasi 5

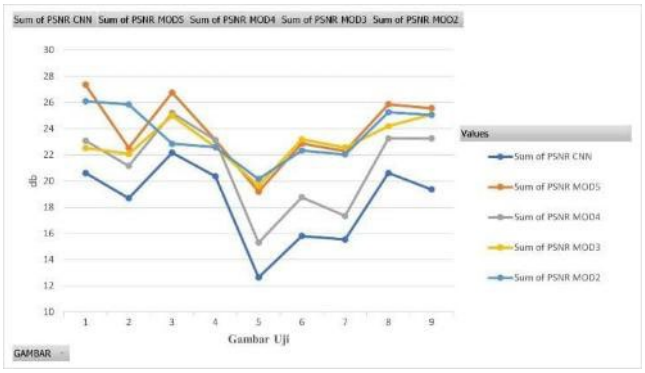
Modifikasi kelima menerapkan arsitektur *Gated MLP* (gMLP) dengan mekanisme *Spatial Gating Unit* (SGU) untuk mereduksi beban komputasi kuadratik pada pemodelan sekuens panjang. Efisiensi struktural ini dicapai melalui operasi *AveragePooling* dengan faktor ukuran 4 untuk melakukan *downsampling* sebelum tahap perataan spasial. Parameter rasio ekspansi *MLP* diatur sebesar 2 untuk memperbesar dimensi saluran asal menjadi dimensi tersembunyi yang direpresentasikan secara laten. Dimensi ini kemudian dipartisi secara simetris; separuh pertama diteruskan sebagai identitas linear, sedangkan separuh kedua ditransformasi melintasi sekuens sebelum akhirnya dikalikan secara *element-wise* untuk mengekstraksi bobot informasi spasial yang paling esensial.

E. Prosedur Pelatihan, Validasi dan Pengujian

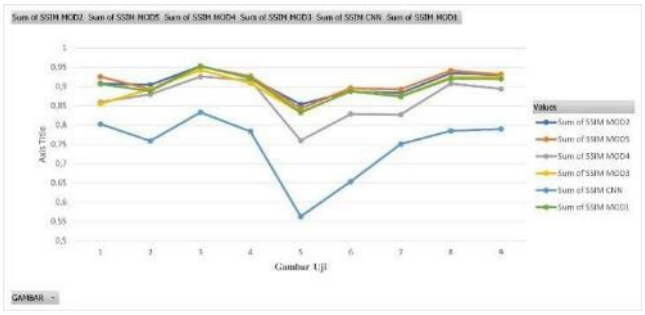
Pelatihan setiap varian model mengikuti prosedur yang sama: data pelatihan (160 citra) diberikan augmentasi dan pra-pemrosesan sesuai desain eksperimen, lalu model dilatih untuk meminimalkan fungsi loss tertentu. Kinerja model dipantau pada set validasi (40 citra) untuk memilih *checkpoint* terbaik dan 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.

IV. HASIL DAN PEMBAHASAN

Gambar 11 dan Gambar 12 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 11. Perbandingan Nilai PSNR untuk Setiap Gambar Uji



Gambar 12. Perbandingan Nilai SSIM untuk Setiap Gambar Uji

Data pada Tabel 3 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 3.
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

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

menggunakan augmentasi sintetis, yang bersama-sama meningkatkan kemampuan model menangkap konteks global dan memperbaiki generalisasi pada dataset terbatas. Hasil pada Tabel 4 mendukung argumen bahwa pemrosesan di domain frekuensi (FNet) beserta augmentasi sintetis efektif meningkatkan rekonstruksi citra dehazing.

Tabel 4.
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 5 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.

Tabel 5.
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 6 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.

Tabel 6.
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 7 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.

Tabel 7.
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 8 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.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Tabel 8.
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

Tabel 9 menunjukan perbandingan hasil rata-rata nilai PSNR dan SSIM penelitian lain terkait image dehazing dengan pengujian penellitian ini. CNN dan Modifikasi 5 adalah hasil terbaik dari penelitian ini

Tabel 9.
Perbandingan rata-rata nilai PSNR dan SSIM pada metode yang berbeda

Model	Hasil Rata-rata	
	PSNR	SSIM
CNN	18.424	0.747
Modifikasi 5	23.95	0.911
AOD-NET	19.6954	0.84478
GCANet	30.23	0.98
MSBDN-DFF	33.79	0.984

Berdasarkan Tabel 9, model CNN sebagai baseline menghasilkan rata-rata PSNR 18,424 dB dan SSIM 0,747, sedangkan Modifikasi 5 sebagai model terbaik dalam penelitian ini mencapai PSNR 23,95 dB dan SSIM 0,911 pada dataset SOTS dengan total 209 citra yang dijalankan pada perangkat AMD Ryzen 3 3250U dengan RAM 8 GB. Peningkatan ini menunjukkan bahwa penerapan mekanisme gMLP-style dengan token-reduction pada bottleneck encoder–decoder mampu memperbaiki kualitas rekonstruksi citra berkabut secara signifikan dibandingkan baseline. Jika dibandingkan dengan AOD-Net yang memperoleh PSNR 19,6954 dB dan SSIM 0,84478, model yang dikembangkan juga menunjukkan performa yang lebih baik dalam mempertahankan detail dan struktur visual citra.

Namun demikian, hasil tersebut masih berada di bawah GCANet dan MSBDN-DFF yang masing-masing mencapai PSNR 30,23 dB / SSIM 0,98 dan PSNR 33,79 dB / SSIM 0,984, yang perlu dipahami dalam konteks penggunaan sumber daya komputasi yang lebih besar serta konfigurasi pelatihan yang lebih kompleks. Secara keseluruhan, hasil ini menegaskan bahwa pendekatan token-mixing pada arsitektur CNN encoder–decoder efektif dalam meningkatkan performa dehazing dan memberikan keseimbangan yang baik antara kualitas hasil dan keterbatasan perangkat komputasi.

V. 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. Penelitian ini menunjukkan kenaikan metrik numerik dan perbaikan perseptual yang nyata: 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.

Walaupun beberapa varian berpotensi menghasilkan keluaran perseptual sangat baik, hasil eksperimen juga memperlihatkan bahwa pendekatan tersebut rentan terhadap varians performa bila hyperparameter dan bobot loss tidak disetel dengan cermat. Temuan ini menekankan pentingnya penyetelan hyperparameter dan pemilihan bobot loss yang teliti untuk merealisasikan keuntungan teoretis metode-metode tersebut.

UCAPAN TERIMAKASIH

Dalam penilitian 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.

REFERENSI

[1] M. Shen, T. Lv, Y. Liu, J. Zhang, and M. Ju, “A Comprehensive Review of Traditional and Deep-Learning-Based Defogging Algorithms,” *Electronics (Switzerland)*, vol. 13, no. 17, Sep. 2024, doi: 10.3390/electronics13173392.

[2] P. V. Deshmukh, A. Kumar, and P. Chakrabarti, “Review of Deep Learning Based Image Dehazing for Autonomous Vehicle,” 2024.

[3] C. Zheng, W. Ying, and Q. Hu, “Comparative analysis of dehazing algorithms on real-world hazy images,” *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-95510-z.

[4] J. Gui *et al.*, “A Comprehensive Survey on Image Dehazing Based on Deep Learning,” 2021.

[5] Kaiming He, Xiaoou Tang, and Jian Sun, *Single image haze removal using dark channel prior*, vol. 33. IEEE, 2011.

[6] H. Ullah *et al.*, “Light-DehazeNet: A Novel Lightweight CNN Architecture for Single Image Dehazing,” *IEEE Transactions on Image Processing*, vol. 30, pp. 8968–8982, 2021, doi: 10.1109/TIP.2021.3116790.

[7] Z. Tu *et al.*, “MAXIM: Multi-Axis MLP for Image Processing.” [Online]. Available: https://github.

[8] A. Kholiq *et al.*, “Enhancing Hazy Image Quality with a Modular CNN Encoder-Decoder,” *ELKOMIKA*, vol. 14, no. 1, pp. 69–83, Jan. 2026, doi: 10.26760/elkomika.

[9] V. Stimper, S. Bauer, R. Ernstorfer, B. Schölkopf, and R. P. Xian, “Multidimensional Contrast Limited Adaptive Histogram Equalization,” *IEEE Access*, vol. 7, pp. 165437–165447, 2019, doi: 10.1109/ACCESS.2019.2952899.

[10] Y. Liu, Q. Dong, C. Zhu, Y. Guo, and F. Wang, “Revitalizing image dehazing in the real world: A high-quality dataset and a customized

method," *Comput. Vis. Media (Beijing).*, vol. 11, no. 4, pp. 833–848, 2025, doi: 10.26599/CVM.2025.9450505.
[11] H. Liu, Z. Dai, D. R. So, and Q. V. Le, "Pay Attention to MLPs," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2105.08050
[12] Z. Chen, Z. He, and Z.-M. Lu, "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention," Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.04805
[13] Tiande Mo, Wai-Yat Chan, Siqian Zheng, and Renhua Yang, "Review of AI Image Enhancement Techniques for In-Vehicle Vision Systems Under Adverse Weather Conditions," *World Electr. Veh. J*, vol. 16, no. 2, Jan. 2025.
[14] S. M. Sopian *et al.*, "Development of a Modified CycleGAN Model with Residual Blocks and Perceptual Loss for Image Dehazing," *Teknika*, vol. 14, no. 2, pp. 232–238, Jul. 2025, doi: 10.34148/teknika.v14i2.1235.
[15] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based Restoration of Images Degraded by Adverse Weather Conditions," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2111.14813
[16] Y. Song, Z. He, H. Qian, and X. Du, "Vision Transformers for Single Image Dehazing," Apr. 2022, doi: 10.1109/TIP.2023.3256763.
[17] Y. Song, Z. He, H. Qian, and X. Du, "Vision Transformers for Single Image Dehazing," *IEEE Transactions on Image Processing*, vol. PP, p. 1, Mar. 2023, doi: 10.1109/TIP.2023.3256763.
[18] S. Satrasupalli, E. Daniel, and S. Guntur, "Single Image Haze Removal Based on Transmission Map Estimation Using Encoder-Decoder Based Deep Learning Architecture," *Optik (Stuttg).*, vol. 248, p. 168197, Oct. 2021, doi: 10.1016/j.ijleo.2021.168197.
[19] H. Zhou, Z. Chen, Y. Liu, Y. Sheng, W. Ren, and H. Xiong, "Physical-priors-guided DehazeFormer," *Knowl. Based. Syst.*, vol. 266, p. 110410, 2023, doi: https://doi.org/10.1016/j.knosys.2023.110410.
[20] S. Okyere-Gyamfi, M. Asante, K. O. Peasah, Y. M. Missah, and V. Akoto-Adjepong, "Contrast limited adaptive histogram equalization (CLAHE) and colour difference histogram (CDH) feature merging capsule network (CCFMCapsNet) for complex image recognition," *PLoS One*, vol. 20, no. 10 October, Oct. 2025, doi: 10.1371/journal.pone.0335393.