

KLASIFIKASI TOKOH WAYANG RAMAYANA MENGGUNAKAN CROSS VISION TRANSFORMER (CROSSViT) DENGAN OPTIMASI ADAM

Lugas Fernanda Wirayudha, Basuki Rahmat, Henni Endah Wahanani
Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur
Jalan Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur 60294, Indonesia
20081010144@student.upnjatim.ac.id

ABSTRAK

Wayang merupakan warisan budaya Indonesia yang diakui UNESCO sejak tahun 2003. Namun, minat generasi muda terhadap wayang terus menurun akibat hambatan bahasa, durasi pertunjukan, serta dominasi media digital modern. Penelitian ini bertujuan mengembangkan model klasifikasi otomatis tokoh Wayang Ramayana menggunakan Cross Vision Transformer (CrossViT) dengan optimasi Adam. Dataset terdiri dari 60 karakter wayang yang diperoleh dari foto langsung, dataset Kaggle, dan pencarian daring. Teknik augmentasi digunakan untuk meningkatkan variasi data, meliputi rotasi acak, flipping, cropping, dan penyesuaian warna. Model CrossViT memanfaatkan patch multi-skala 16×16 dan 32×32 untuk menangkap detail ornamen wayang secara menyeluruh. Hasil penelitian menunjukkan akurasi pelatihan 97,56% dan akurasi validasi 89,12%. Precision rata-rata mencapai 0,90 dan recall 0,88. Sebanyak 18 karakter berhasil diklasifikasi sempurna (F1-score 1,00), sementara beberapa karakter seperti Pratalamariam mengalami kesalahan klasifikasi akibat kemiripan visual. Penelitian ini membuktikan potensi CrossViT dalam upaya digitalisasi dan pelestarian wayang, serta membuka ruang pengembangan lebih lanjut terutama untuk karakter dengan fitur ambigu.

Kata kunci : *Wayang Ramayana, Cross Vision Transformer, Adam Optimization, klasifikasi citra, deep learning.*

1. PENDAHULUAN

Wayang merupakan salah satu warisan budaya Nusantara yang memiliki nilai filosofi dan estetika tinggi, serta telah diakui UNESCO sebagai Warisan Budaya Takbenda pada 7 November 2003 [1]. Seni pedalangan tidak hanya berfungsi sebagai hiburan, tetapi juga sebagai media pemaknaan nilai moral dan spiritual masyarakat Jawa. Proses akulturasi budaya Hindu yang terjadi pada masa klasik membuat kisah-kisah seperti Ramayana dan Mahabharata diadopsi, disesuaikan, dan ditransformasi melalui perspektif budaya lokal sehingga melahirkan karakter visual wayang yang khas dan berbeda dari versi India [2]. Meskipun memiliki nilai historis dan budaya yang kuat, wayang mengalami tantangan serius pada era digital. Survei menunjukkan bahwa 47,2% responden muda bersikap netral terhadap wayang dan 28,5% menyatakan tidak tertarik, dipengaruhi oleh hambatan bahasa Jawa klasik, durasi pertunjukan yang panjang, serta dominasi hiburan digital modern [3]. Kondisi ini menyebabkan kesenjangan apresiasi budaya antargenerasi dan menuntut adanya inovasi pelestarian yang relevan dengan kebutuhan generasi digital native.

Salah satu permasalahan dalam memperkenalkan wayang kepada generasi muda adalah kompleksitas visual setiap tokohnya, seperti detail ornamen, bentuk wajah, serta proporsi tubuh yang sering kali mirip satu sama lain. Tantangan ini membuka peluang pemanfaatan teknologi kecerdasan buatan untuk membantu proses pengenalan karakter wayang secara otomatis. Cross Vision Transformer (CrossViT) merupakan salah satu arsitektur deep learning yang mampu mengolah citra melalui representasi multi-skala sehingga dapat menangkap detail visual lokal sekaligus struktur global pada gambar [4].

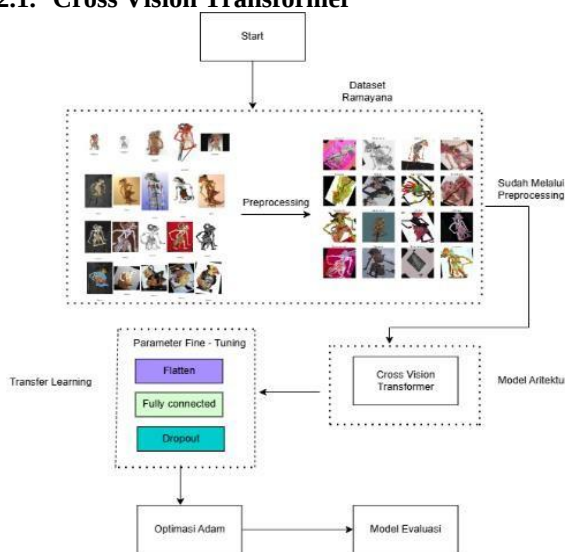
Kemampuannya dipadukan dengan optimizer Adam yang memiliki mekanisme adaptive learning rate sehingga menyediakan pembaruan parameter yang stabil dan efisien selama pelatihan model [5]. Dengan menggabungkan kedua metode tersebut, penelitian ini berupaya mengembangkan sistem klasifikasi otomatis untuk mengenali 60 karakter Wayang Ramayana. Tujuan penelitian meliputi implementasi CrossViT dengan optimasi Adam untuk klasifikasi citra wayang serta evaluasi performa model melalui metrik akurasi, precision, recall, dan F1-score. Penyelesaian masalah dilakukan melalui tahapan pengumpulan dan preprocessing dataset, augmentasi data, perancangan model CrossViT, pelatihan menggunakan Adam, evaluasi performa, dan deployment model dalam bentuk aplikasi berbasis web.

2. TINJAUAN PUSTAKA

Penelitian terdahulu terkait pemanfaatan model transformer untuk klasifikasi citra menunjukkan perkembangan yang signifikan. Studi TreeFormers mengimplementasikan Vision Transformer dengan optimasi Adam dan berhasil meningkatkan akurasi klasifikasi citra satelit dari 72,0% menjadi 72,9%, sekaligus melampaui performa CNN sebesar 9,9% [6]. Penelitian Chen et al. memperkenalkan Cross Vision Transformer yang memanfaatkan patch multi-skala dan cross-attention sehingga meningkatkan akurasi hingga 2% dibandingkan DeiT pada dataset ImageNet1K dengan komputasi yang efisien [7]. Swin Transformer kemudian menghadirkan pendekatan hierarchical attention dengan mekanisme shifted windows, dan dengan dukungan optimizer Adam berhasil mencapai akurasi 87,3% pada ImageNet-1K serta performa unggul pada deteksi objek dan

segmentasi [8]. Pendekatan lain seperti Transformer in Transformer (TNT) juga menunjukkan efektivitas pemrosesan patch internal untuk menangkap detail yang lebih halus, menghasilkan akurasi top-1 sebesar 81,5% [9]. Penelitian mengenai objek budaya tradisional turut mendukung efektivitas algoritma Adam; studi terkait klasifikasi motif kain endek menunjukkan bahwa Adam menghasilkan akurasi 33% lebih tinggi dibandingkan optimizer SGD [10]. Berdasarkan berbagai penelitian tersebut, metode CrossViT dengan optimasi Adam dinilai relevan untuk menganalisis citra wayang yang sarat detail visual, serta menjadi dasar kuat bagi pengembangan sistem klasifikasi karakter Wayang Ramayana.

2.1. Cross Vision Transformer

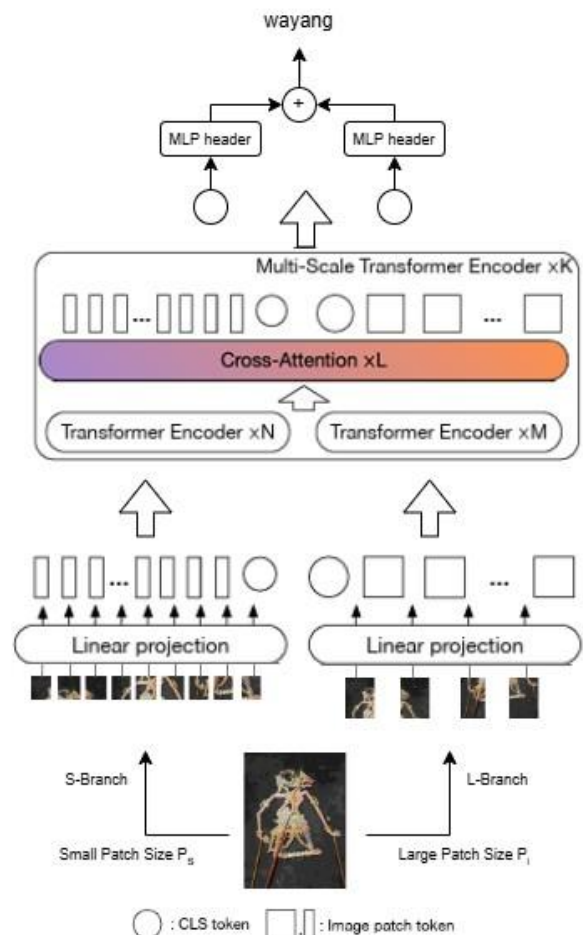


Gambar 1. Arsitektur CVT

Cross Vision Transformer merupakan arsitektur pemrosesan visual yang mengoptimalkan analisis multi-skala melalui dua komponen utama, yaitu *Cross-scale Embedding Layer (CEL)* untuk penggabungan fitur lintas skala dan *Long Short Distance Attention (LSDA)* yang memproses hubungan spasial jarak dekat dan jarak jauh secara terpisah. Arsitektur ini terdiri dari K lapisan encoder paralel dengan dua jalur patch patch kecil (P_s) dan patch besar (P_l) masing-masing menggunakan lapisan transformer N dan M untuk efisiensi komputasi. Setiap blok transformer dilengkapi dengan jaringan FFN (MLP dua lapis), aktivasi GELU, dan Layer Normalization. Integrasi dua jalur melalui *cross-attention* pada token CLS memberikan pemahaman visual menyeluruh dari detail halus hingga struktur global sambil tetap efisien [11]

2.2. Vision Transformer

Sebelum memasuki encoder, gambar dipecah menjadi patch dan direpresentasikan sebagai token. Input awal Vision Transformer terdiri dari token CLS, token patch gambar, dan embedding posisi untuk mempertahankan informasi spasial.



Gambar 2. Arsitektur VT

Pemrosesan pada setiap lapisan transformer melibatkan dua tahap utama[12]:

- Multi-Head Self-Attention (MSA)* untuk menangkap hubungan antar patch dalam satu jalur,
- Feed Forward Network (FFN)* untuk memperkaya representasi fitur.

Perbedaan mendasar ViT dengan CNN terletak pada representasi akhir: CNN mengambil *average pooling*, sedangkan ViT menggunakan token CLS sebagai representasi keseluruhan citra .

2.3. Data Augmentation

Data augmentation merupakan proses memperkaya data dengan memodifikasi citra asli. Teknik augmentasi yang digunakan meliputi rotasi acak hingga 30°, *horizontal flip*, *vertical flip*, *random cropping*, serta penyesuaian warna. Augmentasi membantu meningkatkan kualitas dan keberagaman data, sekaligus mengatasi foto wayang yang terlalu gelap, terang, berisik, atau buram agar dapat dikenali model[12].

2.4. Preprocessing Citra

Preprocessing bertujuan meningkatkan kualitas citra sebelum analisis atau pengenalan objek[8]. Proses ini meliputi normalisasi piksel, *resize* citra,

transformasi label menggunakan one-hot encoding, penentuan batch size, dan *shuffle* data untuk distribusi acak. Tahapan ini memastikan setiap citra memenuhi standar input model[10].

2.5. Algoritma Adam

Optimasi Adam menggabungkan *adaptive learning rate* dan momentum. Adam menghitung momen pertama (mt) dan momen kedua (vt) dari gradien untuk melakukan pembaruan parameter yang lebih presisi[13]. Adam menyesuaikan *learning rate* berdasarkan karakteristik gradien sehingga mempercepat konvergensi pada area tertentu dan memperlambat pembaruan pada area yang memerlukan detail lebih halus

Rumus :

$$mt = \beta_1 mt-1 + (1 - \beta_1) gt \quad (1)$$

Penjelasan:

- gt adalah gradien pada iterasi ke-t
- mt adalah estimasi rata-rata gradien (mirip momentum)
- β_1 mengontrol seberapa besar pengaruh gradien sebelumnya

Momentum ini membantu mengurangi osilasi dan membuat arah pembaruan lebih stabil.

2.6. Optimasi dalam Deep Learning

Optimasi merupakan proses mengatur parameter model melalui *gradient descent*. Variasi optimasi meliputi[14]:

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini-Batch Gradient Descent

Teknik lain seperti momentum, Nesterov Accelerated Gradient, regularisasi L1/L2, dan *dropout* diterapkan untuk mencegah overfitting[13]. Tantangan seperti *saddle points*, *vanishing gradient*, dan *exploding gradient* dapat diatasi dengan batch normalization dan inisialisasi bobot seperti He dan Xavier.

2.7. Evaluasi Kinerja Model

Evaluasi klasifikasi citra menggunakan metrik Accuracy, Precision, Recall, dan F1-score. *Confusion matrix* menggambarkan hubungan antara prediksi dan nilai aktual, mencakup[14]:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

Rumus evaluasi yang digunakan meliputi :

Accuracy adalah tingkat ketepatan proporsi label yang diprediksi dengan benar. Rumus accuracy adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

Precision adalah jumlah prediksi benar pada kelas tertentu dibandingkan seluruh prediksi yang dibuat untuk kelas tersebut. Rumus precision adalah:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall adalah jumlah data pada kelas tertentu yang berhasil diprediksi dengan benar oleh model. Rumus recall adalah:

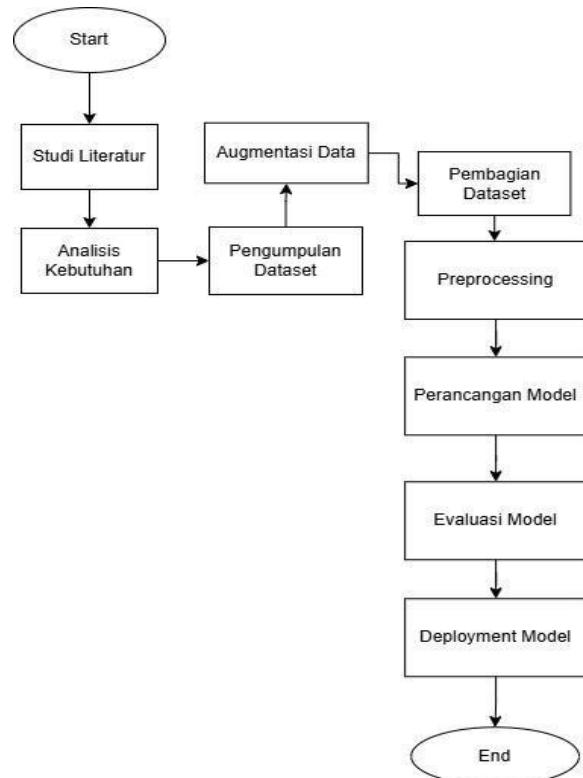
$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1-Score merupakan kombinasi dari precision dan recall dalam bentuk harmonic mean. Rumus F1-score sebagai berikut[15]:

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

3. METODE PENELITIAN

3.1. Alur Penelitian

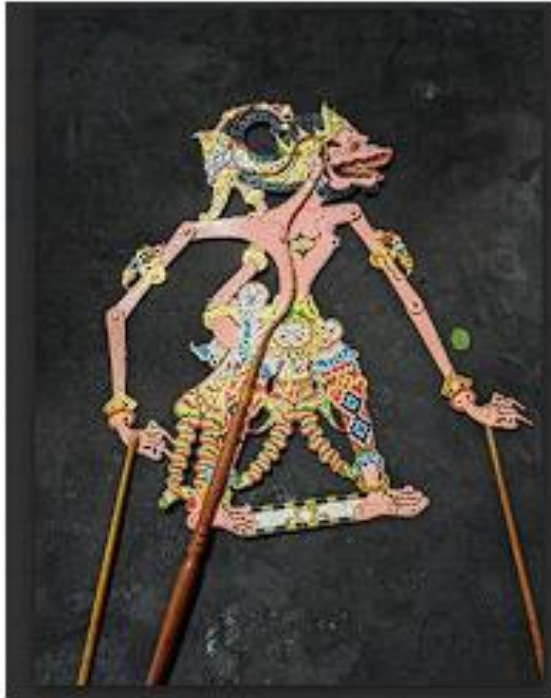


Gambar 3. Alur Penelitian

Penelitian ini menggunakan pendekatan sistematis untuk mengklasifikasikan 60 karakter Wayang Ramayana dengan memanfaatkan metode Cross Vision Transformer (CrossViT) yang dioptimalkan menggunakan Adam. Alur penelitian dimulai dari tahap pengumpulan data, dilanjutkan dengan preprocessing dan augmentasi citra, kemudian implementasi arsitektur CrossViT untuk proses pelatihan model, dan diakhiri dengan evaluasi kinerja menggunakan metrik klasifikasi standar. Secara keseluruhan, seluruh rangkaian proses disusun secara terstruktur dengan tujuan menghasilkan model klasifikasi yang optimal dan akurat. Tahap awal penelitian dilakukan dengan menyusun alur penelitian sebagaimana digambarkan dalam flowchart pada gambar berikut. Alur ini mencakup proses mulai dari input data hingga tahap evaluasi akhir model. Proses dimulai dengan pengumpulan dataset karakter Wayang Ramayana, kemudian data yang diperoleh diproses melalui tahapan preprocessing untuk

menyesuaikan format citra sesuai kebutuhan model. Setelah itu dilakukan augmentasi data untuk memperbanyak variasi citra sehingga model dapat mengatasi perbedaan pencahayaan, orientasi, maupun kualitas gambar. Citra yang telah diproses kemudian masuk ke tahap implementasi CrossViT, di mana model melakukan ekstraksi fitur multi-skala sebelum melalui proses pelatihan menggunakan optimizer Adam. Tahap akhir penelitian berupa evaluasi performa untuk mengukur sejauh mana model mampu mengklasifikasikan karakter wayang secara akurat.

3.2. Pengumpulan dan Pengolahan Dataset



Gambar 4. Dataset Wayang

Dataset yang digunakan dalam penelitian ini terdiri dari 60 karakter Wayang Ramayana. Karakter tersebut dikumpulkan dari tiga sumber utama. Sumber pertama berasal dari pemotretan langsung terhadap artefak wayang, sehingga memberikan citra dengan kualitas yang konsisten terhadap objek fisik. Sumber kedua berasal dari dataset pendukung yang diperoleh melalui platform Kaggle, yang berfungsi sebagai pelengkap dataset agar variasinya lebih luas. Sumber ketiga diperoleh dari pencarian daring secara selektif yang bertujuan untuk mengakomodasi perbedaan bentuk visual wayang dari berbagai daerah, seperti perbedaan gaya sungging, warna, atau proporsi tubuh, yang secara budaya memang bervariasi. Setelah semua gambar terkumpul, proses preprocessing dilakukan untuk memastikan seluruh data memiliki format yang seragam. Langkah preprocessing ini mencakup normalisasi piksel agar berada dalam rentang yang sesuai untuk pemrosesan neural network, mengubah ukuran citra agar kompatibel dengan input model CrossViT, serta melakukan encoding label menggunakan one-hot encoding agar dapat digunakan

dalam klasifikasi multi-kelas. Dataset juga dibagi menjadi tiga bagian, yaitu data latih, data validasi, dan data uji untuk memastikan bahwa hasil evaluasi model bersifat objektif dan tidak bias.

3.3. Augmentasi Data.

Tahap augmentasi dilakukan untuk meningkatkan keberagaman data sehingga model tidak hanya mampu mengenali citra yang seragam, tetapi juga mampu beradaptasi terhadap kondisi pencahayaan, pose, atau orientasi yang berbeda. Augmentasi dalam penelitian ini meliputi rotasi acak hingga 30 derajat untuk mensimulasikan kemungkinan variasi orientasi citra, flipping secara horizontal maupun vertikal untuk mensimulasikan posisi mirroring atau pantulan, serta random cropping sebagai pendekatan untuk mensimulasikan pemotongan objek pada citra. Selain itu, beberapa citra mengalami modifikasi warna untuk menanggulangi kondisi data yang terlalu gelap, terlalu terang, atau memiliki noise tertentu. Dengan demikian, augmentasi membantu model untuk tetap robust dan tidak overfitting pada pola tertentu. Penggunaan augmentasi terbukti penting dalam penelitian ini karena citra wayang yang dikumpulkan memiliki kualitas visual yang sangat beragam.



Gambar 5. Data Augmentasi

3.4. Implementasi Arsitektur Cross Vision Transformer (CrossViT)

Pada tahap ini, seluruh citra yang telah melalui preprocessing dan augmentasi kemudian diproses melalui arsitektur Cross Vision Transformer. CrossViT memanfaatkan dua jenis patch visual, yaitu patch kecil (16×16) dan patch besar (32×32), yang masing-masing memproses representasi berbeda dari gambar. Patch kecil digunakan untuk menangkap detail lokal seperti ornamen wayang, pola kain, dan bentuk wajah. Sementara patch besar digunakan untuk

memahami bentuk global karakter seperti proporsi tubuh, posisi tangan, dan bentuk kepala.

Kedua patch tersebut kemudian diproses melalui jalur encoder transformer yang terpisah. Setiap jalur encoder terdiri dari beberapa lapisan yang memiliki Multi-Head Self Attention (MSA), Layer Normalization, serta Feed Forward Network (FFN) dengan aktivasi GELU. Representasi yang dihasilkan dari kedua jalur patch ini kemudian disatukan menggunakan mekanisme cross-attention pada token CLS sehingga model dapat menggabungkan pemahaman detail halus dan struktur global secara bersamaan. Hasil penggabungan fitur ini kemudian diteruskan ke classification head untuk menentukan kelas dari 60 karakter wayang.

3.5. Proses Pelatihan Model

Model dilatih menggunakan optimizer Adam dengan parameter yang telah ditentukan, termasuk learning rate sebesar 0.0001, jumlah epoch sebanyak 10, serta fungsi loss categorical crossentropy. Selama pelatihan, model mempelajari pola visual karakter wayang dari hasil ekstraksi fitur multi-skala. Early stopping juga diterapkan pada pelatihan untuk menghentikan proses ketika performa model pada data validasi tidak meningkat lagi sehingga mencegah terjadinya overfitting. Proses pelatihan ini berlangsung secara bertahap di mana model terus memperbaiki bobot internal berdasarkan gradien error. Kombinasi CrossViT dan Adam memungkinkan model mendapatkan pemahaman ruang visual yang seimbang antara detail kecil dan struktur global, sehingga mampu membedakan karakter wayang yang mirip sekalipun.

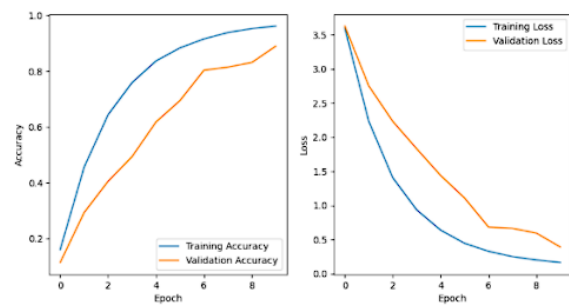
3.6. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai sejauh mana model mampu mengklasifikasikan 60 karakter Wayang Ramayana dengan benar. Evaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Accuracy mengukur proporsi prediksi yang benar dibandingkan keseluruhan data. Precision mengukur ketepatan model dalam memberikan prediksi pada kelas tertentu, sedangkan recall menghitung kemampuan model dalam mengenali kelas yang benar. F1-score merupakan kombinasi dari precision dan recall yang memberikan penilaian seimbang terhadap performa model.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini menunjukkan bahwa model Cross Vision Transformer (CrossViT) mampu melakukan klasifikasi terhadap 60 karakter Wayang Ramayana dengan tingkat akurasi yang tinggi. Proses pelatihan model menghasilkan grafik akurasi dan loss yang menunjukkan perbedaan jelas antara data latih dan validasi, menggambarkan bagaimana model belajar secara bertahap hingga mencapai performa stabil pada beberapa epoch terakhir. Sepanjang proses

pelatihan, nilai akurasi data latih meningkat secara signifikan dan mencapai titik kesetimbangan, sedangkan akurasi validasi menunjukkan pola kenaikan yang konsisten hingga mendekati nilai optimal.



Gambar 6. Grafik Akurasi

Model CrossViT yang digunakan dalam penelitian ini mampu mengenali pola visual berupa bentuk wajah, ornamen, dan ciri khas karakter wayang secara efektif. Hasil pelatihan menunjukkan bahwa model mencapai akurasi pelatihan sebesar 97,56% dan akurasi validasi sebesar 89,12%, dengan nilai loss masing-masing 0,1605 untuk pelatihan dan 0,4255 untuk validasi. Nilai akurasi yang tinggi tersebut menunjukkan bahwa CrossViT mampu menangkap representasi fitur multi-skala yang relevan, sehingga detail ornamen wayang yang rumit dapat dipetakan dengan baik. Selain akurasi, penelitian ini juga menyertakan analisis Classification Report yang menampilkan nilai precision, recall, dan F1-score untuk masing-masing karakter wayang. Hasil analisis menunjukkan bahwa terdapat 18 karakter yang berhasil diklasifikasikan secara sempurna dengan nilai F1-score 1.00. Karakter-karakter seperti Gareng, Petruk, Janulen, Arigibas, serta sejumlah karakter Kapi lainnya menunjukkan performa yang tinggi karena memiliki ciri visual yang lebih khas dan mudah dikenali oleh model. Hal ini mengindikasikan bahwa CrossViT sangat efektif dalam membedakan karakter yang memiliki perbedaan bentuk dan ornamen yang jelas. Namun demikian, tidak semua karakter dapat diklasifikasikan dengan akurasi sempurna. Beberapa karakter seperti Pratalamariam, Resi Subali, dan Lesmana menunjukkan nilai F1-score yang lebih rendah, masing-masing berada pada kisaran 0.59 – 0.63. Nilai ini menunjukkan adanya kesulitan model dalam membedakan karakter dengan kemiripan visual tinggi atau karakter yang memiliki kualitas citra lebih bervariasi. Misalnya, karakter Pratalamariam memiliki ornamen yang mirip dengan beberapa wayang lain dari kelompok Resi, sehingga model sering melakukan misklasifikasi antar-karakter tersebut. Variasi pencahayaan, sudut pengambilan gambar, dan kualitas citra juga turut memengaruhi keberhasilan model dalam mengenali karakter yang kompleks.

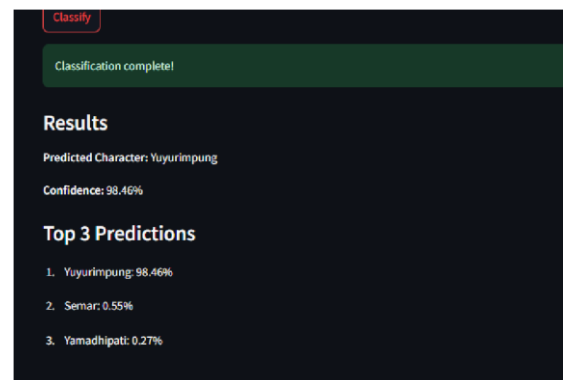
Anggada	0.85	0.88	0.86	32
Anjani	1.00	0.62	0.77	32
Anoman	0.97	0.97	0.97	32
Bagong	0.97	1.00	0.98	32
Batara Narada	1.00	0.91	0.95	32
Batara Surya	0.91	0.94	0.92	32
Batara brahma	1.00	0.72	0.84	32
Batara guru	0.94	1.00	0.97	32
Batara Indra	0.95	0.66	0.78	32
Bukbis	0.86	0.75	0.88	32
Danapati	0.63	1.00	0.77	32
Dasarata	0.87	0.68	0.76	19
Gareng	1.00	1.00	1.00	32
Janaka	0.94	0.97	0.95	32
Jandiserana	1.00	1.00	1.00	6
Kalanakrica	0.73	0.76	0.74	21
Kapi Arimenda	1.00	1.00	1.00	43
Kapi Janulen	1.00	1.00	1.00	6
Kapi anala	0.98	0.98	0.98	55
Kapi anggini	1.00	0.92	0.96	12
Kapi anila	0.89	0.89	0.89	27
Kapi arigibas	1.00	1.00	1.00	6
Kapi cucakrawun	0.86	1.00	0.92	12
Kapi jembawan	0.89	0.94	0.91	17
Kapi suweda	1.00	1.00	1.00	43
Kapi trigangga	0.74	1.00	0.85	14
Kekayi	1.00	1.00	1.00	37
Kumbakarna	0.83	0.89	0.86	28
Lembu suro	1.00	0.63	0.77	27
Lesmana	0.84	0.50	0.63	32
Maheso suro	1.00	0.85	0.92	55
Mbilung	0.97	0.91	0.94	32
Nujung ngalengka	1.00	1.00	1.00	6
Petruk	1.00	1.00	1.00	32
Prahasta	0.56	0.92	0.70	24
Pratalamariam	0.46	0.85	0.59	13
Rahwana	0.82	0.74	0.78	31
Rama	0.92	0.58	0.71	19
Rama bergawa	0.90	1.00	0.95	28
Ranendra	1.00	0.50	0.67	6
Resi Gautama	1.00	1.00	1.00	24
Resi Mayangkara	1.00	0.70	0.82	20
Resi Subali	0.43	0.95	0.59	21
Resi Waliki	0.86	1.00	0.92	6
Sang Hyang Wenang	0.91	0.95	0.93	22
Saupakenaka	0.73	0.62	0.67	13
Semar	0.94	1.00	0.97	32
Shinta	1.00	0.75	0.86	20
Sridenta	0.91	1.00	0.95	31
Sugriwa	0.82	0.90	0.86	31
Sukrasana	0.95	0.90	0.93	21
Temboro	0.93	0.93	0.93	14
Togog	0.91	0.97	0.94	32
Trijaka	0.96	1.00	0.98	43
Wibisana	1.00	0.85	0.92	26
Wil kataksiwi	0.88	0.93	0.90	15
Windardi	1.00	1.00	1.00	37
Yamadhipati	0.96	0.84	0.90	31
accuracy			0.89	1552
macro avg	0.90	0.88	0.88	1552
weighted avg	0.91	0.89	0.89	1552

Gambar 7. Classification Report

Dari keseluruhan hasil evaluasi, dapat disimpulkan bahwa CrossViT memberikan performa yang sangat baik dalam klasifikasi citra wayang, terutama karena pendekatan multi-skala mampu menangkap fitur visual lokal dan global secara bersamaan. Meski demikian, tantangan tetap muncul pada kelas-kelas dengan kemiripan visual tinggi atau kualitas citra yang kurang konsisten. Hal ini menjadi indikasi bahwa peningkatan dataset, seperti penambahan citra dengan variasi pose, pencahayaan, dan latar belakang yang lebih beragam, dapat

membantu meningkatkan akurasi model pada penelitian selanjutnya.

Selain itu, model juga telah berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit. Konversi model ke format TensorFlow Lite memungkinkan aplikasi berjalan secara ringan dan cepat sehingga pengguna dapat mengunggah gambar wayang dan langsung mendapatkan hasil prediksi. Tampilan antarmuka aplikasi yang sederhana memudahkan pengguna dalam melakukan uji coba identifikasi karakter wayang secara mandiri, sehingga penelitian ini tidak hanya menghasilkan model klasifikasi, tetapi juga memberikan kontribusi praktis berupa alat bantu digital untuk pengenalan budaya wayang.



Gambar 8. Deployment

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan metode Cross Vision Transformer (CrossViT) dengan optimasi Adam untuk melakukan klasifikasi 60 karakter Wayang Ramayana. Berdasarkan hasil pelatihan dan evaluasi, CrossViT terbukti mampu mengekstraksi fitur visual multi-skala dengan sangat baik sehingga karakter wayang dengan detail ornamen yang kompleks, variasi bentuk wajah, serta perbedaan warna dapat dikenali dengan akurasi tinggi. Model ini menghasilkan akurasi pelatihan sebesar 97,56% dan akurasi validasi sebesar 89,12%, menunjukkan bahwa model mempunyai kemampuan generalisasi yang kuat terhadap data baru. Analisis lebih lanjut melalui Classification Report dan Confusion Matrix menunjukkan bahwa sebagian besar karakter wayang dapat diidentifikasi secara akurat. Terdapat 18 karakter yang mencapai nilai F1-score sempurna (1.00), menandakan model tidak melakukan kesalahan prediksi pada karakter tersebut. Namun demikian, masih ditemukan misklasifikasi pada karakter yang memiliki kemiripan visual tinggi, seperti Pratalamariam, Resi Subali, dan Lesmana. Kemiripan bentuk ornamen dan ragam visual menjadi faktor utama penyebab kesalahan ini. Selain itu, penelitian ini juga berhasil mengembangkan aplikasi berbasis web menggunakan Streamlit yang telah diintegrasikan dengan model TensorFlow Lite. Aplikasi ini memungkinkan pengguna mengunggah gambar wayang dan memperoleh hasil klasifikasi secara cepat,

sehingga berpotensi menjadi media edukasi yang menarik untuk mengenalkan kembali budaya wayang kepada masyarakat, terutama generasi muda.

Penelitian ini masih memiliki beberapa ruang pengembangan di masa mendatang. Pertama, diperlukan penambahan variasi dataset khususnya pada karakter yang memiliki tingkat kemiripan visual tinggi agar model dapat membedakan fitur-fitur kecil yang menjadi pembeda antar karakter. Kedua, peningkatan teknik augmentasi serta penyempurnaan parameter pelatihan dapat meningkatkan performa model dan mengurangi kesalahan klasifikasi. Ketiga, aplikasi web yang dihasilkan dapat diperluas fungsinya menjadi platform edukasi, misalnya dengan menambahkan penjelasan identitas karakter, cerita singkat, serta visualisasi 3D atau animasi wayang untuk menarik minat pengguna. Dengan pengembangan lebih lanjut, penelitian ini dapat memberikan kontribusi lebih luas terhadap upaya pelestarian budaya wayang melalui integrasi teknologi modern.

DAFTAR PUSTAKA

- [1] S. Soeprapto and D. Jirzanah, "Transformasi Nilai-Nilai dan Alam Pemikiran Wayang bagi Masa Depan Jati Diri Bangsa Indonesia," 2018.
- [2] T. Widiyanto, S. Negeri, and N. Ngawi, "Dekonstruksi Mitos Wayang Ramayana dan Baratayuda dan Nilai-Nilai Pendidikan Karakter dalam Novel Kitab Omong Kosong Karya Seno Gumira Ajidarma dan Perang Karya Putu Wijaya," 2023. [Online]. Available: <http://journal.unesa.ac.id/index.php/paramasastra>
- [3] M. N. R. Pradipa and L. N. H. Syafitri, "Nilai Filosofis dalam Kesenian Wayang Kulit bagi Pembentukan Identitas Kultural Generasi Z Muslim Indonesia," *Ulu-muddin: Jurnal Ilmu-ilmu Keislaman*, vol. 14, no. 5, pp. 16–26, 2024, doi: 10.2207/jjws.91.328.
- [4] H. Simaniduruk, M. E. Setiyawati, Y. H. Masyhuri, D. Khairunnisa, and R. F. Aulia, "Fenomena Perkembangan Wayang Menuju Indonesia Emas 2045 di Museum Wayang Jakarta," 2024, doi: 10.37817/ikraith-humaniora.
- [5] W. Wang et al., "CrossFormer: A Versatile Vision Transformer Hinging on Cross-scale Attention," Jul. 2021. [Online]. Available: <http://arxiv.org/abs/2108.00154>
- [6] C.-F. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," Mar. 2021. [Online]. Available: <http://arxiv.org/abs/2103.14899>
- [7] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," Mar. 2021. [Online]. Available: <http://arxiv.org/abs/2103.14030>
- [8] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in Transformer," Feb. 2021. [Online]. Available: <http://arxiv.org/abs/2103.00112>
- [9] U. Ochuba, "TreeFormers -- An Exploration of Vision Transformers for Deforestation Driver Classification," May 2024. [Online]. Available: <http://arxiv.org/abs/2405.15989>
- [10] A. K. Mulugeta, D. P. Sharma, and A. H. Mesfin, "Deep Learning for Medicinal Plant Species Classification and Recognition: A Systematic Review," 2023, *Frontiers Media SA*, doi: 10.3389/fpls.2023.1286088.
- [11] M. S. I. Musyaffa, N. Yudistira, M. A. Rahman, and J. Batoro, "IndoHerb: Indonesia Medicinal Plants Recognition using Transfer Learning and Deep Learning," Aug. 2023. [Online]. Available: <http://arxiv.org/abs/2308.01604>
- [12] P. Wicaksono, P. A. Aryaguna, and W. Lazuardi, "Benthic Habitat Mapping Model and Cross Validation Using Machine-Learning Classification Algorithms," *Remote Sens. (Basel)*, vol. 11, no. 11, Jun. 2019, doi: 10.3390/rs11111279.
- [13] W. T. Guna, "Endek Classification Based on GLCM Using Artificial Neural Networks with Adam Optimization," 2020.
- [14] T. Widiyanto, H. J. Waluyo, Suyitno, and S. A. Sayuti, "Myth Deconstruction of Wayang Ramayana and Baratayuda in the Novel Kitab Omong Kosong by Seno Gumira Ajidarma and Perang by Putu Wijaya," *Rupkatha Journal on Interdisciplinary Studies in Humanities*, vol. 12, no. 1, Jan. 2020, doi: 10.21659/rupkatha.v12n1.33.
- [15] S. R. R. Hadi and A. Z. Mansoor, "The Potential of Wayang Ramayana R.A. Kosasih Comics as Media (Komplementer) of Junior High School Learning in Central Java," *Demandia: Desain Komunikasi Visual Manajemen Desain dan Periklanan*, vol. 6, no. 2, p. 317, Sep. 2021, doi: 10.25124/demandia.v6i2.3623.