Populer: Jurnal Penelitian Mahasiswa
Volume. 5 Nomor. 3 September 2026
E-ISSN: 2963-5306, P-ISSN: 2962-116X, Hal. 161-170
DOI: https://doi.org/10.58192/populer.v5i3.5516
Tersedia: https://journal.unimar-amni.ac.id/index.php/Populer

Implementasi *Vision Transformer* (VIT) dan K-Means Clustering dalam Klasifikasi Gaya Visual Poster untuk Otomatisasi Ekstraksi Skema Warna Digital

Muhammadun Hanif[1*], Bagus Satrio Waluyo Poetro[2], Ghufron[3]
[1-3]Universitas Islam Sultan Agung Semarang, Indonesia
Email: muhammadunhanif39@gmail.com [1], bagusswp@unissula.ac.id [2], ghufron@unissula.ac.id [3]
**Penulis Korespondensi: muhammadunhanif39@gmail.com*

Abstract*. The purpose of this study is to implement the Vision Transformer (ViT) architecture and K-Means Clustering algorithm to automate fine-grained visual style classification of posters and the extraction of digital color schemes. MSMEs and beginner digital creators often face challenges in determining harmonious color schemes for promotional poster designs, leading to a time-consuming trial-and-error design process. The research utilized a custom dataset of 1,011 poster images scraped from the Behance platform, evenly distributed into three visual style classes: Bauhaus, Swiss Style, and Minimalist. The data preprocessing employed a dual-pipeline approach, where color extraction was optimized through SLIC Superpixel segmentation within the CIELab* perceptual color space to significantly reduce computational complexity without degrading the boundaries of original visual elements. The evaluation results demonstrated that the ViT model performed exceptionally well, achieving an accuracy of 79% and a Macro F1-Score of 0.80. In the color extraction module, the K-Means algorithm successfully formed five highly separated dominant palette clusters with an average Silhouette Score of 0.7698. The perceptual color accuracy validation recorded an average ΔE^* of 4.7648, proving that the extracted hexadecimal (HEX) palette codes operate well within the acceptable limits of human visual deviation.*

Keywords: *Color Scheme; Graphic Design; K-Means Clustering; Poster; Vision Transformer.*

Abstrak. Tujuan dari penelitian ini adalah untuk mengimplementasikan arsitektur *Vision Transformer* (ViT) dan algoritma *K-Means Clustering* guna mengotomatisasi klasifikasi gaya visual poster pada tingkat halus (*fine-grained*) serta ekstraksi skema warna digital. Pelaku UMKM dan kreator digital pemula sering kali menghadapi kendala dalam menentukan skema warna yang harmonis untuk desain poster promosi, sehingga proses desain terpaksa dilakukan melalui *trial and error* yang memakan waktu. Data yang digunakan dalam penelitian ini merupakan himpunan data mandiri sebanyak 1.011 citra poster dari *platform* Behance yang terbagi rata ke dalam tiga kelas gaya visual: *Bauhaus, Swiss Style, dan Minimalist*. Prapemrosesan data menerapkan pendekatan jalur ganda, di mana ekstraksi warna dioptimalkan melalui segmentasi SLIC *Superpixel* di dalam ruang warna perseptual CIELab* guna mereduksi kompleksitas beban komputasi secara signifikan tanpa merusak batas elemen visual aslinya. Hasil evaluasi menunjukkan bahwa model ViT berkinerja sangat baik dengan tingkat akurasi mencapai 79% dan rata-rata harmonik makro (*Macro F1-Score*) sebesar 0,80. Pada modul ekstraksi warna, algoritma K-Means sukses membentuk lima klaster palet dominan yang sangat terpisah dengan rata-rata *Silhouette Score* sebesar 0,7698. Validasi akurasi warna perseptual mencatatkan rata-rata ΔE^* sebesar 4,7648, yang membuktikan bahwa kode palet *hexadecimal* (HEX) hasil ekstraksi beroperasi dalam batas kewajaran deviasi visual manusia.

Kata Kunci: Desain Grafis; *K-Means Clustering*; Poster; Skema Warna; *Vision Transformer*.

1. LATAR BELAKANG

Di era digitalisasi, UMKM dituntut melakukan transformasi ke arah pemasaran digital, di mana desain grafis seperti poster promosi memiliki peran krusial sebagai ujung tombak komunikasi visual. Komposisi elemen visual serta tata letak yang terstruktur terbukti dapat mengarahkan aliran visual audiens, menarik perhatian, dan meningkatkan efektivitas penyampaian pesan, namun menciptakan desain yang sukses menuntut pemahaman mendalam terhadap prinsip estetika (Huang et al., 2023). Keterbatasan keterampilan ini membuat pelaku UMKM maupun kreator digital pemula sering kali kesulitan mengekstraksi secara teknis

Naskah Masuk: 25 Juni 2026; Revisi: 16 Agustus 2026; Diterima: 18 September 2026; Terbit: 24 September 2026

mengapa sebuah komposisi visual harmonis. Akibatnya, pemilihan palet warna terpaksa dilakukan melalui proses *trial and error* yang memakan waktu, yang merupakan hal wajar mengingat sebagian besar desainer menganggap proses pemilihan warna sebagai tahapan yang sangat sulit dan bergantung pada pengalaman (Chen et al., 2021). Ketiadaan pengalaman empiris inilah yang membuat ekstraksi warna berbasis data (*data-driven*) dari gambar referensi diakui sebagai metode efektif untuk menghasilkan palet yang harmonis secara perseptual (Liu et al., 2024).

Untuk menjembatani kesenjangan keterampilan tersebut, teknologi kecerdasan buatan dalam ranah *Graphic Design Intelligence* dapat dimanfaatkan sebagai asisten desain yang mengotomatisasi proses ekstraksi gaya dan warna (Huang et al., 2023). Mengidentifikasi gaya visual pada poster merupakan tugas klasifikasi tingkat halus (*Fine-Grained Visual Classification*) karena gaya seni kontemporer sering berbagi elemen desain serupa yang membuat batas antar-gaya menjadi ambigu (Wei et al., 2022). Oleh karena itu, arsitektur *Vision Transformer* (ViT) (Dosovitskiy et al., 2021) diusulkan karena mekanisme *multi-head self-attention* miliknya mampu menangkap hubungan fitur secara global dan melokalisasi wilayah yang paling informatif tanpa memerlukan pengubahan arsitektur struktural (Conde et al., 2021).

Lebih lanjut, agar ekstraksi warna berjalan efisien dan tidak terbebani oleh analisis tingkat piksel, algoritma *K-Means Clustering* diintegrasikan dengan tahap pra-pemrosesan segmentasi *Superpixel SLIC* yang terbukti secara drastis menurunkan waktu komputasi sekaligus mempertahankan batas struktural citra asli (Frackiewicz et al., 2023). Agar hasil ekstraksi tersebut selaras dengan sistem persepsi visual mata manusia, seluruh proses kuantisasi tersebut idealnya dieksekusi di dalam ruang warna CIELab* (Gao et al., 2025).

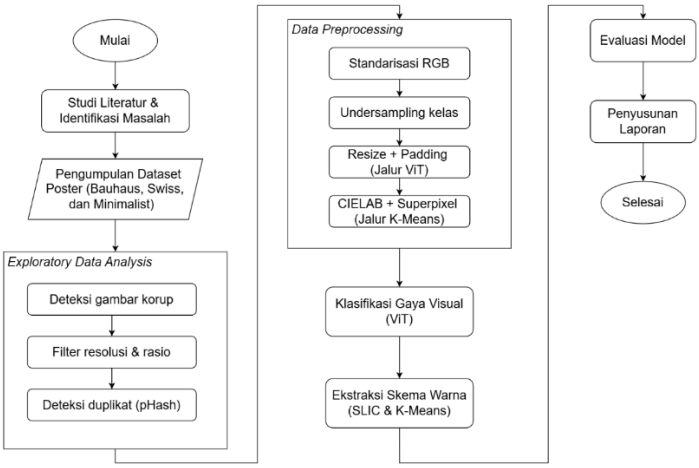
2. KAJIAN TEORITIS

Klasifikasi tingkat halus (*fine-grained classification*) merupakan bidang dalam visi komputer yang bertujuan untuk membedakan kategori spesifik yang berada di bawah satu kelas induk yang sama, di mana tantangan utamanya adalah tingginya variansi intra-kelas dan sangat rendahnya variansi antar-kelas (Wei et al., 2022). Dalam ranah desain poster, tugas ini memiliki kompleksitas yang jauh lebih tinggi karena gaya visual kontemporer seperti *Bauhaus, Swiss Style,* dan *Minimalist* sering kali menggunakan elemen penyusun yang serupa, seperti tata letak berbasis *grid*, bentuk geometris, dan dominasi ruang negatif (*white space*) (Hsu et al., 2023). Untuk mengatasi tantangan tersebut, arsitektur *Vision Transformer* (ViT) terbukti mengungguli model konvensional berkat mekanisme *self-attention* yang mampu menangkap korelasi fitur

estetika secara global (Dosovitskiy et al., 2021). Berbeda dengan CNN yang mengandalkan bidang pandang lokal, ViT memproses gambar secara keseluruhan sebagai urutan bagian-bagian kecil yang disebut *patches*, yang kemudian diproyeksikan ke dalam ruang vektor linear untuk mendapatkan *patch embeddings* (Dosovitskiy et al., 2021). Mekanisme *Multi-Head Self-Attention* pada ViT memungkinkan model untuk menangkap hubungan antar-elemen desain yang tersebar secara acak di seluruh kanvas tanpa memerlukan arsitektur deteksi objek tambahan (Conde et al., 2021).

Dalam konteks otomatisasi warna, algoritma *K-Means Clustering* diakui sebagai metode yang andal untuk tugas kuantisasi warna, yakni mereduksi ribuan palet warna asli menjadi sejumlah titik pusat warna (*centroid*) dominan yang representatif (Frackiewicz et al., 2022). Untuk menghindari beban komputasi yang masif dari pemrosesan tingkat piksel secara tradisional, teknik *superpixel* seperti SLIC (*Simple Linear Iterative Clustering*) diimplementasikan untuk mengelompokkan piksel yang berdekatan dan homogen menjadi satu unit bermakna, sehingga mereduksi redundansi data citra tanpa menghilangkan batas struktural objek (Frackiewicz et al., 2022). Agar klasterisasi warna tersebut selaras dengan mata manusia, perhitungan jarak kemiripan warna dieksekusi di dalam ruang warna perseptual CIELab* (Gao et al., 2025). Ruang warna CIELab* beroperasi pada tiga dimensi utama (L^*, a^*, b^*) yang bersifat independen terhadap perangkat dan mencakup seluruh rentang spektrum warna yang dapat dilihat oleh manusia, sehingga menjamin ekstraksi harmoni desain yang objektif dan presisi (Gao et al., 2025).

3. METODE PENELITIAN



Gambar 1. Diagram Alir Metodologi Penelitian Klasifikasi Gaya Visual Poster.

Penelitian ini menggunakan pendekatan eksperimental komputasional dengan membangun sistem pemrosesan dua tahap yang mengintegrasikan arsitektur *Vision Transformer* (ViT) (Dosovitskiy et al., 2021) dan algoritma *K-Means Clustering*. Data penelitian berupa himpunan data mandiri (*custom dataset*) yang dikumpulkan melalui ekstraksi data (*scraping*) dari platform portofolio desain profesional Behance, berfokus pada tiga kelas gaya visual: *Bauhaus*, *Swiss Style*, dan *Minimalist*. Validasi penentuan kelas (*ground truth*) diterapkan menggunakan metode *Primary Self-Labeling*, yakni identifikasi yang bersumber langsung dari *tagging* dan deskripsi kreator aslinya. Data mentah yang terkumpul sebanyak 1.102 citra disaring melalui tahap *Exploratory Data Analysis* (EDA) yang meliputi deteksi integritas berkas untuk membuang gambar korup, filter resolusi dan batas rasio ekstrem, serta penghapusan citra duplikat menggunakan *Perceptual Hashing* (pHash).

Tahap pra-pemrosesan data (*Data Preprocessing*) dirancang menggunakan pendekatan jalur ganda (*dual-pipeline*). Setelah seluruh citra distandarisasi ke format RGB (mengeliminasi kanal *Alpha*) dan diseimbangkan (*undersampling*) menjadi total 1.011 sampel (337 gambar per kelas), jalur pemrosesan dipecah. Pada jalur klasifikasi, citra diubah ukurannya menjadi resolusi 224×224 piksel menggunakan teknik penambahan batas tepi (*padding*) agar rasio aspek asli poster tidak terdistorsi, lalu dikonversi menjadi representasi *Tensor* dengan normalisasi statistik. Pada jalur ekstraksi warna, citra dikonversi ke ruang warna perseptual CIELab* dan disegmentasi untuk membentuk *Superpixel SLIC*. Dataset kemudian dibagi menggunakan teknik *stratified sampling* (80% data latih, 10% data validasi, dan 10% data uji) dan dimuat ke dalam model secara dinamis (*on-the-fly*).

Evaluasi keandalan sistem dilakukan dengan mengukur performa metrik pada masing-masing arsitektur. Kinerja model ViT dievaluasi menggunakan *Confusion Matrix* yang menghasilkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* (Wang et al., 2026). Di ranah klasifikasi tingkat halus (*fine-grained*), *F1-Score* menjadi indikator krusial untuk menilai keseimbangan performa model secara objektif di tengah tingginya kemiripan fitur antar-kelas (Wei et al., 2022).

Sementara itu, kualitas pengelompokan warna dari K-Means dievaluasi secara matematis untuk memastikan titik warna yang dihasilkan benar-benar representatif menggunakan parameter *Silhouette Score* dan *Inertia Score* (Kösesoy et al., 2025). Algoritma ekstraksi K-Means dikonfigurasi untuk menarik lima klaster warna dominan karena format lima warna terbukti empiris mewakili harmoni visual karya (Sharma et al., 2023), memiliki keseimbangan fungsi dan estetika (Gao et al., 2025), serta mendukung penerapan prinsip proporsi 60-30-10 (Garrison et al., 2022). Keakuratan warna palet hasil komputasi kemudian

divalidasi terhadap citra aslinya menggunakan rumus perbedaan warna perseptual ΔE^* standar CIEDE2000 (Han et al., 2022).

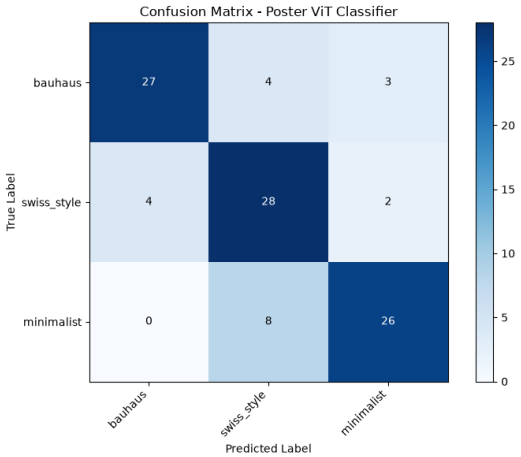
4. HASIL DAN PEMBAHASAN

Berdasarkan serangkaian tahapan pra-pemrosesan data dan pengujian komputasional yang telah dilakukan, implementasi sistem otomatisasi kecerdasan buatan ini menghasilkan beberapa luaran utama. Temuan tersebut secara spesifik mencakup analisis kinerja arsitektur *Vision Transformer* dalam mengenali gaya desain poster, evaluasi kualitas algoritma *K-Means Clustering* dalam mengekstraksi skema warna dominan, serta hasil integrasi keseluruhan model tersebut ke dalam sebuah antarmuka aplikasi interaktif. Secara lebih rinci, hasil dan pembahasan dari penelitian ini diuraikan sebagai berikut:

Hasil Klasifikasi Gaya Visual (Vision Transformer)

Tahap klasifikasi gaya visual menggunakan arsitektur *Vision Transformer* (ViT) yang telah disesuaikan (*fine-tuning*) dari model prapelatihan *vit-base-patch16-224* menunjukkan performa komputasi yang sangat memuaskan. Pada proses pelatihan, kurva validasi mencapai titik paling optimal pada iterasi (*epoch*) ke-4 dengan akurasi validasi 80,20%, sebelum akhirnya dihentikan secara paksa oleh mekanisme proteksi *early stopping* pada iterasi ke-13 guna mencegah degradasi performa akibat *overfitting*.

Hasil pengujian model pada himpunan data uji (*test set*) yang berjumlah 102 citra poster membuktikan keandalan sistem dengan tingkat akurasi keseluruhan sebesar 79% dan rata-rata harmonik makro (*Macro F1-Score*) sebesar 0,80. Secara spesifik, gaya *Bauhaus* meraih *F1-Score* tertinggi (0,83) dengan nilai *Precision* 0,87, gaya *Swiss Style* memiliki tingkat sensitivitas (*Recall*) tertinggi (0,82), sementara gaya *Minimalist* mencatat nilai *Precision* 0,84.

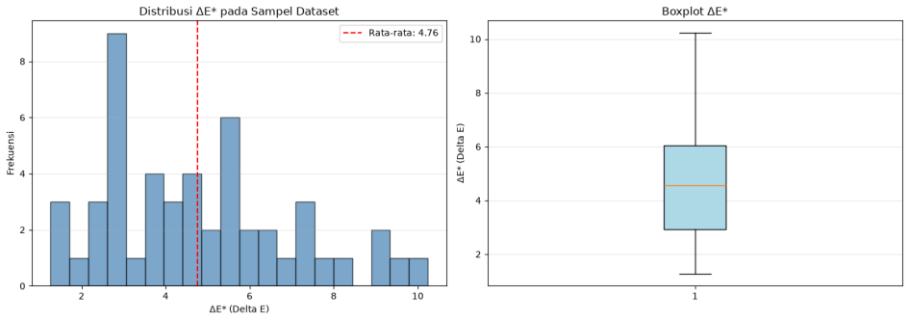


Gambar 2. Confusion Matrix Model Vision Transformer (ViT) dalam Klasifikasi Gaya Visual Poster.

Analisis pada *Confusion Matrix* memperlihatkan bahwa pemusatan prediksi mayoritas berada pada diagonal utama yang menandakan keberhasilan klasifikasi, dengan kesalahan prediksi silang (*misclassification*) tertinggi terjadi pada 8 poster *Minimalist* yang disalahartikan sebagai *Swiss Style*. Kesalahan prediksi antar-gaya desain yang memiliki kemiripan tata letak ini merupakan tantangan utama yang diakui secara akademis dalam analisis citra tingkat halus atau *Fine-Grained Image Analysis* (FGIA) (Conde et al., 2021; Wei et al., 2022). Khususnya pada domain seni dan desain, pengenalan gaya visual sangat kompleks karena batas antar-gaya yang seringkali ambigu dan terjadinya fusi gaya (Wang et al., 2026). Kendati demikian, pencapaian akurasi 79% pada sistem ini secara signifikan melampaui rentang performa model *baseline* pada dataset pengenalan gaya seni yang umumnya hanya berkisar di angka 62% hingga 70% (Wang et al., 2026). Tingginya performa ini membuktikan bahwa mekanisme *self-attention* pada arsitektur ViT sangat tangguh untuk mengekstraksi wilayah diskriminatif pada tugas klasifikasi berbutir halus (Conde et al., 2021; Dosovitskiy et al., 2021; He et al., 2022).

Ekstraksi Skema Warna dan Analisis Klastering

Pada modul ekstraksi warna, kombinasi pra-pemrosesan *SLIC Superpixel* dan algoritma *K-Means* sukses mengekstraksi lima palet warna dominan dari setiap citra poster ($K = 5$). Evaluasi kuantitatif terhadap kualitas pemisahan klaster algoritma *K-Means* mencatatkan rata-rata *Silhouette Score* sebesar 0,7698 dan *Inertia Score* sebesar 11.090,23. Merujuk pada standar analisis klaster, skor *Silhouette* yang melampaui angka 0,7 mengindikasikan bahwa struktur klaster warna yang dihasilkan sangat kuat, padat secara internal, dan terpisah dengan sangat baik tanpa adanya tumpang tindih yang signifikan (Yulisasih et al., 2024).



Gambar 3. Distribusi dan Persebaran Nilai ΔE* pada Dataset Poster.

Validasi akurasi palet warna digital terhadap gambar aslinya diukur secara objektif menggunakan metrik perbedaan warna perseptual CIEDE2000 (ΔE^*) untuk menghindari bias subjektivitas visual (Kösesoy et al., 2025). Hasil komputasi menunjukkan rata-rata nilai ΔE^* sebesar 4,7648. Berdasarkan pedoman standar persepsi manusia, rentang nilai $2 < \Delta E^* < 10$ menginterpretasikan bahwa perbedaan palet warna hasil ekstraksi dengan warna aslinya dapat

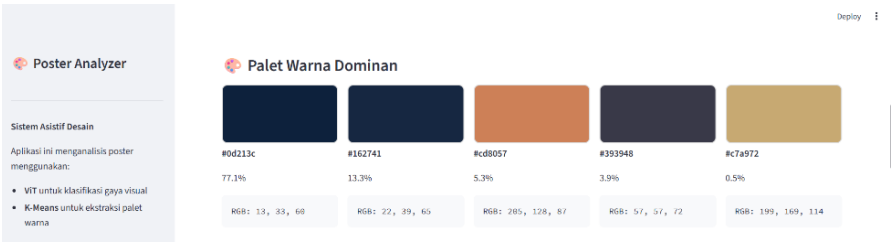
terlihat jelas sekilas (*perceptible at a glance*), namun secara esensial masih lebih mirip dengan aslinya daripada berbeda (Cherepkova et al., 2024). Mengingat toleransi bias alami mata manusia sangat konsisten jika berada di bawah angka 5 Delta-E 2000 (Cherepkova et al., 2024), hasil komputasi ini membuktikan bahwa palet warna (kode HEX) telah beroperasi secara akurat di dalam batas kewajaran deviasi visual manusia.

Lebih lanjut, analisis statistik terhadap keseluruhan dataset (1.011 poster) memperlihatkan bahwa distribusi frekuensi proporsi warna memiliki pola yang condong ke kanan (*right-skewed*). Hierarki proporsinya sangat terstruktur: Klaster 1 sebagai warna latar belakang mendominasi kanvas rata-rata sebesar 56,6%, yang kemudian menurun pada warna pendamping di Klaster 2 (21,2%) dan Klaster 3 (11,6%), hingga menyisakan elemen aksen minor di Klaster 5 sebesar 3,7%. Temuan ini secara empiris sejalan dengan prinsip dasar estetika aturan 60-30-10, di mana desain yang harmonis umumnya mengalokasikan sekitar 60% area untuk warna dominan, 30% untuk warna sekunder, dan 10% untuk aksen (Garrison et al., 2022).

Implementasi Sistem sebagai Asisten Desain Cerdas



Gambar 4. Antarmuka Sistem Analisis Poster Berbasis AI untuk Klasifikasi Gaya Visual.



Gambar 5. Hasil Ekstraksi Palet Warna Dominan pada Poster Menggunakan K-Means.

Seluruh arsitektur pemrosesan tersebut berhasil diimplementasikan ke dalam sebuah purwarupa aplikasi berbasis web interaktif menggunakan kerangka kerja *Streamlit*. Sistem menyajikan hasil analisis yang transparan dengan merender label prediksi gaya visual berserta

distribusi metrik probabilitasnya (*confidence score*), serta blok cuplikan warna (*color swatch*) dari palet hasil ekstraksi. Informasi teknis yang meliputi kode heksadesimal (HEX), nilai komposisi RGB, dan persentase dominasi disajikan secara terstruktur di dalam tabel interaktif, menjadikan aplikasi ini sebuah sistem asistif cerdas yang praktis, objektif, dan siap diimplementasikan oleh kreator ke dalam alur kerja desain mereka.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa arsitektur *Vision Transformer* (ViT) berhasil diimplementasikan secara efektif untuk mengklasifikasikan gaya visual poster pada tingkat halus (*fine-grained*) dengan pencapaian akurasi sebesar 79% dan rata-rata harmonik makro (*Macro F1-Score*) sebesar 0,80. Pengintegrasian algoritma *K-Means Clustering* dengan segmentasi *SLIC Superpixel* pada ruang warna CIELab* juga sukses mengotomatisasi ekstraksi skema warna digital secara akurat. Hal ini dibuktikan dengan rata-rata *Silhouette Score* mencapai 0,7698 yang menandakan pemisahan klaster yang sangat baik, serta nilai deviasi jarak warna perseptual (ΔE^*) sebesar 4,7648 yang terbukti beroperasi secara wajar di dalam batas deviasi penglihatan mata manusia. Keseluruhan model tersebut telah berhasil diimplementasikan ke dalam purwarupa aplikasi berbasis web menggunakan kerangka kerja *Streamlit* sebagai asisten desain yang cerdas, transparan, dan berstandar industri.

Adapun saran untuk pengembangan penelitian selanjutnya adalah memperbesar volume dataset citra dan memperluas cakupan kategori gaya desain grafis kontemporer, seperti gaya *Retro*, *Cyberpunk*, atau *Corporate Design*, guna memperkuat kemampuan generalisasi arsitektur ViT. Selain itu, disarankan untuk mengintegrasikan algoritma deteksi objek atau *Optical Character Recognition* (OCR) agar sistem mampu mengekstrak dan memisahkan palet warna secara lebih spesifik berdasarkan peran elemennya di dalam kanvas, seperti membedakan antara warna teks dan warna latar belakang. Terakhir, sistem asistif desain ini sangat direkomendasikan untuk diunggah ke peladen komputasi awan (*Cloud Server*) komersial agar dapat diakses secara daring, stabil, dan berkesinambungan oleh para pelaku UMKM maupun kreator digital.

DAFTAR REFERENSI

Chen, Y. et al. (2021) "Investigation of designers' colour selection process," *Color Research and Application*, 46(3), hal. 557–565. Tersedia pada: https://doi.org/10.1002/col.22631.

Cherepkova, O., Amirshahi, S.A. dan Pedersen, M. (2024) "Individual Contrast Preferences in Natural Images," *Journal of Imaging*, 10(1). Tersedia pada: https://doi.org/10.3390/jimaging10010025.

Conde, M. V. dan Turgutlu, K. (2021) "Exploring Vision Transformers for Fine-grained Classification," (iii). Tersedia pada: http://arxiv.org/abs/2106.10587.

Dosovitskiy, A. et al. (2021) "an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale," *ICLR 2021 - 9th International Conference on Learning Representations* [Preprint].

Frackiewicz, M. dan Palus, H. (2022) "Efficient Color Quantization Using Superpixels," *Sensors*, 22(16). Tersedia pada: https://doi.org/10.3390/s22166043.

Frackiewicz, M., Palus, H. dan Prandzioch, D. (2023) "Superpixel-Based PSO Algorithms for Color Image Quantization," *Sensors*, 23(3). Tersedia pada: https://doi.org/10.3390/s23031108.

Gao, Y., Liang, J. dan Yang, J. (2025) "Color Palette Generation From Digital Images: A Review," *Color Research and Application*, 50(3), hal. 250–265. Tersedia pada: https://doi.org/10.1002/col.22975.

Garrison, L. dan Bruckner, S. (2022) "Considering best practices in color palettes for molecular visualizations," *Journal of integrative bioinformatics*, 19(2), hal. 1–14. Tersedia pada: https://doi.org/10.1515/jib-2022-0016.

Han, A., Kim, J. dan Ahn, J. (2022) "Color Trend Analysis using Machine Learning with Fashion Collection Images," *Clothing and Textiles Research Journal*, 40(4), hal. 308–324. Tersedia pada: https://doi.org/10.1177/0887302X21995948.

He, J. et al. (2022) "TransFG: A Transformer Architecture for Fine-Grained Recognition," *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, 36, hal. 1174–1182. Tersedia pada: https://doi.org/10.1609/aaai.v36i1.19967.

Hsu, H.Y. et al. (2023) "PosterLayout: A New Benchmark and Approach for Content-Aware Visual-Textual Presentation Layout," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2023-June, hal. 6018–6026. Tersedia pada: https://doi.org/10.1109/CVPR52729.2023.00583.

Huang, D. et al. (2023) "A Survey for Graphic Design Intelligence." Tersedia pada: http://arxiv.org/abs/2309.01371.

Kösesoy, M.B. dan Yilmaz, S. (2025) "A Novel Color Difference-Based Method for Palette Extraction and Evaluation Using Images of Birds," *IEEE Access*, 13(January), hal. 52270–52283. Tersedia pada: https://doi.org/10.1109/ACCESS.2025.3553252.

Liu, S. et al. (2024) "Image-Driven Harmonious Color Palette Generation for Diverse Information Visualization," *IEEE Transactions on Visualization and Computer Graphics*, 30(7), hal. 3089–3103. Tersedia pada: https://doi.org/10.1109/TVCG.2022.3226218.

Sharma, S., Tandukar, J. dan Bista, R. (2023) "Generating Harmonious Colors through the Combination of n-Grams and K-means," *Journal of Computing Theories and Applications*, 1(2), hal. 140–150. Tersedia pada: https://doi.org/10.33633/jcta.v1i2.9470.

Wang, M. et al. (2026) "An art style classification network integrating contrastive learning and counterfactual attention," *PLOS One*, 21(6), hal. e0343954. Tersedia pada: https://doi.org/10.1371/journal.pone.0343954.

Wei, X.S. et al. (2022) "Fine-Grained Image Analysis With Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), hal. 8927–8948. Tersedia pada: https://doi.org/10.1109/TPAMI.2021.3126648.

Yulisasih, B.N. et al. (2024) "Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation," *ILKOM Jurnal Ilmiah*, 16(3), hal. 330–342. Tersedia pada: https://doi.org/10.33096/ilkom.v16i3.2325.330-342.