

IMPLEMENTASI MODEL DeiT UNTUK MEMBEDAKAN GAMBAR BUATAN AI DAN MANUSIA PADA ILUSTRASI ANIMASI 2D

Ibnu Taimiyah Erwin^{1*}; Abdul Latief Arda²; Imran Taufik³; Muhammad Erwin Rosyadi. S⁴; Hilyatul Auliyah Erwin⁵

Sistem Komputer^{1,2,3}
Universitas Handayani Makassar, Makassar, Indonesia^{1,2,3}
<https://handayani.ac.id>^{1,2,3}
ibnuneer17@gmail.com^{1*}, abdullatief@handayani.ac.id², imran_taufik@handayani.ac.id³

Rekam Medis dan Informasi Kesehatan⁴
Sekolah Tinggi Ilmu Kesehatan Panakkukang Makassar, Makassar, Indonesia⁴;
<https://stikespanakkukang.ac.id>⁴
erwinrosyadi@gmail.com⁴

Teknik Informatika⁵
Universitas Dipa Makassar, Makassar, Indonesia⁵
<https://site.undipa.ac.id>⁵
auiyahapadule@gmail.com⁵
(* Corresponding Author



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract— *The development of artificial intelligence (AI) has influenced various fields, including art and visual design. AI Generative Art, which mimics human styles, has sparked debates on originality, artistic value, as well as legal and ethical challenges. Therefore, methods are needed to distinguish between AI-generated and human-made images, particularly in 2D animation illustrations. This study proposes the use of Data-efficient Image Transformers (DeiT) for image classification. Two models tested are DeiT Base and DeiT Tiny, using a dataset of 6,000 images equally divided between AI and human categories. The dataset is split into training (70%), validation (15%), and testing (15%). Experimental results show that DeiT Base achieves over 95% accuracy with fast convergence and optimal loss function stability. Meanwhile, DeiT Tiny attains around 93% accuracy, being more computationally efficient despite requiring more epochs for stability. Compared to previous models using a larger dataset (11,000 images per category) but achieving only 80% accuracy, DeiT performs better in both accuracy and computational efficiency, even with a smaller dataset. In conclusion, DeiT is effective for classifying 2D animation images. DeiT Base excels in accuracy and convergence speed, while DeiT Tiny is more resource-efficient, making it an ideal choice for environments with computational constraints.*

Keywords: *artificial intelligence, deit, image classification, transformers, 2D animation.*

Abstrak— Perkembangan kecerdasan buatan (AI) telah memengaruhi berbagai bidang, termasuk seni dan desain visual. AI Generative Art, yang meniru gaya manusia, memicu perdebatan tentang orisinalitas, nilai artistik, serta tantangan hukum dan etis. Oleh karena itu, diperlukan metode untuk membedakan gambar AI-generated dan human-made, terutama dalam ilustrasi animasi 2D. Penelitian ini mengusulkan penggunaan Data-efficient Image Transformers (DeiT) untuk klasifikasi gambar. Dua model yang diuji adalah DeiT Base dan DeiT Tiny, menggunakan dataset 6.000 gambar, terbagi rata dalam kategori AI dan manusia. Dataset ini dibagi menjadi pelatihan (70%), validasi (15%), dan pengujian (15%). Hasil eksperimen menunjukkan DeiT Base mencapai akurasi lebih dari 95% dengan konvergensi cepat dan stabilitas loss function optimal. Sementara itu, DeiT Tiny memperoleh akurasi sekitar 93%, lebih efisien secara komputasi meski membutuhkan lebih banyak epoch untuk stabilitas. Dibandingkan model sebelumnya dengan dataset lebih besar (11.000 gambar per kategori) yang hanya mencapai 80% akurasi,

DeiT lebih unggul dalam akurasi dan efisiensi komputasi meskipun menggunakan dataset lebih kecil. Kesimpulannya, DeiT efektif untuk klasifikasi gambar animasi 2D. DeiT Base unggul dalam akurasi dan kecepatan konvergensi, sedangkan DeiT Tiny lebih hemat sumber daya, menjadikannya pilihan ideal untuk lingkungan dengan keterbatasan komputasi.

Kata kunci: kecerdasan buatan, *deit*, klasifikasi gambar, *transformer*, animasi 2D.

PENDAHULUAN

Perkembangan kecerdasan buatan (AI) telah memberikan dampak yang signifikan dalam berbagai bidang, termasuk seni dan desain. AI kini mampu menghasilkan gambar yang dikenal sebagai *AI Generative Art*, yang dihasilkan oleh algoritma berdasarkan data visual tertentu (Giorgetti, 2024). Karya-karya ini sering kali meniru gaya manusia, sehingga menimbulkan perdebatan mengenai orisinalitas dan nilai artistik (Bellaiche et al., 2023). Penggunaan AI generatif dalam seni gambar menimbulkan dilema etika seperti disinformasi, produksi konten berkualitas rendah, dan hambatan komunikasi antar-pemangku kepentingan (Vyas, 2022). Perkembangan AI dalam seni dan desain meningkatkan efisiensi kreatif, tetapi berpotensi mendegradasi nilai humanis dan memicu ambiguitas hak cipta (Amankwah-Amoah et al. 2024). *Generative Adversarial Networks* (GANs) dan model transformasi seperti ViT telah digunakan untuk menghasilkan gambar animasi 2D (Huang et al. 2023). Namun, karya AI sering kehilangan elemen emosional yang melekat pada karya manusia (Samo et al. 2023). Salah satu contoh yang mencolok adalah saat sebuah karya seni yang dihasilkan oleh AI memenangkan penghargaan dalam ajang *Colorado State Fair*, yang memicu diskusi mengenai apakah karya seni yang diciptakan oleh AI dapat dianggap setara dengan karya manusia.

Di tengah diskusi tersebut, muncul tantangan yaitu bagaimana cara membedakan gambar yang dibuat oleh AI dari gambar yang dibuat oleh manusia, terutama dalam ilustrasi animasi 2D. Karya seni berbasis AI cenderung kehilangan kedalaman emosional dan intensi subjektif yang lahir dari pengalaman hidup manusia, meskipun secara teknis mampu meniru estetika (Johnson & Patel, 2024). Oleh karena itu, diperlukan metode yang efektif untuk mengklasifikasikan dan membedakan antara gambar buatan AI dan gambar buatan manusia guna melindungi hak cipta dan orisinalitas karya seni tersebut.

Dalam bidang pengenalan gambar, *Vision Transformer* (ViT) telah muncul sebagai inovasi baru yang merevolusi pendekatan dalam klasifikasi gambar. ViT menggunakan mekanisme *self-attention* yang awalnya dikembangkan untuk pemrosesan bahasa alami, tetapi diadaptasi untuk

memproses gambar dengan membagi gambar menjadi *patch* kecil yang dianggap sebagai token. Dengan demikian, ViT dapat mengatasi keterbatasan arsitektur konvolusi tradisional yang sering kali memerlukan *inductive bias* lokal seperti translasi dan kesetaraan spasial (Dosovitskiy et al., 2021).

Sebelumnya, salah satu metode yang digunakan untuk membedakan gambar buatan AI dan manusia adalah BEiT (*Bidirectional Encoder Representations from Image Transformers*), sebuah varian dari *Vision Transformer* (ViT). BEiT telah menunjukkan performa yang baik dalam tugas-tugas klasifikasi gambar, termasuk dalam membedakan gambar yang dihasilkan oleh AI dan manusia. BEiT menggunakan mekanisme *pre-training* berbasis *transformer* untuk representasi gambar (Prawiratama et al., 2024).

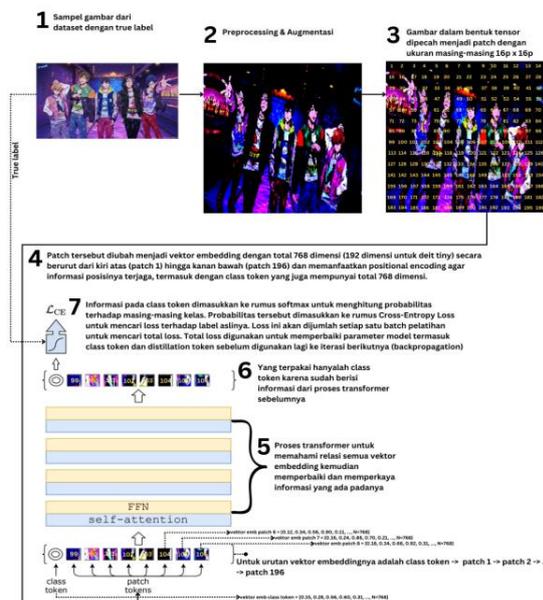
Penelitian tersebut menggunakan *dataset* yang lebih besar, yaitu 11.000 gambar di masing-masing kategori (*AI-generated* dan *human-made*), dengan total 22.000 gambar, dan mencapai akurasi 80% menggunakan model BEiT. Meskipun demikian, BEiT bukan satu-satunya varian dari ViT. Ada varian lain yang dikenal sebagai *Data-efficient Image Transformers* (DeiT), yang menawarkan keunggulan dalam efisiensi data dan komputasi (Touvron et al., 2021).

Selain itu, DeiT juga memiliki fleksibilitas melalui beberapa varian model, seperti *DeiT Base* dan *DeiT Tiny*, yang dapat disesuaikan dengan kebutuhan komputasi tanpa mengorbankan akurasi. Meskipun begitu, model DeiT belum diterapkan secara luas dalam klasifikasi gambar antara AI dan manusia pada ilustrasi animasi 2D. Hal ini juga sejalan dengan temuan (Ha et al., 2024) yang menyoroti potensi *transformer* dalam membedakan gambar AI dan manusia

Oleh karena itu, penelitian ini bertujuan untuk menyelidiki bagaimana DeiT dapat berfungsi sebagai model yang efektif dan akurat untuk klasifikasi gambar buatan AI dan manusia dalam ilustrasi animasi 2D. Penelitian ini juga bertujuan untuk mengeksplorasi varian DeiT dan menemukan model yang paling optimal untuk menyelesaikan tugas klasifikasi ini.

BAHAN DAN METODE

Berdasarkan Gambar 1, alur kerja sistem ini terdiri dari beberapa tahapan mulai dari *preprocessing* data hingga model mampu menghasilkan label prediksi. Sampel yang digunakan baik untuk sampel gambar maupun nilai dari suatu variabel yang nantinya akan digunakan untuk melakukan perhitungan dari implementasi alur kerja sistem ini adalah sampel yang langsung bersumber dari *training*, *validating*, dan *testing* yang dihasilkan oleh model pada sistem ini.



Sumber: (Hasil Penelitian, 2024)
Gambar 1. Alur Kerja Sistem

Dalam penelitian ini, metode yang digunakan meliputi beberapa tahapan penting mulai dari pengumpulan *dataset* hingga analisis hasil. Pada awalnya, *dataset* yang terdiri dari gambar-gambar buatan AI dan gambar buatan manusia dalam format ilustrasi animasi 2D dikumpulkan. Sumber data ini berasal dari berbagai platform digital seperti Safebooru dan Zerochan, serta melalui proses pembuatan manual melalui aplikasi *generate* gambar AI. Koleksi ini kemudian dibagi menjadi dua kategori utama, yaitu Gambar Buatan AI dan Gambar Buatan Manusia, dengan masing-masing kategori berisi 3000 gambar. Distribusi *dataset* dilakukan dengan proporsi 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian, guna memastikan kecukupan data dalam setiap fase eksperimen. Contoh dari gambar buatan AI dan buatan manusia bisa dilihat pada Gambar 2 hingga Gambar 5.

Gambar 2 dan Gambar 3 merupakan contoh dari gambar yang dibuat oleh AI, sedangkan Gambar 4 dan Gambar 5 merupakan contoh gambar yang dibuat oleh manusia.



Sumber: (ZeroChan, 2024c)
Gambar 2. Contoh Gambar Buatan AI



Sumber: (ZeroChan, 2024a)
Gambar 3. Contoh Gambar Buatan AI



Sumber: (ZeroChan, 2024b)
Gambar 4. Contoh Gambar Buatan Manusia



Sumber: (Safebooru, 2024)

Gambar 5. Contoh Gambar Buatan Manusia

Selanjutnya, tahap *preprocessing* dan augmentasi data dilakukan sebelum proses pelatihan. Teknik augmentasi yang digunakan meliputi *resize*, *flip horizontal*, penyesuaian kecerahan/kontras, serta *shift scale rotate* yang diimplementasikan menggunakan pustaka *Albumentations*. Augmentasi ini bertujuan untuk meningkatkan keragaman visual dalam *dataset* pelatihan sehingga model dapat belajar mengenali berbagai pola dengan lebih efektif.

Untuk fase pelatihan, arsitektur *Data-efficient Image Transformers* (DeiT) diadopsi. Dua varian dari DeiT, yaitu *DeiT base* dan *DeiT tiny*, dipilih untuk dievaluasi dalam penelitian ini. Kedua model ini dilatih menggunakan *dataset* yang telah diproses, dengan fokus pada perbandingan efisiensi komputasi dan akurasi yang dihasilkan oleh masing-masing model.

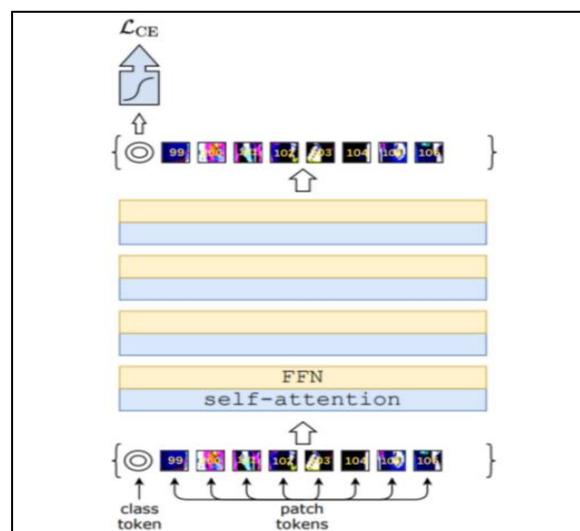
Proses pelatihan dan validasi model melibatkan pemantauan metrik akurasi dan *loss*, baik untuk set pelatihan maupun validasi. Setelah fase pelatihan, model-model tersebut diuji menggunakan subset pengujian untuk menilai kinerja mereka. Metrik evaluasi yang digunakan meliputi akurasi, *confusion matrix*, dan laporan klasifikasi. Selain itu, visualisasi dari metrik *loss* dan akurasi juga disajikan untuk setiap *epoch*, memberikan wawasan lebih lanjut tentang dinamika pembelajaran model (Kang et al., 2024).

Analisis hasil merupakan tahap terakhir dari penelitian ini. Evaluasi kinerja model dilakukan untuk menentukan varian DeiT yang paling efektif dan efisien dalam membedakan antara gambar buatan AI dan gambar buatan manusia. Perbandingan antara DeiT *base* dan DeiT *tiny* akan

difokuskan pada aspek akurasi dan efisiensi komputasi, memberikan rekomendasi model yang paling sesuai untuk implementasi lebih lanjut dalam konteks serupa.

HASIL DAN PEMBAHASAN

Berdasarkan Gambar 6 yaitu arsitektur DeiT *Base*, setiap *patch* gambar diubah menjadi vektor *embedding* melalui operasi linear yang menghasilkan vektor 768 dimensi (Bbouzidi et al. 2024). Proses melibatkan perkalian vektor RGB yang dinormalisasi dengan matriks bobot dan penambahan bias. Setelah itu, *class token* dan *positional embedding* ditambahkan untuk memberikan informasi posisi dalam *transformer*.



Sumber: (Hasil Penelitian, 2024)

Gambar6 . Arsitektur DeiT

Selanjutnya, proses dalam *transformer* meliputi *Multi-Head Self-Attention* (MHA) yang mengolah *embedding* melalui perhitungan *query*, *key*, dan *value*. *Attention score* dihitung menggunakan *dot product* dan dinormalisasi dengan fungsi *softmax* untuk menghasilkan bobot *attention*. *Output* dari *self-attention* kemudian diolah melalui *Feed-Forward Neural Network* (FFN) yang terdiri dari dua lapisan *fully connected* untuk mengolah dan mengembalikan dimensi *embedding* ke ukuran semula.

Setelah melalui normalisasi lapisan, *class token* yang telah diproses digunakan untuk klasifikasi melalui *fully connected layer* yang membedakan gambar menjadi *AI-generated* atau *Human-made*. Proses klasifikasi ini menggunakan fungsi *softmax* untuk mengonversi *logit* menjadi probabilitas.

Dalam fase pelatihan, *Cross-Entropy Loss* digunakan untuk mengukur kesalahan prediksi, dan

model dioptimalkan melalui *backpropagation* dan AdamW yang merupakan *optimizer* dengan *adaptive learning rates* dan *weight decay*. Proses ini membantu dalam mengurangi *overfitting* dan mempercepat konvergensi model.

Secara keseluruhan, proses ini menunjukkan bagaimana DeiT Base mengintegrasikan komponen-komponen *neural network* untuk mengolah dan mengklasifikasikan gambar dengan efektif. Berikut hasil dari beberapa pengujian fungsionalitas dari sistem yang dibuat pada Gambar 7 hingga 14.



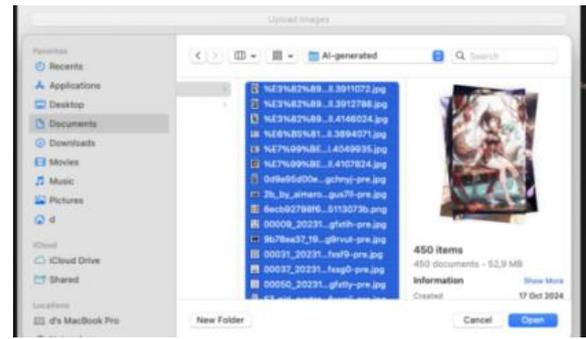
Sumber: (Hasil Penelitian, 2024)
Gambar 7. Menerapkan Augmentasi Pada Gambar Latih

```
Augmented sample images saved in outputs/batch_10_samples, with original filenames saved in
Logits, probabilities, loss, total loss, dan accuracy disimpan di outputs/batch_10_logits.p
Epoch 1/10: 100%
Epoch [1/10] - Loss: 0.2631, Accuracy: 0.8862, Time: 495.82s
Validation - Average Loss: 0.1773, Average Accuracy: 0.9333
Best model saved at epoch 1 with validation accuracy 0.9333
Starting Epoch 2/10...
Epoch 2/10: 24%
```

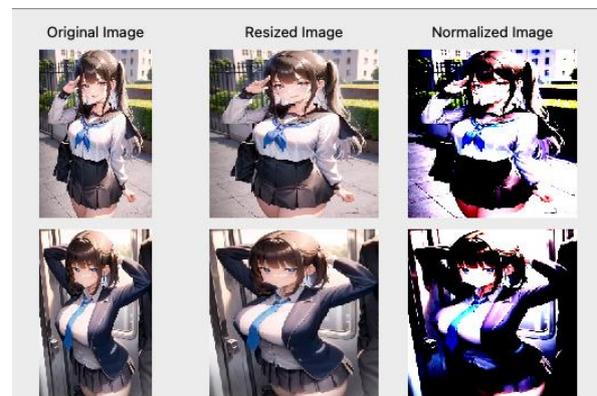
Sumber: (Hasil Penelitian, 2024)
Gambar 8. Melakukan Iterasi Pelatihan

```
Best model saved at epoch 1 with validation accuracy 0.9333
Starting Epoch 2/10...
```

Sumber: (Hasil Penelitian, 2024)
Gambar 9. Simpan Model Terbaik Berdasarkan Akurasi Validasi



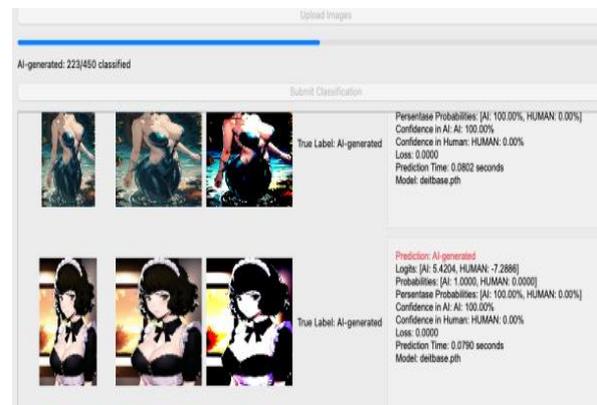
Sumber: (Hasil Penelitian, 2024)
Gambar 10. Pengguna Mengunggah Gambar Melalui Antarmuka Pada Fase Pengujian



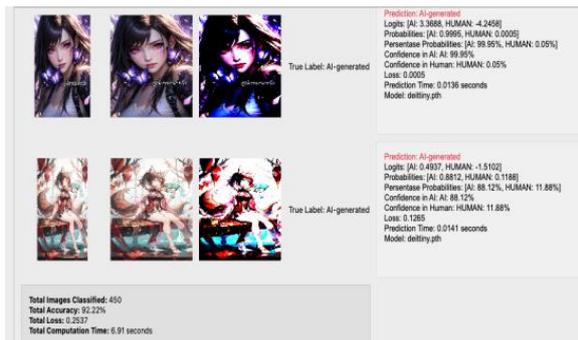
Sumber: (Hasil Penelitian, 2024)
Gambar 11. Resize dan Normalisasi Gambar Yang Diunggah Pada Fase Pengujian

```
deitbase.pth
✓ deittiny.pth
Model deittiny.pth loaded and ready for evaluation.
Model deitbase.pth loaded and ready for evaluation.
Model deittiny.pth loaded and ready for evaluation.
```

Sumber: (Hasil Penelitian, 2024)
Gambar 12. Memuat Model Yang Dipilih Pada Fase Pengujian

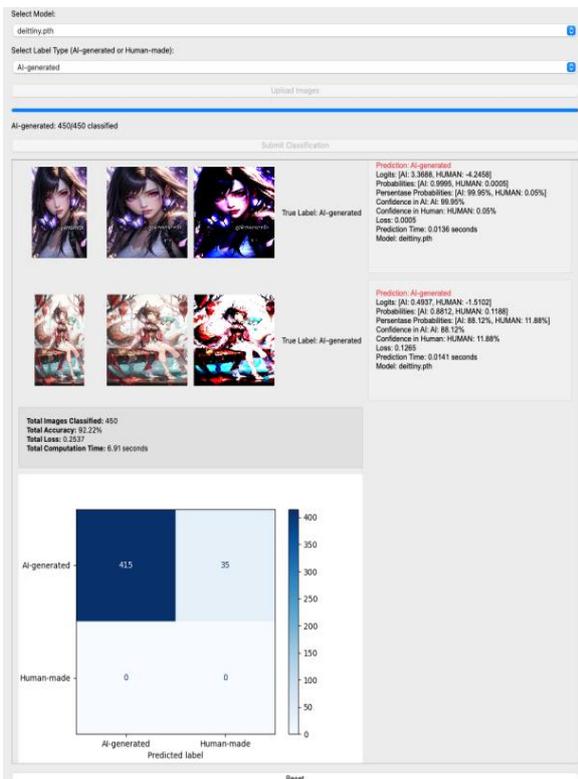


Sumber: (Hasil Penelitian, 2024)
Gambar 13. Progress Model Memprediksi Label Gambar Pada Fase Pengujian



Sumber: (Hasil Penelitian, 2024)
 Gambar 14. Menampilkan Hasil Klasifikasi Di UI Pada Fase Pengujian

Pada Gambar 15 menampilkan antarmuka sistem yang digunakan pada fase pengujian, model yang telah melalui proses pelatihan dan validasi diuji dengan menggunakan *dataset* yang sepenuhnya baru, yang disebut *dataset testing*. Uji coba ini bertujuan untuk mengukur kinerja model secara objektif pada data yang belum pernah digunakan selama fase pelatihan dan validasi, yang krusial untuk memverifikasi kemampuan generalisasi model (Maleki et al. 2020).



Sumber: (Hasil Penelitian, 2024)
 Gambar 15. Antarmuka Sistem

Dua model berbasis *Transformer* dari DeiT, yaitu DeiT *Base* dan DeiT *Tiny*, diimplementasikan untuk mengklasifikasikan dua jenis gambar: yang

dihasilkan oleh AI dan yang dibuat oleh manusia. Hasil klasifikasi ini kemudian dibandingkan dengan label yang sebenarnya menggunakan *confusion matrix* untuk memberikan gambaran tentang performa model dalam kondisi yang mirip dengan aplikasi nyata. Evaluasi pada *dataset testing* yang belum pernah dilihat sebelumnya memungkinkan penilaian terhadap kemampuan generalisasi model serta memastikan bahwa model tidak mengalami *overfitting* terhadap data pelatihan dan validasi.

Hasil klasifikasi untuk Model DeiT *Base* menunjukkan bahwa pada *dataset AI-generated*, dari total 450 gambar yang diuji, model mencapai akurasi sebesar 94.00% dengan 423 gambar diklasifikasikan dengan benar, sementara 27 gambar salah diklasifikasikan sebagai *Human-made*, dan *loss* yang dicatat adalah 0.1980. Pada *dataset Human-made*, model mencapai akurasi 96.44% dengan 434 gambar diklasifikasikan dengan benar dan 16 salah diklasifikasikan sebagai *AI-generated*, dengan *loss* sebesar 0.0722. Sementara itu, Model DeiT *Tiny* pada *dataset AI-generated* mencapai akurasi 92.22% dengan 415 gambar diklasifikasikan dengan benar dan 35 gambar salah diklasifikasikan sebagai *Human-made*, dengan *loss* sebesar 0.2537. Pada *dataset Human-made*, model mencapai akurasi 94.44% dengan 425 gambar diklasifikasikan dengan benar dan 25 salah diklasifikasikan sebagai *AI-generated*, dengan *loss* sebesar 0.2147.

Dari hasil tersebut, terlihat bahwa kedua model lebih sering membuat kesalahan dalam mendeteksi gambar *AI-generated* yang salah diklasifikasikan sebagai *Human-made*, menunjukkan bahwa gambar *AI-generated* lebih sulit untuk dideteksi secara akurat. Secara keseluruhan, klasifikasi gambar *Human-made* lebih baik, dengan kesalahan yang lebih sedikit dibandingkan dengan *AI-generated*. Namun, baik DeiT *Base* maupun DeiT *Tiny* masih menunjukkan beberapa kesalahan dalam mengklasifikasikan gambar *Human-made* sebagai *AI-generated*, menandakan bahwa masih ada ruang untuk peningkatan dalam pengembangan model lebih lanjut. Secara keseluruhan, penelitian ini menganalisis performa dua arsitektur model, yakni DeiT *Base* dan DeiT *Tiny*, dalam mengklasifikasikan gambar *AI-generated* dan *Human-made*.

Berdasarkan Tabel 2 dan 3, evaluasi dilakukan menggunakan *confusion matrix* yang mengkategorikan hasil prediksi ke dalam empat metrik utama: *True Positive* (TP), *False Negative* (FN), *True Negative* (TN), dan *False Positive* (FP).

Tabel 2. *Confussion Matrix* DeiT Base

<i>Confusion Matrix</i>	<i>True Positive</i>	<i>True Negative</i>	<i>False Positive</i>	<i>False Negative</i>
<i>Training</i>	2091	2088	9	12
<i>Validating</i>	427	432	23	18
<i>Testing</i>	423	434	27	16

Sumber: (Hasil Penelitian, 2024)

Tabel 3. *Confussion Matrix* DeiT Tiny

<i>Confusion Matrix</i>	<i>True Positive</i>	<i>True Negative</i>	<i>False Positive</i>	<i>False Negative</i>
<i>Training</i>	2060	2062	40	38
<i>Validating</i>	417	412	33	38
<i>Testing</i>	415	425	35	25

Sumber: (Hasil Penelitian, 2024)

Berdasarkan hasil analisis *classification report* pada Tabel 4 yang diperoleh dari model DeiT Base, model ini menunjukkan performa yang sangat baik pada semua metrik evaluasi, baik pada tahap pelatihan, validasi, maupun pengujian. Pada *dataset* pengujian, model ini berhasil mencapai akurasi sebesar 95,2%, dengan *precision* dan *recall* untuk gambar *AI-generated* masing-masing 0,94 dan 0,96, serta untuk gambar *Human-made* masing-masing 0,94 dan 0,94. Hasil *F1-Score* untuk kedua kategori ini juga menunjukkan keseimbangan yang sangat baik, masing-masing mencapai 0,952. Meskipun memiliki waktu pelatihan yang lebih lama (226,334 detik), model ini menunjukkan konvergensi yang lebih cepat dan stabilitas *loss function* yang baik, dengan *loss* sebesar 0,017 pada tahap pelatihan.

Tabel 4. *Classification Report* DeiT Base

<i>Metrics</i>	<i>Training</i>	<i>Validating</i>	<i>Testing</i>
<i>Precision (AI-generated)</i>	0.996	0.949	0.94
<i>Recall (AI-generated)</i>	0.994	0.959	0.964
<i>F1-Score (AI-generated)</i>	0.995	0.954	0.952
<i>Support (AI-generated)</i>	2100	450	450
<i>Precision (Human-made)</i>	0.994	0.96	0.941
<i>Recall (Human-made)</i>	0.996	0.949	0.941
<i>F1-Score (Human-made)</i>	0.995	0.954	0.952
<i>Support (Human-made)</i>	2100	450	450
<i>Accuracy</i>	0.995	0.954	0.952
<i>Loss</i>	0.017	0.186	0.135
<i>Time</i>		226.334	38.55

Sumber: (Hasil Penelitian, 2024)

Sementara itu pada Tabel 5, model DeiT Tiny, meskipun lebih efisien dalam waktu pelatihan (66,168 detik), menunjukkan sedikit penurunan dalam performa dibandingkan dengan DeiT Base. Pada pengujian, akurasi model ini tercatat 93,3%, dengan *precision* dan *recall* untuk gambar *AI-generated* masing-masing 0,92 dan 0,94, serta untuk gambar *Human-made* masing-masing 0,92 dan 0,92. *F1-Score* untuk kedua kategori ini lebih rendah

dibandingkan DeiT Base, yaitu 0,933, dan *loss* pada pengujian lebih tinggi (0,234) dibandingkan DeiT Base.

Tabel 5. *Classification Report* DeiT Tiny

<i>Metrics</i>	<i>Training</i>	<i>Validating</i>	<i>Testing</i>
<i>Precision (AI-generated)</i>	0.981	0.927	0.922
<i>Recall (AI-generated)</i>	0.982	0.915	0.944
<i>F1-Score (AI-generated)</i>	0.981	0.92	0.933
<i>Support (AI-generated)</i>	2100	450	450
<i>Precision (Human-made)</i>	0.981	0.926	0.924
<i>Recall (Human-made)</i>	0.982	0.915	0.922
<i>F1-Score (Human-made)</i>	0.981	0.92	0.933
<i>Support (Human-made)</i>	2100	450	450
<i>Accuracy</i>	0.981	0.921	0.933
<i>Loss</i>	0.054	0.269	0.234
<i>Time</i>		66.168	7.78

Sumber: (Hasil Penelitian, 2024)

Dibandingkan dengan penelitian sebelumnya yang menggunakan arsitektur BEiT dengan akurasi 80%, penelitian ini menunjukkan bahwa model DeiT, baik Base maupun Tiny, memberikan hasil yang lebih optimal dalam hal akurasi dan efisiensi komputasi. Dalam penelitian sebelumnya, meskipun model BEiT mencapai skor F1 yang tinggi dan menunjukkan keseimbangan yang baik antara *precision* dan *recall*, hasil akurasi yang dicapai hanya 80%, yang lebih rendah dibandingkan dengan hasil yang diperoleh dari kedua model DeiT. Oleh karena itu, arsitektur DeiT terbukti lebih unggul dalam hal performa klasifikasi gambar animasi 2D, dengan DeiT Base memberikan hasil yang lebih akurat dan stabil, sementara DeiT Tiny menawarkan efisiensi komputasi yang lebih baik, meskipun dengan sedikit penurunan dalam metrik evaluasi lainnya.

Model dalam penelitian ini dilatih menggunakan *dataset* yang dibagi menjadi tiga subset utama: pelatihan, validasi, dan pengujian. Setiap subset ini diproses menggunakan kedua varian model DeiT untuk mendapatkan prediksi klasifikasi. Hasil prediksi tersebut kemudian dibandingkan dengan data aktual (*ground truth*) menggunakan fungsi *confusion_matrix* dari *library sklearn.metrics*, yang memberikan *insight* mendalam mengenai performa kedua model.

Pada subset pelatihan, DeiT Base menunjukkan performa yang sangat baik dengan *True Positive* (TP) sebesar 2091, yang mengindikasikan akurasi tinggi dalam mendeteksi gambar yang dihasilkan oleh AI. *True Negative* (TN) yang dicapai adalah 2088, menunjukkan kemampuan superior model dalam mengidentifikasi gambar buatan manusia. Model ini hanya memiliki sedikit kesalahan dengan *False*

Positive (FP) sebanyak 9 dan *False Negative* (FN) sebanyak 12. Sementara itu, *DeiT Tiny* menunjukkan performa yang sedikit lebih rendah dengan TP 2060, TN 2062, FP 40, dan FN 38, yang menandakan variasi kecil dalam kemampuan deteksi yang tepat.

Evaluasi pada *dataset* validasi mengungkapkan bahwa *DeiT Base* kembali menunjukkan konsistensi yang lebih superior dengan TP 427, TN 432, FP 23, dan FN 18. Di sisi lain, *DeiT Tiny* mencatat TP 417, TN 412, FP 33, dan FN 38, menunjukkan bahwa *DeiT Base* lebih unggul dalam generalisasi model.

Analisis *dataset* pengujian menunjukkan bahwa *DeiT Base* mempertahankan performa optimalnya dengan TP 423, TN 434, FP 27, dan FN 16. *DeiT Tiny*, meskipun menunjukkan hasil yang baik, tetap suboptimal dengan TP 415, TN 425, FP 35, dan FN 25. Dalam semua metrik evaluasi, *DeiT Base* menunjukkan superioritas statistik, khususnya terlihat dari tingkat *False Positive* dan *False Negative* yang secara konsisten lebih rendah dibandingkan dengan *DeiT Tiny*.

Berdasarkan analisis terhadap ketiga subset data, model *DeiT Base* menunjukkan performa yang lebih baik dalam klasifikasi gambar *AI-generated* dan *human-made* dibandingkan dengan *DeiT Tiny*. Model ini tidak hanya menghasilkan tingkat akurasi yang lebih tinggi, tetapi juga menunjukkan *error rate* yang lebih rendah, yang mengindikasikan potensi reliabilitas yang lebih baik untuk implementasi praktis. Meskipun demikian, perlu dicatat bahwa hasil ini terbatas pada *dataset* animasi 2D dengan dua kategori gambar *AI-generated* dan *human-made* sehingga generalisasi hasil untuk kategori gambar lain atau jenis *dataset* yang berbeda mungkin memerlukan penelitian lebih lanjut.

DeiT Base juga menunjukkan keunggulan dalam hal presisi dan daya ingat, terutama pada fase validasi dan pengujian, meskipun metrik evaluasi mengindikasikan bahwa perbedaannya dengan *DeiT Tiny* tidak terlalu besar. Hal ini menunjukkan bahwa meskipun *DeiT Base* lebih akurat, performa *DeiT Tiny* masih cukup solid, dengan keunggulan dalam efisiensi komputasi, yang menjadikannya lebih cocok untuk aplikasi di lingkungan dengan keterbatasan sumber daya.

Meskipun *DeiT Base* lebih andal dalam hal akurasi dan kesalahan yang lebih rendah, penelitian ini menghadapi beberapa batasan, antara lain ukuran *dataset* yang terbatas (6.000 gambar) dan fokus pada klasifikasi gambar animasi 2D dalam dua kategori spesifik. Oleh karena itu, hasil yang diperoleh mungkin tidak sepenuhnya dapat diterapkan pada *dataset* atau domain gambar yang

lebih luas tanpa pengujian lebih lanjut (Khan et al., 2024).

KESIMPULAN

Dalam hal konvergensi, *DeiT Base* menunjukkan penurunan *loss function* yang konsisten dan stabil, menandakan bahwa model telah mempelajari pola data secara efektif tanpa *overfitting*. Sementara itu, *DeiT Tiny*, meskipun menunjukkan penurunan *loss* selama pelatihan, belum mencapai konvergensi yang optimal, yang terlihat dari fluktuasi pada *loss function* dan akurasi yang lebih rendah. Hal ini menunjukkan bahwa *DeiT Tiny* mungkin memerlukan penyesuaian lebih lanjut untuk mencapai konvergensi yang lebih baik.

Secara keseluruhan, *DeiT Base* lebih unggul dalam mendeteksi gambar *AI-generated* dan *Human-made*, sementara *DeiT Tiny*, meskipun efisien, menunjukkan kinerja yang lebih rendah dalam hal akurasi. Hasil ini menegaskan bahwa *DeiT Base* merupakan pilihan yang lebih baik untuk aplikasi yang membutuhkan akurasi tinggi, sedangkan *DeiT Tiny* cocok untuk aplikasi yang mengutamakan efisiensi komputasi dengan beberapa kompromi pada akurasi. Penelitian terbaru menunjukkan bahwa model berbasis *transformer* seperti AniGAN (Li et al., 2022) atau teknik augmentasi berbasis perhatian (Huang et al., 2023) dapat meningkatkan kemampuan deteksi. Namun, akurasi masih menjadi tantangan (Ha et al., 2024).

DeiT Base lebih cocok digunakan dalam skenario di mana akurasi tinggi menjadi prioritas utama, seperti dalam aplikasi industri yang membutuhkan klasifikasi gambar *AI-generated* dan *Human-made* dengan tingkat kepercayaan tinggi. Namun, jika efisiensi komputasi lebih penting, seperti pada aplikasi *real-time* atau yang berjalan pada perangkat dengan sumber daya terbatas, *DeiT Tiny* bisa menjadi pilihan tepat, meskipun akurasi mungkin tidak setinggi *DeiT Base*. Untuk meningkatkan performa *DeiT Tiny*, disarankan melakukan *fine-tuning* lebih lanjut atau menggunakan teknik *knowledge distillation* dari *DeiT Base* untuk mentransfer pengetahuan model yang lebih besar ke model yang lebih kecil, sehingga akurasi dapat ditingkatkan tanpa mengorbankan kecepatan komputasi.

REFERENSI

Amankwah-Amoah, J., Abdalla, S., Mogaji, E., Elbanna, A., & Dwivedi, Y. K. (2024). The impending disruption of creative industries by generative AI: Opportunities, challenges, and research agenda. *International Journal of*

- Information Management*, Vol. 79.
<https://doi.org/10.1016/j.ijinfomgt.2024.10.2759>
- Bbouzidi, S., Hcini, G., Jdey, I., & Drira, F. (2024). *Convolutional Neural Networks and Vision Transformers for Fashion MNIST Classification: A Literature Review*. Retrieved from <https://arxiv.org/abs/2406.03478>
- Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., ... Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, 8(1). <https://doi.org/10.1186/s41235-023-00499-6>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2021). AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE. *ICLR 2021 - 9th International Conference on Learning Representations*.
- Giorgetti, I. (2024). Exploring the Intersection of Art, Human Creativity, and AI. *Itinera*. <https://doi.org/10.54103/2039-9251/27841>
- Ha, A. Y. J., Passananti, J., Bhaskar, R., Shan, S., Southen, R., Zheng, H., & Zhao, B. Y. (2024). Organic or Diffused: Can We Distinguish Human Art from AI-generated Images? *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, 4822–4836. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3658644.3670306>
- Huang, Z., Xie, H., Fukusato, T., & Miyata, K. (2023). AniFaceDrawing: Anime Portrait Exploration during Your Sketching. *Proceedings - SIGGRAPH 2023 Conference Papers*. <https://doi.org/10.1145/3588432.3591548>
- Kang, K., Setlur, A., Ghosh, D., Steinhardt, J., Tomlin, C., Levine, S., & Kumar, A. (2024). *What Do Learning Dynamics Reveal About Generalization in LLM Reasoning?* Retrieved from <https://arxiv.org/abs/2411.07681>
- Khan, F. F., Kim, D., Jha, D., Mohamed, Y., Chang, H. H., Elgammal, A., ... Elhoseiny, M. (2024). AI Art Neural Constellation: Revealing the Collective and Contrastive State of AI-Generated and Human Art. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 7470–7478. <https://doi.org/10.1109/CVPRW63382.2024.00742>
- Li, B., Zhu, Y., Wang, Y., Lin, C. W., Ghanem, B., & Shen, L. (2022). AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation. *IEEE Transactions on Multimedia*, 24. <https://doi.org/10.1109/TMM.2021.3113786>
- Maleki, F., Muthukrishnan, N., Ovens, K., Md, C., & Forghani, R. (2020). Machine Learning Algorithm Validation. *Neuroimaging Clinics of North America*, 30, 433–445. <https://doi.org/10.1016/j.nic.2020.08.004>
- Prawiratama, R. A., Sumarno, & Kautsar, I. A. (2024). RANCANG BANGUN APLIKASI UJI KEMIRIPAN GAMBAR AI GENERATIVE DAN GAMBAR BUATAN TANGAN MENGGUNAKAN METODE DEEP LEARNING. *TEKINKOM*. Retrieved from <https://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/1192/599>
- Safebooru. (2024). Safebooru / original.
- Samo, A., & Highhouse, S. (2023). Artificial Intelligence and Art: Identifying the Aesthetic Judgment Factors That Distinguish Human and Machine-Generated Artwork. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000570>
- Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. *International Conference on Machine Learning*, 10347–10357.
- Vyas, B. (2022). Ethical Implications of Generative AI in Art and the Media. *International Journal For Multidisciplinary Research*, 4, 1–11.
- ZeroChan. (2024a). NovelAI - AI Art - ZeroChan Anime Image Board.
- ZeroChan. (2024b). Official Art - ZeroChan Anime Image Board.
- ZeroChan. (2024c). StableDiffusion - AI Art - ZeroChan Anime Image Board.