

# Performa *Backbone* ResNet50 dan MobileNetV2 pada DeeplabV3+ untuk Segmentasi Karakter Komik Lokal

I Gede Bintang Arya Budaya<sup>1</sup>, Gede Angga Pradipta<sup>2</sup>

Program Studi Teknologi Informasi<sup>1</sup>, Program Studi Sistem Informasi<sup>2</sup>

Institut Teknologi dan Bisnis STIKOM Bali

Denpasar, Indonesia

e-mail: <sup>1</sup>bintang@stikom-bali.ac.id, <sup>2</sup>angga\_pradipta@stikom-bali.ac.id

Diajukan: ....; Direvisi: ....; Diterima: ....

## Abstrak

*Dalam analisa dokumen digital, khususnya komik, pengenalan karakter menjadi salah satu langkah penting terkait potensi pengaplikasiannya di industri, seperti analisa konten, perlindungan hak cipta, dan penerapannya sebagai perangkat pintar interaktif. Penelitian ini bertujuan untuk menerapkan teknik analisa dokumen digital pada komik dengan cara menguji potensi penerapan teknik segmentasi semantik dalam mengenali karakter pada komik lokal. Dua arsitektur DeepLabV3+ dengan backbone ResNet50 dan MobileNetV2 diimplementasikan untuk menganalisis performa pada dataset kecil. Tujuan penelitian ini adalah untuk memahami karakteristik karakter komik lokal dan kontribusi pada perkembangan teknik analisis dokumen digital. Namun, hasil prediksi pada kedua arsitektur menunjukkan tingkat akurasi yang belum memuaskan, terutama dalam mengenali karakter pada kasus penelitian ini. Meskipun demikian, model dengan backbone MobileNetV2 menunjukkan performa yang lebih baik karena mampu mengenali lebih banyak kelas daripada ResNet50 yang cenderung lebih overfitting. Penelitian ini menyimpulkan bahwa pengenalan karakter dalam komik lokal merupakan tantangan yang memerlukan pendekatan lebih cermat dalam mengatasi variasi dan kompleksitas gambar. Hasil penelitian ini dapat menjadi landasan bagi penelitian selanjutnya, seperti penambahan dataset, penyesuaian hyperparameter dan teknik menangani kelas yang tidak seimbang, serta penerapan teknik untuk memisahkan mengabaikan kelas yang tidak digunakan. Diharapkan penelitian ini memberikan wawasan baru bagi pengembangan teknologi analisis dokumen digital dan penerapan dalam konteks komik lokal di masa depan.*

**Kata kunci:** pretrained model, segmentasi citra, dataset citra komik

## Abstract

*In digital document analysis, particularly in comics, character recognition is a crucial step with potential applications in various industries, such as content analysis, copyright protection, and its integration into interactive smart devices. This research aims to apply digital document analysis techniques to comics by testing the potential application of semantic segmentation in character recognition within local comics. Two DeepLabV3+ architectures with ResNet50 and MobileNetV2 backbones are implemented to analyze their performance on a small dataset. The objective of this study is to understand the characteristics of local comic characters and their contribution to the advancement of digital document analysis techniques. However, the prediction results of both architectures show unsatisfactory accuracy, especially in recognizing characters in this particular research case. Nevertheless, the model with MobileNetV2 backbone demonstrates better performance as it can recognize more classes compared to ResNet50, which tends to be more prone to overfitting. This research concludes that character recognition in local comics is a challenging task that requires more careful approaches to address image variations and complexity. The findings of this study can serve as a basis for future research, such as dataset expansion, hyperparameter adjustments, and techniques to handle class imbalance, as well as the implementation of techniques to ignore unused classes. The research is expected to provide new insights into the development of digital document analysis technology and its application in the context of local comics in the future.*

**Keywords:** pretrained model, image segmentation, comic image dataset

## 1. Pendahuluan

Komik merupakan media yang digunakan untuk menyampaikan suatu ide dengan dukungan visual grafis dan berisi teks yang disusun secara berurutan. Secara global, komik sudah menjadi suatu kebudayaan populer, negara seperti Amerika Serikat, Jepang, China dan Korea Selatan menjadi negara yang memiliki industri komik terbesar [1]. Industri komik lokal di Indonesia juga telah mengalami perkembangan yang signifikan dalam beberapa tahun terakhir [2]. Bersamaan dengan semakin tingginya penggunaan internet dan *platform* media sosial untuk berbagi konten gambar yang membuat komik digital semakin mudah diakses dan semakin banyak peminat [3], [4]. Dalam hal penggambaran, para kreator komik di Indonesia banyak yang menciptakan karakter komik mereka secara original, namun ada juga kreator yang membuat karakter komik yang mereferensikan pada kebudayaan lokal seperti pada tokoh pewayangan. Karakter Delem dan Sangut adalah salah satu contoh dari karakter yang mereferensikan secara langsung dari pengetahuan kebudayaan lokal. Mereka merupakan tokoh punakawan dari salah satu kisah pewayangan dan telah menjadi ikon budaya yang populer, khususnya di Provinsi Bali. Kedua karakter tersebut sering digunakan untuk menyampaikan suatu pesan atau kisah dalam suatu bentuk komik strip kepada masyarakat umum [5], [6].

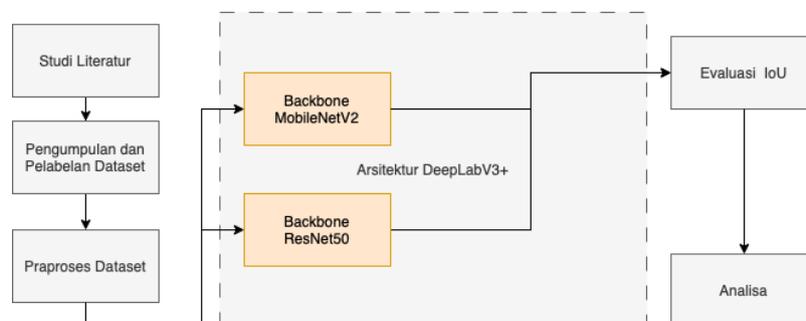
Dalam analisa dokumen digital, khususnya komik, pengenalan karakter menjadi salah satu langkah penting terkait potensi pengaplikasiannya di industri, seperti analisa konten, perlindungan hak cipta, dan penerapannya sebagai perangkat pintar interaktif. Penelitian [7] menerapkan teknik analisis dokumen digital untuk mengetahui jenis objek yang ada dalam dataset komik tersebut yaitu manusia, panel, dan balon teks, dan memang tidak berfokus pada pengenalan karakter yang ada. Penelitian [8] melakukan analisa yang berbeda yaitu segmentasi untuk merekognisi mata, mulut, serta balon teks dan penelitian [9] mencoba merekognisi aksi atau pose dari suatu karakter komik. Hal ini menunjukkan bagaimana analisa dokumen komik digital merupakan sebuah pekerjaan yang besar.

Tantangan utama adalah adanya berbagai ragam gaya dan skenario cerita sehingga dalam proses penggambaran komik tersebut akan menghasilkan representasi gambar akhir yang lebih rumit dan berbeda. Berdasarkan hal tersebut dapat disimpulkan bahwa setiap karakter dari setiap kreator memiliki keunikan tersendiri yang menyebabkan model rekognisi yang dikembangkan, tidak bisa secara penuh mengenali setiap informasi dalam komik digital tersebut. Ditambah karakter Delem dan Sangut ini sendiri bisa dibilang sebagai sebuah karakter unik yang memang belum ada peneliti yang menggunakan sebagai sebuah studi kasus dataset untuk rekognisi karakter. Hal ini mendasari diperlukannya suatu teknik untuk mengenali setiap karakter tersebut dalam berbagai skenario penggambaran, khususnya bagi komik lokal yang mereferensikan pada tokoh dan kisah kebudayaan lokal yang jarang diketahui oleh masyarakat umum.

Pendekatan yang digunakan untuk merekognisi suatu karakter dari komik adalah dengan menggunakan teknik *deep learning* untuk segmentasi semantik. DeepLabv3+ [8], [10] adalah arsitektur segmentasi semantik yang terbukti memiliki performa yang baik dalam berbagai masalah segmentasi. Dalam penelitian ini, arsitektur DeeplabV3+ akan dibuat dengan menggunakan dengan *backbone* ResNet50 dan MobileNetV2, beberapa penelitian menunjukkan performa baik dari *backbone* ResNet50 [11], [12] dan terkait MobileNetV2 yang mampu meningkatkan akurasi segmentasi gambar [13], [14].

## 2. Metode Penelitian

Adapun metode penelitian segmentasi karakter komik lokal tokoh Delem dan Sangut ini dapat dilihat pada Gambar 1 berikut.



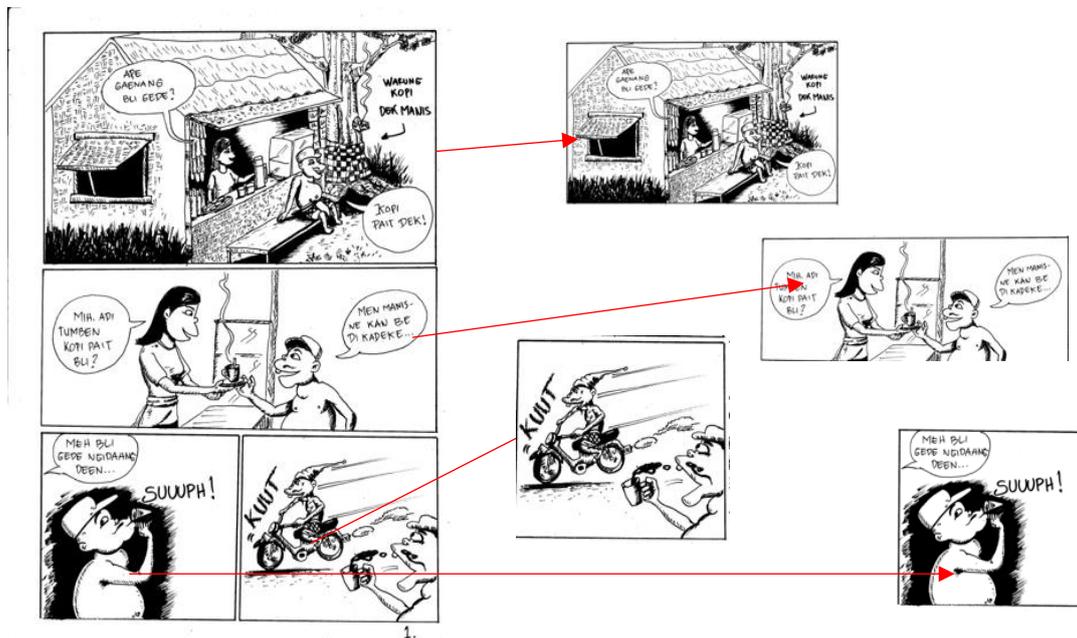
Gambar 1. Metode penelitian

2.1. Studi Literatur

Pada tahap ini peneliti melaksanakan studi literatur khususnya mengenai perkembangan komik digital serta bagaimana peran kecerdasan buatan dalam mendukung industri komik saat ini.

2.2. Pengumpulan Dataset dan Pelabelan Dataset

Pada tahap ini dikumpulkan sumber – sumber komik digital yang berkaitan dengan kasus penelitian ini. Dalam penelitian ini, memfokuskan pada penerapan teknik segmentasi gambar karakter komik loka dimana tokoh yang difokuskan adalah tokoh Delem dan Sangut. Sumber dataset berasal dari salah satu kreator komik lokal Komang Roy Prismayudi dengan judul “Delem Jak Sangut”. Komik ini selanjutnya dibagi berdasarkan panel. Adapun total panel yang dihasilkan sebanyak 126 panel yang menjadi sumber dataset utama. Gambar 2 menampilkan proses untuk merubah halaman komik menjadi panel sebagai dataset untuk kasus segmentasi karakter komik tokoh Delem dan Sangut ini.



Gambar 2. Proses membentuk panel sebagai dataset

Dataset yang terbentuk selanjutnya akan dilaksanakan proses pelabelan. Dalam kasus penelitian ini, akan dilaksanakan pelabelan dataset untuk membentuk *ground truth* karakter Delem dan Sangut, pelabelan terdiri dari 3 kelas, pertama kelas Delem, kedua kelas Sangut, dan yang ketiga adalah *background*, setiap panel memiliki pasangan *ground truth*nya. Tabel 1 menampilkan kode warna untuk proses pelabelan dataset sebagai *ground truth*. Total panel yang akhirnya dapat digunakan berdasarkan studi kasus ini adalah 89 panel.

Tabel 1. Kode warna kelas karakter.

No	Kelas	Kode Warna RGB
1	Background	[0,0,0]
2	Delem	[112,112,112]
3	Sangut	[34,34,34]

2.3. Praproses Dataset

Pada tahap ini dilaksanakan praproses dataset yang bertujuan untuk melakukan standarisasi dataset serta melaksanakan proses augmentasi dataset untuk meningkatkan variabilitas dari dataset [15]. Standarisasi yang dimaksud adalah memastikan warna pada pelabelan kelas yang digunakan sudah sama dan merubah ukuran gambar menjadi ukuran yang ditentukan, dalam kasus ini ukuran yang ditentukan

sebagai input data adalah 256 x 256. Dalam proses augmentasi dataset dilakukan dengan cara melakukan transformasi dataset melalui proses *flipping* secara vertikal dan horizontal serta *random rotation*. Satu gambar diaugmentasi menjadi 8 gambar sehingga total dari 89 gambar menjadi 712 gambar augmentasi. Pada tahap ini pula dilaksanakan normalisasi untuk setiap input gambar dan proses pembagian dataset menjadi data latih sebanyak 70% dan data uji sebanyak 30%.

#### 2.4. Pembuatan Model dengan Arsitektur DeepLabV3+

Pembuatan model segmentasi pada dataset, dimana pada tahap ini mengimplementasikan dua arsitektur jaringan saraf untuk proses segmentasi semantik yaitu DeepLabv3+ [10]. Model dibuat dengan menggunakan 2 *pretrained* model sebagai *backbone* yaitu ResNet50 [11], [12] dan MobileNetV2 [13], [14]. Pengaturan parameter adalah *batch size* sebesar 8 agar komputasi yang digunakan lebih rendah, dan pengaturan *epoch* standar 100 dengan penerapan metode *early stopping* jika tidak ada peningkatan performa model.

#### 2.5. Evaluasi

Pada tahap ini akan menganalisa performa dari arsitektur DeeplabV3+ dengan *backbone* ResNet50 dan MobileNetV2. Matriks evaluasi yang digunakan untuk menilai hasil segmentasi adalah *Intersection over Union* (IoU) dan juga berdasarkan pengamatan dari nilai *accuracy* dan *loss* hasil pelatihan dan pengujian. Rumus 1 menampilkan formula dari IoU.

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \dots\dots\dots (1)$$

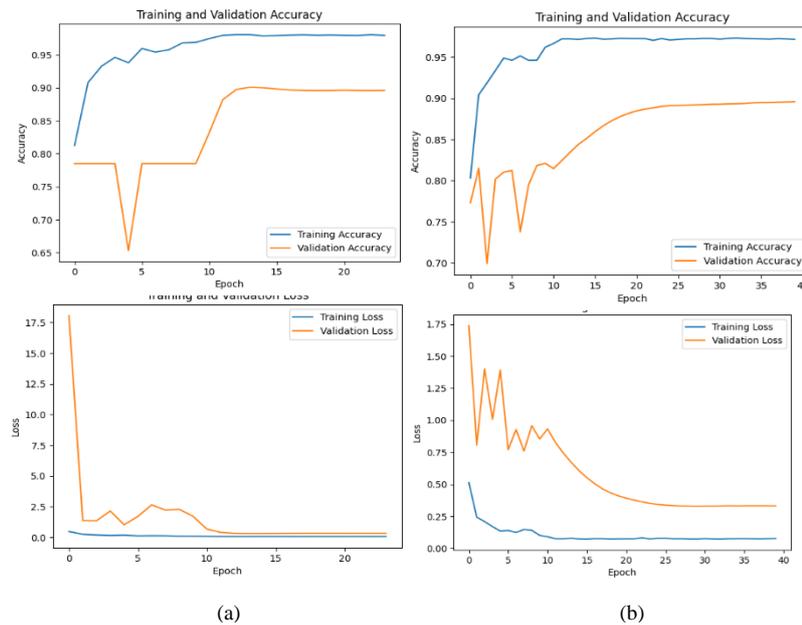
#### 2.6. Analisa

Pada tahap ini akan dilaksanakan analisa terkait performa dari arsitektur DeeplabV3+ dengan *backbone ResNet50* dan *MobileNetV2*.

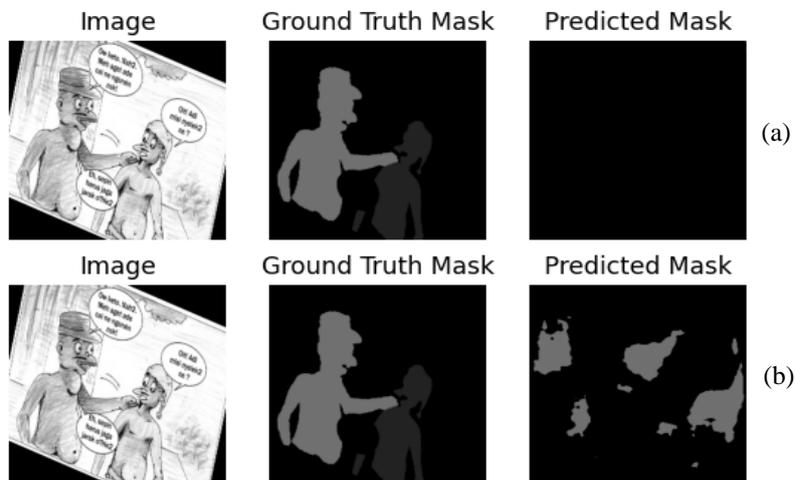
### 3. Hasil dan Pembahasan

Pada Pembuatan model memiliki pengaturan parameter yang sama, Pada Gambar 3 juga menampilkan grafik hasil pelatihan terkait nilai *accuracy* dan *loss*. Pada pelatihan dengan ResNet50, model berhenti pada *epoch* 20 dikarenakan tidak ada peningkatan performa yang signifikan setelah itu. Hal ini dapat mengindikasikan bahwa model ResNet50 mencapai batas kemampuan dalam mempelajari pola-pola yang ada dalam dataset pelatihan pada *epoch* 20. Pada pelatihan dengan MobileNetV2, model berhenti pada *epoch* 40, yang berarti model MobileNetV2 mungkin memerlukan lebih banyak iterasi untuk mencapai tingkat performa yang optimal. ResNet50 memiliki kompleksitas yang lebih besar dibandingkan dengan MobileNetV2 yang lebih ringan.

Kapasitas model yang lebih besar pada ResNet50 dapat menyebabkan *overfitting*, terutama jika jumlah data pelatihan yang terbatas. Ini bisa menyebabkan model ResNet50 lebih cenderung untuk memprediksi kelas mayoritas dalam kasus ini adalah kelas *background* karena kebanyakan data pelatihan adalah kelas tersebut. MobileNetV2, dengan kapasitas yang lebih kecil, dapat lebih general dalam menangkap pola yang berbeda dalam data dan menghasilkan prediksi dengan variasi warna kelas yang lebih banyak seperti yang terlihat pada Gambar 4. Hasil *mean* IoU untuk model ResNet50 sebesar 0,731 dan pada model MobileNetV2 memiliki *mean* IoU sebesar 0,7394, nilai yang dihasilkan terlihat tinggi namun berdasarkan hasil percobaan, masih model masih belum mampu belajar.



Gambar 3 Grafik *accuracy* dan *loss* pada *training* dan *validation* (a) ResNet50 (b) MobileNetV2



Gambar 4 Sampel hasil prediksi dengan kedua model (a) ResNet50 (b) MobileNetV2

#### 4. Kesimpulan

Penelitian ini bertujuan untuk mengetahui potensi penerapan teknik segmentasi untuk kasus segmentasi karakter pada sebuah tokoh komik lokal. Penelitian ini menerapkan arsitektur DeeplabV3+ dengan *backbone* ResNet50 dan MobileNetV2 dan membahas tentang performa dari arsitektur tersebut dalam kasus dataset kecil yang digunakan. Berdasarkan penelitian yang dilaksanakan, diketahui bahwa model dengan arsitektur DeeplabV3+ dengan kedua arsitektur tersebut masih belum mampu mengenali karakter secara benar, khususnya dengan *backbone* ResNet50 yang cenderung *overfitting*. Secara performa dapat diketahui bahwa model dengan arsitektur DeeplabV3+ dengan *backbone* MobileNetV2 memiliki hasil yang lebih baik karena sudah lebih mampu mempelajari kelas selain kelas mayoritas (kelas *background*). Harapannya penelitian ini dapat menjadi salah satu sumber referensi percobaan untuk kedepannya, dimana dari beberapa pengembangan masih dapat dilaksanakan seperti dalam pengaturan

*hyperparameter* dan pengaturan agar kelas lebih seimbang, seperti mengabaikan nilai kelas *background* untuk dipelajari oleh jaringan.

#### Daftar Pustaka

- [1] A. F. Lestari and I. Irwansyah, "Line Webtoon sebagai Industri Komik Digital," *SOURCE: Jurnal Ilmu Komunikasi*, vol. 6, no. 2, pp. 134–148, 2020.
- [2] B. S. Ramadhan, "Kajian Industri Komik Daring Indonesia: Studi Komik Tahilalats," *JSRW (Jurnal Senirupa Warna)*, vol. 8, no. 1, pp. 2–18, 2020.
- [3] G. L. A. K. Putra and G. P. P. A. Yasa, "Komik sebagai sarana komunikasi promosi dalam media sosial," *Jurnal nawala visual*, vol. 1, no. 1, pp. 1–8, 2019.
- [4] A. Resha, "The Blue Age of Comic Books," *Inks: The Journal of the Comics Studies Society*, vol. 4, no. 1, pp. 66–81, 2020.
- [5] N. N. S. Witari, "Pengaruh Tokoh Punakawan Sangut dan Delem Terhadap Pemahaman Pembaca Kartun Editorial Sangut Delem di Harian Bali Post," *Prasi: Jurnal Bahasa, Seni, dan Pengajarannya*, vol. 7, no. 13, 2012.
- [6] N. N.-S. Witari, "NILAI BUDAYA BALI DALAM KARTUN EDITORIAL SANGUT DELEM PADA KORAN BALI POST," *Prasi: Jurnal Bahasa, Seni, dan Pengajarannya*, vol. 11, no. 01, 2016.
- [7] R. J. H. Watanabe, "Complex Character Retrieval from Comics using Deep Learning," 2019.
- [8] A. Terauchi, N. Mori, and M. Ueno, "Analysis of Partial Semantic Segmentation for Images of Four-Scene Comics," in *Distributed Computing and Artificial Intelligence, 17th International Conference*, Springer, 2021, pp. 51–59.
- [9] L. Wang and C. Ma, "Research on Character Action Recognition of Digital Comics," *Procedia Comput Sci*, vol. 208, pp. 286–292, 2022.
- [10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [11] Q. Liu, Y. Dong, Y. Pei, L. Zheng, and L. Zhang, "Long and short-range relevance context network for semantic segmentation," *Complex & Intelligent Systems*, pp. 1–16, 2023.
- [12] R. L. Garcia, P. N. Happ, and R. Q. Feitosa, "Large Scale Semantic Segmentation of Virtual Environments to Facilitate Corrosion Management," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 465–470, 2021.
- [13] M. Cai *et al.*, "Image segmentation method for sweetgum leaf spots based on an improved DeeplabV3+ network," *Forests*, vol. 13, no. 12, p. 2095, 2022.
- [14] C. Chen, C. Wang, B. Liu, C. He, L. Cong, and S. Wan, "Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [15] M. Elgendi *et al.*, "The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective," *Front Med (Lausanne)*, vol. 8, p. 629134, 2021.