BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

Evaluasi Kinerja U-Net ResNet34 dan MDSBN: Studi Komparatif untuk Segmentasi Naskah Kuno Indonesia

Rino Zakharia, Budi Nugroho*, Eka Prakarsa Mandyartha

Fakultas Ilmu Komputer, Prodi Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia
Email: [1]rino.zakharia14@gmail.com, [2,*]budinugroho.if@upnjatim.ac.id, [3]eka_prakarsa.fik@upnjatim.ac.id
Email Penulis Korespondensi: budinugroho.if@upnjatim.ac.id

Abstrak−Digitalisasi dokumen kuno merupakan langkah penting dalam melestarikan informasi historis dan budaya, tetapi citra yang dihasilkan sering mengalami degradasi berupa noda, tekstur latar belakang tidak merata, tinta memudar, dan kontras rendah sehingga pemisahan teks dari latar belakang menjadi sulit. Penelitian ini membandingkan U-Net ResNet34 dan Modified Deep Semantic Binarization Network (MDSBN) untuk segmentasi naskah kuno Indonesia. Dataset terdiri atas manuskrip lontar Bali, naskah Sunda, dan citra dokumen kuno tambahan dari Wikimedia Commons. Eksperimen dilakukan melalui pencarian *learning rate* dan analisis sensitivitas *batch size*, kemudian kinerja model dievaluasi menggunakan Dice Coefficient, Intersection over Union (IoU), precision, recall, dan Root Mean Squared Error (RMSE). Penelitian ini berkontribusi melalui evaluasi terkontrol kedua arsitektur menggunakan dataset, tahapan preprocessing, fungsi loss, metrik evaluasi, dan lingkungan komputasi yang konsisten sehingga perbedaan performa dapat dianalisis secara lebih objektif. Hasil menunjukkan bahwa U-Net ResNet34 memperoleh performa terbaik pada *learning rate* 5e-5 dan *batch size* 16, dengan Dice pengujian sebesar 0,79338 dan IoU sebesar 0,65752. MDSBN memperoleh performa terbaik pada *learning rate* 1e-6 dan *batch size* 32, dengan Dice pengujian sebesar 0,75338 dan IoU sebesar 0,60433. Selain menghasilkan tingkat tumpang tindih segmentasi yang lebih tinggi, U-Net ResNet34 menunjukkan pemisahan teks dan latar belakang yang lebih konsisten pada citra dengan tekstur dan degradasi kompleks. Hasil tersebut mengindikasikan bahwa ekstraksi fitur hierarkis melalui encoder residual membantu model mempertahankan pola goresan teks sekaligus membedakannya dari gangguan latar belakang, sehingga U-Net ResNet34 lebih sesuai untuk dataset naskah kuno Indonesia yang digunakan.

Keywords: Segmentasi Dokumen Kuno; Pembelajaran Mendalam; MDSBN; U-Net ResNet34; Naskah Indonesia; Binarisasi Citra

Abstrak−The digitization of ancient documents is an important step in preserving historical and cultural information. However, the resulting images often suffer from degradation, such as stains, uneven background textures, faded ink, and low contrast, making text-background separation difficult. This study compares two deep learning architectures, namely U-Net ResNet34 and the Modified Deep Semantic Binarization Network (MDSBN), for the segmentation of Indonesian ancient documents. The dataset consists of Balinese palm-leaf manuscripts, Sundanese manuscripts, and additional ancient document images obtained from Wikimedia Commons. The experiments were conducted through a learning rate search and batch size sensitivity analysis, and the models were evaluated using the Dice Coefficient, Intersection over Union (IoU), Precision, Recall, and Root Mean Squared Error (RMSE). This study contributes through a controlled evaluation of both architectures using a consistent dataset, preprocessing pipeline, loss function, evaluation metrics, and computational environment, enabling performance differences to be analyzed more objectively. The results show that U-Net ResNet34 achieved its best performance using a learning rate of 5e-5 and a batch size of 16, with a test Dice score of 0.79338 and a test IoU score of 0.65752. It outperformed MDSBN, which achieved its best performance using a learning rate of 1e-6 and a batch size of 32, with a test Dice score of 0.75338 and a test IoU score of 0.60433. The functional advantage of U-Net ResNet34 is associated with the ability of its residual encoder to extract hierarchical features from complex textures and degradation patterns, while its skip connections help preserve the spatial details of thin text strokes. These characteristics make U-Net ResNet34 more adaptive to variations in degradation within the Indonesian ancient document dataset than the more compact MDSBN architecture.

Keywords: Ancient Document Segmentation; Deep Learning; MDSBN; U-Net ResNet34; Indonesian Manuscript; Image Binarization

1. PENDAHULUAN

Dokumen kuno merupakan salah satu bentuk warisan budaya yang memiliki nilai historis, linguistik, dan ilmiah. Digitalisasi menjadi langkah penting dalam melestarikan informasi tersebut [1], [2]. Namun, citra yang dihasilkan sering mengalami gangguan visual berupa noda, warna media yang tidak merata, tinta memudar, bayangan, tekstur latar belakang kompleks, dan kerusakan fisik. Kondisi tersebut menyulitkan pemisahan teks dari latar belakang sehingga diperlukan metode segmentasi yang mampu menangani variasi degradasi secara akurat [1], [3].

Salah satu tahap penting dalam analisis dokumen adalah segmentasi atau binarisasi citra, yaitu proses mengubah citra masukan menjadi citra biner yang merepresentasikan teks dan latar belakang. Kualitas hasil binarisasi dapat memengaruhi tahap analisis lanjutan, seperti ekstraksi karakter, *optical character recognition*, dan preservasi digital [4]. Tantangan utama muncul karena dokumen kuno umumnya memiliki degradasi kompleks yang tidak dapat ditangani secara konsisten oleh metode *thresholding* sederhana [5], [6], [7].

Metode binarisasi klasik seperti Otsu, Niblack, Sauvola, dan Wolf banyak digunakan karena sederhana dan tidak membutuhkan proses pelatihan. Namun, metode global cenderung kurang stabil pada dokumen berkualitas rendah, sedangkan metode lokal sangat dipengaruhi oleh ukuran jendela dan karakteristik goresan teks [5], [8], [9]. Penelitian [6], [7] menegaskan bahwa tidak terdapat satu metode *thresholding* yang konsisten optimal untuk seluruh jenis citra dokumen. Perkembangan *deep learning* kemudian menghasilkan pendekatan seperti SauvolaNet [10], yang mengadaptasi prinsip Sauvola ke dalam jaringan yang dapat dilatih secara *end-to-end*, serta metode Tensmeyer dan Martinez [11], yang memformulasikan binarisasi sebagai klasifikasi piksel berbasis *Fully Convolutional Network* [12], [13], [14]. Meskipun demikian, kajian lanjutan menunjukkan bahwa binarisasi dokumen tetap menjadi permasalahan terbuka karena perbedaan karakteristik degradasi, dataset, dan protokol evaluasi [6], [7], [15].

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

Arsitektur U-Net dengan struktur *encoder-decoder* dan *skip connection* dikenal efektif dalam mempertahankan informasi spasial pada tugas segmentasi citra [16], [17], [18], [19]. Yuadi dkk. membandingkan sepuluh metode binarisasi klasik dengan U-Net ResNet34 pada manuskrip lontar Bali dan menunjukkan bahwa pendekatan *deep learning* mampu memberikan performa lebih baik [20]. Di sisi lain, Modified Deep Semantic Binarization Network (MDSBN) dikembangkan untuk menangani degradasi pada manuskrip lontar Malayalam dan menunjukkan kemampuan dalam mempertahankan detail teks pada dokumen yang mengalami degradasi berat [21]. MDSBN relevan untuk diuji karena memiliki kesamaan konteks objek berupa manuskrip lontar, tetapi kesesuaiannya terhadap naskah kuno Indonesia dengan variasi aksara, media, tekstur, dan degradasi yang berbeda masih belum banyak dikaji [8], [22]. Selain itu, perbandingan model perlu dilakukan menggunakan dataset dan skema evaluasi yang konsisten karena performanya dapat dipengaruhi oleh karakteristik data dan protokol pengujian [23].

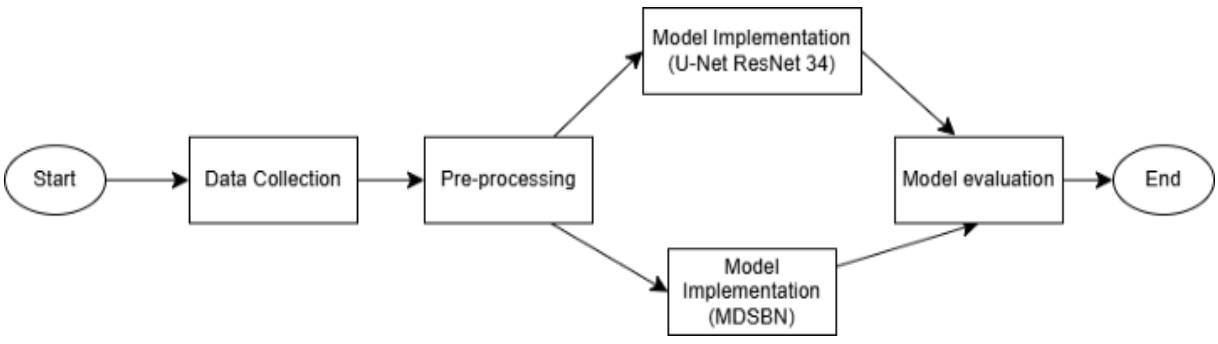
Meskipun berbagai penelitian telah menunjukkan kemajuan dalam binarisasi dokumen terdegradasi, beberapa kesenjangan penelitian masih ditemukan. Sukesh dkk. [23] menunjukkan bahwa berbagai metode binarisasi berbasis deep learning dapat menghasilkan performa yang sangat bervariasi ketika dievaluasi menggunakan dataset dan protokol yang berbeda, sehingga sulit menentukan model yang benar-benar unggul tanpa skema evaluasi yang konsisten. Selain itu, hasil evaluasi mereka menunjukkan bahwa tidak terdapat satu model yang secara konsisten menjadi terbaik pada seluruh dataset pengujian. Xiong dkk. [5] mengembangkan kerangka binarisasi untuk dokumen historis terdegradasi, tetapi penelitian tersebut berfokus pada pengembangan satu metode tanpa mengevaluasi perbedaan performa antararsitektur deep learning. Selanjutnya, Yuadi dkk. [20] menunjukkan bahwa U-Net ResNet34 mampu mengungguli berbagai metode binarisasi klasik pada manuskrip lontar Bali, namun belum membandingkannya dengan arsitektur deep learning lain dalam kondisi eksperimen yang sama. Di sisi lain, Nair dan Shobha Rani [21] mengembangkan MDSBN yang efektif pada manuskrip lontar Malayalam, tetapi kemampuan generalisasinya pada naskah kuno Indonesia dengan karakteristik aksara, tekstur media, dan pola degradasi yang berbeda masih belum diketahui. Oleh karena itu, belum terdapat penelitian yang membandingkan U-Net ResNet34 dan MDSBN secara langsung pada naskah kuno Indonesia menggunakan dataset, tahapan preprocessing, dan skema evaluasi yang konsisten. Akibatnya, pengaruh perbedaan karakteristik arsitektur kedua model terhadap performa segmentasi dan kebutuhan komputasi pada konteks naskah kuno Indonesia masih belum diketahui.

Berdasarkan kesenjangan tersebut, penelitian ini melakukan evaluasi terkontrol terhadap U-Net ResNet34 dan MDSBN pada naskah kuno Indonesia. U-Net ResNet34 dipilih karena memanfaatkan encoder residual untuk ekstraksi fitur hierarkis, sedangkan MDSBN dirancang untuk mempertahankan detail teks pada manuskrip terdegradasi. Oleh karena itu, penelitian ini berkontribusi dalam tiga aspek. Pertama, penelitian ini menyajikan perbandingan langsung antara U-Net ResNet34 dan MDSBN pada tugas segmentasi naskah kuno Indonesia yang hingga saat ini masih jarang dilaporkan. Kedua, penelitian ini menerapkan evaluasi terkontrol menggunakan pembagian data, tahapan preprocessing, fungsi loss, metrik evaluasi, dan lingkungan komputasi yang konsisten sehingga perbedaan performa dapat dianalisis secara lebih objektif. Ketiga, penelitian ini melengkapi evaluasi tersebut dengan analisis sensitivitas learning rate dan batch size untuk mengkaji pengaruh hyperparameter terhadap kedua arsitektur. Berdasarkan kontribusi tersebut, penelitian ini bertujuan membandingkan kinerja kedua model berdasarkan Dice, Intersection over Union (IoU), precision, recall, Root Mean Squared Error (RMSE), serta kebutuhan komputasinya untuk menentukan model yang lebih sesuai dalam mendukung segmentasi dan preservasi digital naskah kuno Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian dimulai dengan pengumpulan dataset citra dokumen kuno dari beberapa sumber yang relevan dengan konteks dokumen historis Indonesia. Citra yang dikumpulkan kemudian diperiksa berdasarkan ketersediaan label ground truth biner. Untuk dataset yang sudah menyediakan mask ground truth, label tersebut digunakan secara langsung. Sementara itu, untuk dataset yang tidak memiliki ground truth, mask biner dibuat melalui proses semi-otomatis yang melibatkan konversi grayscale, reduksi noise, peningkatan kontras, binarisasi adaptif, post-processing, dan koreksi manual. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



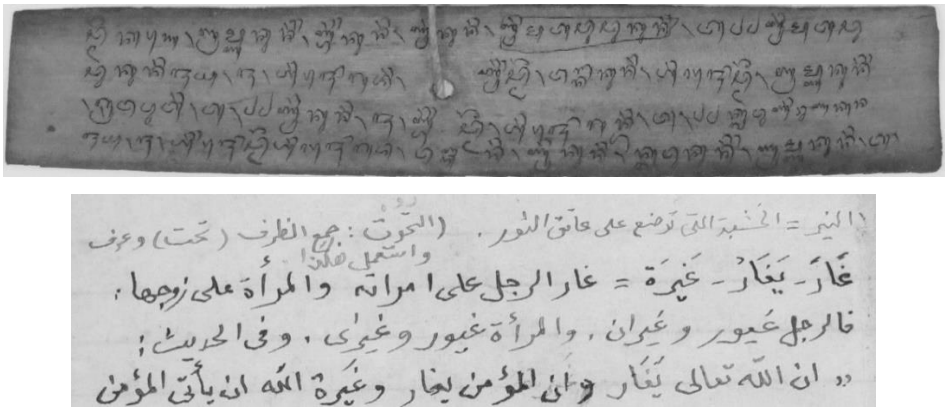
Gambar 1. Alur penelitian

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

Setelah mask ground truth disiapkan, seluruh citra dokumen beserta mask pasangannya diproses ke dalam format masukan yang konsisten. Citra dinormalisasi, diberi padding, dan dipotong menjadi patch non-overlapping berukuran 256 × 256 piksel. Patch tersebut kemudian digunakan sebagai data masukan untuk kedua model segmentasi. Keluaran yang dihasilkan oleh model dibandingkan dengan mask ground truth menggunakan beberapa metrik evaluasi, yaitu Intersection over Union (IoU), Dice Coefficient, Precision, Recall, dan Root Mean Squared Error (RMSE). Selain performa segmentasi, efisiensi komputasi juga dievaluasi dengan mempertimbangkan jumlah parameter model, waktu pelatihan, penggunaan memori, waktu inferensi, dan penggunaan memori inferensi dari masing-masing konfigurasi terbaik tiap model.

2.2 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari tiga sumber utama, yaitu AMADI LONTAR, subset Sunda dari ICFHR 2018, dan Wikimedia Commons. Sumber-sumber tersebut dipilih karena memuat citra naskah kuno atau historis yang relevan dengan tujuan penelitian ini, sebagaimana ditunjukkan pada Gambar 2. Dataset AMADI LONTAR berisi citra manuskrip lontar Bali dan menyediakan mask ground truth biner yang memisahkan area teks dari latar belakang [24]. Dataset ini dianggap sesuai untuk segmentasi dokumen berbasis deep learning karena label yang tersedia dapat langsung digunakan untuk pelatihan dan evaluasi secara supervised.



Gambar 2. Contoh dataset yang digunakan dalam penelitian

Dataset kedua adalah subset Sunda dari ICFHR 2018 [25], [26]. Dataset ini berisi citra manuskrip lontar Sunda kuno dari Kabuyutan Ciburuy, Garut. Serupa dengan AMADI LONTAR, dataset ini juga menyediakan mask ground truth biner, sehingga sesuai untuk evaluasi model segmentasi. Selain kedua dataset tersebut, penelitian ini menggunakan citra dokumen kuno tambahan yang diperoleh dari Wikimedia Commons. Data tambahan tersebut dipilih karena file tersedia di bawah lisensi CC0 1.0 Universal Public Domain Dedication, sehingga dapat digunakan secara legal untuk keperluan penelitian. Dua sumber naskah dari Wikimedia Commons yang digunakan, yaitu WM ID 003 0003 – Kakawin Sutasoma dan WM ID 010 0011 – Da'wa. Dari Kakawin Sutasoma, digunakan 10 citra, sedangkan dari Da'wa digunakan 86 citra. Citra tambahan tersebut disertakan untuk meningkatkan keragaman karakteristik dokumen dalam dataset, terutama dari segi gaya aksara, media penulisan, tekstur latar belakang, dan pola degradasi.

Tabel 1. Distribusi dataset

Dataset	Data Split	Total
AMADI + kakawin	86 training, 12 validation, 12 testing	110 images
Sundanese	43 training, 9 validation, 9 testing	61 images
Da'wa	60 training, 13 validation, 13 testing	86 images

Pembagian data seperti yang ditunjukkan pada Tabel 1 dilakukan pada tingkat citra sebelum tahap cropping. Hal ini dilakukan untuk mencegah patch yang berasal dari citra yang sama muncul pada subset yang berbeda, karena dapat menyebabkan kebocoran data antara data training, validation, dan testing. AMADI dan Kakawin digabungkan ke dalam satu domain karena kedua dataset memiliki karakteristik bahan manuskrip yang serupa dan digunakan untuk meningkatkan variasi data. Dataset Sundanese dan Da'wa dibagi menggunakan rasio mendekati 70:15:15 agar subset validation dan testing tidak terlalu kecil.

2.3 Pre-processing

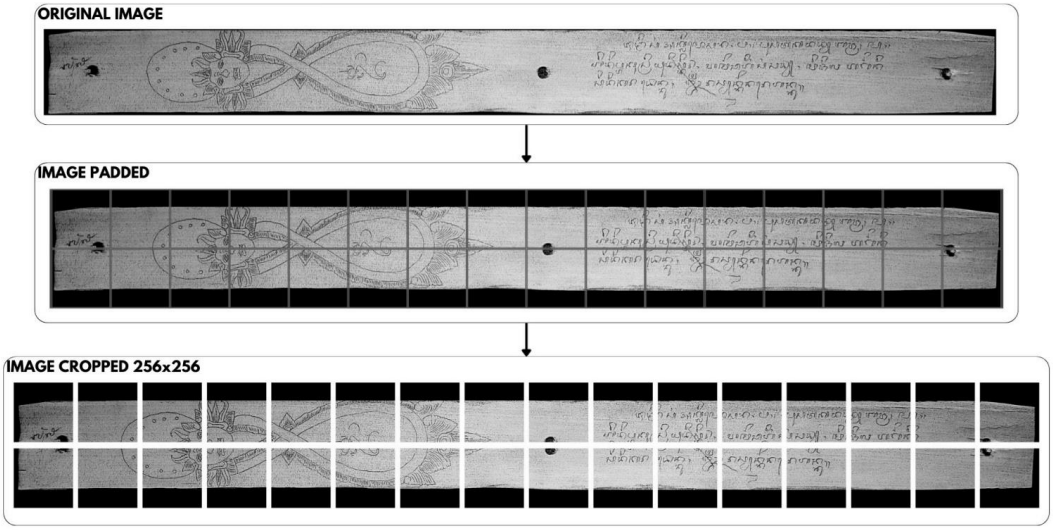
Tahap pre-processing dilakukan untuk menyiapkan citra dokumen dan mask pasangannya sebelum digunakan dalam pelatihan dan pengujian model. Tahap ini terdiri atas persiapan ground truth untuk dataset yang tidak menyediakan label biner, diikuti dengan normalisasi dan penyesuaian ukuran masukan. Untuk dataset AMADI LONTAR dan Sundanese, mask ground truth biner yang tersedia digunakan secara langsung. Sementara itu, dataset Wikimedia Commons tidak menyertakan mask ground truth; oleh karena itu, label dibuat melalui proses semi-otomatis yang terdiri atas konversi grayscale, median filtering, peningkatan kontras berbasis CLAHE, binarisasi Sauvola, post-processing, dan koreksi

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

manual. Konversi grayscale digunakan untuk menyederhanakan representasi citra dengan berfokus pada perbedaan intensitas piksel, sedangkan median filter diterapkan untuk mengurangi noise kecil tanpa merusak tepi goresan karakter secara signifikan. CLAHE kemudian digunakan untuk meningkatkan kontras lokal, terutama pada teks yang samar dan area dokumen dengan pencahayaan yang tidak merata.

Secara umum, tahapan seperti perubahan representasi warna, filtering, peningkatan kontras, dan penyesuaian ukuran citra merupakan bagian dari proses dasar pengolahan citra digital yang bertujuan untuk menyiapkan citra agar lebih sesuai untuk tahap analisis berikutnya [3]. Dalam konteks computer vision, penyesuaian representasi citra dan pemrosesan awal juga dapat membantu model dalam memperoleh masukan yang lebih konsisten sebelum proses ekstraksi fitur dan segmentasi dilakukan [1], [3].

Setelah peningkatan kontras, binarisasi Sauvola diterapkan untuk menghasilkan mask biner awal. Metode ini dipilih karena dapat menentukan threshold adaptif berdasarkan karakteristik lokal citra, sehingga sesuai untuk citra dokumen kuno terdegradasi dengan latar belakang tidak seragam, noda, dan variasi intensitas teks. Hasil binarisasi awal kemudian disempurnakan menggunakan operasi post-processing, termasuk small object removal, small hole removal, dan connected component filtering. Operasi tersebut bertujuan untuk menghapus komponen noise yang tidak relevan, sekaligus mempertahankan area yang merepresentasikan goresan karakter. Selanjutnya, seluruh mask diperiksa dan dikoreksi secara manual menggunakan aplikasi GIMP dengan mengacu pada citra asli. Koreksi dilakukan dengan menghapus area latar belakang yang masih terdeteksi sebagai teks serta menambahkan kembali goresan teks yang tidak terdeteksi oleh proses otomatis.



Gambar 3. Pemotongan citra menjadi 256 × 256 [20].

Setelah mask ground truth tersedia, seluruh citra dan mask disesuaikan agar sesuai dengan format masukan yang dibutuhkan oleh model. Nilai piksel dinormalisasi dari rentang 0–255 menjadi 0–1 untuk mendukung pelatihan yang lebih stabil dan konvergensi yang lebih cepat. Padding kemudian diterapkan agar dimensi citra dapat dibagi secara merata oleh ukuran patch sebesar 256 piksel, sehingga teks yang berada di dekat batas citra tidak terhapus selama proses cropping. Terakhir, setiap citra yang telah diberi padding beserta mask pasangannya dipotong menjadi patch non-overlapping berukuran 256 × 256 piksel. Ukuran masukan yang digunakan dalam penelitian ini adalah 256 × 256 × 3, sedangkan keluarannya adalah mask biner berukuran 256 × 256 × 1 yang merepresentasikan area teks foreground dan background. Proses detail pemotongan citra ditunjukkan pada Gambar 3.

2.4 Implementasi Model U-Net ResNet34

Model pertama yang digunakan dalam penelitian ini adalah U-Net dengan encoder ResNet34. Arsitektur ini menggabungkan kemampuan ekstraksi fitur dari ResNet34 dengan struktur decoder U-Net. Penggunaan ResNet34 sebagai encoder didukung oleh konsep residual learning, yaitu penggunaan shortcut connection untuk membantu aliran informasi dan mempermudah pelatihan jaringan yang lebih dalam [27]. Konfigurasi lengkap tiap layer ditunjukkan pada Tabel 2. Encoder bertanggung jawab untuk mengekstraksi fitur hierarkis dari citra dokumen masukan, sedangkan decoder secara bertahap mengembalikan resolusi spasial untuk menghasilkan mask segmentasi tingkat piksel. Encoder ResNet34 terdiri atas lapisan konvolusi awal, max pooling, dan empat tahap encoder residual dengan masing-masing 3, 4, 6, dan 3 residual block. Setiap residual block menerapkan dua operasi konvolusi 3 × 3 dengan batch normalization dan aktivasi ReLU, diikuti shortcut connection untuk mempertahankan aliran informasi selama pelatihan.

Pada bagian decoder, bilinear upsampling digunakan untuk meningkatkan resolusi spasial feature map. Fitur yang telah di-upsampling digabungkan dengan fitur encoder yang sesuai melalui skip connection. Skip connection ini memungkinkan model menggabungkan informasi semantik tingkat tinggi dari lapisan yang lebih dalam dengan informasi spasial tingkat rendah dari lapisan encoder yang lebih awal [17], [18]. Setelah setiap penggabungan, fitur gabungan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

diproses menggunakan dua blok konvolusi yang terdiri atas convolution, batch normalization, dan aktivasi ReLU. Lapisan akhir menggunakan konvolusi 1 × 1 untuk mengurangi channel fitur menjadi keluaran satu channel. Karena proses pelatihan menggunakan BCEWithLogitsLoss, model menghasilkan raw logits sebagai keluaran, dan fungsi sigmoid hanya diterapkan selama evaluasi untuk mengubah logits menjadi nilai probabilitas.

Tabel 2. Konfigurasi layer U-Net ResNet34

Layer	Output size	Filters	Key Operations
Input	256 × 256 × 3	-	Input image patch
Conv0	128 × 128 × 64	64	Conv2D 7 × 7, stride 2 + Batch Normalization + ReLU
MaxPool	64 × 64 × 64	-	MaxPool 3 × 3, stride 2
Encoder1	64 × 64 × 64	64	3 × Residual Blocks
Encoder2	32 × 32 × 128	128	4 × Residual Blocks
Encoder3	16 × 16 × 256	256	6 × Residual Blocks
Encoder4	8 × 8 × 512	512	3 × Residual Blocks
Decoder1	16 × 16 × 256	256	Bilinear UpSample + [Encoder4, Encoder3] + DoubleConv
Decoder2	32 × 32 × 128	128	Bilinear UpSample + [Decoder1, Encoder2] + DoubleConv
Decoder3	64 × 64 × 64	64	Bilinear UpSample + [Decoder2, Encoder1] + DoubleConv
Decoder4	128 × 128 × 32	32	Bilinear UpSample + [Decoder3, Conv0] + DoubleConv
Decoder5	256 × 256 × 16	16	Bilinear UpSample + DoubleConv
Output	256 × 256 × 1	1	Conv2D 1 × 1 producing raw logits

2.5 Implementasi Modified Deep Semantic Binarization Network

Model kedua yang digunakan dalam penelitian ini adalah Modified Deep Semantic Binarization Network (MDSBN). Model ini mengikuti struktur encoder-decoder dengan skip connection, tetapi dirancang lebih ringkas dibandingkan model U-Net ResNet34. Konfigurasi lengkap ditunjukkan pada Tabel 3. Dalam implementasi ini, encoder dimulai dengan stem layer yang menerapkan konvolusi 3 × 3, batch normalization, dan aktivasi ReLU untuk menghasilkan feature map awal dengan 64 channel tanpa mengurangi resolusi spasial. Encoder kemudian terdiri atas empat tahap downsampling. Setiap encoder block menerapkan konvolusi 3 × 3 yang diikuti oleh batch normalization, aktivasi ReLU, dan max pooling 2 × 2. Proses ini secara bertahap mengurangi resolusi spasial sekaligus meningkatkan jumlah channel fitur dari 64 menjadi 1024.

Decoder MDSBN terdiri atas empat tahap upsampling. Berbeda dengan decoder U-Net ResNet34 yang menggunakan bilinear upsampling, decoder MDSBN menggunakan transposed convolution untuk meningkatkan resolusi spasial. Setiap feature map yang telah di-upsampling digabungkan dengan feature map yang sesuai dari encoder melalui skip connection. Setelah penggabungan, feature map gabungan diproses menggunakan konvolusi 3 × 3, batch normalization, dan aktivasi ReLU. Lapisan akhir menerapkan konvolusi 1 × 1 untuk menghasilkan keluaran segmentasi satu channel. Serupa dengan model U-Net ResNet34, keluaran MDSBN direpresentasikan sebagai raw logits karena BCEWithLogitsLoss digunakan selama pelatihan.

Tabel 3. Konfigurasi layer Modified Deep Semantic Binarization Network

Layer	Output size	Filters	Key Operations
Input	256 × 256 × 3	-	Input image patch
Stem	256 × 256 × 64	64	Conv2D 3 × 3 + Batch Normalization + ReLU
Encoder1 / S1	128 × 128 × 128	128	Conv2D 3 × 3 + Batch Normalization + ReLU + MaxPool 2 × 2
Encoder2 / S2	64 × 64 × 256	256	Conv2D 3 × 3 + Batch Normalization + ReLU + MaxPool 2 × 2
Encoder3 / S3	32 × 32 × 512	512	Conv2D 3 × 3 + Batch Normalization + ReLU + MaxPool 2 × 2
Encoder4 / S4	16 × 16 × 1024	1024	Conv2D 3 × 3 + Batch Normalization + ReLU + MaxPool 2 × 2
Decoder1 / D1	32 × 32 × 512	512	Transposed Conv2D + [S4, S3] + Conv2D 3 × 3 + Batch Normalization + ReLU
Decoder2 / D2	64 × 64 × 256	256	Transposed Conv2D + [D1, S2] + Conv2D 3 × 3 + Batch Normalization + ReLU
Decoder3 / D3	128 × 128 × 128	128	Transposed Conv2D + [D2, S1] + Conv2D 3 × 3 + Batch Normalization + ReLU
Decoder4 / D4	256 × 256 × 64	64	Transposed Conv2D + [D3, Stem] + Conv2D 3 × 3 + Batch Normalization + ReLU
Output	256 × 256 × 1	1	Conv2D 1 × 1 producing raw logits

2.6 Skenario Eksperimen

Skenario eksperimen dirancang sebagai analisis sensitivitas hyperparameter untuk mengevaluasi pengaruh learning rate dan batch size terhadap kinerja setiap model. Untuk memastikan perbandingan yang adil, beberapa komponen eksperimen dibuat sama di seluruh skenario, termasuk optimizer, loss function, preprocessing citra, skema pembagian data, ukuran citra masukan, dan metrik evaluasi. Optimizer yang digunakan dalam penelitian ini adalah Adam, sedangkan loss function

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

yang digunakan adalah BCEWithLogitsLoss. Loss function ini dipilih karena menggabungkan operasi sigmoid dan perhitungan binary cross-entropy dalam formulasi yang stabil secara numerik. Oleh karena itu, selama pelatihan, keluaran model diperlakukan sebagai logits, sedangkan aktivasi sigmoid diterapkan setelahnya ketika mengubah prediksi menjadi probability map untuk evaluasi. Pelatihan model deep learning pada dasarnya dilakukan dengan mengoptimalkan parameter model berdasarkan fungsi objektif tertentu, sehingga pemilihan loss function, optimizer, learning rate, batch size, dan jumlah epoch dapat memengaruhi proses konvergensi model [28]. Jumlah maksimum epoch pelatihan ditetapkan sebesar 100. Untuk menghindari pelatihan yang tidak diperlukan ketika model tidak lagi menunjukkan peningkatan performa, early stopping diterapkan. Mekanisme early stopping menggunakan nilai patience sebesar 15 epoch, minimum delta sebesar 0.0001, dan periode warm-up sebesar 10 epoch. Periode warm-up digunakan agar early stopping tidak terpicu terlalu awal selama tahap awal pelatihan. Detail parameter pelatihan ditulis pada Tabel 4.

Tabel 4. Parameter pelatihan

Parameter	Value	Deskripsi
Input size	256 × 256 × 3	Ukuran patch masukan
Output size	256 × 256 × 1	Keluaran mask biner
Batch size	16	Batch size default
Epochs	100	Jumlah epoch maksimum
Optimizer	Adam	Digunakan untuk kedua model
Loss function	BCEWithLogitsLoss	Menggabungkan sigmoid dan binary cross-entropy dalam formulasi yang stabil secara numerik
Patience	15 epochs	Patience early stopping
Minimum delta	0.0001	Ambang batas peningkatan minimum
Warm-up	10 epochs	Periode warm-up early stopping

Eksperimen dibagi menjadi dua tahap. Tahap pertama dilakukan untuk menentukan learning rate terbaik untuk setiap model. Pada tahap ini, batch size ditetapkan sebesar 16, sedangkan learning rate divariasikan menjadi lima nilai, yaitu 1e-4, 5e-5, 1e-5, 5e-6, dan 1e-6. Karena dua model dievaluasi, tahap ini terdiri atas sepuluh skenario utama. Setiap skenario dilatih satu kali hingga mencapai epoch maksimum atau berhenti lebih awal apabila kriteria early stopping terpenuhi. Tahap kedua dilakukan untuk mengevaluasi pengaruh batch size terhadap performa model. Pada tahap ini, learning rate yang digunakan untuk setiap model adalah learning rate terbaik yang diperoleh dari tahap pertama. Batch size yang diuji adalah 4, 8, dan 32. Batch size 16 tidak diuji kembali pada tahap kedua karena sudah digunakan sebagai batch size default pada tahap pencarian learning rate. Oleh karena itu, tahap kedua terdiri atas enam skenario tambahan, yang terdiri atas dua model dan tiga variasi batch size. Detail tiap skenario eksperimen ditulis pada Tabel 5.

Tabel 5. Skenario eksperimen

Stage	Parameter	Scheme	Deskripsi
1	Learning rate	1e-4, 5e-5, 1e-5, 5e-6, 1e-6	Setiap model dilatih menggunakan lima nilai learning rate
2	Batch size	4, 8, 32	Setiap model dilatih menggunakan learning rate terbaik dari Tahap 1

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pencarian Learning Rate

Eksperimen pertama dilakukan untuk menentukan learning rate yang paling sesuai untuk setiap model segmentasi. Pada tahap ini, batch size ditetapkan sebesar 16 sehingga pengaruh learning rate dapat diamati secara lebih terkontrol. Lima nilai learning rate dievaluasi, yaitu 1e-4, 5e-5, 1e-5, 5e-6, dan 1e-6. Setiap learning rate diuji pada U-Net ResNet34 dan Modified Deep Semantic Binarization Network (MDSBN). Checkpoint terbaik pada setiap skenario dipilih berdasarkan skor Dice validasi tertinggi, sedangkan IoU validasi digunakan sebagai kriteria sekunder ketika skor Dice berdekatan. Data test tidak digunakan untuk menentukan learning rate terbaik, tetapi hanya untuk mengevaluasi checkpoint yang terpilih setelah pelatihan.

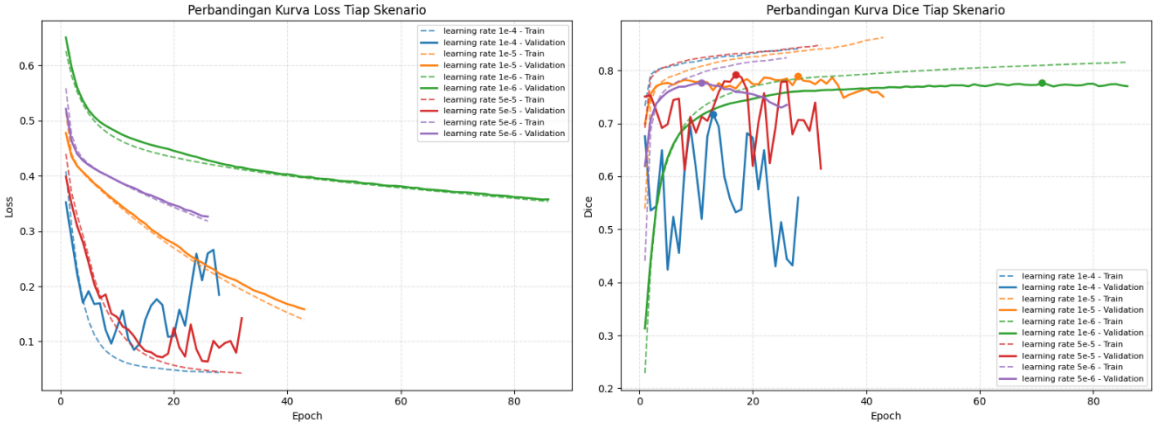
Tabel 6. Hasil pencarian learning rate

Model	Learning rate	Best epoch	Val Dice	Val IoU	Test Dice	Test IoU
U-Net ResNet34	1e-4	13	0.71807	0.56015	0.71582	0.55741
U-Net ResNet34	5e-5	17	0.79233	0.65609	0.79338	0.65752
U-Net ResNet34	1e-5	28	0.78864	0.65104	0.80185	0.66924
U-Net ResNet34	5e-6	11	0.77633	0.63442	0.79049	0.65356
U-Net ResNet34	1e-6	71	0.77635	0.63446	0.78872	0.65114
MDSBN	1e-4	7	0.58417	0.41260	0.60309	0.43173
MDSBN	5e-5	16	0.63357	0.46366	0.64209	0.47285

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

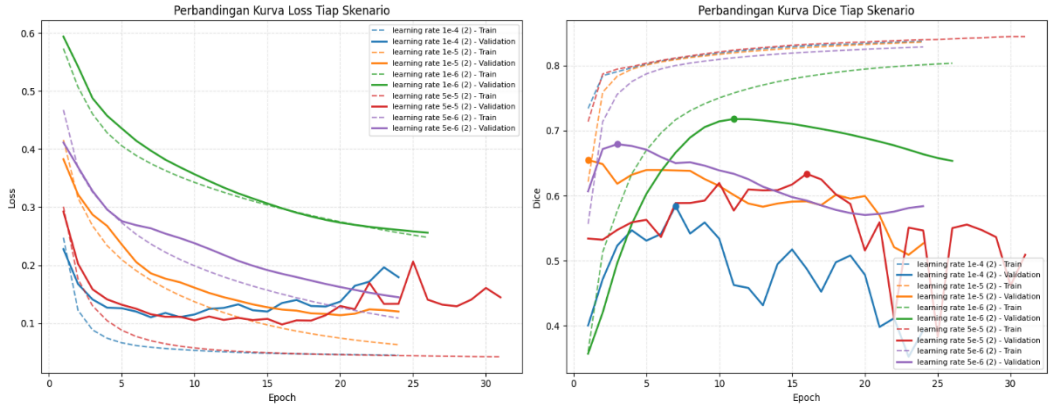
Model	Learning rate	Best epoch	Val Dice	Val IoU	Test Dice	Test IoU
MDSBN	1e-5	1	0.65493	0.48691	0.68106	0.51637
MDSBN	5e-6	3	0.67910	0.51412	0.70872	0.54886
MDSBN	1e-6	11	0.71785	0.55988	0.73962	0.58682

Seperti ditunjukkan pada Tabel 6, U-Net ResNet34 mencapai performa validasi terbaik menggunakan learning rate 5e-5, dengan Dice validasi sebesar 0,79233 dan IoU validasi sebesar 0,65609. Konfigurasi ini juga menghasilkan Dice test sebesar 0,79338 dan IoU test sebesar 0,65752. Meskipun learning rate 1e-5 menghasilkan Dice dan IoU test yang sedikit lebih tinggi, konfigurasi tersebut tidak dipilih karena pemilihan model didasarkan pada performa validasi. Kurva pada Gambar 4 menunjukkan bahwa pada learning rate 5e-5, training loss cenderung menurun dan training Dice meningkat selama pelatihan, sedangkan validation loss menurun pada tahap awal sebelum mengalami fluktuasi. Validation Dice mencapai nilai tertinggi pada epoch ke-17, tetapi setelah epoch tersebut tidak lagi menunjukkan peningkatan yang konsisten meskipun performa training masih meningkat. Pola ini menunjukkan bahwa model telah mencapai kondisi relatif stabil di sekitar checkpoint terbaik dan mulai menunjukkan kecenderungan overfitting pada epoch berikutnya, sehingga checkpoint epoch ke-17 dipilih untuk evaluasi akhir.



Gambar 4. Kurva skenario learning rate U-Net Resnet34

Untuk MDSBN performa validasi terbaik diperoleh menggunakan learning rate 1e-6, dengan Dice validasi sebesar 0,71785 dan IoU validasi sebesar 0,55988. Konfigurasi ini juga mencapai Dice test sebesar 0,73962 dan IoU test sebesar 0,58682. Seperti ditunjukkan pada Gambar 5, penggunaan learning rate 1e-6 menghasilkan penurunan training loss secara bertahap dan peningkatan training Dice yang konsisten. Validation Dice meningkat hingga mencapai nilai tertinggi pada epoch ke-11, tetapi kemudian menurun secara bertahap, sementara training Dice masih terus meningkat. Perbedaan pola antara kurva training dan validation tersebut menunjukkan bahwa pelatihan setelah epoch ke-11 tidak lagi meningkatkan kemampuan generalisasi dan mulai mengarah pada overfitting. Oleh karena itu, checkpoint pada epoch ke-11 dipilih sebagai model terbaik, sedangkan pola penurunan loss yang lebih bertahap menunjukkan bahwa MDSBN memerlukan pembaruan parameter yang lebih kecil untuk mencapai performa optimal pada dataset penelitian ini.



Gambar 5. Kurva skenario learning rate MDSBN

Hasil tersebut juga menunjukkan bahwa learning rate tertinggi tidak selalu menghasilkan performa terbaik. Pada U-Net ResNet34, learning rate 1e-4 menghasilkan Dice dan IoU validasi yang lebih rendah dibandingkan dengan 5e-5 dan 1e-5, yang menunjukkan bahwa langkah pembaruan yang lebih besar mungkin tidak optimal untuk dataset ini. Sementara itu, learning rate yang sangat kecil seperti 1e-6 membutuhkan lebih banyak epoch untuk mencapai checkpoint terbaik, sebagaimana ditunjukkan oleh best epoch sebesar 71. Pada MDSBN, learning rate 1e-6 memberikan hasil terbaik,

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

sedangkan learning rate yang lebih besar menghasilkan skor validasi yang lebih rendah. Oleh karena itu, learning rate terbaik untuk setiap model dipilih berdasarkan Dice validasi dan IoU validasi, kemudian learning rate terpilih tersebut digunakan pada eksperimen berikutnya untuk mengevaluasi pengaruh batch size.

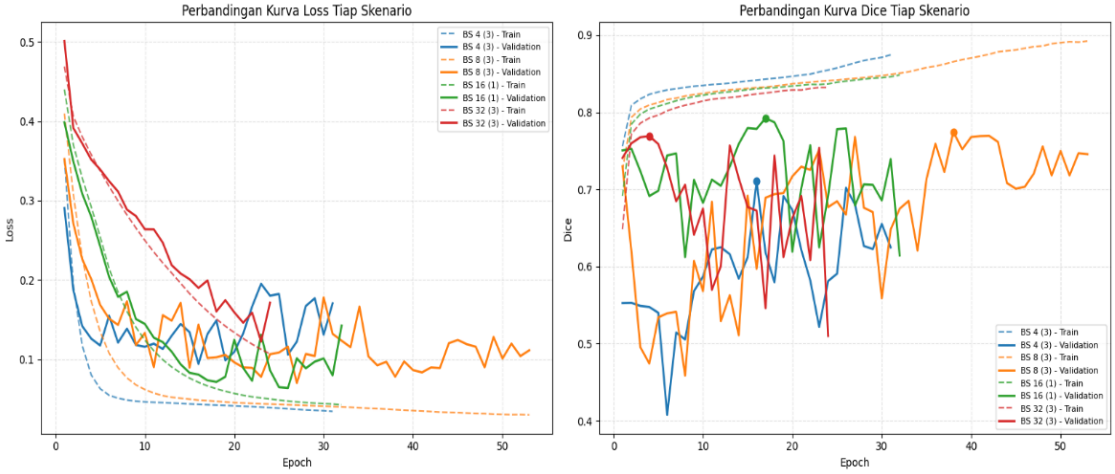
3.2 Hasil Batch Size

Eksperimen kedua dilakukan untuk mengevaluasi pengaruh batch size terhadap performa segmentasi setiap model. Pada tahap ini, learning rate ditetapkan menggunakan learning rate terbaik yang diperoleh dari eksperimen sebelumnya. U-Net ResNet34 dilatih menggunakan learning rate 5e-5, sedangkan MDSBN dilatih menggunakan learning rate 1e-6. Batch size yang diuji adalah 4, 8, dan 32. Batch size 16 tidak diulang pada tahap ini karena telah digunakan sebagai batch size default pada pencarian learning rate. Oleh karena itu, hasil dari batch size 16 dimasukkan sebagai referensi baseline untuk perbandingan.

Tabel 7. Hasil batch size

Model	Learning rate	Batch Size	Best epoch	Val Dice	Val IoU	Test Dice	Test IoU
U-Net ResNet34	5e-5	4	16	0.71030	0.55075	0.70017	0.53867
U-Net ResNet34	5e-5	8	38	0.77470	0.63225	0.77745	0.63592
U-Net ResNet34	5e-5	16	17	0.79233	0.65609	0.79338	0.65752
U-Net ResNet34	5e-5	32	4	0.76917	0.62492	0.78322	0.64369
MDSBN	1e-6	4	3	0.68511	0.52104	0.71621	0.55788
MDSBN	1e-6	8	7	0.69112	0.52803	0.72432	0.56779
MDSBN	1e-6	16	11	0.71785	0.55988	0.73962	0.58682
MDSBN	1e-6	32	23	0.73201	0.57730	0.75338	0.60433

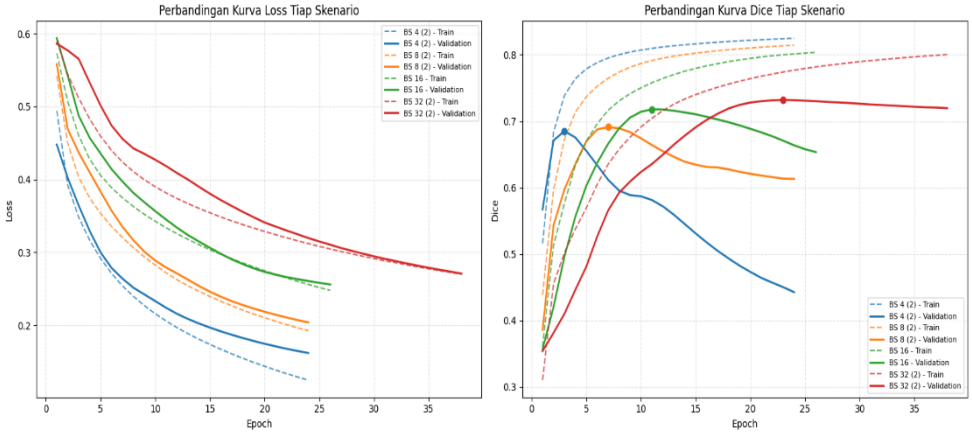
Seperti ditunjukkan pada Tabel 7, batch size memengaruhi performa U-Net ResNet34, dengan hasil validasi terbaik diperoleh menggunakan batch size 16. Konfigurasi tersebut menghasilkan Dice validasi sebesar 0,79233 dan IoU validasi sebesar 0,65609, serta mencapai checkpoint terbaik pada epoch ke-17. Kurva pada Gambar 6 menunjukkan bahwa training loss pada seluruh konfigurasi cenderung menurun dan training Dice meningkat selama pelatihan. Namun, kurva validasi menunjukkan pola yang berbeda. Batch size 4, 8, dan 32 menghasilkan validation Dice yang lebih berfluktuasi dan mencapai nilai terbaik pada epoch yang berbeda, sedangkan batch size 16 mencapai Dice validasi tertinggi pada epoch ke-17. Setelah titik tersebut, peningkatan training Dice tidak lagi diikuti oleh peningkatan validation Dice yang konsisten, sehingga menunjukkan kecenderungan overfitting. Oleh karena itu, batch size 16 dipilih karena memberikan kemampuan generalisasi terbaik pada data validasi, bukan berdasarkan performa data test atau penurunan training loss semata.



Gambar 6. Kurva skenario batch size U-Net Resnet34

Untuk MDSBN, performa validasi meningkat secara bertahap ketika batch size diperbesar. Dice validasi terendah diperoleh menggunakan batch size 4, sedangkan Dice dan IoU validasi tertinggi diperoleh menggunakan batch size 32, masing-masing sebesar 0,73201 dan 0,57730. Seperti ditunjukkan pada Gambar 7, seluruh konfigurasi mengalami penurunan training dan validation loss serta peningkatan Dice pada tahap awal pelatihan. Namun, batch size 4, 8, dan 16 mencapai puncak Dice validasi lebih awal, kemudian mengalami penurunan meskipun training Dice masih meningkat. Sebaliknya, batch size 32 menunjukkan peningkatan validation Dice yang lebih bertahap hingga mencapai checkpoint terbaik pada epoch ke-23, sebelum mengalami sedikit penurunan pada epoch berikutnya. Hasil tersebut menunjukkan bahwa, pada learning rate 1e-6 dan dataset penelitian ini, batch size 32 menghasilkan proses optimasi dan kemampuan generalisasi yang lebih baik bagi MDSBN, kemungkinan karena jumlah sampel yang lebih besar dalam setiap pembaruan membantu menghasilkan estimasi gradien dan statistik batch normalization yang lebih stabil.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236



Gambar 7. Kurva skenario batch size MDSBN

Peningkatan performa MDSBN pada batch size yang lebih besar dapat dikaitkan dengan proses pembaruan parameter yang menjadi lebih stabil karena gradien dihitung dari lebih banyak sampel pada setiap iterasi. Selain itu, MDSBN menggunakan batch normalization pada beberapa tahap encoder dan decoder, sehingga batch size yang lebih besar dapat menghasilkan estimasi statistik batch yang lebih stabil selama pelatihan. Kondisi ini kemungkinan membantu MDSBN yang dilatih dari awal dalam mempelajari pola teks dan latar belakang yang bervariasi. Namun, hasil tersebut tidak menunjukkan bahwa batch size 32 selalu menjadi kebutuhan MDSBN, melainkan merupakan konfigurasi terbaik pada learning rate 1e-6 dan dataset yang digunakan dalam penelitian ini. Berbeda dengan MDSBN, U-Net ResNet34 mencapai performa terbaik pada batch size 16. Batch size 32 tidak lagi meningkatkan performanya, yang menunjukkan bahwa peningkatan ukuran batch tidak selalu memberikan generalisasi yang lebih baik. Batch size yang terlalu besar menghasilkan jumlah pembaruan parameter yang lebih sedikit dalam setiap epoch dan dapat mengurangi variasi gradien yang sebelumnya membantu proses optimasi.

3.3 Perbandingan Akhir antara U-Net ResNet34 dan MDSBN

Setelah eksperimen learning rate dan batch size selesai dilakukan, konfigurasi terbaik dari setiap model dipilih berdasarkan skor Dice validasi tertinggi, dengan IoU validasi digunakan sebagai kriteria sekunder. Data test tidak digunakan selama pemilihan model dan hanya dievaluasi setelah checkpoint terbaik ditentukan. Strategi evaluasi ini diterapkan untuk mencegah kebocoran data dan memastikan bahwa performa test yang dilaporkan merepresentasikan kemampuan generalisasi setiap model.

Evaluasi akhir dilakukan menggunakan beberapa metrik segmentasi tingkat piksel, termasuk Dice Coefficient, Intersection over Union (IoU), Precision, Recall, dan Root Mean Squared Error (RMSE). Dice dan IoU digunakan untuk mengukur overlap antara mask prediksi dan ground truth. Precision digunakan untuk mengevaluasi seberapa akurat model memprediksi piksel foreground, sedangkan Recall digunakan untuk mengukur seberapa banyak piksel foreground sebenarnya yang berhasil dideteksi. RMSE digunakan untuk mengukur kesalahan prediksi tingkat piksel, dengan nilai yang lebih rendah menunjukkan bahwa mask prediksi lebih dekat dengan ground truth.

Tabel 8. Perbandingan akhir konfigurasi model terbaik.

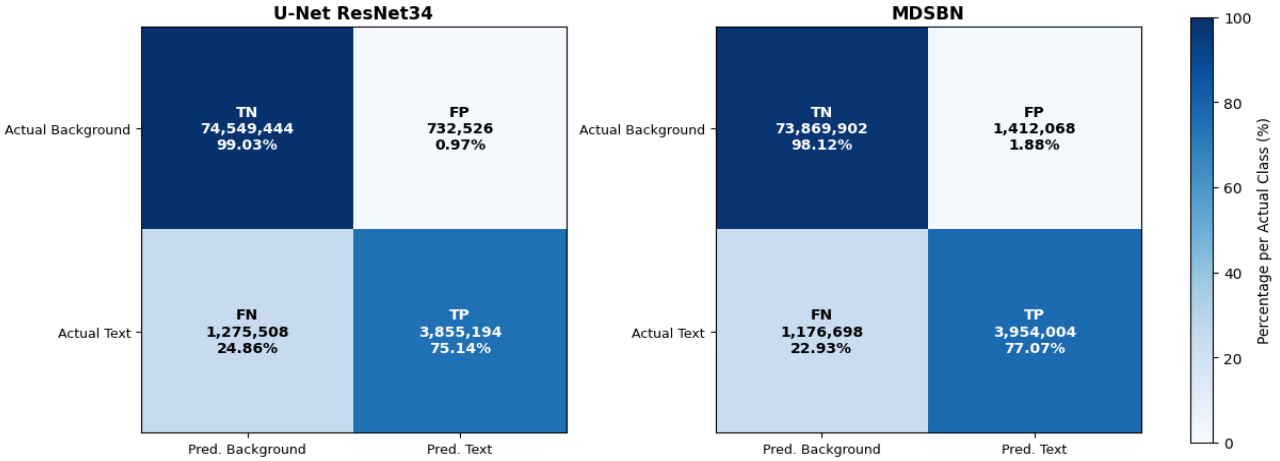
Model	Learning rate	Batch Size	Best epoch	Test Precision	Test Recall	Test RMSE	Test Dice	Test IoU
U-Net ResNet34	5e-5	16	17	0.84033	0.75140	0.15802	0.79338	0.65752
MDSBN	1e-6	32	23	0.73685	0.77066	0.17943	0.75338	0.60433

Hasil pada Tabel 8 menunjukkan bahwa U-Net ResNet34 mencapai performa segmentasi keseluruhan yang lebih baik dibandingkan MDSBN. U-Net ResNet34 memperoleh skor validasi dan test yang lebih tinggi pada Dice dan IoU, yang menunjukkan bahwa mask prediksinya memiliki overlap yang lebih kuat dengan mask ground truth. Karena Dice dan IoU secara langsung mengukur overlap segmentasi, hasil ini menunjukkan bahwa U-Net ResNet34 lebih efektif dalam memisahkan area teks foreground dari background pada citra dokumen kuno Indonesia.

Precision dan Recall memberikan informasi tambahan mengenai jenis kesalahan segmentasi yang dihasilkan oleh setiap model. Nilai Precision yang lebih tinggi menunjukkan bahwa lebih sedikit piksel background yang salah diprediksi sebagai teks, sedangkan nilai Recall yang lebih tinggi menunjukkan bahwa lebih sedikit piksel teks sebenarnya yang terlewat. Oleh karena itu, kedua metrik ini harus diinterpretasikan secara bersamaan. Jika satu model mencapai Recall yang lebih tinggi tetapi Precision yang lebih rendah, model tersebut dapat mendeteksi lebih banyak piksel teks tetapi juga menghasilkan lebih banyak false positive. Sebaliknya, jika suatu model mencapai Precision yang lebih tinggi tetapi Recall yang lebih rendah, model tersebut mungkin lebih konservatif dan melewatkan beberapa goresan teks yang tipis atau terdegradasi. RMSE melengkapi metrik tersebut dengan menunjukkan kesalahan tingkat piksel antara mask prediksi dan ground truth.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

Untuk menganalisis lebih lanjut perilaku klasifikasi setiap model, confusion matrix tingkat piksel dibuat untuk konfigurasi terbaik U-Net ResNet34 dan MDSBN. Confusion matrix terdiri atas True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP). Dalam penelitian ini, TP merepresentasikan piksel teks yang terdeteksi dengan benar, TN merepresentasikan piksel background yang terdeteksi dengan benar, FP merepresentasikan piksel background yang salah diklasifikasikan sebagai teks, dan FN merepresentasikan piksel teks yang salah diklasifikasikan sebagai background.



Gambar 8. Confusion matrix tingkat piksel dari model U-Net ResNet34 dan MDSBN terbaik.

Confusion matrix pada Gambar 8 seharusnya diinterpretasikan bersama dengan Precision dan Recall. Nilai FP yang lebih kecil menunjukkan segmentasi background yang lebih bersih, sedangkan nilai FN yang lebih kecil menunjukkan pelestarian goresan teks yang lebih baik. Karena citra dokumen kuno sering kali memiliki lebih banyak piksel background dibandingkan piksel teks, jumlah mentah piksel TN dapat jauh lebih besar dibandingkan komponen lainnya.

3.4 Profil Komputasi Konfigurasi Terbaik

Profil komputasi dianalisis menggunakan konfigurasi terbaik dari setiap model yang dipilih dari eksperimen sebelumnya. U-Net ResNet34 menggunakan learning rate 5e-5 dan batch size 16, sedangkan MDSBN menggunakan learning rate 1e-6 dan batch size 32. Aspek komputasi yang dievaluasi meliputi jumlah parameter yang dapat dilatih, jumlah epoch yang dijalankan sebelum early stopping, penggunaan peak GPU memory selama pelatihan, waktu inferensi per citra, dan penggunaan peak GPU memory selama inferensi. Dengan demikian, hasil pada bagian ini menunjukkan kebutuhan komputasi aktual ketika masing-masing model dijalankan menggunakan konfigurasi terbaiknya, bukan perbandingan arsitektur pada konfigurasi batch size atau learning rate yang identik.

Tabel 9. Profil komputasi konfigurasi terbaik masing-masing model.

Model	Parameters	Epochs Ran	Best Epoch	Peak Train GPU Memory (MB)	Inference Time (ms/image)	Peak Inference GPU Memory (MB)
U-Net ResNet34	24,436,241	32	17	1,273.62	2.46	1,246.81
MDSBN	15,327,681	38	23	6,750.50	10.12	4,315.17

Seperti ditunjukkan pada Tabel 9, U-Net ResNet34 memiliki jumlah parameter yang lebih besar dibandingkan MDSBN. U-Net ResNet34 terdiri atas 24.436.241 parameter, sedangkan MDSBN terdiri atas 15.327.681 parameter. Jumlah parameter merupakan karakteristik arsitektur dan tidak dipengaruhi oleh learning rate maupun batch size. Meskipun memiliki parameter yang lebih sedikit, MDSBN mencatat penggunaan peak GPU memory yang lebih tinggi pada konfigurasi terbaiknya. Pada proses pelatihan, hasil tersebut perlu diinterpretasikan secara hati-hati karena U-Net ResNet34 menggunakan batch size 16, sedangkan MDSBN menggunakan batch size 32. Selain ukuran batch, kebutuhan memori juga dipengaruhi oleh dimensi feature map, jumlah aktivasi antara yang disimpan, dan operasi yang digunakan pada bagian decoder. Oleh karena itu, peak GPU memory pelatihan pada Tabel 9 menggambarkan kebutuhan komputasi aktual dari konfigurasi terbaik masing-masing model dan tidak sepenuhnya mengisolasi pengaruh arsitektur

Pada pengujian inferensi dengan ukuran masukan, batch size inferensi, perangkat, dan prosedur pengukuran yang sama, U-Net ResNet34 mencatat waktu inferensi sebesar 2,46 ms per citra, sedangkan MDSBN membutuhkan 10,12 ms per citra. Hasil tersebut menunjukkan bahwa U-Net ResNet34 memiliki waktu pemrosesan yang lebih cepat dalam implementasi penelitian ini. Perbedaan batch size yang digunakan selama pelatihan tidak memengaruhi perbandingan waktu inferensi per citra karena pengukuran inferensi dilakukan menggunakan protokol yang sama.

Berdasarkan konfigurasi terbaik masing-masing model, U-Net ResNet34 memberikan keseimbangan yang lebih baik antara performa segmentasi dan kebutuhan komputasi. Model tersebut memperoleh metrik validasi dan test yang

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 1828-1840
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1236

lebih tinggi, serta mencatat waktu inferensi dan peak GPU memory inferensi yang lebih rendah pada protokol pengujian yang sama. Namun, perbandingan peak GPU memory selama pelatihan hanya merepresentasikan kebutuhan aktual dari konfigurasi terbaik masing-masing model karena batch size yang digunakan berbeda. Oleh karena itu, hasil tersebut tidak dimaksudkan untuk menyimpulkan efisiensi intrinsik arsitektur secara penuh, tetapi untuk menunjukkan kebutuhan komputasi ketika setiap model dijalankan menggunakan konfigurasi terbaiknya. Dalam implementasi penelitian ini, biaya inferensi MDSBN yang lebih tinggi dapat menjadi pertimbangan pada penerapan yang membutuhkan pemrosesan cepat atau memiliki keterbatasan memori GPU.

3.5 Hasil Segmentasi Kualitatif

Selain evaluasi kuantitatif, analisis kualitatif dilakukan untuk mengamati secara visual hasil segmentasi yang dihasilkan oleh kedua model. Analisis ini penting karena metrik numerik mungkin tidak sepenuhnya menggambarkan kualitas visual dari mask prediksi, terutama pada citra dokumen kuno dengan goresan tipis, tekstur latar belakang yang tidak merata, noda, atau pola tulisan yang terdegradasi. Oleh karena itu, beberapa sampel test dipilih untuk membandingkan citra asli, mask ground truth, prediksi U-Net ResNet34, dan prediksi MDSBN.

Gambar 9. Hasil visualisasi.

Gambar 9 menyajikan hasil segmentasi representatif dari data test. Kolom pertama berisi citra dokumen asli, kolom kedua menunjukkan mask ground truth, kolom ketiga menunjukkan prediksi yang dihasilkan oleh U-Net ResNet34, dan kolom keempat menunjukkan prediksi yang dihasilkan oleh MDSBN. Susunan ini memungkinkan perbandingan langsung antara kedua model dalam mempertahankan goresan teks dan menekan noise background.

Berdasarkan hasil kuantitatif, U-Net ResNet34 diharapkan menghasilkan mask segmentasi yang lebih konsisten, terutama dalam memisahkan area teks dari background yang kompleks. Skor Dice dan IoU yang lebih kuat menunjukkan bahwa area foreground yang diprediksi memiliki overlap yang lebih dekat dengan mask ground truth. Secara visual, hal ini seharusnya tampak sebagai goresan karakter yang lebih lengkap dan lebih sedikit area teks yang hilang. Encoder residual pada U-Net ResNet34 dapat membantu model menangkap fitur teks-background yang lebih dalam, sedangkan decoder dan skip connection membantu mengembalikan detail spasial yang diperlukan untuk segmentasi tingkat piksel. MDSBN juga menghasilkan hasil segmentasi yang cukup baik, terutama setelah menggunakan batch size yang lebih besar. Namun, nilai Dice dan IoU yang lebih rendah menunjukkan bahwa prediksinya mungkin mengandung lebih banyak kesalahan segmentasi dibandingkan U-Net ResNet34. Kesalahan ini dapat muncul sebagai hilangnya goresan tipis, karakter yang terfragmentasi, atau noise background yang masih tersisa pada beberapa area terdegradasi.

3.6 Analisis Hasil dan Perbandingan Literatur

Hasil penelitian menunjukkan bahwa U-Net ResNet34 memberikan performa segmentasi yang lebih baik dibandingkan MDSBN pada dataset naskah kuno Indonesia. Keunggulan ini diduga berkaitan dengan kemampuan encoder ResNet34 dalam mengekstraksi fitur hierarkis dari berbagai pola degradasi, sementara skip connection membantu mempertahankan detail spasial pada goresan teks. Temuan tersebut juga didukung oleh hasil visual yang menunjukkan bentuk karakter yang lebih konsisten dibandingkan MDSBN.

Temuan ini memperkuat hasil penelitian Yuadi dkk. [20] yang menunjukkan bahwa U-Net ResNet34 mampu mengungguli berbagai metode binarisasi klasik pada manuskrip lontar Bali dengan mencapai Dice sebesar 0,986. Pada penelitian tersebut, keunggulan U-Net ResNet34 dikaitkan dengan kemampuannya dalam memisahkan teks dari latar belakang yang mengalami degradasi. Hasil penelitian ini menunjukkan bahwa performa yang baik dari U-Net ResNet34 tidak hanya terlihat pada manuskrip lontar Bali, tetapi juga tetap dipertahankan pada dataset yang lebih beragam yang mencakup naskah lontar Bali, naskah Sunda, dan citra dokumen kuno dari sumber lain. Meskipun nilai Dice dan IoU yang diperoleh pada penelitian ini lebih rendah dibandingkan hasil yang dilaporkan Yuadi dkk., perbedaan tersebut dapat dipengaruhi oleh tingkat keragaman dataset yang lebih tinggi, termasuk variasi aksara, media penulisan, tekstur latar belakang, dan pola degradasi yang meningkatkan kompleksitas proses segmentasi.

Hasil penelitian juga menunjukkan bahwa MDSBN tetap mampu menghasilkan performa segmentasi yang kompetitif. Temuan ini sejalan dengan penelitian Nair dan Shobha Rani [21] yang melaporkan bahwa MDSBN mampu mengungguli U-Net dan SauvolaNet pada manuskrip lontar Malayalam yang mengalami degradasi seperti pencahayaan tidak merata, noda, dan tinta yang menyebar. Namun, berbeda dengan hasil yang diperoleh pada manuskrip Malayalam, MDSBN pada penelitian ini masih berada di bawah performa U-Net ResNet34. Perbedaan tersebut mengindikasikan bahwa efektivitas suatu arsitektur dapat dipengaruhi oleh karakteristik domain data yang digunakan. Dataset pada

penelitian ini mencakup variasi aksara, media, dan pola degradasi yang berbeda dari dataset yang digunakan oleh Nair dan Shobha Rani, sehingga kemampuan generalisasi MDSBN terhadap naskah kuno Indonesia masih menghadapi tantangan. Selain itu, hasil analisis sensitivitas menunjukkan bahwa MDSBN lebih sensitif terhadap perubahan learning rate dan batch size dibandingkan U-Net ResNet34, yang mengindikasikan bahwa performa optimal model ini lebih bergantung pada pemilihan hyperparameter yang tepat.

Temuan penelitian ini sejalan dengan kajian Sukesh dkk. [23] yang menekankan pentingnya evaluasi menggunakan protokol yang konsisten untuk memperoleh perbandingan performa yang lebih objektif antar model binarisasi. Oleh karena itu, penelitian ini menerapkan pembagian data, tahapan preprocessing, fungsi loss, metrik evaluasi, dan lingkungan komputasi yang sama untuk kedua model guna menghasilkan perbandingan yang lebih objektif. Dengan pendekatan tersebut, perbedaan performa yang diperoleh lebih merefleksikan karakteristik masing-masing arsitektur daripada perbedaan prosedur eksperimen. Selain itu, hasil penelitian ini juga berkaitan dengan temuan Xiong dkk. [5] yang menunjukkan bahwa keberhasilan binarisasi dokumen terdegradasi sangat dipengaruhi oleh kemampuan metode dalam mempertahankan konektivitas goresan teks dan membedakan teks dari gangguan latar belakang. Pada penelitian ini, U-Net ResNet34 menghasilkan bentuk karakter yang lebih mendekati ground truth dan memperoleh nilai Dice serta IoU yang lebih tinggi dibandingkan MDSBN, yang mengindikasikan kemampuan yang lebih baik dalam mempertahankan informasi tekstual pada dokumen yang mengalami degradasi kompleks.

Selain performa segmentasi, profil komputasi menunjukkan bahwa U-Net ResNet34 menghasilkan waktu inferensi dan penggunaan memori inferensi yang lebih rendah dibandingkan MDSBN meskipun memiliki jumlah parameter yang lebih besar. Namun, karena kedua model dievaluasi menggunakan konfigurasi terbaik yang berbeda, hasil tersebut tidak dimaksudkan untuk menyimpulkan efisiensi intrinsik arsitektur secara langsung, melainkan menggambarkan kebutuhan sumber daya aktual pada kondisi performa terbaik masing-masing model.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa U-Net ResNet34 memberikan hasil segmentasi yang lebih baik dibandingkan MDSBN pada dokumen kuno Indonesia. Dengan konfigurasi terbaik berupa learning rate 5e-5 dan batch size 16, U-Net ResNet34 memperoleh Dice test sebesar 0,79338 dan IoU test sebesar 0,65752, sedangkan MDSBN dengan learning rate 1e-6 dan batch size 32 memperoleh Dice test sebesar 0,75338 dan IoU test sebesar 0,60433. Selain menghasilkan perbandingan performa yang objektif melalui protokol evaluasi yang konsisten, penelitian ini juga menunjukkan bahwa sensitivitas terhadap learning rate dan batch size berbeda pada masing-masing arsitektur. Temuan tersebut memberikan dasar empiris dalam pemilihan model dan konfigurasi pelatihan untuk segmentasi naskah kuno Indonesia. U-Net ResNet34 juga menghasilkan waktu inferensi yang lebih rendah pada implementasi penelitian ini, meskipun profil komputasi kedua model diukur menggunakan konfigurasi terbaik yang berbeda sehingga tidak sepenuhnya merepresentasikan perbandingan arsitektur dalam kondisi identik. Secara praktis, kemampuan U-Net ResNet34 dalam menghasilkan mask dengan overlap yang lebih baik dan menekan kesalahan segmentasi dapat mendukung peningkatan keterbacaan dokumen, penyediaan masukan yang lebih bersih bagi proses OCR atau transliterasi, serta pengembangan sistem pengarsipan dan preservasi digital naskah kuno Indonesia. MDSBN tetap menunjukkan kemampuan segmentasi yang kompetitif, tetapi lebih sensitif terhadap pemilihan learning rate dan batch size. Keterbatasan penelitian ini terletak pada jumlah dan keragaman dataset yang masih terbatas serta perbedaan karakteristik antarjenis manuskrip. Penelitian selanjutnya perlu mengevaluasi generalisasi model melalui pengujian lintas dataset, menerapkan transfer learning dan domain adaptation untuk menangani perbedaan aksara, media, dan pola degradasi, serta mengembangkan augmentasi yang meniru noda, tinta memudar, tekstur tidak merata, dan kerusakan fisik dokumen secara lebih realistis.

REFERENCES

[1] Z. Ziran, M. Mecella, and S. Marinai, “AI-Driven Enhancement Of Historical Documents,” *Int. J. Digit. Libr.*, vol. 26, p. 22, 2025, doi: 10.1007/s00799-025-00430-y.
[2] S. Münster *et al.*, “Artificial Intelligence For Digital Heritage Innovation: Setting Up An R\&D Agenda For Europe,” *Heritage*, vol. 7, no. 2, pp. 794–816, 2024, doi: 10.3390/heritage7020038.
[3] Y. Zhou, S. Zuo, Z. Yang, J. He, J. Shi, and R. Zhang, “A Review Of Document Image Enhancement Based On Document Degradation Problem,” *Appl. Sci.*, vol. 13, no. 13, p. 7855, 2023, doi: 10.3390/app13137855.
[4] S. Khamekhem Jemni, M. A. Souibgui, Y. Kessentini, and A. Fornés, “Enhance To Read Better: A Multi-Task Adversarial Network For Handwritten Document Image Enhancement,” *Pattern Recognit.*, vol. 123, p. 108370, 2022, doi: 10.1016/j.patcog.2021.108370.
[5] W. Xiong, L. Zhou, L. Yue, L. Li, and S. Wang, “An Enhanced Binarization Framework for Degraded Historical Document Images,” *EURASIP Journal on Image and Video Processing*, vol. 2021, no. 1, art. no. 13, 2021, doi: 10.1186/s13640-021-00556-4.
[6] Z. Yang, S. Zuo, Y. Zhou, J. He, and J. Shi, “A Review Of Document Binarization: Main Techniques, New Challenges, And Trends,” *Electronics*, vol. 13, no. 7, p. 1394, 2024, doi: 10.3390/electronics13071394.
[7] B. Bataineh *et al.*, “A Comprehensive Review On Document Image Binarization,” *J. Imaging*, vol. 11, no. 5, p. 133, 2025, doi: 10.3390/jimaging11050133.
[8] R. D. Lins, R. Bernardino, R. da Silva Barboza, and R. C. De Oliveira, “Using Paper Texture For Choosing A Suitable Algorithm For Scanned Document Image Binarization,” *J. Imaging*, vol. 8, no. 10, p. 272, 2022, doi: 10.3390/jimaging8100272.
[9] Q.-V. Dang and G.-S. Lee, “Document Image Binarization With Stroke Boundary Feature Guided Network,” *IEEE Access*, vol.

9, pp. 36924–36936, 2021, doi: 10.1109/ACCESS.2021.3062904.
[10] D. Li, Y. Wu, and Y. Zhou, "SauvolaNet: Learning Adaptive Sauvola Network for Degraded Document Binarization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12824 LNCS, no. 189, pp. 538–553, 2021, doi: 10.1007/978-3-030-86337-1_36.
[11] C. Tensmeyer and T. Martinez, "Document Image Binarization with Fully Convolutional Neural Networks," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 1, pp. 99–104, 2017, doi: 10.1109/ICDAR.2017.25.
[12] S. Suh, J. Kim, P. Lukowicz, and Y. O. Lee, "Two-Stage Generative Adversarial Networks For Binarization Of Color Document Images," *Pattern Recognit.*, vol. 130, p. 108810, 2022, doi: 10.1016/j.patcog.2022.108810.
[13] Y. Guo, C. Ji, X. Zheng, Q. Wang, and X. Luo, "Multi-Scale Multi-Attention Network For Moiré Document Image Binarization," *Signal Process. Image Commun.*, vol. 90, p. 116046, 2021, doi: 10.1016/j.image.2020.116046.
[14] L. Zhang, K. Wang, and Y. Wan, "An Efficient Transformer--CNN Network For Document Image Binarization," *Electronics*, vol. 13, no. 12, p. 2243, 2024, doi: 10.3390/electronics13122243.
[15] Z. Du and C. He, "Nonlinear Diffusion System For Simultaneous Restoration And Binarization Of Degraded Document Images," *Comput. \& Math. with Appl.*, vol. 153, pp. 237–248, 2024, doi: 10.1016/j.camwa.2023.11.033.
[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Lecture Notes in Computer Science, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28..
[17] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020..
[18] N. Detsikas, N. Mitianoudis, and N. Papamarkos, "A Dilated MultiRes Visual Attention U-Net For Historical Document Image Binarization," *Signal Process. Image Commun.*, vol. 122, p. 117102, 2024, doi: 10.1016/j.image.2024.117102.
[19] S. Kang, B. K. Iwana, and S. Uchida, "Complex Image Processing With Less Data—Document Image Binarization By Integrating Multiple Pre-Trained U-Net Modules," *Pattern Recognit.*, vol. 109, no. 109, p. 107577, 2021, doi: 10.1016/j.patcog.2020.107577.
[20] I. Yuadi et al., "A Benchmark Study of Classical and U-Net ResNet34 Methods for Binarization of Balinese Palm Leaf Manuscripts," *Heritage*, vol. 8, no. 8, art. no. 337, 2025, doi: 10.3390/heritage8080337.
[21] B. Nair B. J. and N. Shobha Rani, "A Modified Deep Semantic Binarization Network for Degradation Removal in Palm Leaf Manuscripts," *Multimedia Tools and Applications*, vol. 83, no. 23, pp. 62937–62969, 2024, doi: 10.1007/s11042-023-18020-y.
[22] F. J. Castellanos, A.-J. Gallego, and J. Calvo-Zaragoza, "Unsupervised Neural Domain Adaptation For Document Image Binarization," *Pattern Recognit.*, vol. 119, p. 108099, 2021, doi: 10.1016/j.patcog.2021.108099.
[23] R. Sukesh, M. Seuret, A. Nicolaou, M. Mayr, and V. Christlein, "A Fair Evaluation of Various Deep Learning-Based Document Image Binarization Approaches," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13237 LNCS, pp. 771–785, 2022, doi: 10.1007/978-3-031-06555-2_52.
[24] M. W. A. Kesiman, J.-C. Burie, J.-M. Ogier, G. N. M. A. Wibawantara, and I. M. G. Sunarya, "AMADI_LontarSet: The First Handwritten Balinese Palm Leaf Manuscripts Dataset," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 168–172, 2016, doi: 10.1109/ICFHR.2016.39..
[25] M. W. A. Kesiman, D. Valy, J.-C. Burie, E. Paulus, M. Suryani, S. Hadi, M. Verleysen, S. Chhun, and J.-M. Ogier, "ICFHR 2018 Competition on Document Image Analysis Tasks for Southeast Asian Palm Leaf Manuscripts," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 483–488, 2018, doi: 10.1109/ICFHR-2018.2018.00090.
[26] M. Suryani, E. Paulus, S. Hadi, U. A. Darsa, and J.-C. Burie, "The Handwritten Sundanese Palm Leaf Manuscript Dataset from 15th Century," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, pp. 796–800, 2017, doi: 10.1109/ICDAR.2017.135.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
[28] B. Bischl *et al.*, "Hyperparameter Optimization: Foundations, Algorithms, Best Practices, And Open Challenges," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 13, no. 2, p. e1484, 2023, doi: 10.1002/widm.1484.