

CHAPTER 1

INTRODUCTION

This chapter presents the fundamental framework of the research concerning the segmentation of Indonesian ancient document images using deep learning techniques. The study focuses on a comparative evaluation of two semantic segmentation architectures, namely U-Net-ResNet34 and the Modified Deep Semantic Binarization Network (MDSBN). The chapter outlines the research background, research questions, objectives, expected contributions, and the scope and limitations of the study. Collectively, these sections establish the context of the research while clarifying its motivation, direction, and significance.

1.1. Background

Historical documents function not only as records of governmental and societal activities but also as valuable multidisciplinary sources that preserve historical knowledge and support the understanding of political, social, economic, and cultural developments [1]. In Indonesia, ancient manuscripts include palm-leaf manuscripts (lontar), paper manuscripts, books, and colonial archives that contain historical, legal, religious, and literary knowledge. The number of these manuscript collections exceeds 11,275 registered ancient manuscripts, while many others remain dispersed among communities and traditional institutions [2]. The risk of document deterioration is very high due to age, humidity, fungi, insects, and limited storage facilities. UNESCO has even included several Indonesian manuscripts, such as Negarakertagama and La Galigo, in the Memory of the World register because of their universal value and vulnerability to loss [3][4]. Without immediate preservation efforts, an important part of Indonesian civilization may be permanently lost.

Manuscript preservation can essentially be carried out through the physical protection of manuscripts and the preservation of the textual content contained within them [5]. In the context of preserving textual content, preservation is conducted through a digitization process. This process includes pre-digitization, digitization, and post-digitization stages, encompassing manuscript selection, high-

resolution scanning, and the storage and management of digital outputs so that they can be accessed without affecting the physical condition of the manuscripts [6]. However, digitization often faces technical challenges. The resulting quality is frequently inconsistent in terms of both accuracy and readability due to the completeness of the material, the physical condition of the document, and technical factors following the digitization process. As explained in [7], quality is an important consideration in digital preservation because it determines the accuracy and consistency of document representation in digital form. These challenges make the textual content difficult to read, both for humans and machines. Research in [8] also emphasizes that digitization without further processing only preserves the digital physical form and does not guarantee the readability of the textual content. Therefore, image processing methods are needed to improve the quality of digital documents so that they can be utilized more effectively.

Among the image processing techniques applied to digital manuscripts, binary segmentation, or binarization, plays a particularly important role. This technique converts grayscale images into black-and-white images with the objective of separating text (foreground) from the background. As emphasized in [9], binarization represents a crucial stage in document image analysis because its performance directly influences the accuracy of subsequent recognition tasks. Various classical methods have long been used, such as Niblack, Sauvola, and Otsu. Although these conventional methods have been widely adopted for document segmentation, their performance decreases considerably when processing manuscripts containing severe degradations such as ink stains, bleed-through, water damage, and highly textured paper surfaces [10]. This limitation has driven the need for more adaptive approaches capable of capturing complex visual pattern variations. Deep learning has become a powerful alternative for addressing this challenge by leveraging neural network architectures capable of learning complex visual patterns from data. One model that can be utilized is U-Net, which was introduced in [11] for medical segmentation and later adapted to image segmentation tasks, including document binarization. The U-Net architecture introduced in [11] employs a convolution-based encoder and a symmetric decoder with skip connections, allowing important details to be preserved throughout the

segmentation process. To overcome degradation issues frequently observed in palm-leaf manuscripts, including stains, ink bleed-through, and non-uniform illumination, the study in [12] introduced the Modified Deep Semantic Binarization Network (MDSBN), a specialized architecture designed for document binarization tasks. This model is a modification of U-Net with fewer parameters, block-based training using 256×256 patches, and a simplified encoder-decoder architecture. Evaluation results on the manuscript datasets showed that MDSBN outperformed Sauvola Net and U-Net, achieving accuracy of up to 90.55% and an F-Measure of 90.44. These findings highlight the effectiveness of MDSBN as a lighter yet superior alternative for the binary segmentation of ancient documents.

On the other hand, the study in [10] proposed a U-Net with a ResNet34 for Balinese palm-leaf manuscript binarization. ResNet utilizes residual connections to maintain gradient flow, thereby enabling more stable training of deeper networks [20]. Among the various ResNet variants, ResNet34 offers a balanced between network depth and resource efficiency. ResNet18 is relatively lighter but has more limited feature representation capacity, whereas ResNet50, ResNet101, and deeper variants involve higher computational complexity [20]. The study compared U-Net-ResNet34 with classical methods such as Otsu and Sauvola. The experimental results demonstrated that U-Net-ResNet34 significantly outperformed the classical methods in terms of Dice, IoU, and SSIM metrics, particularly in preserving text readability under textured background conditions.

Based on the above discussion, this study aims to conduct a direct comparison between MDSBN, which is a modification of U-Net, and U-Net-ResNet34 for the binarization of degraded Indonesian ancient documents. This approach was chosen because there remains a gap in understanding the extent to which U-Net-ResNet34, as a pretrained encoder-based model, can provide advantages over a modified architecture such as MDSBN in handling ancient document degradation. In addition, both methods share the same baseline model, namely U-Net. Both models have also been proven effective for the task of binary segmentation of ancient documents. Therefore, this study is expected to provide a practical contribution in the form of recommendations regarding the most suitable binary segmentation model for the digital preservation of Indonesian ancient manuscripts.

1.2. Research Questions

Based on the problems identified in the research background, the following research questions are formulated to guide the scope and focus of this study.

1. How does the performance of U-Net-ResNet34 compare with the MDSBN in the binarization of degraded Indonesian ancient documents/manuscripts?
2. How significant are the differences in computational requirements (number of parameters, inference time, and memory usage) between U-Net-ResNet34 and MDSBN in the context of ancient document segmentation?
3. Are there specific conditions under which MDSBN, with its lighter complexity, can achieve results comparable to or even surpass those of U-Net-ResNet34, making it more practical for digital preservation applications?

1.3. Research Objectives

This study aims to investigate and compare the effectiveness of U-Net-ResNet34 and MDSBN in segmenting degraded Indonesian ancient document images. The comparison focuses not only on segmentation quality but also on the computational demands associated with each architecture. The specific objectives of the study are outlined as follows:

1. To measure and compare the performance of both models using IoU, Dice Coefficient, Precision, Recall, and RMSE.
2. To measure and compare computational efficiency, including the number of parameters, inference time, and memory usage.
3. To provide recommendations regarding the most suitable model for the digital preservation of Indonesian ancient documents based on the trade-off between segmentation accuracy and computational efficiency.

1.4. Research Benefits

The benefits of this study are described as expected contributions that may provide positive impacts, both theoretically and practically, for relevant stakeholders, as follows:

1. To enrich the literature in the field of Indonesian ancient document binarization through a comparison of U-Net-ResNet34 and MDSBN.
2. To assist libraries, archives, and preservation institutions in selecting optimal solutions for the digitization of degraded manuscripts.
3. To provide technical data regarding model parameters and computational requirements, enabling model deployment in environments with varying computational resources.
4. To support efforts in preserving Indonesia's cultural heritage and manuscript literacy so that ancient manuscripts remain accessible to present and future generations through effective digital methods.

1.5. Research Scope and Limitations

The scope and limitations of this study are established to delineate the extent of the research, identify its boundaries, and ensure that the analysis is conducted within a well-defined context. They are as follows:

1. This study only compares two architectures: U-Net-ResNet34 and the Modified Deep Semantic Binarization Network (MDSBN).
2. The dataset is limited to Indonesian ancient documents/manuscripts available in digital form.
3. The input image resolution and block size (256×256) are standardized across both models to ensure a fair comparison.
4. Performance evaluation is focused on binary segmentation.
5. Efficiency analysis is limited to the number of parameters, inference time, and memory usage.