Handwritten Optical Character Recognition (OCR) on Sundanese Number Script using Residual Neural Networks (ResNet)

1st Faishal Zharfan
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
18119002@telecom.stei.itb.ac.id

2nd Niken Dwi Carolina Putri
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
18119006@telecom.stei.itb.ac.id

***Abstract*—Sundanese script is one of scripts that is being used in Indonesia, particularly in Sunda areas such as West Java and Banten province. OCR or Object Character Recognition is a technology which enables computers to recognize certain objects and classify it to a class or category. OCR is often used to recognize characters in scripts. Nowadays, OCR is often trained using deep learning method. In this paper, we present an OCR system which can recognize Sundanese number scripts using deep learning algorithm which is ResNet or known as Residual Networks. We compare the performance of two ResNet architectures namely ResNet-18 and ResNet-34. We found that ResNet-34 achieves better performance, in terms of accuracy, compare to ResNet-18.**

***Index Terms*—Sundanese script, character classification, deep learning, residual networks**

I. Introduction

Sundanese script is one of scripts which is often be used in Indonesia, particularly in Sunda areas such as West Java and Banten province. This script has been used by Sundanese since 14th century, thus this script is strictly related to the existence of Sundanese people, in other words, very crucial. This script often be used on tourism places, street name signs, and various government websites. However, Sundanese people sometimes find it difficult to read the script since this script is never used in formal education anymore. This script consists of five vowels called aksara swara, 25 letters called aksara ngalagena with 13 diacritic called rarangken and 10 numbers. The numbers script is shown in figure 1.

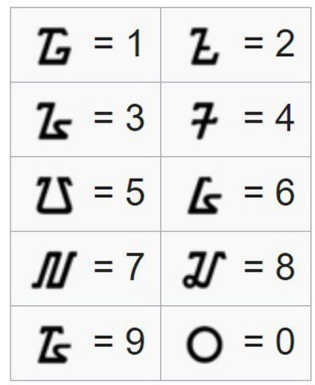


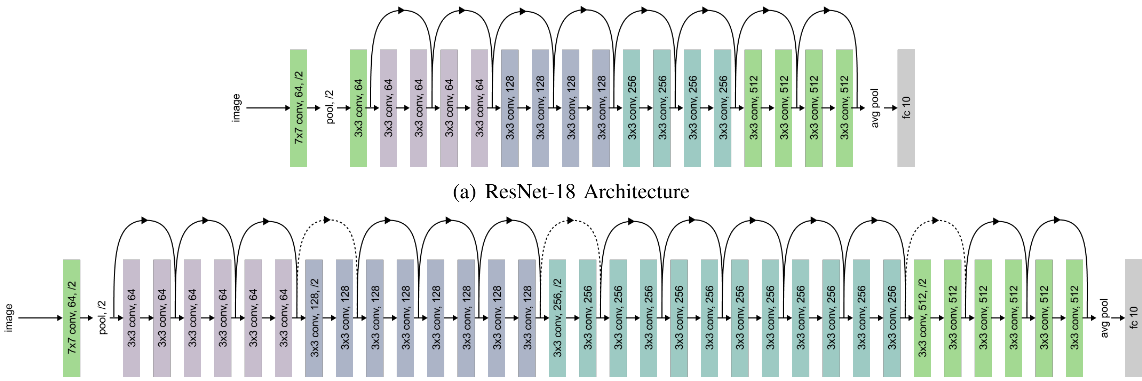
Fig. 1. Sundanese number script

Optical Character Recognition (OCR) is a tool we use to recognize Sundanese characters. Taufiq thesis proposed handwritten Sundanese script character recognition using an SVM classifier with an accuracy of 90.83% [1]. Taufiq said that the system he made still has a problem when recognizing a character which has some similarities with other characters. Erik's thesis introduced an OCR system which is able to recognize handwritten Sundanese script using Artificial Neural Network Radial Basis Function (RBF) which resulted in 95.9% for testing accuracy and 96.5% for training accuracy [2].

Neural networks have been proposed to create more accurate OCRs. Residual network (ResNet), a popular neural network architecture, has shown the ability to recognize objects. The advantages of ResNet are that ResNet has the ability solve the vanishing gradient problem as well as improve the conventional Convolutional Neural Network (CNN) performance. Additionally, ResNet doesn't require significant computational resources compared to other deep learning architectures [3].

According to Suprava [4] who compared Deep CNN and ResNet for Devanagari characters recognition, ResNet needs only 40 epochs for achieving 98.67% of validation accuracy, meanwhile Deep CNN needs more than 100 epochs for achieving such accuracy.

This paper presents an OCR which recognizes handwritten Sundanese number script using a deep neural network called ResNet. Numbers were selected rather than letters because the numbers require low computational overhead in order to perform well. Firsthand experience, indicated a model of 10 classes is easier to implement compared to a model of 355 classes. The second section presents the brief explanation of ResNet. In section three, simulation and training setup are presented. In fourth section, the result and evaluation are presented. Lastly, the last two sections present the discussion and conclusion, respectively.



(a) ResNet-18 Architecture

(b) ResNet-34 Architecture

Fig. 2. ResNet Architecture

II. RESNET SYSTEM MODEL

Residual network (ResNet) is a popular model of neural network architecture. Resnet was first introduced in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun [3]. ResNet has won first place ILSVRC 2015 competition in ImageNet classification and localization and COCO detection and segmentation.

ResNet is used to mitigate the complexity of deep learning networks. To solve a complex problem, deep convolutional neural networks add additional layers to get accurate results. When more layers are added, the problem of degradation occurs. Degradation may results from vanishing or exploding gradients, optimization function or network foundation. ResNet address this problem with deep residual learning framework.

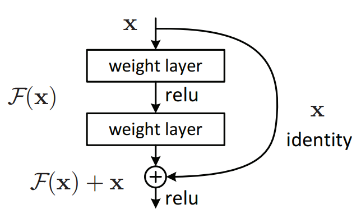


Fig. 3. Residual Block [3]

The deep residual learning framework is made of residual block that is shown in Figure 3. This block has a direct connection which skips some layers, called shortcut connection. The residual block specifically allowed these layers to suit a residual mapping rather than assuming that each few stacked levels would directly fit a desired underlying mapping. We denote H(x) as the desired underlying mapping and let the stacked nonlinear layers fit another mapping of:

$$\mathcal{F}(x) := \mathcal{H}(x) - x \tag{1}$$

which recast into:

$$\mathcal{H}(x) := \mathcal{F}(x) + x \tag{2}$$

With shortcut connection, ResNet alleviates the issue of vanishing gradient by creating a shortcut for gradient to flow through and learn the identity functions.

ResNet has several architectures such as: ResNet-34, ResNet-50, ResNet 101, and ResNet-152. ResNet-34 was the first architecture introduced which was inspired by VGG neural networks (VGG-16 and VGG-19). The main difference for each of the architectures were the amount of layers used. We tried ResNet-18 and ResNet-34 as our ResNet model for handwritten OCR on Sundanese number script.

III. FRAMEWORK

A. ResNet Architecture

TABLE I
PARAMETER COMPARISON

Category	ResNet-18	ResNet-34
Trainable Parameter	11,181,642	21,289,802
Non-trainable Parameter	0	0
Parameter Layers	18	34

The ResNet architecture used in this paper are ResNet-18 and ResNet-34 as shown in Figure 2. The parameter difference of each architecture is shown in Table I.

B. Dataset

The handwritten Sundanese number training and validation dataset was taken from three people. Handwriting is made of:

1) Digital handwriting, the digital handwriting was collected with drawing tablet.
2) Pre-processed image, the pre-processed image was collected by taking a picture of character written on paper. The image is processed by setting image color to gray scale, adjusting black and white contrast, and cropping individual characters.
3) Augmented image, the augmented image was collected by adding background, noise and skew to the digital handwriting.

Collected data was resized to 100×100 then divided into training and validation data. Total data collected for training was **3,897 images** and total data collected for validation was **979 images**. The training batch is shown in Figure 4.

The testing data was taken from five people, collected with the same method as the digital handwriting. Total data collected for testing was **350 images**.

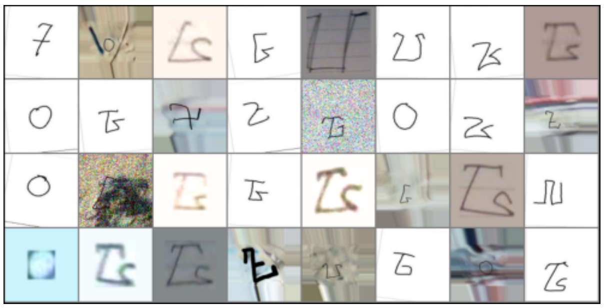


Fig. 4. Training Batch

IV. Result and Evaluation

A. *Training Process*

TABLE II
Complexity Comparison

Category	ResNet-18	ResNet-34
Runtime (s)	66.25	90.99

To begin with, we perform the training for both ResNet-18 and ResNet-34. The training itself is fully implemented using pytorch library helped by cuda to speed up the process. We choose 15 to be the number of epochs since we believe it is sufficient to achieve the desired accuracy. As for the validation, we are using 20% of data.

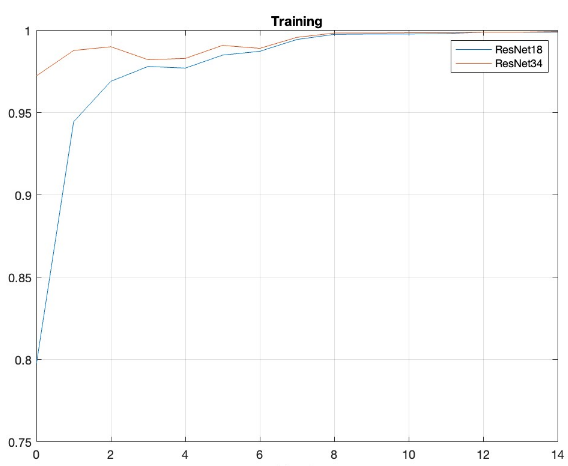


Fig. 5. Training Accuracies

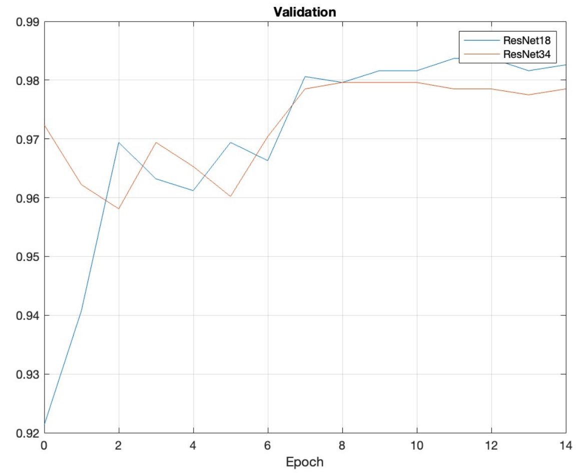


Fig. 6. Validation Accuracies

After performing the training and validation, we plot the accuracy of training and validation per epoch, as shown in figure 5 and 6, respectively. From the figure 5, it can be inferred that ResNet-34 outperforms ResNet-18 in every iteration. At first epoch, ResNet-34 achieves higher accuracy compared to ResNet-34 with 97.25% of accuracy, meanwhile ResNet-18 only achieves 79.75% of accuracy. However, we find that ResNet-34 takes longer to be trained than ResNet-18 as shown in table II.

B. *Evaluation*

TABLE III
ResNet-18 Misclassification Summary

Number Misclassification	Percentage
$1 \rightarrow 0$	20%
$2 \rightarrow 5$	8.5%
$5 \rightarrow 8$	8.5%
$9 \rightarrow 0$	8.5%
$9 \rightarrow 6$	5.7%
$* \rightarrow$ Bg	6%

[a]We only consider the number which gets misclassified more than 1.

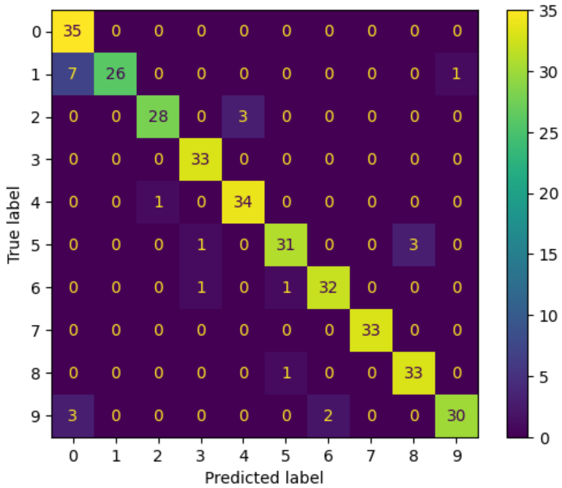
TABLE IV
ResNet-34 Misclassification Summary

Number Misclassification	Percentage
$0 \rightarrow 1$	8.5%
$2 \rightarrow 4$	5.7%
$5 \rightarrow 8$	5.7%
$6 \rightarrow 3$	8.5%
$* \rightarrow$ Bg	1.7%

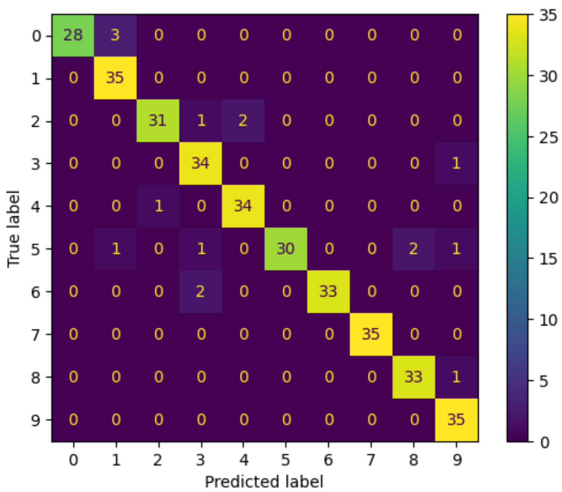
[a]We only consider the number which gets misclassified more than 1.

The evaluation was performed using the test dataset which was collected by using a crowd sourcing method. The 350 data, 35 images for each class, was collected from five people which have never written Sundanese script before.

We therefore believed this method was sufficient to test the robustness of our models.



(a) Confusion matrix of ResNet-18



(b) Confusion matrix of ResNet-34

Fig. 7. Confusion Matrix

Figure 7(a) and 7(b) show the confusion matrix of ResNet-18 and ResNet-34, respectively. As for ResNet-18, we obtained 90% of accuracy and for ResNet-34, we obtained 93.7% of accuracy. We found that some characters were misclassified among the characters which have similar characteristics as shown in table III and IV. Nevertheless, ResNet-34 outperforms ResNet-18 in terms of accuracy. We also find that some characters are misclassified as ”background” or ”nothing”. For ResNet-18, 21 characters were misclassified as background, meanwhile in ResNet-34, 6 characters were classified as background, confirming that ResNet-34 is more robust than ResNet-18.

V. Discussion

As mentioned, the accuracy of ResNet-34 is better than ResNet-18 with 93.7% of accuracy compared to 90%. From table III, there is a misclassification in which 1 gets classified as 0, however this problem has been solved by using ResNet-34. We believe that feature extraction in ResNet-18 is not quite good for this case according to the data shown in the misclassification summary. If we compare the result from table III and IV, we can confirm that ResNet-34 has tackled the problem for number 5 being classified to 8. The number of characters which were classified as background are dramatically reduced in ResNet-34, this is because the granularity of ResNet-34 itself allows feature extraction to perform optimally.

VI. Conclusion

In this paper, we described an OCR system using two ResNet architectures: ResNet-18 and ResNet-34. The accuracy of ResNet-34 outperforms ResNet-18 with 93.7% of accuracy compared to 90%. However, ResNet-34 consumes 137% more resources than ResNet-18. If scientific accuracy is not a concern, ResNet-18 still satisfies. If better feature extraction can be achieved, ResNet-18 could compete with ResNet-34.

References

[1] M. Taufiqurahman, “Handwritten sundanese script character recognition using hog feature and svm classifier,”

[2] E. F. Malik, “Analisis dan implementasi optical character recognition (ocr) menggunakan jaringan syaraf tiruan (jst) untuk pengenalan aksara sunda baku,”

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.

[4] S. Patnaik, S. Kumari, and S. Das Mahapatra, “Comparison of deep cnn and resnet for handwritten devanagari character recognition,” in *2020 IEEE 1st International Conference for Convergence in Engineering (ICCE)*, pp. 235–238, 2020.