Tourist Attraction Classification for Supporting Thoughtful Indonesia Program using Siamese Neural Networks

Tita Karlita, Feri Afrianto, Nana Ramadijanti, Achmad Basuki, Ulima Inas Shabrina, Andro Aprila Adiputra, Muhammad Dzalhaqi
Department of Informatics and Computer Engineering
Electronic Engineering Polytechnic Institute of Surabaya, Surabaya, Indonesia
{tita; nana; basuki}@pens.ac.id, {feriafrianto; uishabrina; androaprila12}@it.student.pens.ac.id, dzalhaqi@ds.student.pens.ac.id

***Abstract*— Tourist attractions both on a local scale in Indonesia and on an international scale are very numerous. Nowadays, more and more information on tourist attractions is represented as images rather than text. Tourists are interested in the specific tourist attraction shown in the picture, do not know the attraction's name, and cannot do a text search to get more information about the attraction in question. Convolutional neural networks (CNNs) perform well on large data sets of images. However, due to the diversity of tourist attractions in Indonesia, not all tourist attractions in Indonesia have a large sample image. So, this paper will discuss adopting one-shot learning with the Siamese network to solve the problem of the availability of a small sample of tourist data. Siamese networks are a type of twin network with two or more identical subnets. The settings and weights are the same for all subnets. The parameters of the Siamese network are modified by operating together in all its subnets. In addition, the Siamese network can learn well even with limited input. This study resulted in an image classification of 102 tourist attractions in Indonesia. With each class, five samples resulted in a validation accuracy of 93%.**

Keywords*— deep learning, siamese neural network; tourist attraction.

I. INTRODUCTION

The COVID-19 pandemic has hit Indonesia's tourism industry and creative economy. In 2020 the number of foreign tourists entering Indonesia experienced a very drastic decline, with the number of tourists only as many as 158,000. Extensive-scale social restrictions and significant immigration closures were reduced in Indonesia. The government's revenue in the tourism sector fell by Rp. 20.7 billion. The Ministry of Tourism and Creative Economy (Kemenparekraf) has developed a strategy to overcome the crisis during the pandemic by changing the brand from Wonderful Indonesia Tourism to Thoughtful Indonesia [1].

Indonesia has many tribes, cultures, races, and religions, and various types of natural beauty can be found. Indonesia has various sectors that can boost the country's foreign exchange. One is the tourism sector, the country's primary foreign exchange source [2]. According to BPS data, in 2020, 1,865 tourist objects are managed by the private sector, local governments manage 556 attractions, the Authority manages 72 attractions, and the central government manages 59 attractions [3]. With the existence of many diverse tourist attractions in Indonesia, but due to the limited knowledge of local and foreign tourists about tourist attractions in Indonesia, it makes tourists confused in recognizing the image of tourist attractions that are not accompanied by additional information or information about the names of these tourist attractions that they meet in magazines or social media.

Several large datasets have resulted in significant advances in object detection and image classification because most datasets are labeled. A typical example is an ImageNet database, which has millions of images that are better for model learning. Convolutional neural networks (CNNs) provide robust performance on large image datasets [4]. With the diversity of tourist attractions in Indonesia, not all tourist attractions in Indonesia have large sample images. So, in this paper, we will discuss adopting one-shot learning with the Siamese neural network to solve the problem of the availability of a small sample of tourist data.

In today's era, people share ideas, photos, videos, and posts with others to maintain their social relationships; and we can find news and information through social networking services [5]. As the number of users connected to networking platforms has increased exponentially, social networking services can be used as the primary data source in various fields. The development of social media services contributes to the increasing amount of information about tourist attractions being represented as images rather than text [6]. As a result, tourists who are interested in a particular tourist spot shown in the image may not know how to perform a text search for more information about the tourist attraction.

One-shot learning is a technique that successfully avoids overfitting by training a model with small data. The idea is developed by people who can learn something from a limited number of examples. The standard machine learning algorithms will be severely overfitted if taught using only a small amount of data. In contrast, our approach is based on a Siamese network that uses a twin convolutional neural network to construct the architecture and share the same parameters. According to empirical evidence, using the Omniglot dataset, one-time image recognition with Siamese convolutional

A. Azizah and E. D. Ariyani (Eds.): iCAST-SS 2022, ASSEHR 719, pp. 645–650, 2022.
https://doi.org/10.2991/978-2-494069-83-1_112

networks achieved a test accuracy of 92.0 percent [4]. Through many pre-processing stages in our system, the tourist attraction samples are converted into a visual representation and fed into a Siamese convolutional neural network. The output sigmoid layer similarity score identifies the determination of the tourist attraction family, which is the fundamental idea behind adopting the Siamese neural network.

This study's main contribution is applying the Siamese neural network for tourist attraction classification using a small data sample. We present an efficient model to overcome the complications of data shortages for unbalanced data sets.

The outline of this paper is structured as follows. Section 2 reviews the material and methods for image classification of Siamese attractions and networks. Section 3 presents the result and discussion. The last section concludes the paper and suggests ideas for future work.

II. The Material and Method

A. Dataset

In this study, the authors collect images from the internet (Pinterest, Shutterstock, Google Search Images, Google Maps Images, 500px, Dreamstime, freepik) and then sort them according to the specifications of the photos to be taken. The photo specifications that will be used are iconic photos, which are photos without obstacles from other objects. In processing image data, images containing unimportant objects will be cropped. Images blocked by trivial objects will be removed and not used in the dataset , which can be seen in Figure 1. The image type used is three-channel jpg (RGB) and saved in the Joint Photographic Experts Group (JPEG) format.

In Figure 2, one class is organized into a single folder that contains five images of various tourist destinations, and in Figure 3, a few samples of pairs of attractions are shown. This dataset's images have a 224x224 resolution. The data for training, validation, and test sets are three subsets that comprise the dataset below.

1) The Training Set

The training set is part of the dataset used in the training process to train the model of a Machine Learning algorithm. In this study, the percentage used in the training set is 60%, with 306 images.

2) The Validation Set

The validation set is part of the training set used in the early stages of testing. In this study, the percentage used in the validation set is 20% with 102 images.

3) The Testing Set

The testing set is part of the dataset used in the testing process to test a model. In this research, the percentage used in the testing set is 20%, with 102 images.

B. Image Pair

The making of the image pair is done after the distribution of the dataset. Making a pair of images is carried out randomly where one of the similar images of a class is paired with a similar one (labeled true or 1) and not similar (labeled false or 0) as in Figure 3, which shows examples of some of the results of the image pair and the label.

After the dataset has been distributed, the image pair is created. As shown in Figure 3, which provides examples of some of the outcomes of the picture pair and the label, making a pair of photos is done at random, pairing one of the similar images of a class with a similar one (labeled true or 1) and not similar (labeled false or 0).

Fig. 1. (A) The image is deleted because an object is blocking (B) The image is kept. After all, there is no object blocking (C) The image must be edited because it contains an unimportant object (D). The result of the image edit is then saved.

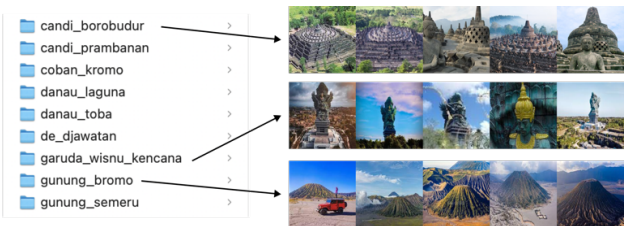

Fig. 2. Some sample dataset images

Fig. 3. Some examples of creating image pairs.

C. Siamese Neural Network

LeCun et al. first proposed the siamese network in the 1990s as a comparative learning challenge of images to

confirm signatures [7]. With the help of two similar sub-networks, this neural network automatically learns expressions as structures for non-linear metric learning. Simply put, two sub-networks can naturally extract a representation of the input pair using information about similarities and dissimilarities [8].

According to the authors, a siamese network is a type of twin framework with two or more identical subnets that provide a complete introduction to its construction and attributes [9], [10]. The settings and weights are the same for all subnets. The parameters of a siamese neural network are modified by operating together across all of its subnets. Moreover, they have shown that the Siamese network can learn well even with limited input [8]–[10].

We will train the model to distinguish between different tourist attractions. For example, tourist attraction A must be distinguished from other tourist attractions. To do this, we will select N random images from class A (for example, for the Tugu Monas class) and pair them with N random images from another class B (for example, for the Lake Toba class). Then, we can repeat this process for all tourist attraction classes. The primary process carried out is to create a Siamese neural network model. The basic principle of the Siamese neural network modeling process is to train the Siamese neural network to produce the best model for good accuracy.

In a Siamese neural network, there are two input layers, each leading to its network, which results in embeddings. The Lambda layer combines them using Euclidean distances, and the combined output is fed to the final network. After getting the distance between the two images, a contrastive loss function will be performed to study the embedding from Euclidean distance with two objectives. Namely, the same tourist spot image produces adjacent embedding in the embedding space, while images of different tourist attractions produce remote embedding in the embedding space. After doing all these processes, a siamese neural network model will be formed to predict or determine whether the tourist attractions are the same (genuine) or different (impostors).

In Siamese neural network training, there is a dataset consisting of train data, data validation, and data testing. Data train is used to train the former model, and data validation is used to validate the model and prevent overfitting. Data testing is used to test the model that has been trained. The model's accuracy can be determined by performing validation using data validation.

Figure 4 illustrates how the system's design for this research's work begins by taking a dataset with five images from each class and associating the pictures of tourist attractions with true or false values. The siamese neural network architecture network input is data from an image pair. After the model has been trained, class predictions are made, and their accuracy is evaluated using performance metrics.

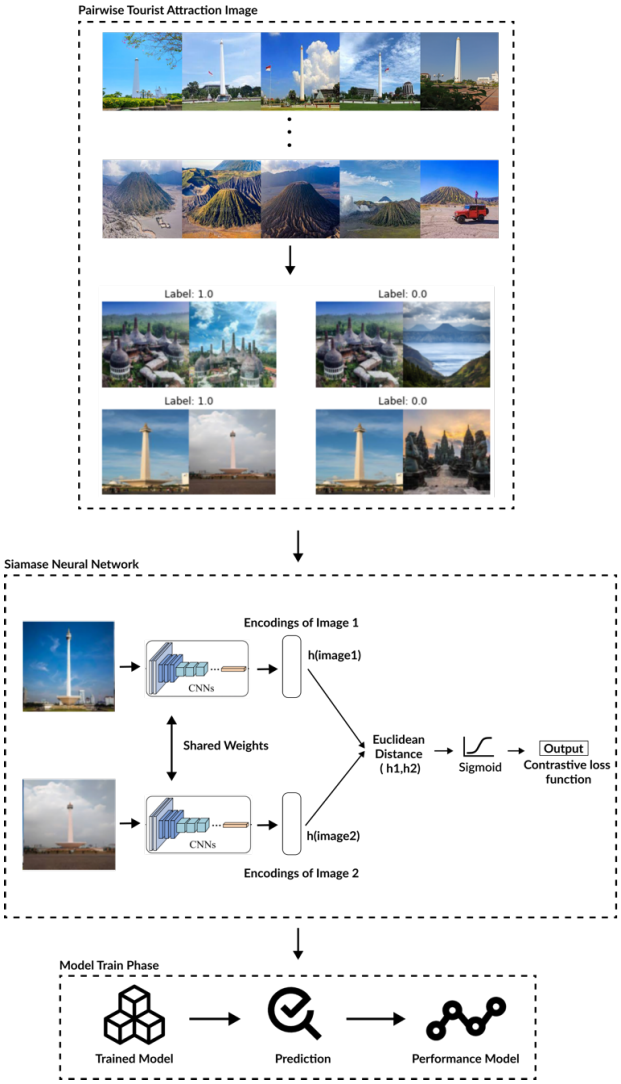

Fig. 4. Siamese Neural Network System Design

III. Results And Discussion

The tourist attractions dataset is tested against several parameters to obtain optimal parameter values at this stage. Parameters that need to be set on the Siamese Neural Network include the number of epochs or many iterations, the image batch size that determines the number of images used in one training process, the number of filters or convolution map outputs, and the size of the convolution kernel. The test results of each scenario can be compared in terms of classification accuracy.

We used a test dataset in this experiment. Every experiment outside the training process uses this dataset. Data for the training, validation, and testing sets are the three subsets that make up the dataset.

A. Comparison of filters size in 1 Convolution Layer

In this test, the epoch value is set to 100 by experimenting with a combination of filter sizes 4, 8, 64, and 128. The optimized result for this model is filter size 4. The smaller the kernel size, the more optimal the results. Experiments carried out with experimental parameters produce Table 1 comparison of filter size results.

TABLE I. COMPARISON OF FILTER SIZE

Filter Size	Validation Loss	Validation Accuracy
4	**0.060**	**91%**
8	0.080	90%
64	0.076	88%
128	0.088	87%

B. Comparison of Kernel Size on 1 Convolution Layer

In this test, the epoch value is set to 100 by experimenting with a combination of kernel sizes 64, 32, 16, and 5. The optimized result for this model is kernel size 64. The larger the kernel size, the more optimal the results. Experiments carried out with experimental parameters produce Table II comparison of kernel size results.

TABLE II. COMPARISON OF KERNEL SIZE

Kernel Size	Validation Loss	Validation Accuracy
64	**0.071**	**91%**
32	0.067	90%
16	0.073	89%
5	0.100	85%

C. Comparison of Kernel Size on 2 Convolution Layer

In this test, the epoch value is set to 100 by experimenting with a combination of filter sizes 8, 16, and 32. The optimized result for this model is a filter size of 16. Experiments carried out with experimental parameters produce Table III comparison of kernel size results.

TABLE III. COMPARISON OF KERNEL SIZE

Kernel Size	Validation Loss	Validation Accuracy
8	0.072	90%
16	**0.054**	**92%**
32	0.067	90%

D. Comparison of Batch Size on 2 Convolution Layer

In this test, the epoch value is set to 100 by experimenting with batch sizes 16, 32, and 64. The optimized result for this model is batch size 16. The smaller the batch size, the more optimal the results. Experiments carried out with experimental parameters produce Table IV comparison of batch size results.

TABLE IV. COMPARISON OF BATCH SIZE

Batch Size	Validation Loss	Validation Accuracy
16	**0.055**	**92%**
32	0.084	88%
64	0.168	89%

E. Comparison of the Number of Epochs

In this test, the experiment uses a combination of epoch sizes of 50, 100, and 200. The optimized result for this model is the epoch size of 100. Experiments carried out with experimental parameters produce Table V comparison of epoch results.

TABLE V. COMPARISON OF EPOCH

Number of Epoch	Validation Loss	Validation Accuracy
50	0.0753	88%
100	**0.0548**	**93%**
200	0.0500	92%

F. Comparison with Deep CNN Methods

In this test, the pre-train model comparison experiment found that the shallow layer VI used had a higher accuracy of 93%.

TABLE VI. COMPARISON WITH DEEP CNN METHODS

Deep CNN Methods	Validation Loss	Validation Accuracy
MobileNet	0.2503	50%
RestNet50	0.1302	87%
Shallow CNN	**0.0548**	**93%**

G. Comparison of the number of datasets

This test is carried out using several different datasets. The dataset consists of 50,100,150 class tourist attractions. Table VII shows that the more loss classes will be smaller with the accuracy level still above 90%.

TABLE VII. COMPARISON OF NUMBER OF DATASETS

Number of class	Validation Loss	Validation Accuracy
50	0.1029	94%
100	0.0768	91%
150	0.0604	93%

The learning process results can be seen in the graphs in Figure 5 and Figure 6.

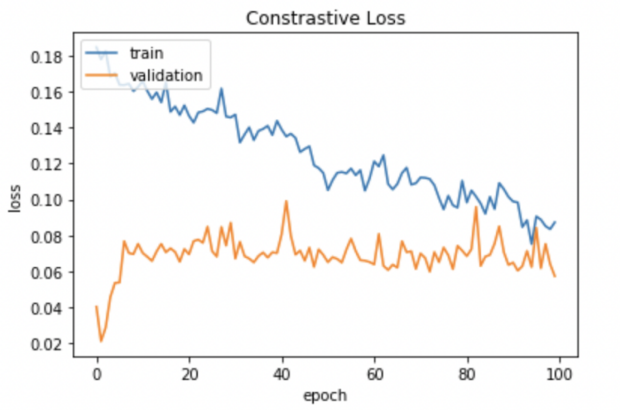

Fig. 5. Loss Chart Results

On the loss value during the training and validation process, the graph explains that the training and testing process is moving down the loss value. The loss value in validation moves down until it reaches a value of 0.0548. The decreasing loss value proves that the model formed has good accuracy because the error value generated in the model is very small.

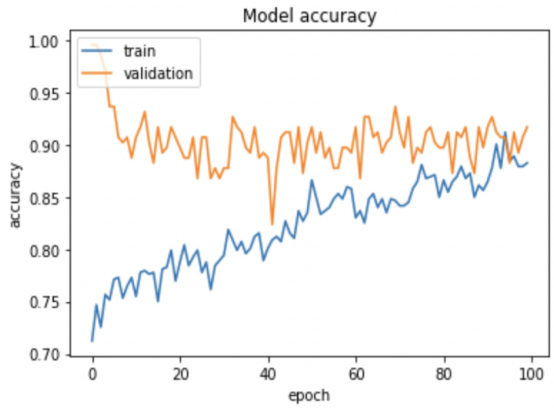

Fig. 6. Accuracy Graph Results

In the training and validation process, the model learned well to get an accuracy of 93% on data validation. This model can be interpreted as learning well to get higher accuracy in data validation, namely data that has never been seen before.

Figure 7 shows the prediction results from testing data where the predicted value indicates the distance between the images if the closer the value approaches 1, and the actual value shows whether the image is 1 class or not if the same class is worth one; otherwise, it will be 0.

Fig. 7. Data Prediction Results

IV. Conclusion

This paper discusses the image classification of tourist attractions using a siamese neural network. Because using a typical convolutional neural network (CNN) does not match the training sample data set that is small in each class. This paper discusses this problem by using a siamese neural network. In this study, we have made a good model with hyperparameter epoch 100, batch size 16, with an accuracy of 93% for testing. Future work will focus on achieving better accuracy with more class data and improving the Siamese network to be applied to attractions with many spots resulting in multiple classes in one attraction.

Acknowledgment

We sincerely thank Knowledge Engineering Laboratory - Politeknik Elektronika Negeri Surabaya for supporting this research by providing high computation devices.

References

[1] Kemenparekraf/Baparekraf RI, “Tren Pariwisata Indonesia di Tengah Pandemi,” 2021. https://kemenparekraf.go.id/ragam-pariwisata/Tren-Pariwisata-Indonesia-di-Tengah-Pandemi (accessed Jun. 20, 2022).

[2] A. A. Rahma, “Potensi Sumber Daya Alam dalam Mengembangkan Sektor Pariwisata Di Indonesia,” *Jurnal Nasional Pariwisata*, vol. 12, no. 1, p. 1, Apr. 2020, doi: 10.22146/jnp.52178.

[3] B. Rahmad and S. Naning Tri, “Statistik Objek Daya Tarik Wisata,” Badan Pusat Statstik, 2020.

[4] S. C. Hsiao, D. Y. Kao, Z. Y. Liu, and R. Tso, “Malware image classification using one-shot learning with siamese networks,” in

Procedia Computer Science, 2019, vol. 159, pp. 1863–1871. doi: 10.1016/j.procs.2019.09.358.

[5] Y. C. Chen, K. M. Yu, T. H. Kao, and H. L. Hsieh, "Deep learning based real-time tourist spots detection and recognition mechanism," *Science Progress*, vol. 104, no. 3_suppl, 2021, doi: 10.1177/00368504211044228.

[6] Y. E. Ozkose, T. A. Yilikoglu, L. Karacan, and A. Erdem, "Finding location of a photograph with deep learning," in 2018 26th *Signal Processing and Communications Applications Conference (SIU),* May 2018, pp. 1–4. doi: 10.1109/SIU.2018.8404530.

[7] J. Bromley et al., "Signature Verification using a 'Siamese' Time Delay Neural Network*," International Journal of Pattern Recognition and Artificial Intelligence,* vol. 7, p. 25, Aug. 1993, doi: 10.1142/S0218001493000339.

[8] J. Gao, Q. Wang, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," in 2017 *IEEE International Conference on Robotics and Automation* (ICRA), 2017, pp. 219–224. doi: 10.1109/ICRA.2017.7989027.

[9] P. Neculoiu, M. Versteegh, and M. Rotaru, Learning Text Similarity with Siamese Recurrent Networks. *In Proceedings of the 1st Workshop on Representation Learning for NLP,* pages 148–157, Berlin, Germany. Association *for Computational Linguistics,* 2016. doi: 10.18653/v1/W16-1617.

[10] G. R. Koch, "Siamese Neural Networks for One-Shot Image Recognition," Proceedings of the 32 nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37, 2015.