Residual Neural Network Model for Detecting Waste Disposing Action in Images

I Made Arsa Suyadnya and Duman Care Khrisne

[1,2]Department of Electrical Engineering, Faculty of Engineering
Udayana University
Bali, Indonesia
arsa.suyadnya@unud.ac.id

Abstract - Waste in general has become a major problem for people around the world. Evidence internationally shows that everyone, or nearly everyone, admits to polluting at some point, with the majority of people littering at least occasionally. This research wants to overcome these problems, by utilizing computer vision and deep learning approaches. This research was conducted to detect the actions carried out by humans in the activities / actions of disposing of waste in an image. This is useful to provide better information for research on better waste disposal behavior than before. We use a Convolutional Neural Network model with a Residual Neural Network architecture to detect the types of activities that objects perform in an image. The result is an artificial neural network model that can label the activities that occur in the input image (scene recognition). This model has been able to carry out the recognition process with an accuracy of 88% with an F1-Score of 0.87.

Index Terms—Action Detection, Residual Neural Network, Scene Recognition, Waste Disposal.

I. INTRODUCTION

Waste in general has become a major problem for people around the world. In the Netherlands, children (ages 12-18) are among the most littering [1]. Evidence both in the UK and internationally shows that everyone, or nearly everyone, admits to polluting at some point, with the majority of people littering at least occasionally [2]. This encourages researchers to solve how to overcome the habit of littering. Results from previous research [3]–[7], researchers have identified that conducting education to change behavior is an effective way to tackle the habit of littering.

Ideally, making effective interventions that target specific behaviors, including littering, requires identifying the determinants of that behavior. However doing this has proven difficult, as it is not yet known what the most important determinants of waste littering behavior are and by what strategies these can best be targeted [1]. This is because there are wide variations between individuals, in the frequency and extent of littering behavior and the types of objects they throw away. Although some groups are more likely to litter than others, there is no evidence that there is a "littering demographic", as no specific group can be identified as responsible for most of the waste [2].

The problem of littering behavior is starting to get a bright spot with the discovery of technology to carry out surveillance. The existence of surveillance cameras [8], [9] is considered to be helpful in monitoring and determining the determinants of behavior and habits of littering. This research is here to help overcome the problems mentioned above, by utilizing computer vision and deep learning approaches. This research was conducted to detect human actions in activities or acts of disposing of waste in digital images; the resulting model is expected to be the basis for recognizing the activity of throwing waste in the video.

II. RELATED WORKS

The research entitled Sequential Deep Learning for Human Action Recognition [10], carried out the recognition of human action using a convolutional neural network that was extended to 3D using the KTH dataset. The result of this research is a model that can study the spatio-temporal features automatically.

In 2014 [11] conducted a research entitled Large-scale Video Classification with Convolutional Neural Networks, which attempted to classify 1 million videos on YouTube into 487 classes. The spatio-temporal network in this research displayed a significant increase in performance compared to the feature-based baseline technique (55.3% versus 63.9%), but had only a very small increase when compared to the single-frame model (59.3% versus 60.9%) and had a significant performance improvement when using the UCF-101 Action Recognition dataset and compared to the base model UCF-101 (43.9% versus 63.3%).

Research [12] used a combined CNN-RNN approach to

analyze human facial expressions, with the result that this approach can outperform the previously applied CNN approach using temporal averages for aggregation.

Ijjina and Mohan in 2016 [13] conducted a research entitled Human action recognition using genetic algorithms and convolutional neural networks using a combined genetic algorithm (GA) and Convolutional Neural Network (CNN) approach. The global search capabilities of the genetic algorithm and the local search capabilities of the gradient descent algorithm are exploited to find a solution closer to the global-optimum. This research shows that combining classifiers generated using genetic algorithms helps improve performance in the recognition process carried out on the UCF50 dataset.

In a research [14] entitled Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition, the process of recognizing human activities and hand movements was carried out using one of the popular RNN models, namely Long Short-Term Memory (LSTM) combined with CNN. The results are good enough to solve the problem of human activity recognition.

Research [15] entitled Human Activity Recognition Using LSTM-RNN Deep Neural Network Architecture, provides an overview of Long Short-Term Memory (LSTM) architectural models that are specifically for human activity recognition applications.

Research [16] shows a review on research with the theme Video Activity Recognition.

From the research that has been discussed previously, it is concluded that using a deep neural network (deep learning) approach is one very good approach to use to build image and video recognition, in this research we want to build an image recognition that can recognize the activity of throwing waste in therein.

III. PROPOSED APPROACH

In general, the model is built through 2 main stages, namely:
1. Build the CNN model with the ResNet architecture and;
2. Conduct training on the CNN Residual Network (ResNet) model in image classification.

A. Build CNN model with ResNet architecture

In general, the ResNet architecture that is built can be seen in Fig. 2. and Fig. 3. From this figure, it can be seen that the main component of ResNet being built is a convolutional block with several residual blocks that carry out the skip process in parallel..

B. Conduct training on the ResNet model

An artificial neural network model that can perform image classification is obtained by conducting training on a previously designed model. To do this, training data is needed in the form of 2 image classes, the first-class image

is an image with the theme or object of people who throw garbage and the second theme image is an image with objects that are not related to trash. The number of training data collected is 480 training data.

IV. RESULT

After determining the deep neural network architecture model, the process of making an application is carried out to form a model architecture and then the model training process is followed. As a training material, the dataset model that has been prepared previously is divided into 3, 10% (49 pictures - we add 1 validation data outside the dataset) taken to validate the remaining 90% of the data in the dataset (432 pictures) are used to conduct model training. The 432-image dataset provided for model training, divided again into 75% training data (324 pictures) and the remaining 25% (108 pictures) used as testing data.

After going through the training process, we get a model that can do labeling on the inputted image, before carrying out the validation process, we first look at the graph of the training process that the model went through. From Fig 1. we can find out information that when the training ended at the 50th epoch the model built in this research had managed to get a fairly good accuracy, higher than 80% (87% to be precise). From the same image we can also see that the training loss is at a fairly small point at below 40% (29% to be precise). For data testing the model built gets the same accuracy at 87% and a smaller loss, namely 22%. This information is sufficient to prove that the model built is good enough for image classification and what is no less important is that the model built can be said to be free from overfitting problems.

Fig. 1. Training Loss and Accuracy on Dataset

After obtaining a trained model, this model is ready for use in image classification. To find out the model's performance, we re-tested the resulting model using the validation data that had been prepared (as many as 49 images). A summary of the results of model validation can be seen in Table I. In this table we can see that the accuracy

of the model reaches 88% with the F1-score of 0.87.

TABLE I
MODEL EVALUATION SUMMARY

class	precision	recall	f1-score	The amount of data
Waste	0.82	1.00	0.90	27
neutral	1.00	0.73	0.84	22
macro avg	0.91	0.86	0.87	49
weighted avg	0.90	0.88	0.87	49
accuracy	0.88			49

From Table I, we can also see that the recall value of the activity of disposing of the waste is quite good, up to 100%, while for the background class (neutral) the precision value reaches 100%. This means that the trash class will be easier to appear, while the neutral class will appear with high precision.

V. CONCLUSION

To build an image classification model we can use the Convolutional Neural Network approach with a Residual Neural Network (ResNet) architecture. In this research, a CNN has been successfully built with a Residual Network 50 Layer (ResNet 50) architecture that can classify waste disposal activities and activities outside of littering (background or neutral).

The model built has a pretty good accuracy of 88% with an F1-Score of 0.87. The model training process has also been shown to be free from overfitting problems.

From this research process, the researcher can convey some suggestions. First, this research uses a baseline version of the activity recognition process in images that can be used to solve the problem of recognizing activities in video. Due to the temporal nature of the video in the next research, we can try the use of DNN such as RNN and LSTM.

Last but not least, this research does not divide the waste disposal activity classes into more specific classes due to the limited dataset, by increasing the quality and quantity of the dataset, it is hoped that the next research can classify waste disposal activities in more detail.

ACKNOWLEDGMENT

REFERENCES

[1] S. Leijdekkers *et al.*, "Effective Interventions on littering behaviour of youngsters What are the ingredients ?," 2015.

[2] Brook Lyndhurst, "Rapid Evidence Review of Littering Behaviour and Anti-Litter Policies," 2013.

[3] R. B. Cialdini, R. R. Reno, and C. A. Kallgren, "A Focus Theory of Normative Conduct : Recycling the Concept of Norms to Reduce Littering in Public Places," *J. Personal. Scocial Psychol.*, vol. 58, no. 6, pp. 1015–1026, 1990.

[4] Y. A. W. De Kort, L. T. Mccalley, and C. J. H. Midden, "Persuasive Trash Cans Activation of Littering Norms by Design," *Environ. Behav.*, vol. 40, no. 6, pp. 870–891, 2016.

[5] R. M. Krauss, J. L. Freedman, and M. Whitcijp, "Studies of Littering," *J. Exp. Scodial Psychol.*, vol. 14, pp. 109–122, 1978.

[6] P. W. Schultz, R. J. Bator, L. B. Large, C. M. Bruni, and J. J. Tabanico, "Littering in Context : Personal and Environmental Predictors of Littering Behavior," *Environ. Behav.*, vol. 45, no. 1, pp. 35–59, 2011, doi: 10.1177/0013916511412179.

[7] C. G. Sibley and J. H. Liu, "DIFFERENTIATING ACTIVE AND PASSIVE LITTERING A Two-Stage Process Model of Littering Behavior in Public Spaces," *Environ. Behav.*, vol. 35, no. 3, pp. 415–433, 2003, doi: 10.1177/0013916503251446.

[8] M. Bateson, R. Robinson, T. Abayomi-cole, J. Greenlees, A. O. Connor, and D. Nettle, "Watching eyes on potential litter can reduce littering : evidence from two field experiments," *PeerJ*, vol. 3, no. e1443, 2015, doi: 10.7717/peerj.1443.

[9] A. M. Jansen, E. Giebels, T. J. L. Van Rompay, and M. Junger, "The Influence of the Presentation of Camera Surveillance on Cheating and Pro-Social Behavior," *Front. Psychol.*, vol. 9, p. 1937, 2018, doi: 10.3389/fpsyg.2018.01937.

[10] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential Deep Learning for Human Action Recognition," in *International workshop on human behavior understanding*, 2011, pp. 29–39.

[11] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-fei, "Large-scale Video Classification with Convolutional Neural Networks," in *The IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[12] S. E. Kahou, V. Michalski, and R. Memisevic, "Recurrent Neural Networks for Emotion Recognition in Video," in *The 2015 ACM on International Conference on Multimodal Interaction*, 2015, pp. 467–474.

[13] E. P. Ijjina and C. K. Mohan, "Human action recognition using genetic algorithms and convolutional neural networks," *Pattern Recognit.*, vol. 59, pp. 199–212, 2016, doi: 10.1016/j.patcog.2016.01.012.

[14] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognit.*, vol. 76, pp. 80–94, 2018, doi: 10.1016/j.patcog.2017.10.033.

[15] S. W. Pienaar and R. Malekian, "Human Activity Recognition Using LSTM-RNN Deep Neural Network Architecture," *arXiv Prepr.*, vol. arXiv, p. 1905.00599, 2019.

[16] I. Rodríguez-moreno, J. M. Martínez-otzeta, B. Sierra, I. Rodriguez, and E. Jauregi, "Video Activity Recognition : State-of-the-Art," *Sensors*, vol. 19, no. 14, p. 3160, 2019, doi: 10.3390/s19143160.
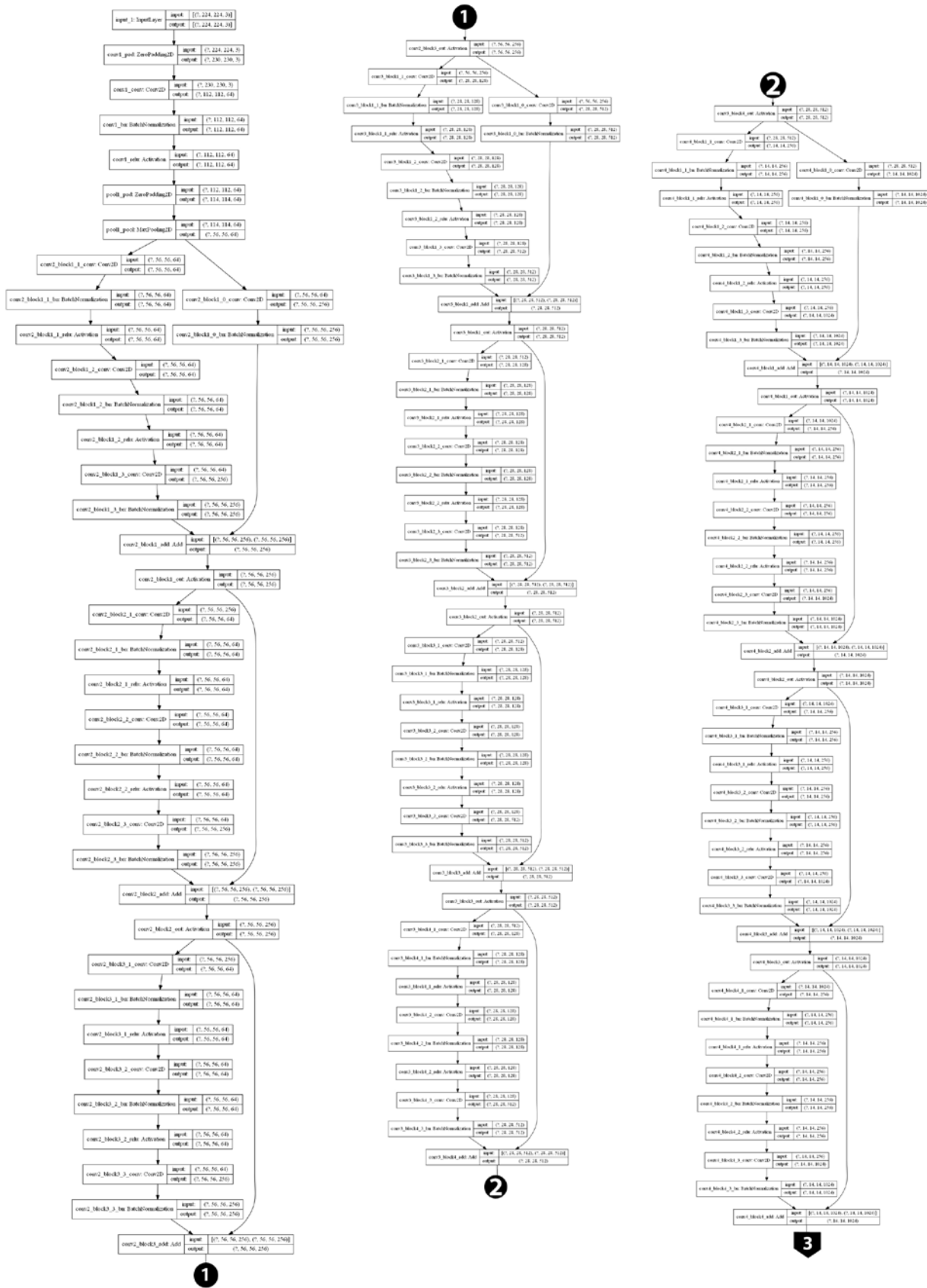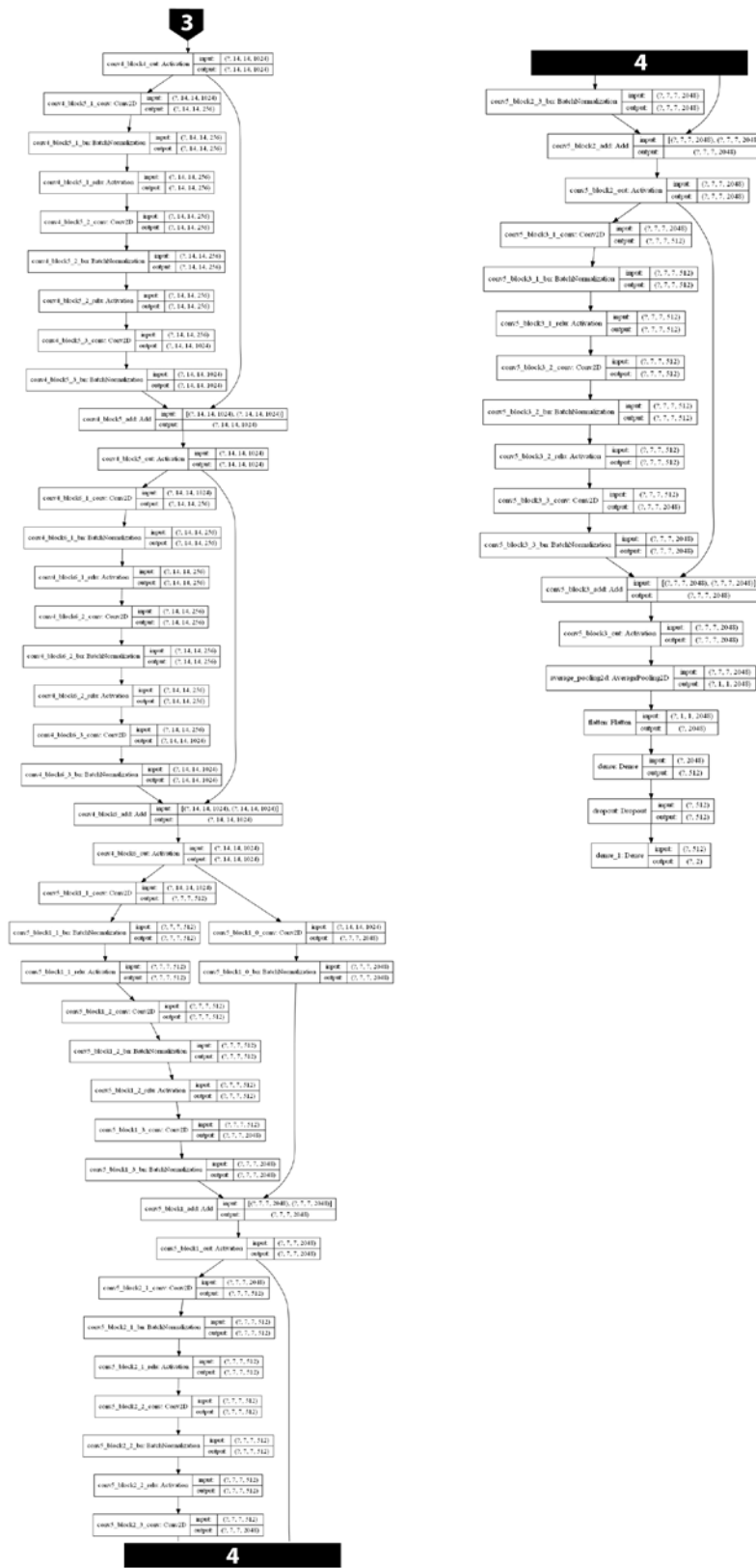
Fig. 2. ResNet50 Architecture (a)

Fig. 3. ResNet50 Architecture (b) continue from (a)