딥 전이 학습을 이용한 인간 행동 분류

닌담 솜사우트 [1], 통운 문마이 [2], 숭타이리엥 [1], 오가화 [1], 이효종 [1]
[1] 전북대학교 공과대학
[2] 시사켓 라자밧 대학교수학교육과, 태국
nindam.somsawut@jbnu.ac.kr, thongoon.m@sskru.ac.th, thaileang@jbnu.ac.kr, salmon2wu@gmail.com, hlee@jbnu.ac.kr

Human Activity Classification Using Deep Transfer Learning

Somsawut Nindam[1], Thong-oon Manmai[2], Thaileang Sung[1], Jiahua Wu[1], Hyo Jong Lee[1]*
[1]Division of Computer Science and Engineering, Jeonbuk National University, Korea
[2]Department of Mathematics Education, Sisaket Rajabhat University, Thailand
*Corresponding author

Abstract

This paper studies human activity image classification using deep transfer learning techniques focused on the inception convolutional neural networks (InceptionV3) model. For this, we used UFC-101 public datasets containing a group of students' behaviors in mathematics classrooms at a school in Thailand. The video dataset contains Play Sitar, Tai Chi, Walking with Dog, and Student Study (our dataset) classes. The experiment was conducted in three phases. First, it extracts an image frame from the video, and a tag is labeled on the frame. Second, it loads the dataset into the inception V3 with transfer learning for image classification of four classes. Lastly, we evaluate the model's accuracy using precision, recall, F1-Score, and confusion matrix. The outcomes of the classifications for the public and our dataset are 1) Play Sitar (precision = 1.0, recall = 1.0, F1 = 1.0), 2), Tai Chi (precision = 1.0, recall = 1.0, F1 = 1.0), 3) Walking with Dog (precision = 1.0, recall = 1.0, F1 = 1.0), and 4) Student Study (precision = 1.0, recall = 1.0, F1 = 1.0), respectively. The results show that the overall accuracy of the classification rate is 100% which states the model is more powerful for learning UCF-101 and our dataset with higher accuracy.

1. Introduction

Artificial intelligence (AI) is evolving and growing to top technology; it is powerful and highly efficient, especially in machine learning (ML) and deep learning (DL). ML and DL are subsections of AI and are very popular in various fields such as industries [1], medical, agricultural, educational, and so on. AI has many parts and differences in the branch, such as computer vision and natural language processing. When referring to computer vision is very interesting to apply a video analysis for detection and classification, such as violence detection and behavior classification. As mentioned, it is good to consider studying a video dataset on human activity classification.

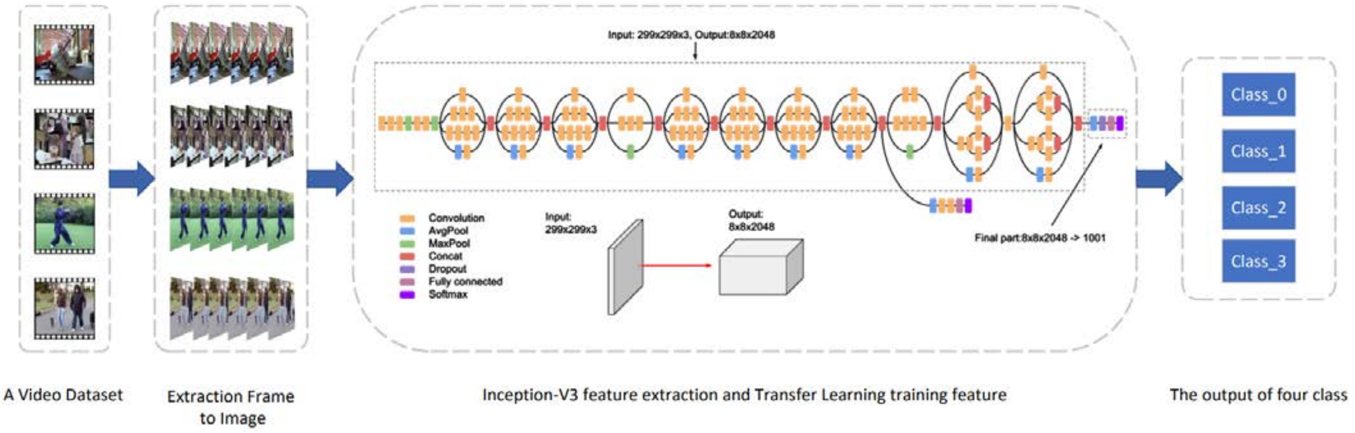
This research applies deep transfer learning methods to classify human activities. To begin with, we collected the public UCF-101 datasets [2], which are Play Sitar, Tai Chi, Walking with Dog, and Student Behavior (our dataset). Each dataset contains different textures, colors, and shapes with various labels. Then we applied a deep transfer learning technique focused on Inception V3 architecture to classify each activity on images. Finally, the model is evaluated on the accuracy of the statistic value for precision, recall, and F1-Score. A confusion matrix is used to calculate the efficiency of the model.

The remaining paper is systematized as follows. Section 2 presents the proposed method used the transfer learning. Section 3 contains the experimental process. The results and analysis are discussed in Section 4. The conclusion is given in Section 5.

2. Proposed Method

2.1 Transfer Learning

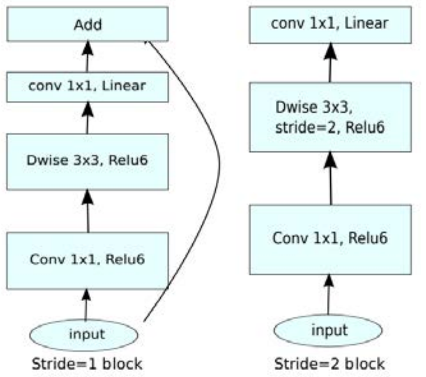
Transfer learning or transfer of learning is primarily proposed to explore how individuals transfer their learning from one context to another similar context (Woodworth and Thorndike, 1901). Transfer learning is usually described as the process and the practical extent to which past experiences affect learning performances in a new situation. That is, a pre-trained model can be transferred to implement a similar task by learning further data distribution and fine-tuning parameters across all layers of the model [3].



(Figure 1) Overview of the proposed framework.

2.2 MobileNetV2

The MobileNetV2 is the next version after the MobileNetV1. A depth-wise separable convolution exists introduced to reduce the complexity of the network. A residual has two types, a block with one stride and another with two block strides used for downsizing, and each block has three layers, the first 1×1 convolution with ReLU6, the second has depth-wise convolution, and the last 1x1 convolution unincluded non-linearity [4].



(Figure 2) MobileNetV2 [5].

2.3 Inception V3

The GoogLeNet is a CNN developed by Google teams in 2014. In the network, the inception network structure is adopted in a new way to reduce the number of network parameters and increase the network depth. Culminate is widely used in image classification tasks in this architecture. As a set of the GoogLeNet is the Inception network structure, the GoogLeNet network is called the Inception network. There are many versions of GoogLeNet such as Inception v1 (2014), Inception v2 (2015), Inception v3 (2015), Inception v4 (2016), and Inception-ResNet (2016). The Inception architecture is a module that typically has three types of convolutions and one maximum pooling with different sizes. The channel is aggregated in the previous layer for the network output after the convolution operation. After that, the nonlinear fusion is performed. This architecture presents the expression of the network, and the adaptability to different scales can be improved [6].

2.4 Evaluation Metrics

In this experiment, we used the confusion matrix to evaluate the performance of the proposed model. The precision metric measures correctly classified positive class samples using Eq. (1). Recall measures the percentage of identified all actual positive samples derived in Eq. (2). And F-measures or F1-Score derived in Eq. (3). are used to analyze the intuitive trade-off between precision and recall so that the constraint can be adjusted [7].

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

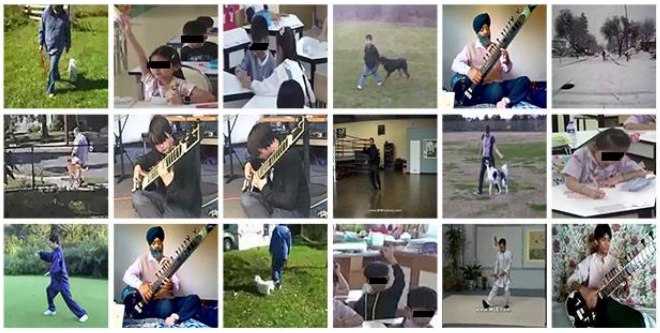
$$F-measure = \frac{(1+\beta)PrecisionRecall}{\beta^2 Precision + Recall} \tag{3}$$

3. Experiment

We collected the UCF-101 public dataset, and our datasets consisted of Playing Sitar, Tai Chi, Walking with Dog, and Student Study. We extracted images from a video frame and selected 2,867 photos randomly. Then we separated them into the train, validate, and test-set. The details are shown in Table 1. After that, the datasets are loaded into the model shown in Figure 1, and the sample dataset is shown in Figure 3.

<Table 1> Collection datasets.

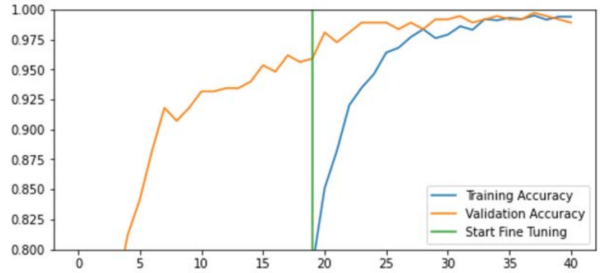
No.	Class Name	Train 70%	Validate 15%	Test 15%	Amount 100%
1	PlayingGuitar	555	119	119	793
2	StudentStudy	452	97	97	646
3	TaiChi	463	99	99	662
4	WalkingWithDog	536	115	115	766
Total/Images					2,867



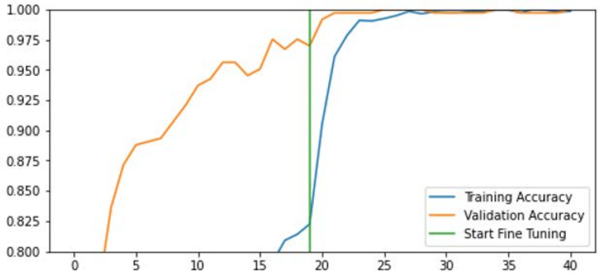
(Figure 3) Sample dataset.

4. Result and Discussion

Using the transfer learning technique, we loaded the datasets (shown in Table 1) into the MobileNetV2 and InceptionV3 models. The training accuracy of the models is shown in Figures 4 and 5. The results of the macro average of Precision, Recall, and F1-Score are shown in Table 2.



(Figure 4) Training result of MobileNetV2.

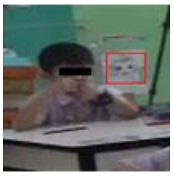


(Figure 5) Training accuracy of InceptionV3.

We randomly loaded an image to predict the model consisting of 430 photos; the result is shown in Table 2.

<Table 2> Macro average of Precision, Recall, and F1-Score.

Model	Precision	Recall	F1-Score
MobileNetV2	0.998	0.997	0.998
InceptionV3	1.000	1.000	1.000



(Figure 6) MobileNetV2 failed to predict.

The actual class was Student Study, but the model predicted Walk with Dog. It happened because one picture behind the boy looked like a dog.

5. Conclusion

The MobileNetV2 obtained the accuracy of Training is 1.0 Validation is 0.98 and Test is 0.98 compared with InceptionV3 can get the accuracy of Training is 0.99, Validation is 1.00 and Test of 1.00, The InceptionV3 is stronger than MobileNetV2. The dataset is wildly different in each class, making the model learn very well of various features.

In the next experiment, we will use only the student's behavior dataset, separating it into four classes with similar features for train and testing in the same environment to find the best accuracy.

Acknowledgment

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2019R1D1A3A03103736 and in part by project for Joint Demand Technology R&D of Regional SMEs funded by Korea Ministry of SMEs and Startups in 2021 (No. S3035805). The data collection of student behavior is funded by the TRSI and the Ministry of Higher Education, Science, Research and Innovation (MHESI) of Thailand (Project-ID: 63673).

Reference

[1] Synced G, Shaoyou (Victor) L, Baorui (Alex) C, Qingyan T, Chenchen (Chain) Z, Chen (Robert) T, Meghan H. (2018). Year of AI: How Global Public Company Adapted to the Wave of AI Transformation: A 2018 Report about Fortune Global 500 Public Company Artificial Intelligence Adaptivity (Kindle Edition). ASIN: B07K91RKVK.

[2] CRCV. (2022). UCF101 - Action Recognition Data Set. https://www.crcv.ucf.edu/data/UCF101.php.

[3] Lin, C., Li, L., Luo, W., Wang, K. C., & Guo, J. (2019). Transfer learning based traffic sign recognition using inception-v3 model. Periodica Polytechnica Transportation Engineering, 47(3), 242-250.

[4] Dertat, A. Review: MobileNetV2—Light Weight Model (Image Classification). Available online: https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification8febb490e61c (accessed on 25 Sep 2022).

[5] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[6] Dong, N., Zhao, L., Wu, C. H., & Chang, J. F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. Applied Soft Computing, 93, 106311.

[7] Tanha, J., Abdi, Y., Samadi, N. et al. Boosting methods for multi-class imbalanced data classification: an experimental review. J Big Data 7, 70 (2020).