Journal of Advanced Agricultural Technologies Vol. 6, No. 3, September 2019

Accuracy Analyses for Detecting Small Creatures Using an OpenCV-Based System with AI for Caffe's Deep Learning Framework

Shinji Kawakura
Laboratory for Future Interdisciplinary Research in Science and Technology, Tokyo Institute of Technology, Kanagawa, Japan
Email: kawakura.s.aa@m.titech.ac.jp

***Abstract*—Agricultural workers want to detect, eliminate, and avoid touching small creatures such as frogs and insects in advance of and during their agricultural work. On the other side, recent researches have suggested diverse countermeasures such as developing robot arm-based machines for harvesting vegetables and pulling up weeds using camera systems; past methods have included monitoring and identifying the positive and negative targets. However, there are not sufficient previous systems for sensing and analyzing the aforementioned small creatures in farmlands. The purpose of this original research is proving the utility of our visual data analysis system based on huge image datasets using Caffe Framework for deep learning using ImageNet 2012, which connects to our program using OpenCV libraries and other outside files. In short, this study selects and executes static visual analyses using AI-based computing by tools concerning deep learning using several hidden layers after obtaining and accumulating field pictures and video data concerning small creatures such as frogs and insects in outdoor farmlands. Additionally, the author calculates the ratio between the sizes of outline of leaves on which small creatures existed as well as that of the targeted small creatures as one original standard for giving a unity to the pictures selected to some extent. Our results confirm the utility of the detection methodologies. In future, these results could contribute to the development of automatic agricultural harvesting robot-systems and to improving the daily work effectiveness of actual manual workers. Furthermore, an automatic system for eliminating small creatures could support the recruitment of agricultural workers.**

***Index Terms*—picture classification using deep learning, small animals in farmlands, ImageNet 2012, Caffe Framework, OpenCV**

I. Introduction

In recent years, agricultural researchers and workers (agri workers) have developed several automatic and mechanical techniques to improve the utility of harvesting robot-systems by enabling them to search for the color and size of vegetables and fruits based on visual data [1]-[4]. These achievements in academic and business fields have already reached sufficient levels to utilize them in outer fields and inner farmlands. Additionally, researchers in the field of agricultural informatics and robotics have proposed various promising methods for improving these tasks. Existing visual analysis methods have focused mainly on vegetables, fruits, weeds, and farmers, including robotic farming systems, and have not focused specifically or sufficiently on small creatures living in the fields [2], [5]-[16]. Therefore, past studies and systems have been insufficient for analyzing small creatures and other objects, such as leaves and stems of vegetables, to determine their colors and textures. However, new technologies continue to be developed, and this study aims to develop a video data analysis system based on huge image datasets of Caffe Framework for deep learning using ImageNet 2012, which connects to our program using OpenCV libraries and other external files. In this study, the target field is situated in a common outer farm, and the targeted creatures are small frogs, mini-tomatoes, and rotten mini-tomatoes.

II. Materials and Methods

In this study, we selected and executed a number of layer-oriented deep learning based analyses to obtain picture data sets. We used the latest Caffe Framework, and various libraries and packages of OpenCV 3.2 for Visual Studio 2015 and 2017.

A. Target

First, we carefully categorized the following small creatures in agri fields according to their physical and visual characteristics (color, shape, size, etc): (1) insects, (2) amphibian and reptiles, (3) small birds, and (4) small mammals. Based on feedback from real farmers, this study targeted (1) small frogs (abbreviated as "frog"; n = 4), (2) vivid-colored mini-tomatoes ("mini-tomato"; n = 5), and (3) utterly rotten mini-tomatoes ("rotten mini-tomato"; n = 4), as shown in Fig. 1.

Manuscript received April 3, 2019; revised August 20, 2019.

doi: 10.18178/joaat.6.3.166-170

Figure 1. Example sets of four-square pictures concerning the target (1) frog on the mini-tomato leaf, (2) medium-sized mini-tomato, and (3) rotten mini-tomato. The left column is wide range (ratio of item in picture ≒ 5.00 %), and the right column shows close range (ratio of item in picture ≒ 50.0 %).

B. Objective of the Analyses

This study selected the following deep learning method using AI and huge open datasets, such as ImageNet 2012, as the target and teacher data (reference data). In light of current academic trends and past results using them, the methodology has adequate universality in the agricultural informatics field.

Fig. 2 shows the flow of the experiment, which comprised (1) obtaining pictures and movie data from the target area farmlands, (2) analyzing the data using the original system utilizing open data on the internet, (3) calculating and comparing charts of the statistical information, and (4) presenting the results to agricultural systems' developers, agri-workers, and managers.

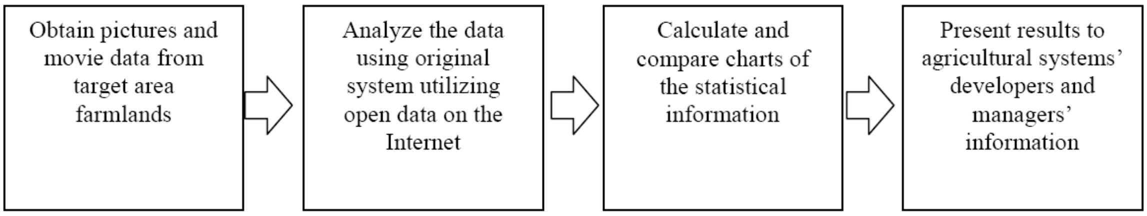

Figure 2. Flow of the experiment.

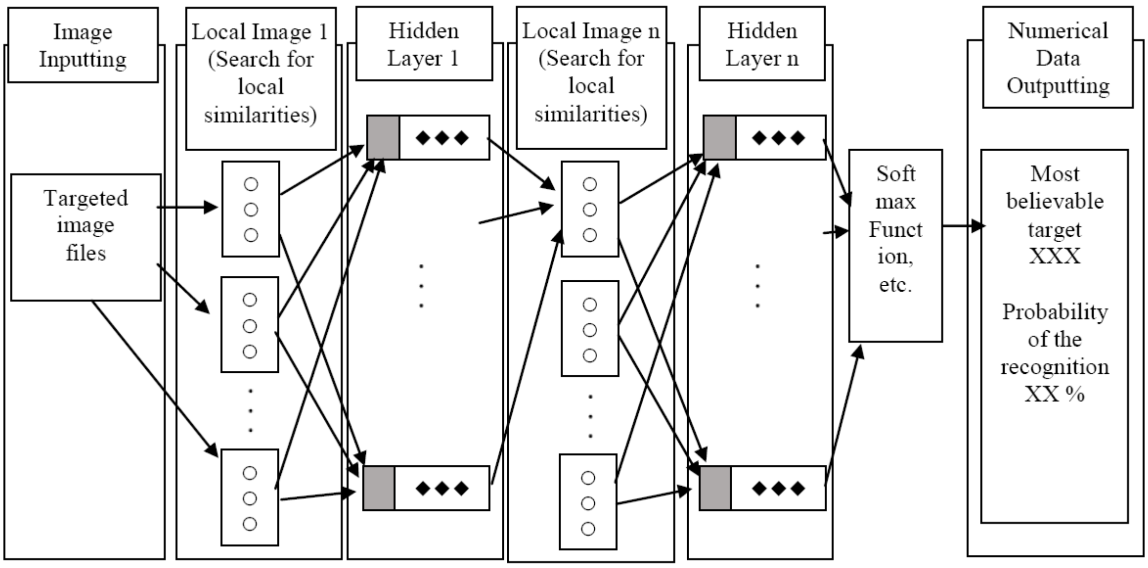

Figure 3. Flowchart of the Caffe framework based deep learning model of the system.

The frogs were 25 mm to 45 mm long, and the mini-tomatoes were 20 mm to 35 mm (although, for computing processes, these sizes were uniformed as 224 px * 224 pixels). However, the complexity of the real, non-trimmed farmlands made it difficult to take measurements, and differences in the responses between individuals impeded understanding of the data. Table I, Fig. 3 and Fig. 4 show a set of items used for the Caffe Framework based analyses, which comprised multiple layers.

As shown in Fig. 3 and Fig. 4, the sequential processing was programmed in Visual C++ in Visual Studio 2015 and 2017, and the ratios of the areas between the main targets and the whole pictures were calculated using an accuracy-assessed original program through OpenCV libraries (see Tables II-IV).

"A "Blob" is a wrapper over the actual data being processed and passed along by Caffe, and also under the hood provides synchronization capability between the CPU and the GPU. Mathematically, a blob is an N-dimensional array stored in a C-contiguous fashion. Caffe stores and communicates data using blobs. Blobs provide a unified memory interface holding data; e.g., batches of images, model parameters, and derivatives for optimization. Blobs conceal the computational and mental overhead of mixed CPU/GPU operation by synchronizing from the CPU host to the GPU device as needed. Memory on the host and device is allocated on demand (lazily) for efficient memory usage. The conventional blob dimensions for batches of image data are number N * channel K * height H * width W. Blob memory is row-major in layout, so the last / rightmost dimension changes fastest. For example, in a 4D blob, the value at index (n, k, h, w) is physically located at index ((n * K + k) * H + h) * W + w. Number / N is the batch size of the data. Batch processing achieves better throughput for communication and device processing. For an ImageNet training batch of 256 images N = 256. Channel / K is the feature dimension e.g. for RGB images K = 3." [17], [18].

TABLE I. DIAGRAM OF CAFFE FRAME-NETWORK BASED ANALYSIS

Layer of analysis	Description
Data	Raw data inputting layer
Conv 1	Convolutional layer 1
Pool 1	Pooling layer
Norm 1	Normalizing layer
Conv 2	Convolutional layer 2
…	...
Pool n	Pooling layer n
Classifier	Classifying layer
Prob	Result outputting layer

Channel numbers and timeline changes in pictures' heights and widths were not calculated.

As of October 2018, there are not enough qualified, defect-standard theories concerning the number of layers and the setting of parameters.

[Pre-processing]
・Initial declarations of header and library files and namespaces

[Main function]
・Declaration of variable numbers

[Inputting open data models using the Internet]
・Inputting Caffe model of bvlc_googlenet onto the system
・Reading class-explanation files of extracted from ImageNet 2012 database

[Inputting captured data into the program]
・Inputting static pictures and frame-data extracted from movie data file, and setting them in prepared arrays for the data
・Resizing the data to the appropriate size
・Changing the data into Blob-style

[Analyses data using deep learning based program system]
・Inputting the data into Blob
・Analyzing and classifying the data using deep learning
・Calculating the class(es) in which output value can be maximized
・Showing the class(es) and each reliabilities (%) concerning the data

[Post-processing]
・Waiting and closing the system.

Figure 4. Flowchart of program for the trials.

III. RESULTS

Table II-IV present and discuss the statistical results ([5]-[10], [12]-[14], [19]-[22]). The items in each column are "Distance between targets and the camera," "The most believable class (Primary candidate of the deduction by AI)," "Percentage of the most believable class (average standard deviation [SD])," and "Ratio of the main target to whole picture (average SD)."

TABLE II. STATISTICAL VALUES OF FROG DATA JUDGED AUTOMATICALLY USING AI AND THE CAFFE MODEL

Distance between targets and camera	The most believable class (Primary candidate of the deduction by AI)	Percentage of the most believable class (Average Standard Deviation (SD))	Ratio of the main target to whole picture (Average SD)
Closest pattern	Tree frog	49.2 % (8.32)	50.0 % (1.69)
Medium distant pattern	Axolotl, mud puppy, ambystoma mexicanum	18.1 % (15.5)	25.5 % (4.64)
Most distant pattern	Head cabbage	7.29 % (2.05)	5.00 % (5.82)
None target (only mini-tomato leaf)	Head cabbage	6.93 % (1.97)	0.00 % (Nothing)

TABLE III. STATISTICAL VALUES OF MINI-TOMATO DATA JUDGED AUTOMATICALLY USING AI AND THE CAFFE MODEL

Distance between targets and camera	The most believable class (Primary candidate of the deduction by AI)	Percentage of the most believable class (Average SD)	Ratio of the main target to whole picture (Average SD)
Closest pattern	Vegetable, veggie, veg	50.9 % (8.19)	50.1 % (1.40)
Medium distant pattern	Solanaceous vegetable	22.3 % (4.77)	25.1 % (5.30)
Most distant pattern	Solanaceous vegetable	12.9 % (4.87)	5.60 % (5.05)
None target (only mini-tomato leaf)	Greens, green, leafy vegetable	10.1 % (1.98)	0.00 % (Nothing)

TABLE IV. STATISTICAL VALUES OF THE ROTTEN MINI-TOMATO DATA JUDGED AUTOMATICALLY USING AI AND THE CAFFE MODEL

Distance between targets and camera	The most believable class (Primary candidate of the deduction by AI)	Percentage of the most believable class (Average SD)	Ratio of the main target to whole picture (Average SD)
Closest pattern	Vegetable, veggie, veg	28.8 % (9.52)	52.8 % (3.49)
Medium distant pattern	Solanaceous vegetable	8.94 % (6.68)	24.5 % (5.20)
Most distant pattern	Solanaceous vegetable	7.61 % (8.91)	5.35 % (8.94)
None target (only mini-tomato leaf)	Greens, green, leafy vegetable	11.1 % (2.51)	0.00 % (Nothing)

IV. DISCUSSION

Considering our future contributions to agricultural systems' developers, workers, and managers, this study obtained significant data in future systems' evolving ways, and confirmed the usefulness of deep learning based analyses.

As Table II-IV show, the cases of the frog, mini-tomatoes, and rotten mini-tomato datasets indicate numerical features related to them. It is difficult to comment on the cases concerning the frogs because their colors and textures are similar to those of common mini-tomato leaves. However, obvious differences were evident between the mini-tomatoes and rotten mini-tomatoes concerning both average values and the SD; we could confirm these tendencies from the analyzed numeric data and video data.

Concretely, for the data of frog's most distant pattern, the value (7.29 %) became closer to the values of "None target" (6.93 %) for feelings, in light of responses from real agri workers and managers. For the most believable data classes of mini-tomato and rotten mini-tomato in Table III and IV, not all items were different. Especially, for the data regarding mini-tomato's upside three items values ("Closest pattern", "Medium distant pattern", and "Most distant pattern") became larger than those of rotten mini-tomato, the author guessed that was because of these vivid-colors and round shapes.

Totally speaking, the author thought that closest pattern concerning above three targets, the most believable classes would be useful to achieve the aims written earlier for actual automatic agricultural judging aims. Concerning the SD values in themselves, it was difficult to comment that the indicator is suit to utilize for judging actual agricultural items, because, all pictures in each category were similar, so the range of almost all SD values except for data concerning frog's medium distant pattern seemed to be statistically limited.

V. CONCLUSION AND FUTURE TASKS

In this study, we constructed and demonstrated a deep learning based visual data analysis system for use at diverse high-tech agricultural work sites. This study demonstrated these accuracies and availabilities, and presented various numerical data, mainly the probabilities of data classifying. The captured picture files were analyzed automatically using the system's appropriate classification with changing picture types and conditions (mini-tomatoes were ripe or rotten, the distance between targets and camera), with consideration for future practical usages. In addition, we described the differences between correctly certified and non-certified data, and we described these features both quantitatively and qualitatively.

Thinking of the histories of ImageNet 2012 and the Caffe Framework, the author should not describe concerning these findings resembled past similar academic achievements. Furthermore, that is also because of the insufficiency of the volume and quality of appreciated samples in agricultural fields. In light of recent high-tech agricultural techniques and farmlands at different sized farmlands, we confirmed the usefulness and scalability of the methodologies. Our future work will aim to provide further confirmation related to the variety of the detected targets and background conditions, the validity, and the system's durability and long-term performance, including the methodologies for automatic agricultural machinery, such as cultivating systems, weeding machines, and cutting rotten vegetables. In addition, patterns from other databases would be helpful in such fields.

We hope that the aforementioned promising methodologies will be widely applied to real working sites to promote the recruitment of workers into agricultural fields. These methods could also prevent farmers from suffering injuries (e.g., being attacked by snakes or insects). Finally, in the long term, these results could be used for automatic systems to help inner and outer farmers revise and improve their work practices.

REFERENCES

[1] Y. Huang, "The advancement of nature-inspired algorithms for agriculture," in *Proc. American Society of Agricultural and Biological Engineers (ASABE) Annual International Meeting*, 2018, p. 1.

[2] S. S. Mehta and T. F. Burks, "Vision-based control of robotic manipulator for citrus harvesting," *Computers and Electronics in Agriculture*, vol. 102, pp. 146-158, January 2014.
[3] A. Shinjo and M. Kudo, "The practical use of IT in agriculture: the movement into high-value-added crops and integrated solutions," *The Journal of the Institute of Electronics, Information, and Communication Engineers*, vol. 96, no. 4, pp. 280-285, April 2013.
[4] S. Han, Q. Zhang, B. Nic, and J. F. Reid, "A guidance directrix approach to vision-based vehicle guidance systems," *Computers and Electronics in Agriculture*, vol. 43, pp. 179-195, January 2004.
[5] N. Zhu, X. Liu, Z. Liu, K. Hu, Y. Wang, J. Tan, and Y. Guo, "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 32-44, July 2018.
[6] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, pp. 1-12, May 2016.
[7] F. A. Guth, S. Ward, and K. P. McDonnell, "Autonomous disease detection in crops using deep learning," *Biosystems and Food Engineering Research Review*, vol. 22, pp. 1-209, May 2017.
[8] K. Ahlin, B. Joffe, A. P. Hu, G. McMurray, and N. Sadegh, "Autonomous leaf picking using deep learning and visual-servoing," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 177-183, October 2016.
[9] E. Cengil, A. Çınar, and E. Özbay, "Image classification with Caffe deep learning framework," in *Proc. IEEE International Conference in Computer Science and Engineering*, 2017, pp. 440-444.
[10] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, and V. Kumar, "Counting apples and oranges with deep learning: A data-driven approach," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 781-788, December 2016.
[11] T. T. Zin, C. N. Phyo, P. Tin, H. Hama, and I. Kobayashi, "Image technology based cow identification system using deep learning," in *Proc. the International Multi Conference of Engineers and Computer Scientists*, March 2018, vol. 1, pp. 320-323.
[12] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299-315, April 2017.
[13] A. Dutta, J. M. Gitahi, P. Ghimire, and R. Mink, "Weed detection in close-range imagery of agricultural fields using neural networks," Publikationen der DGPF, Band 27, pp. 633-645, 2018.
[14] F. J. Rodríguez, A. García, P. J. Pardo, F. Chávez, and R. M. Luque-Baena, "Study and classification of plum varieties using image analysis and deep learning techniques," *Progress in Artificial Intelligence*, vol. 7, no. 2, pp. 119-127, January 2018.
[15] E. Okafor, R. Smit, L. Schomaker, and M. Wiering, "Operational data augmentation in classifying single aerial images of animals," in *Proc. IEEE International Conference on Intelligent Systems and Applications*, 2017, pp. 354-360.
[16] R. Wang, J. Zhang, W. Dong, J. Yu, C. J. Xie, R. Li, and H. Chen, "A crop pests image classification algorithm based on deep convolutional neural network," *Telkomnika*, vol. 15, no. 3, pp. 1239-1246, August 2017.
[17] Y. Jia and E. Shelhamer. (September 2018). Blobs, layers, and nets: Anatomy of a Caffe model. Caffe Berkeleyvision. [Online]. Available: http://caffe.berkeleyvision.org
[18] Berkeley Artificial Intelligence Research Lab. (September 2018). Berkeley Vision and Learning Center (BVLC). [Online]. Available: https://bair.berkeley.edu/978-4-274-22114-9
[19] T. Kan, *Excel de Manabu Toukei Kaiseki Nyuumon*, Tokyo, Japan: Ohmsha, 2013.
[20] N. Toyama and M. Tsujitani, *Jissen R Toukei Bunseki*, Tokyo, Japan: Ohmsha, 2015.
[21] N. Nagata and S. Toyosawa, *Jissen OpenCV 2.4*, Tokyo, Japan: CUTT SYSTEM, 2009.
[22] N. Nagata and S. Toyosawa, *Jissen OpenCV 3 for C++*, Tokyo, Japan: CUTT SYSTEM, 2017.

Shinji Kawakur was born on July 14, 1978. He received his Ph.D. in Environmentology from the University of Tokyo in 2015 (Meguro-ku, Tokyo, Japan); B.A in control system engineering from Tokyo Institute of Technology in 2003; and M.A. in human-factors engineering from Tokyo Institute of Technology in 2005 (Meguro-ku, Tokyo, Japan). His past career was about system engineering, researching for private companies. His research interest includes Development and Verification of Wearable Sensing Systems with Real-time Spoken Commands for Outdoor Agricultural Workers.