Deep-learning Object Detection for Resource Recycling

Yeong-Lin Lai¹, Yeong-Kang Lai²*, Syuan-Yu Shih¹, Chun-Yi Zheng¹, and Ting-Hsueh Chuang¹

¹Department of Mechatronics Engineering, National Chung Hsing University, 402204 Taichung, Taiwan
Email: yklai@dragon.nchu.edu.tw

Abstract. Recent years have seen a growing concern over global warming, as well as environmental pollution and protection issues. Resource recycling helps the effective reduction of greenhouse gases and environmental pollution, and improves the quality of life for many people. This paper proposes a deep-learning object detection system for resource recycling. The resource recycling of the objects including paper cups, plastic bottles, and aluminum cans was conducted by artificial intelligence. Single shot multibox detector (SSD) and faster region-based convolutional neural network (Faster R-CNN) models were utilized for the training of the deep-learning object detection. With regard to data set images and training time, the accuracy, training steps, and loss function of the SSD and Faster R-CNN models were studied. The accuracy and loss characteristics of the deep-learning object detection system for resource recycling were demonstrated. The system exhibits good potential for the applications of resource recycling and environmental protection.

1. Introduction

Awareness of and concern for environmental issues have grown consistently over the last few decades, and the current global warming trend is widely acknowledged as particularly significant. Frequent extreme weather events around the world have attracted everyone’s attention. Climate change is a direct result of global warming. Signs of rapid climate change are compelling, such as global temperature rise, warming oceans, declining arctic sea ice, glacial retreat, decreased snow cover, shrinking ice sheets, sea level rise, extreme climate events, and ocean acidification. The scientific evidence for the warming of the climate system is unequivocal [1-14]. Since the industrial revolution, the use of energy has caused a considerable increase in greenhouse gas emissions, and the global climate change derived from greenhouse gas emissions has become increasingly serious [13,15]. Therefore, greenhouse gas reduction is a necessary measure in the fight against climate change [16]. In modern society, people are very aware of global warming as well as the environmental pollution and protections are important issues that the people pay much attention to. The relationships between global warming and resource issues, such as resource exploitation, garbage incineration, and waste landfilling have been discussed and studied [17-23], and it has been established that the separate collection of recyclable waste is an extremely crucial procedure in waste management [17-20,23] for the reduction of greenhouse gases.

As it stands, the separate collection of waste requires heavy manpower. However, the utilization of convenient mechanical treatment techniques can increase the recovery of valuable materials [23]. The employment of image processing and automation technology [24-28] in the separate collection of recyclable waste can thus save labor and processing time.
Deep learning is a relatively new, but rapidly growing technology. Several networks, such as single shot multibox detector (SSD) [29] and faster region-based convolutional neural network (Faster R-CNN) [30], have been reported for many applications. In this paper, a deep-learning image object detection technology for resource recycling is developed. The SSD and Faster R-CNN models are studied for the training of the resource-recycling deep-learning. A self-taken data set is used. The object detection of the system is related to many factors, including the deep learning neural network, object types, background settings, field brightness, detection angles, and so on.

2. Networks

The proposed deep-learning object detection system for resource recycling is carried out using both the SSD and Faster R-CNN models. The SSD network consists of two parts. The first part of the SSD network architecture is feature extraction, which is a convolutional layer from other networks such as visual geometry group (VGG), inception (GoogLeNet), etc. The output of these networks is passed to the detection section. The second part of the SSD network architecture is detection, which consists of a series of detection modules. The detection module contains the bounding box generation, classification, and localization calibration functions, as shown in Fig. 1. Each detection module has its own output that helps the final output of the network. After the detection process, the size of the feature map is reduced.

The loss function of the SSD network consists of two parts [29]. The first part is related to classification. The confidence loss shown in Eq. 1 is the loss of classification predictions. For positive match predictions, the loss is penalized according to the confidence score of the corresponding classification. For negative match predictions, the loss is penalized in accordance with a confidence score of classification “0”. The classification “0” indicates that the objects are not detected. It is calculated as a softmax loss on the confidence $y$, classification score, of multiple classifications. $N$ is the number of matched default boxes.

$$
L_{conf}(x, y) = - \sum_{i \in Pos} x_i^p \log(y_i^p) - \sum_{i \in Neg} \log(y_i^0)
$$

**Figure 1.** Detection module of SSD network.
\[
\hat{y}_i^p = \frac{\exp(y_i^p)}{\sum_p \exp(y_i^p)}
\]

The second part of the loss function is related to localization. Weighted smooth L1 function is used for the localization loss. The localization loss between the prediction box, \( l \), and the ground truth box, \( g \), is defined as the smooth L1 loss:

\[
L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smoothL1}(l_i^m - g_i^m)
\]

\[
\begin{align*}
\bar{g}_{j}^{cx} &= \frac{(g_{j}^{cx} - d_{i}^{cx})}{d_{i}^{w}}, \\
\bar{g}_{j}^{cy} &= \frac{(g_{j}^{cy} - d_{i}^{cy})}{d_{i}^{h}}
\end{align*}
\]

where \( cx \) and \( cy \) are the offsets to the default bounding box, \( d \), of width, \( w \), and height, \( h \).

The loss of Faster RCNN is mainly divided into the loss of RPN and Fast RCNN [30]. Both losses include classification loss and regression loss. The normalized classification loss is

\[
\frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*)
\]

where \( N_{cls} \) is the total number of anchor for classification, \( p_i \) is the predicted probability of the anchor \( i \), and \( L_{cls}(p_i, p_i^*) \) is the classification loss.

The normalized regression loss is

\[
\lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(q_i, q_i^*)
\]

where \( \lambda \) is the balancing parameter, \( N_{reg} \) is the total number of anchor for regression, \( q_i \) is the vector representing the predicted bounding box, and \( L_{reg}(q_i, q_i^*) \) is the regression loss.

Faster RCNN uses convolution layers to extract feature maps. The entire loss function of the Faster RCNN network is

\[
L(\{p_i\}, \{q_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(q_i, q_i^*)
\]

3. Results and discussion

This system used an NVIDIA-GTX1080Ti graphics processing unit (GPU), and linux Ubuntu16.04 to conduct experiments. In addition, compute unified device architecture (CUDA) version 9.0, NVIDIA CUDA® deep neural network library (cuDNN) version 7.0.5, Python version 3.5, and TensorFlow GPU version 1.12.0 were used. The setting procedure of software employed in this study is shown in Fig. 2. The object detection flow chart of this study is shown in Fig. 3.
The classification data were the images of three classifications including paper cups, PET bottles, and aluminum cans. Fig. 4 shows one of the data set images of aluminum cans. Each data set consisted of 1503 images. In order to train the system to recognize whether the object images were paper cups,
PET bottles, or aluminum cans, it was necessary to label the objects in the image, as shown in Fig. 5. The purpose was to make the system learn to recognize objects. The xml extension file stored the data of the labeled objects. The image sizes of photos, location coordinates, and item names of labeled objects could be inspected in the file. The files with the extension file name csv were converted from the xml file. The file contents contained information, including the file names, widths and heights of images, labeled classifications, and coordinate positions \((x_{\text{min}}, y_{\text{min}}, x_{\text{max}}, y_{\text{max}})\) of labels.

The system read the converted csv file data and performed the training of learning procedures on the basis of the data. The epoch, loss, and training time results were recorded. The trained files were then converted into the validation data files. Finally, the trained model were obtained.

A total of 4509 classification images for paper cups, PET bottles, and aluminum cans were used. There were 1503 data set images for each classification. For each classification, 1353 images were selected for the training data set, and the remaining 150 images of each classification were utilized for the test data set.

The deep-learning object detection system training using the SSD model took 19 hr, 49 min, and 53 sec for an epoch of 268,600. The loss value was 0.9244. The validation data set using the SSD model for the deep-learning object detection system consisted of 33 images of PET bottles, 33 images of...
paper cups, and 33 images of aluminum cans. The system correctly recognized all PET bottles and aluminum cans, and correctly recognized 26 images of paper cups in the validation data set. An accuracy of 92% was obtained.

While the deep-learning object detection system was trained using the Faster RCNN model for an epoch of 160,500, over 20 hr, 59 min, and 52 sec. The loss value was 0.04388. The validation data set using the Faster RCNN model consisted of 33 images of PET bottles, 33 images of paper cups, and 33 images of aluminum cans. The system correctly recognized all PET bottles and aluminum cans, and correctly recognized 30 images of paper cups in the validation data sets. The object detection results of the system using the Faster RCNN model are shown in Fig. 6. The system exhibited an accuracy of 97%. The characteristics of the deep-learning object detection system using the SSD and Faster RCNN models are listed in Table 1. In the training time close to 20 hours, the system employing the Faster RCNN model showed achieved superior performance than that using the SSD model in terms of loss and accuracy.

![Object detection results using the Faster RCNN model for (a) paper cups, (b) PET bottles, and (c) aluminum cans.](image)

**Table 1.** Characteristics of deep-learning object detection using the SSD and Faster RCNN models.

| Item     | SSD            | Faster RCNN    |
|----------|----------------|----------------|
| Training time | 19 hr, 49 min, and 53 sec | 20 hr, 59 min, and 52 sec |
| Loss     | 0.9244         | 0.04388        |
| Epoch    | 268,600        | 160,500        |
| Accuracy | 92%            | 97%            |

4. Conclusion

This study developed a deep-learning object detection system for resource recycling. The object detection of paper cups, plastic bottles, and aluminum cans was carried out using the SSD and Faster RCNN models. The loss values of the characteristics of the system using the SSD and Faster RCNN models were 0.9244 and 0.04388, respectively. The accuracy values for the SSD and Faster RCNN models were 92% and 97%, respectively. The network architecture of the SSD model is simpler than that of the Faster RCNN model. As far as the loss and accuracy are concerned, the system with the Faster RCNN model exhibited better characteristics than that with the SSD model. The system thus demonstrated the effective results for the object detection of resource recycling, and had considerable potential for environmental protection and pollution reduction.
5. Acknowledgment

This work was supported in part by the Ministry of Science and Technology of Taiwan, R.O.C. under Contracts MOST 108-2622-E-018-001-CC3, MOST 108-2221-E-018-017, and MOST 108-2218-E-005-010.

6. References

[1] B.D. Santer, K.E. Taylor, T.M.L. Wigley, T.C. Johns, P.D. Jones, D.J. Karoly, J.F.B. Mitchell, A.H. Oort, J.E. Penner, V. Ramaswamy, M.D. Schwarzkopf, R.J. Stouffer, S. Tett, A search for human influences on the thermal structure of the atmosphere, Nature. 382 (1996) 39-46.

[2] G.C. Hegerl, H. von Storch, K. Hasselmann, B.D. Santer, U. Cubasch, P.D. Jones, Detecting greenhouse-gas-induced climate change with an optimal fingerprint method, J. Clim. 9 (1996) 2281-2306.

[3] V. Ramaswamy, M.D. Schwarzkopf, W.J. Randel, B.D. Santer, B.J. Soden, G.L. Stenchikov, Anthropogenic and natural influences in the evolution of lower stratospheric cooling, Science. 311 (2006) 1138-1141.

[4] B.D. Santer, M.F. Wehner, T.M.L. Wigley, R. Sausen, G.A. Meehl, K.E. Taylor, C. Ammann, J. Arblaster, W.M. Washington, J.S. Boyle, W. Brüggemann, Contributions of anthropogenic and natural forcing to recent tropopause height changes, Science. 301 (2003) 479-483.

[5] M. Hulme, On the origin of ‘the greenhouse effect’: John Tyndall’s 1859 interrogation of nature, Weather. 64 (2009) 121-123.

[6] H. Rodhe, R. Charlson, E. Crawford, Svante arrenius and the greenhouse effect, Ambio. 26 (1997) 2-5.

[7] National Research Council, Surface temperature reconstructions for the last 2,000 years, National Academies Press, 2007.

[8] NASA GISS: NASA News & Feature Releases: NASA, NOAA Data Show 2016 Warmest Year on Record Globally. Information on https://www.giss.nasa.gov/research/news/20170118/.

[9] S. Levitus, J.I. Antonov, T.P. Boyer, O.K. Baranova, H.E. Garcia, R.A. Locarnini, A.V. Mishonov, J.R. Reagan, D. Seidov, E. Yarosh, M.M. Zweng, NCEI ocean heat content, temperature anomalies, salinity anomalies, thermosteric sea level anomalies, halosteric sea level anomalies, and total steric sea level anomalies from 1955 to present calculated from in situ oceanographic subsurface profile data, Natl. Centres Environ. Inf. Dataset (2017) DOI: 10.7289/V53F4MVP.

[10] R. Hock, A. Bliss, B. Marzeion, R.H. Giesen, Y. Hirabayashi, M. Huss, V. Radic, A.B.A. Slangen, GlacierMIP-A model intercomparison of global-scale glacier mass-balance models and projections, J. Glaciol. 65 (2019) 453-467.

[11] M. Zemp, M. Huss, E. Thibert, N. Eckert, R. McNabb, J. Huber, M. Barandun, H. Machguth, S. U. Nussbaumer, I. Gärnter-Roer, L. Thomson, F. Paul, F. Maussion, S. Kutuzov, J.G. Cogley, Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016, Nature 568 (2019) 382-386.

[12] M. Zemp, H. Frey, I. Gärnter-Roer, S.U. Nussbaumer, M. Hoelzle, F. Paul, W. Haeberli, F. Denzinger, A.P. Ahlstrom, B. Anderson, S. Bajracharya, C. Baroni, L.N. Braun, B.E. Cáceres, G. Casassa, G. Cobos, L.R. Dávila, H. Delgado Granados, N.M. Demuth, L. Espizua, A. Fischer, K. Fujita, B. Gadek, A. Ghazanfar, J.O. Hagen, P. Holmlund, N. Karimi, Z. Li, M. Pelto, P. Pite, V.V. Popovnin, C.A. Portocarrero, R. Prinz, C.V. Sangewar, I. Severskiy, O. Sigurdsson, A. Soruco, R. Usualiev, C. Vincent, Historically unprecedented global glacier decline in the early 21st century. J. Glaciol. 61 (2015) 745-762.

[13] R.S. Nerem, B.D. Beckley, J.T. Fasullo, B.D. Hamlington, D. Masters, G.T. Mitchum, Climate-change–driven accelerated sea-level rise detected in the altimeter era, Proc. Natl. Acad. Sci. U.S.A. 115 (2018) 2022-2025.

[14] SOTC: Sea Ice | National Snow and Ice Data Center. Information on https://nsidc.org/cryosphere/sotc/sea_ice.html.
[15] G.H. Roe, M.B. Baker, F. Herla, Centennial glacier retreat as categorical evidence of regional climate change, Nat. Geosci. 10 (2017) 95-99.

[16] Y. Gao, X. Gao, X. Zhang, The 2 °C global temperature target and the evolution of the long-term goal of addressing climate change—from the United Nations framework convention on climate change to the Paris agreement, Engineering. 3 (2017) 272-278.

[17] R. Cremiato, M.L. Mastellone, C. Tagliaferri, L. Zaccariello, P. Lettieri, Environmental impact of municipal solid waste management using life cycle assessment: The effect of anaerobic digestion, materials recovery and secondary fuels production, Renew Energy. 124 (2018) 180-188.

[18] M. Haupt, E. Waser, J.C. Würmli, S. Hellweg, Is there an environmentally optimal separate collection rate?, Waste Manag. 77 (2018) 220-224.

[19] P.S. Calabrò, Greenhouse gases emission from municipal waste management: The role of separate collection, Waste Manag. 29 (2009) 2178-2187.

[20] C. Wünsch, F.G. Simon, The reduction of greenhouse gas emissions through the source-separated collection of household waste in Germany, in: R. Maletz, C. Dornack, L. Ziyang (Eds.), Source Separation and Recycling: Implementation and Benefits for a Circular Economy. The Handbook of Environmental Chemistry, Springer International Publishing, Cham, Switzerland, 2018, pp. 269-287.

[21] T. Astrup, T. Fruegaard, T.H. Christensen, Recycling of plastic: accounting of greenhouse gases and global warming contributions, Waste Manag. Res. 27 (2009) 763-772 DOI:10.1177/0734242X09345868.

[22] A. Damgaard, A.W. Larsen, T.H. Christensen, Recycling of metals: accounting of greenhouse gases and global warming contributions, Waste Manag. Res. 27 (2009) 773-780 DOI:10.1177/0734242X09346838.

[23] A. Hadzic, N. Voca, S. Golubic, Life-cycle assessment of solid-waste management in city of Zagreb, Croatia, J. Mater. Cycles Waste Manag. 20 (2018) 1286-1298.

[24] C.Y. Hsu, Y.L. Lai, C.C. Chen, Y.T. Lee, Y.K. Tseng, Y.K. Lai, C.Y. Zheng, H.C. Jheng, Time sequence image analysis of positron emission tomography using wavelet transformation, Technol. Heal. Care 24 (2016) S393-S400 DOI:10.3233/THC-151105.

[25] J.K. Park, B.K. Kwon, J.H. Park, D.J. Kang, Machine learning-based imaging system for surface defect inspection, Int. J. Precis. Eng. Manuf. - Green Technol. 3 (2016) 303-310.

[26] L. Fan, Image processing algorithm of Hartmann method aberration automatic measurement system with tensor product model, Eurasip J. Image Video Process. 2019, (2019) 1-9.

[27] C.F.J. Kuo, W.C. Lo, Y.R. Huang, H.Y. Tsai, C.L. Lee, H.C. Wu, Automated defect inspection system for CMOS image sensor with micro multi-layer non-spherical lens module, J. Manuf. Syst. 45 (2017) 248-259.

[28] Z. Qiao, Y. Li, Z. Wu, J. Kou, Automatic puncture system based on NIR image and ultrasonic image, MATEC Web Conf. 108 (2017) 1-5 DOI:10.1051/matecconf/201710815002.

[29] W. Liu, D. Angelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, SSD: Single shot multibox detector, in: B. Leibe, J.Matas, N. Sebe, M. Welling (Eds.), Computer Vision -- ECCV 2016. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer International Publishing, Cham, Switzerland, 2016, pp. 21-37.

[30] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 1137-1149.