17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

Cut and paste curriculum learning with hard negative mining for point-of-sale systems

Jaechul Kim†, Xiaoyan Dai†, Yisan Hsieh†, Hiroki Tanimoto†, Hironobu Fujiyoshi‡
†Advanced Technology Research Institue, Minatomirai Research Center
Kyocera Corporation, Yokohama Japan, ‡Chubu University, Aichi Japan
{jaechul.kim.yb, xiaoyan.dai.cy, yisan.hsieh.ke, hiroki.tanimoto.hs}
@kyocera.jp†, fujiyoshi@isc.chubu.ac.jp‡

Abstract

Although point-of-sale (POS) systems generally use barcodes, progress in automation in recent years has come to require real-time performance. Since these systems use machine learning models to detect products from images, the models need to be retrained frequently to support the continual release of new products. Thus, methods for efficiently training a model from a limited amount of data are needed. Curriculum learning was developed to achieve this kind of efficient machine learning. However, curriculum learning in general has the problem that early learning progress is slow. Therefore, we developed a new curriculum learning method using hard negative mining to boost the learning progress. This method provides a remarkable learning effect through simple cut and paste. We test our method on various test data, and the proposed method is found to achieve better performance at the same learning epoch compared with conventional cut and paste methods. We expect our method to contribute to the realization of real-time and easy-to-operate POS systems.

1 Introduction

Over the past few years, many companies have attempted to develop artificial intelligence (AI) point-of-sale (POS) systems because of reduced working population and longer queue times. The problem with training models for AI POS is the cost of procuring the data. The addition of new items in stock would likely require thousands of diverse images with varied backgrounds and viewpoints to be curated and annotated with boxes. In order to commercialize AI POS systems, the data cost problem needs to be solved. Therefore, we attempted to solve this problem using an efficient cut and paste learning method. Current cut and paste object detection algorithms focus on how to obtain cut data and create paste data well[11], and how to generate cut data automatically in an unsupervised manner[10]. No efficient cut and paste learning methods have yet been proposed. In this work, we propose an approach to training models more efficiently by cut and paste curriculum learning using hard negative mining (HNM). Our proposed method offers an outstanding learning effect compared with conventional methods. We prepared various real POS evaluation data to test the proposed method, which achieved better performance than did the naïve cut and paste approach in the same learning epoch time. Our main contributions are as follows:

1. We propose cut and paste curriculum learning that is able to guide model training to convergence faster and achieve better minima.

2. We combined curriculum learning and HNM using generated evaluation data by cut and paste with the result of more efficient model training.

3. We evaluate our method using various real evaluation data sets.

2 Related work

Many methods have been proposed for improving the performance of training models with localizable object features[15, 2, 13]. Because cut and paste approach generate data with input corruptions and occulusions with other objects, this method can be said to be one of the above methods. Moreover, cut and paste offers advantages over other methods. One advantage is that it enables creation of data in various environments close to the real image using real cut data and a background image. Another is that it makes it possible to apply different processes to the objects and background image. For the background, this consists of adding noise data that are likely to be real for reducing false detection, and for objects, this consists of various augmentations. Just using simple cut and paste method well, we achieved good performance. However, AI POS systems are difficult to achieve good enough performance because of learning time constraints and limited data. Our method employs curriculum learning. The more difficult the problem, the more effective the curriculum learning is.

2.1 Cut and paste

The easiest and most straightforward approach to cut and paste was taken by Rao and Zhang[9]. They

simply took an object detection data set (VOC07 and VOC12), cut out objects according to their ground truth labels, and pasted them onto images with different backgrounds. Even this simple approach improved the performance of object detection. A similar but slightly less naïve approach to cutting and pasting was introduced by Dwibedi et al[3]., who used the same basic idea but with segmentation masks instead of just placing whole bounding boxes. We use the Rao and Zhang method for comparison with the proposed method (Figure 1).

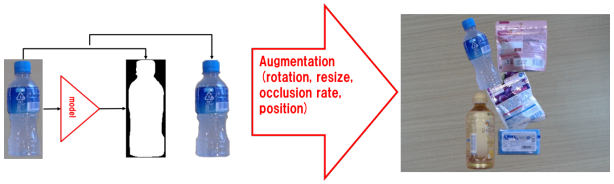

Figure 1. Conventional cut and paste method, in which objects are cut out according to their ground truth labels, and after random data augmentation are pasted onto a background image.

2.2 Curriculum learning

Curriculum learning is a concept that is borrowed from education systems in which students are presented with easier concepts first with the difficult gradually increasing. In FlowNet 2.0[1], simpler training data are fed into the network first, followed by more difficult data sets. This approach can help to expedite the training process while achieving better minima.

3 Methodology

Previous methods perform HNM by finding the false negative patch during model training[8, 7, 4]. Generally, this approach is greatly influenced by the quality and quantity of the prepared evaluation data set. To overcome this disadvantage, we propose cut and paste HNM. By employing cut and paste, we are able to generate as many evaluation data as are needed. In addition, the generated data are directly evaluated on the trained model, with false negative evaluation data acting as hard negative data for training the next epoch of the model. Further, in order for the model to learn the data more easily and efficiently, we propose cut and paste curriculum learning with HNM. This approach combines cut and paste HNM with cut and paste curriculum learning. In cut and paste curriculum learning, the model is trained using a training data set that gets more difficult depending on given cut and paste parameters (Figure 3). During cut and paste curriculum learning, we execute cut and paste HNM using an evaluation data set with the same level of difficulty (Figure 5).

3.1 Cut and paste curriculum learning

Cut and paste learning methods offer many parameters for generating data. Figure 1 shows how data are generated by cut and paste with various parameters. In order to apply curriculum learning to cut and paste, we need to design cut and paste parameters that can be adjusted for more difficult data (Figure 2, 3). In this

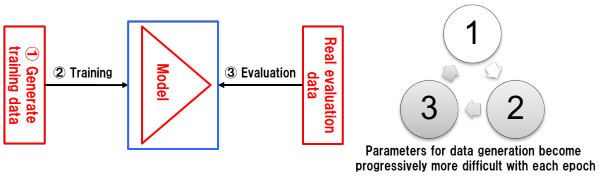

Figure 2. To apply curriculum learning to cut and paste, we set cut and paste parameters that get progressively more difficult with each epoch.

paper, we use the number of objects and lighting conditions as parameters for setting cut and paste difficulty. More difficult training data have a higher number of objects and worse lighting conditions (Figure 3, 4). In the case of conventional cut and paste learning, these parameters are set randomly.

Figure 3. All odd epoch data are shown by blue bars and consist of normal lighting condition data. All even epoch data are shown by orange bars and consist of poor lighting condition data. The number of objects in the data increases every 2 epochs from 1 to 10. The number of train data in each epoch is the same, with 10000 items of data.

3.2 Cut and paste HNM

The HNM method has been applied only to false negative patch in the past because of the limited evaluation data. Cut and paste can generate as many evaluation data as needed (Figure 4), which makes data-driven HNM possible. Moreover, this approach is extremely easy and requires no human effort to obtain a variety of evaluation data.

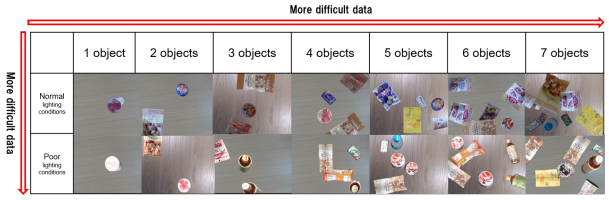

Figure 4. Example generated data set for curriculum learning. The data becomes more difficult the higher the number of objects in the data and the worse the lighting conditions.

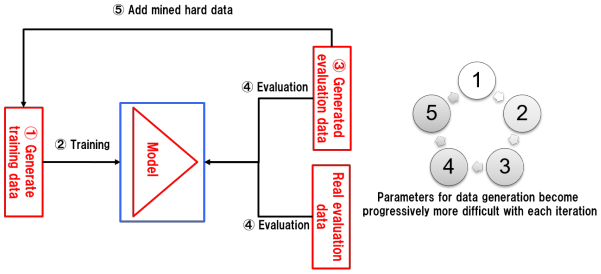

Figure 5. During cut and paste curriculum learning, hard negative data are mined from the generated evaluation data set for use in the model for training the next epoch.

4 Experiment

In cut and paste curriculum learning, we defined difficult data as having a higher number of objects in the data and worse lighting conditions. In order to compare the effectiveness of our proposed method with the conventional method, we prepared various difficult real evaluation data sets containing a number of objects varied up to 7 and normal and poor lighting conditions (Figure 6). As a result, we confirmed that our proposed method achieves better mean average precision(mAP) performance.

4.1 Experimental setup

The total number of data items in the real evaluation data set is 1401 and the number of classes is 20 (Figure 6). We evaluate the model using this data set for comparing performance (Figure 7). We have tested our method with Single Shot MultiBox Detector(SSD)[14]. We use the batch size of 32 on two RTX2080, the initial learning rate is 0.001 and SGD optimizer with the momentum of 0.9 and the weight decay of 0.0005.

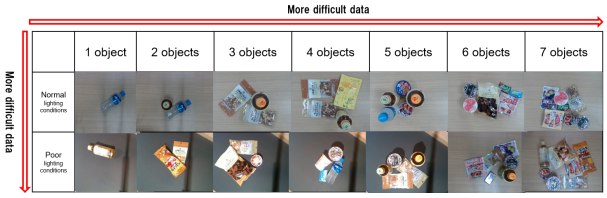

Figure 6. Real evaluation data set containing different numbers of objects and lighting conditions. The higher the number of objects and the worse the lighting conditions, the more difficult the data. Cut data also consist of poor and normal lighting conditions for training and evaluation data sets by the cut and paste method.

4.2 Experimental metric

mAP is a well-known metric for detection tasks that is used by many detection challenges[5, 6, 12].

$$mAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q} \tag{1}$$

where Q is the number of queries in the set and $AveP(q)$ is the average precision (AP) for a given query, q.

$$AP@n = \frac{1}{GTP} \sum_{k}^{n} P@k \times rel@k \tag{2}$$

where GTP refers to the total number of ground truth positives, n refers to the total number of documents of interest, $P@k$ refers to the precision@k, and $rel@k$ is a relevance function. The relevance function is an indicator function that equals 1 if the document at rank k is relevant and equals 0 otherwise.

4.3 Result

We compared our proposed method with conventional cut and paste in terms of the best mAP during model training. The proposed method performed better than did the conventional method (Figure 7).

5 Ablation study

In order to show the effectiveness of our proposed cut and paste curriculum learning with HNM, we performed an ablation study comparing our method against the conventional method. Although the learning times of the cut and paste method and cut and paste curriculum learning were the same (Figure 8), the curriculum learning approach gaves better performance than did simple cut and paste (Figure 9). Our proposed method, cut and paste curriculum learning

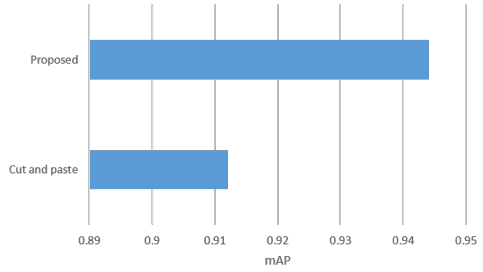

Figure 7. Our proposed method achieves better performance than does the conventional method.

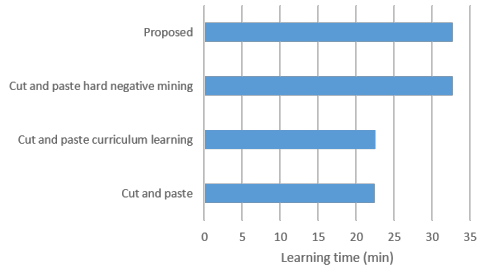

Figure 8. The average learning times of cut and paste and curriculum cut and paste learning were the same, as were the learning times of the proposed method and cut and paste HNM. Cut and paste curriculum learning with HNM and cut and paste HNM took longer because of the evaluation processing time.

with HNM, performed better compared with cut and paste curriculum learning (Figure 9). However, the learning time was longer owing to the evaluation process not required by other methods (Figure 5). An experiment was conducted to evaluate the performance of the proposed method according to the difficulty of the evaluation data set. Figure 10 shows the performance of each method for the each levels of validation data set difficulty (Figure 6). We evaluated each data set in terms of the trained model. Odd validation data sets consisted of data with normal lighting conditions and even validation data sets consisted of data with poor lighting conditions. The results confirm that the proposed method performs well over a variety of data difficulty conditions (Figure 10). As Figure 10 shows, the results for the even numbered epochs exhibited a large deviation owing to the existence more difficult data even within the same poor lighting conditions. As Figure 6 shows, the data in the second row and second column lost many more features compared with data in the second row and the sixth column owing to light reflections.

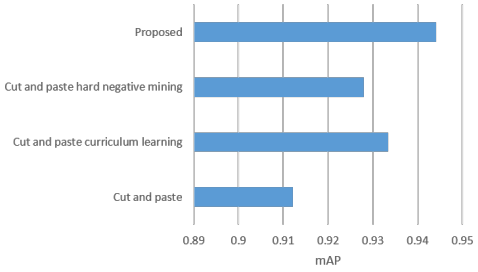

Figure 9. Results of each cut and paste method. Cut and paste curriculum learning with HNM had better performance compared with the other methods.

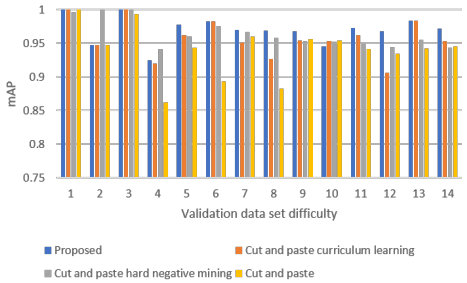

Figure 10. Various evaluation data sets were prepared. The evaluation data were divided into difficulty depending on the number of objects and lighting conditions. The higher the number of objects, the more difficult the evaluation data set. Odd validation data sets consisted of data with normal lighting conditions and even validation data sets consisted of data with poor lighting conditions.

6 Conclusion

In this work, we proposed cut and paste curriculum learning with HNM to enable training of models more efficiently compared with conventional cut and paste learning. In addition, we studied the effectiveness of cut and paste curriculum learning and HNM with cut and paste with various real POS evaluation data sets. Our study reveals that curriculum learning and HNM effectively train the model by the cut and paste approach, and experimental results show that cut and paste curriculum learning with HNM achieves the best performance.

References

[1] Eddy I. Nikolaus M. Tonmoy S. Margret K. Alexey D. Thomas B. Flownet 2.0: Evolution of optical flow estimation with deep networks. *CVPR*, 2017.

[2] Zhang H. Cisse M. Dauphin Y. N. Lopez Paz D. mixup: Beyond empirical risk minimization. *ICLR*, 2018.

[3] Hebert M. Dwibedi D, Misra I. Cut, paste and learn: Surprisingly easy synthesis for instance detection. *ICCV*, 2017.

[4] Jin S.Y. RoyChowdhury A. Jiang H. Singh A. Prasad A. Chakraborty D. Learned-Miller E. Unsupervised hard example mining from videos for improved object detection. *ECCV*, 2018.

[5] Eslami S. van Gool L. Williams C. Winn J. Zisserman A. Everingham, M. The pascal visual object classes challenge: A retrospective. *IJCV*, 2015.

[6] Van Gool L. Williams C.K.I. Winn J. Zisserman A Everingham, M. The pascal visual object classes (voc) challenge. *IJCV*, 2010.

[7] Shrivastava A. Abhinav G. Ross G. Training region-based object detectors with online hard example mining. *CVPR*, 2016.

[8] Bucher M. Stephane H. Frederic J. Hard negative mining for metric learning based zero-shot classification. *ECCV*, 2016.

[9] Rao J. Zhang J. Cut and paste: Generate artificial labels for object detection. *ICVIP*, 2017.

[10] Remez T. Huang J. Brown M. Learning to segment via cut and paste. *ECCV*, 2018.

[11] Koturwar S. Shiraishi S. Iwamoto K. Robust. Multi object detection basedon data augmentation with realistic image synthesis for point of sale automation. *AAAI*, 2019.

[12] Deng J. Su H. Krause J. Satheesh S. Ma S. Huang-Z. Karpathy A. Khosla A. Bernstein M. Berg A. Fei-Fei L. Russakovsky, O. Imagenet large scale visual recognition challenge. *IJCV*, 2015.

[13] DeVries T. Taylor G. W. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

[14] D. Erhan C. Szegedy W. Liu, D. Anguelov and S. Reed. Ssd: single shot multibox detector. *CoRR abs/1512.02325*, 2015.

[15] Yun S. Han D. Oh S. Chun S. Choe J. Yoo Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. *ICCV*, 2019.