ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 19-04-04

j.inst.Korean.electr.electron.eng.Vol.23,No.4,1134∼1139,December 2019
http://dx.doi.org/10.7471/ikeee.2019.23.4.1134

KNN-Based Automatic Cropping for Improved Threat Object Recognition in X-Ray Security Images

Joanna Kazzandra Dumagpi*★, Woo-Young Jung*, Yong-Jin Jeong*

Abstract

One of the most important applications of computer vision algorithms is the detection of threat objects in x-ray security images. However, in the practical setting, this task is complicated by two properties inherent to the dataset, namely, the problem of class imbalance and visual complexity. In our previous work, we resolved the class imbalance problem by using a GAN-based anomaly detection to balance out the bias induced by training a classification model on a non-practical dataset. In this paper, we propose a new method to alleviate the visual complexity problem by using a KNN-based automatic cropping algorithm to remove distracting and irrelevant information from the x-ray images. We use the cropped images as inputs to our current model. Empirical results show substantial improvement to our model, e.g. about 3% in the practical dataset, thus further outperforming previous approaches, which is very critical for security-based applications.

Key words : *KNN, automatic cropping, x-ray security images, threat object recognition, anomaly detection*

Ⅰ. Introduction

Advances in computer vision algorithms over the years have led to the realization of several practical applications especially for security systems industries where detection of threat objects is of utmost importance. For this reason, several studies have been done on the automatic detection of threats in x-ray security images [1-4]. However, previous approaches were designed under ideal conditions that do not reflect the practical setting. For instance, Fig. 1 shows sample images from the then-largest publicly available x-ray security image dataset [5]. This dataset only consists of more than 19, 400 single energy x-ray images, which are also carefully taken in a laboratory ensuring that the objects of interest are in focus. Yet, security systems nowadays commonly use multi-energy x-rays that allow pseudo-coloring. Moreover, prior studies do not consider a realistic distribution between images with threats and those without.

Fig. 1. Sample images from the GDXray dataset [5].

* Dept. of Electronics and Communications Engineering, Kwangwoon University
★ Corresponding author
E-mail : jkmdumagpi@kw.ac.kr, Tel : +82-02-940-5551
※ Acknowledgment
This work was supported by Kwangwoon University and by the MISP Korea, under the National Program for Excellence in SW(2017-0-00096) supervised by IITP.
Manuscript received Oct. 30, 2019; revised Nov. 27, 2019; accepted Dec 10, 2019.

Consequently, Miao et al. [6] recently released the now-largest publicly available x-ray security image dataset, consisting of more than 1 million x-ray images, which closely resembles real-world x-ray security images, as shown in Fig. 2. They also proposed an approach to improve classification in a practical dataset by refining the images through a hierarchal approach. Further analysis of this new dataset reveals two important aspects of threat detection in the practical setting. First, the data distribution of x-ray security images is highly skewed towards images that do not contain threats. This is also known as the class imbalance problem. Second, x-ray security images are visually complicated because of overlapping, heavy clutter, and variations in scale, viewpoint, etc. The first aspect makes a traditional classifier more biased to always predict the majority class. This is a problem especially when the misclassification cost of the minority class is considerably higher than the majority class, as is the case for threat object classification [7]. discusses the methods that can be used to combat this issue. In our previous work, we attempted to solve this issue by implementing a unique ensemble of generative and discriminative models, shown in Fig. 3. The GAN-based anomaly detector is used to suppress the surge of false positives in the practical dataset, which is a dataset resembling the realistic distribution, induced by training our classifier on an non-practical dataset, consisting of an unrealistic balance between images with and without threats [8]. The second aspect causes the model to be distracted by irrelevant information. Fig. 4 shows that an x-ray image without too much clutter gives stronger information about the target object compared to a more cluttered x-ray image. This study is focused on alleviating the effects of the second aspect.

Fig. 2. Sample images from the SIXray dataset [6].

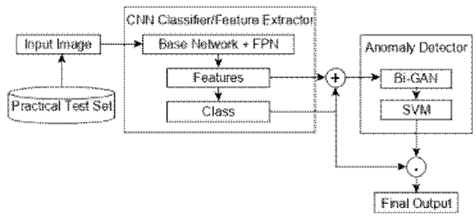

Fig. 3. Our classification network [8].

Photo cropping is commonly defined as the process of ridding an image of unnecessary details by cutting out the outer portion of the target areas [9]. Previous approaches have successfully automated this process for natural images [10-11]. For x-ray images, Imura et al. [12] used adaptive binarization to automatically crop chest x-rays. They set a threshold value based on the analysis of the frequency distribution of the pixel intensities in a local area, which is unique to each x-ray image. In this paper, we propose to use automatic cropping to visually improve the x-ray images. We implement KNN to emphasize the target objects and tightly crop around the isolated area. By constraining the images, we also reduce the information loss incurred when the inputs are pre-processed prior to training. We improve our previous work by using the cropped images as inputs to our model. Empirical results show that our model performs better with cropped input images and significantly outperforms previous approaches.

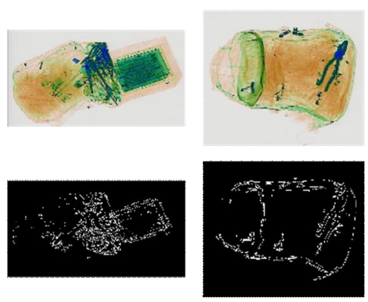

Fig. 4. Edge detection on a (right) more cluttered x-ray image versus (left) on a less cluttered image.

II. Proposed Method

We found from our previous work that our model gives confident predictions when the input image does not consist of too much air space. On the other hand, the model is uncertain when the input consists of too much air space. Air space is the portion of the image void of any object. Table 1 shows the confidence scores of our model for the test images shown in Fig. 5. Confidence scores tells of how certain our model is about the presence of a particular threat, wherein higher values mean higher certainty. Images A and B are sampled from the negative images, which are images that do not contain threats, while C and D are sampled from the positive images, which are images that contain at least one threat. Though both C and D contain pliers and wrench, the air space in D have weakened the model's confidence.

Table 1. Class confidence scores for sample images in Fig. 5.

Image	Gun	Knife	Wrench	Pliers	Scissors
A	1.54e-07	1.97e-06	2.70e-06	4.60e-03	3.10e-06
B	1.26e-08	1.38e-07	9.46e-10	7.58e-06	5.14e-08
C	9.96e-05	1.53e-06	**9.99e-01**	**1.00e+00**	2.44e-07
D	3.43e-06	2.13e-06	**2.65e-02**	**9.47e-02**	1.27e-06

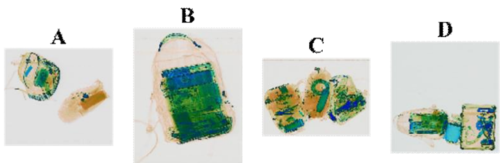

Fig. 5. Sample of test images from the SIXray dataset.

Aside from unnecessary air space filling up most of the image, much of the relevant information is also lost when transforming such images to be suitable as inputs to the CNN classifier. For instance, the default input size of a ResNet model [13] is 224x 224. Fig. 6 demonstrates how transforming images with too much air space compare to images that have less air space. It can be observed that images with too much air space tend to be more warped, hence distorting information. Also, the air space contains no information about the target objects but consumes most of the space, thus distracting the model from obtaining relevant information from the image. From this observation, we hypothesize that cropping the air space out of the images would improve the current performance of our classification model.

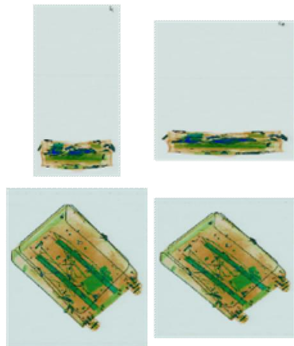

Fig. 6. Transforming an image with too much air space (upper) versus one with less air space (lower).

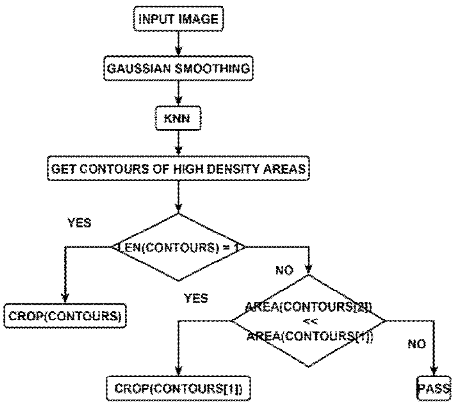

Fig. 7. The flow of KNN-based automatic cropping algorithm.

X-ray images are unique in that the intensity level of each pixel allows for a rough estimate of the material characteristics or, in the case of overlapping objects, how many stacks of objects are present [14]. Denser objects tend to have higher pixel intensity. Metal objects, such as guns, knives, wrench, pliers, and scissors, are denser compared to non-metals, such as clothes, papers, and food. Since the target threats are mostly metals, we can separate them from the rest of the image by observing the pixel intensity levels. Once isolated, we can tightly crop around the areas and disregard the information outside the boundaries that are more likely irrelevant to the task at hand. Fig. 7 shows the flow of the automatic cropping algorithm.

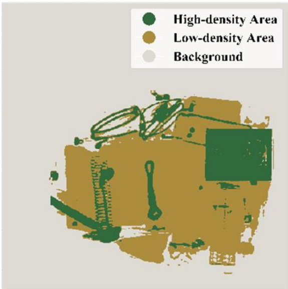

Fig. 8. Pixels grouped into three classes using KNN.

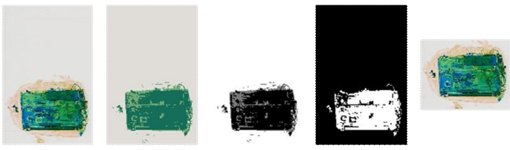

Fig. 9. Crop transformation progression (from left to right).

KNN is used to separate the image into three parts namely, background, low-density areas, and high-density areas, as illustrated in Fig. 8. We get the contours of the high-density areas and sort them based on decreasing area size. Before the final crop, we consider three possible cases. First, there is only one region of high-density pixels and we tightly crop around the region immediately. Second, there is more than one region of high-intensity pixels, but one region is significantly larger than the other region. In such case, we crop around the largest region and assume that the other regions are artifacts or noise. Finally, when there is more than one high-density region and the area of the regions do not differ significantly, then we do not crop the image so as not to risk removing important information. Fig. 9 shows the transformation that the image undergoes during automatic cropping.

III. Experiment Results

We evaluate the effect of automatic cropping using our previous model [8] and the SIXray dataset [6]. SIXray is divided into three subsets that exhibit varying levels of class imbalance. We test on all the subsets using ResNet50 [13] as the base model. Moreover, we compare our results to [6] and to our previous work [8]. To our knowledge, these are the only published works so far that has tackled the issue of threat classification in x-ray images in a practical setting.

For the evaluation metric, we adapt the mean average precision (mAP) that is used in the PascalVOC image classification task [15]. This metric is commonly used to account for the class imbalance by calculating the mean of the average precisions computed for each class. Higher mAP means better performance. This experiment is conducted on a single machine with Intel Xeon processor and NVIDIA Tesla T4 GPU using Pytorch [16] as the deep learning framework. Since our model is already optimized to handle class imbalance, we expect that feeding constrained images will further improve the model performance.

Table 2. Classification mAP (%) for SIXRay10.

Method	Gun	Knife	Wrench	Pliers	Scissors	mean
Plain CNN [13]	90.64	87.82	63.62	84.8	57.35	76.85
CNN+CHR [6]	87.55	86.38	69.12	85.72	60.91	77.94
CNN+GBAD [8]	92.3	84.76	75.48	84.22	65.18	80.39
Proposed method	90.86	89.42	78.02	83.36	66.09	**81.55**

Table 3. Classification mAP (%) for SIXRay100.

Method	Gun	Knife	Wrench	Pliers	Scissors	mean
Plain CNN [13]	84.75	77.92	28.49	50.53	19.39	52.22
CNN+CHR [6]	82.64	79.6	41.19	58.02	27.89	57.87
CNN+GBAD [8]	82.85	77.66	46.05	52.94	33.21	58.54
Proposed method	83.68	78.36	50.12	52.85	42.12	**61.43**

Table 4. Classification mAP (%) for SIXRay1000.

Method	Gun	Knife	Wrench	Pliers	Scissors	mean
Plain CNN [13]	74.19	59.82	16.03	16.59	2.87	33.90
CNN+CHR [6]	73.43	61.32	18.88	12.32	19.03	37.00
CNN+GBAD [8]	72.59	70.02	18.45	20.82	21.01	40.58
Proposed method	74.87	71.38	16.41	17.08	25.99	**41.14**

Table 2 shows the comparison of the model performances in the SIXray10 subset. This subset is the least imbalanced subset with positive images comprising only 10% of the entire subset. Consequently, Table 3 also demonstrate further improvements in the SIXRay100 subset, considered to be the one that closely emulates the distribution in the practical setting [6], with the positive images representing 1% of the entire subset. Although the performance of our model for individual categories is generally improved, it can be observed that there is significant improvement for the scissors' category. This is because scissors, compared to other targets, are relatively smaller and can be easily lost under heavy clutter, which is only worsened by the presence of huge air space. By cropping around the potential locations of scissors, we help the model gain stronger information about the target, thereby increasing its predictive power for the category. Finally, Table 4 also shows further improvement for the SIXRay1000 subset, which exhibits extreme case of class imbalance where the positive images are just 0.1% of the entire subset. The average throughput of the algorithm is calculated to be 21 images per second.

Ⅳ. Conclusion

In this paper, we identified two important aspects of practical x-ray security image datasets namely, the class imbalance problem and visual complexity. We propose to use a KNN-based automatic cropping algorithm to alleviate the problems caused by the visual complexity brought by too much irrelevant information in the input image. This is an extension of our previous work wherein we have dealt with the class imbalance problem. Empirical results show that using cropped images further improved the performance of our model, thereby expanding the performance gap between our approach and the previous approaches, which is very critical for such security-focused practical applications. This is only a steppingstone for solving the visual complexity in x-ray images. Further research includes developing more sophisticated transformation and manipulation algorithms to push the performance of our current model.

References

[1] V. Riffo and D. Mery, "Automated Detection of Threat Objects Using Adapted Implicit Shape Model," *IEEE Trans. Syst., Man, Cybern. Syst.,* vol.46, no.4, 2016. DOI: 10.1109/TSMC.2015.2439233

[2] D. Mery, E. Svec and M. Arias, "Object Recognition in X-ray testing Using Adaptive Sparse Representations," *Journal of Nondestructive Evaluation,* Vol.35, No.45, pp.9, 2016.
DOI: 10.1007/s10921-016-0362-8

[3] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra and S. Banarjee, "Modern Computer Vision Techniques for X-Ray Testing in Baggage Inspection," *IEEE Trans. Syst., Man., Cybern, Syst.,* vol.47, no.4, pp.682-692, 2017.
DOI: 10.1109/TSMC.2016.2628381

[4] S. Ackay, M. E. Kundegorski, C. G. Willcocks and T. P. Breckon, "Using Deep Convolutional Neural Network Architectures for Object Classification and Detection within X-Ray Baggage Security Imagery," *IEEE Trans. Inf. Forensics Security,* vol.13, no.9, pp.2203-2215, 2018.
DOI: 10.1109/TIFS.2018.2812196

[5] D. Mery, V. Riffo and U. Zscherpel et al., "GDXray: The Database of X-ray Images for Nondestructive Testing," *J Nondestruct Eval,*

2015. DOI: 10.1007/s10921-015-0315-7
[6] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao and Q. Ye, "SIXray: A Large-scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit (CVPR),* 2019.
[7] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.,* vol.21, no.9, pp.1263-1284, 2009.
[8] J. K. Dumagpi, W. Y. Jung and Y. J. Jeong, "A New GAN-based Anomaly Detection (GBAD) Approach for Mult7i-Threat Object Classification on Large-Scale X-Ray Security Images," *IEICE Trans. Inf. & Syst.,* vol.E103-D, no.2,. 2020.
[9] J. Chen, G. Bai, S. Liang, and Z. Li, "Automatic Image Cropping: A Computational Complexity Study," *IEEE Conf. Comput. Vis. Pattern Recognit (CVPR)*, 2016.
DOI: 10.1109/CVPR.2016.61
[10] M. Nishiyama, T. Okabe, Y. Sato and I. Sato, "Sensation-based photo cropping," *Proc. 17th ACM Int. Conf. Multimedia*, 2009.
DOI: 10.1145/1631272.1631384
[11] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," *Proc. SIGCHI Conf. Human Factors Computing Systems,* 2006.
DOI: 10.1145/1124772.1124886
[12] M. Imura, Y. Tabata, R. Ishigaki, Y. Kuroda, Y. Uranishi and O. Oshiro, "Automatic Cropping Method for Chest Radiographs Based on Adaptive Binarization," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 2013. DOI: 10.1109/EMBC.2013.6611042.
[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.
[14] D. Mery, "Computer Vision for X-Ray Testing," Springer International Publishing, 2015.
DOI: 10.1109/CVPR.2016.90
[15] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classificaiton (VOC) Challenge," *Int. J. Comput. Vis.,* vol.88, no.2, pp.303-338, 2010.
DOI: 10.1007/s11263-009-0275-4
[16] A. Paszke, S. Gross, S. Chintala, G. et al., "Automatic Differentiation in Pytorch," *in NIPS Autodiff Workshop,* 2017.

BIOGRAPHY

Joanna Kazzandra Dumagpi (Member)

2017 : BS degree in Electronics Engineering, Ateneo de Davao University
2018~ : Course of MS degree in Electronics and Comm. Engineering, Kwangwoon Univ.

Woo-Young Jung (Member)

1991 : BS degree in Electronics and Comm. Engineering, Kwangwoon Univ.
1996 : MS degree in Electronics and Comm. Engineering, Kwangwoon Univ.
Current : PhD degree in Elec. and Comm Engineering, Kwangwoon Univ.
2012~current : CTO, Innovative Research and Analysis, Inc., Philippines
2013~current : CTO, EPSOLUTE Co. Ltd., Korea

Yong-Jin Jeong (Member)

1983 : BS degree in Control and Instrumentation Engineering, Seoul National University
1995 : MS, PhD degree in Electronics and Computer Engineering, University of Massachusetts, Amherst
1983~1989 : Research Engineer, ETRI
1995~1999 : Chief Research, Samsung Electronics
1999~current : Professor, Dept. of Electronics and Communications Engineering, Kwangwoon Univ.