Accuracy Consistency of Object Detection With Contrast Reduction by Pixel Value Limitation

Kein Yamada*, Takahiro Shindo†, Taiju Watanabe‡, and Hiroshi Watanabe§

*School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan
E-mail: stslm738.ymd@toki.waseda.jp

†Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan
E-mail: taka_s0265@ruri.waseda.jp

‡Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan
E-mail: lvpurin@fuji.waseda.jp

§Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan
E-mail: hiroshi.watanabe@waseda.jp

***Abstract*—This paper provides a technique of retaining the accuracy in object detection while reducing the entropy of the image. For the past few years, the topic of using artificial intelligence (AI) for image and video processing has been very prominent. With AI gradually being more viable for image and video tasks, computer vision is becoming more fitting for such data analysis. If the entropy of an image is small, images can be expressed in simple form, increasing efficiency for transmitting and saving files. As one approach, we attempted and found an effective method of entropy reduction for object detection. In our method, the contrast of the image is changed to fit in a certain pixel value range, resulting in the repression of entropy. Ultimately, an image with less entropy is achieved while keeping the accuracy of object detection consistent.**

***Keywords*—image processing, object detection, contrast reduction, YOLOv5, entropy reduction**

I. Introduction

Artificial intelligence (AI) has been showing growth and popularity in many tasks of image and video processing. Following the improvement of AI for image and video recognition, computer vision is also becoming more capable of image and video processing. As a result, the combination of AI with object detection [1] has been trending in recent years. There are many real-life applications of object detection such as surveillance system [2], face recognition [3], autonomous driving [4], medical anomaly detection [5], animal classification [6], and more. Thus, images and videos are starting to be more applicable for these purposes.

Currently, most of these detection tasks utilize clear images and videos with high quality information content. Since these media files are presented in high definition (HD), there is a room of improvement for transmission. For instance, images with high quality information content directly means to have large entropy values. As a possible refinement, we can attempt to decrease the image entropy. If the information of images can be reduced to satisfy the AI detection tasks, an efficient transmission can be achieved. This is because we can have a simpler expression of bitmap for images, due to a less complex binary data expression. Furthermore, it has been revealed by Choi and Bajić that videos required for computer vision can be smaller, compared to ones for human [7]. In other words, Choi and Bajić proved that videos do not have to be provided in HD for computer vision related tasks. Since low quality videos can be practical for video processing, theoretically, computer vision should be able to process images with small information content as well. By reducing the information of the image, we attempt to restrain the entropy.

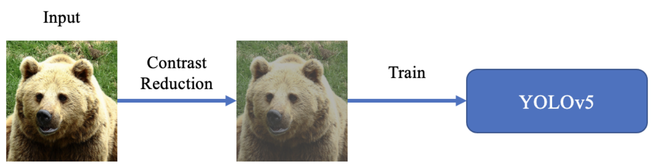

Fig. 1. Overview of contrast reduction and training process. Images that are processed by contrast reduction are used for validation as well. The image of the bear used in this figure is from the COCO2017 Val dataset, provided by Common Objects in Context (COCO) [11].

Entropy in image processing measures the level of disorder within an image, which is determined by the frequency of each pixel value. In this research, the entropy is suppressed by placing a limit on pixel values that could potentially appear. By narrowing down the pixel values to a certain range, the contrast of the image is reduced. Hence, we decided to adopt the alternation of pixel value range as an entropy reduction method, which will be denoted as contrast reduction in this paper.

II. Related Work

YOLOv5 [8] is a modified object detection model, belonging to the You Only Look Once (YOLO) family [9]. YOLOv5 is built to have a quick process of training and

979-8-3503-4017-4/23/$31.00 ©2023 IEEE

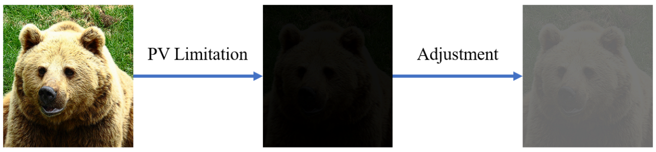

Fig. 2. An overview of our proposed contrast reduction method. In this figure, contrast reduction is being applied to the original image of a bear. The pixel value range is 16 in this example.

is highly suitable for users that insist on installing computer vision technologies to their proposed mechanism. The model is provided with 5 different pre-trained weights or training checkpoints called nano (YOLOv5n), small (YOLOv5s), medium (YOLOv5m), large (YOLOv5l), and extra-large (YOLOv5x). If one insists on acquiring the result quickly, regardless of its accuracy and precision, YOLOv5n should be used since it has the smallest parameter among the 5 weights, giving the shortest detection process. On the other hand, if the emphasis is placed more on the detection evaluation than the processing time, YOLOv5x with the largest parameter should be considered.

In terms of model structure, YOLOv5 consists of 3 components called Backbone, Neck, and Head, making it a single-stage object detector. As a feature of single-stage object detector, the neural network is passed through once and predicts the object's coordinates, size, and classification simultaneously. YOLOv5 utilizes the same head as the one configured for YOLOv3 [10]. However, YOLOv5 has a completely different backbone and neck. For the backbone, YOLOv5 has the CSP-Darknet53, which is an upgrade of the Darknet53 backbone from YOLOv3, where Cross Stage Partial network strategy is added to make the inference speed better. For the neck, YOLOv5 uses a modified version of Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet). Due to this change, the speed of the network is raised, making YOLOv5 a reasonable object detection model to wield.

III. Proposed Method

In this section, we propose a method of lowering the entropy by reducing the contrast of the image. In this research, images are treated as 3 channel images with a bit depth of 8. Since these images have a bit depth of 8, pixel values range from 0 to 255. Depending on an assigned pixel value, a pixel depicts a certain intensity of the color shade. A collection of these pixel values compose the digital image, which acts as a form of information content. The process of contrast reduction requires the modification in pixel values. Contrast reduction method is expressed by the following equation,

$$O(w,h) = \left\lfloor \frac{\alpha}{256} I(w,h) \right\rfloor + \left\lfloor \frac{256-\alpha}{256WH} \sum_{w=1}^{W} \sum_{h=1}^{H} I(w,h) \right\rfloor \quad (1)$$

where the values are changed by the limitation of the pixel value range. I(w, h) and O(w, h) represent the pixel values of the input image and output image. Symbol α refers to the target range of the pixel value. For example, in the case of limiting the pixel value range to 128, α should be equal to 128 as well. W and H stand for the width and height of the image. The portion $\lfloor \frac{\alpha}{256} I(w,h) \rfloor$ corresponds to the pixel value limitation step. This is done by reducing the image pixel value inputs by a certain scale factor, which is dependent on the target pixel value range. The other portion $\left\lfloor \frac{256-\alpha}{256WH} \sum_{w=1}^{W} \sum_{h=1}^{H} I(w,h) \right\rfloor$ corresponds to the adjustment of shifting the range by applying the mean. Moreover, this expression aids in adjusting the limit range to optimally use the pixel values close to the original image. This is evident, especially for the example shown in Fig. 2. Truncation is applied to both of these portions to avoid the inaccuracy of range limitation. If the two portions of the sum is not truncated, it is very likely to get an image resembled by values slightly outside of the pixel value range.

Equation (1) is actually a simplified version of the original equation that is utilized for single-channeled images, such as grayscale images. The equation for images with multiple channels is

$$O(w,h) = \left\lfloor \frac{\alpha}{256} I(c,w,h) \right\rfloor + \left\lfloor \frac{256-\alpha}{256CWH} \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} I(c,w,h) \right\rfloor \quad (2)$$

where c represents the selected channel (R, G, or B for RGB images) and C is the total number of channels, which will be 3 for this research. With (2), an image can be successfully converted to have a smaller range of pixel values.

The range downscaling directly correlates to the reduction of entropy within the images. This is because the randomness of the image's colors per pixel gets smaller with range limitation. For instance, if the range limit is shrunk to express 128 pixel values, usable colors for image representation is halved. With fewer colors used for expression, the color intensity or the shade within the images become less dynamic. Therefore, the texture of an image will be more simplified, resulting in an entropy decrease.

IV. Experiment

A. Evaluation method

To validate our proposed method, we examined the relationship between the accuracy of object detection and the entropy of the images. In Fig. 1, the overview of the experiment is demonstrated to give a clear idea about the experimental flow. The datasets used for training and validation are COCO 2017 Train Images and COCO2017 Val Images [11]. As mentioned in the proposed method, images of the datasets are converted by contrast reduction to pixel value ranges of 256, 128, 64, and 32, as shown in Fig. 3. For more evaluation data, we also converted the images to pixel value ranges of 16, 8, and 4. These specific numbers are selected for the range limitations because they correspond to bits expression of

(a) (b) (c) (d)

Fig. 3. Visualization of images in different pixel value range, (a) range of 256; (b) range of 128; (c) range of 64; (d) range of 32.

8bits, 7bits, 6bits, 5bits, 4bits, 3bits, and 2bits. For the object detection model, YOLOv5 is utilized for detecting objects in the COCO2017 Train and COCO2017 Val dataset. As for the weight, we used the pre-trained weight YOLOv5m, taking into account that this weight provides a moderate detection accuracy and processing time. Additionally, images of the converted datasets are utilized for the process of object detection. For each pixel value range, YOLOv5m has been trained on 100 epochs. After completing the train on YOLOv5m, the output results are examined and the best values are used for comparison with different ranges of pixel values. For the evaluation, we will depict the mean Average Precision (mAP) values, split into mAP50-95, mAP50, and mAP75. The entropy of an image is calculated by the concept of Shannon entropy. The equation for entropy is represented as

$$H = -\sum_{i=0}^{255} p_i \log_2 p_i. \tag{3}$$

In the equation above, H is the entropy of an image given as an output. Letter i refers to a specific pixel value or the shade of the color used in the image. Furthermore, p_i refers to the probability at which the pixel value appears in a specific image. Since (3) only works for images with single channel, the RGB images are converted to grayscale images before the calculation. Using (3), the entropy is calculated for each image in the validation dataset. Since the COCO2017 Val dataset includes 5000 images, the mean of 5000 entropy values is used for comparison and measurement of the results.

B. Results and discussion

In Table I, results of the mAP values and average entropy are presented with the corresponding pixel value range, abbreviated as PVR. Default is equivalent to pixel value range of 256, which is the maximum range of color shades in our experiment. This will also be the benchmark for comparing the obtained results of other pixel value range. Starting with the average entropy, it can be observed that the value decreases along with the stricter limits on the pixel value range. Comparing the default to the pixel value range of 32, we achieve a 40.6% decrease in the average entropy. Results also show that entropy decreased by 77.9% from default to pixel value range of 4. Looking at the accuracy evaluation, Table I indicates that values for mAP50-95 stayed consistent throughout the fall in pixel value range, until the range of 32. Afterwards, the values for mAP50-95 keep decreasing to a point where the accuracy is insignificant, despite the low entropy. Similarly, values for mAP50 and mAP75 held great stability in accuracy, barely changing in numbers while the pixel value range is above or equal to 32. Comparing the mAP values of pixel value range 32 to the default benchmark, mAP50-95 decreased by 0.45%, mAP50 decreased by 0.31%, and mAP75 decreased by 0.62%. While there is a slight decrease, the fall in mAP values are all under 1%, which is acceptable for accuracy consistency. Our experiment successfully conveyed that the pixel value range of an image can be scaled down to 32, while providing an object detection performance relatively close to the original. Hence, the output results of Table I revealed that object detection accuracy could be maintained, despite the application of contrast reduction to reduce the image entropy.

TABLE I
OBJECT DETECTION ACCURACY PERFORMANCE AND ENTROPY OF EACH PIXEL VALUE RANGE

PVR	Entropy	mAP50-95	mAP50	mAP75
Default	7.216	0.445	0.637	0.484
128	6.232	0.447	0.640	0.485
64	5.251	0.447	0.639	0.484
32	**4.285**	**0.443**	**0.635**	**0.481**
16	3.346	0.438	0.630	0.474
8	2.448	0.423	0.614	0.456
4	1.594	0.385	0.573	0.411

V. CONCLUSION

The outcomes of this research solidifies the use of contrast reduction as a reliable entropy reduction method for omitting excess textures within the image without greatly decreasing the performance of object detection. Our research also confirms that the pixel value range of images can be dropped to 32 for object detection. Therefore, contrast reduction is effective for computer vision, especially if the pixel value range is within 256 to 32 for RGB images with a maximum of 256 shades per channel. Hence, the entropy of an image can be decreased to be applied for computer vison-based image recognition.

For future research, we plan on recovering the images processed by contrast reduction to their original form. We see this as a possibility because some colors are preserved for images, depending on the pixel value range. If image restoration works, this research will become an application for human and machine purposes. For an alternate plan, testing this experiment on video file formats is also notable. Video files use a different color space from image files, the common one being YCbCr. Theoretically, if the YCbCR color space images provide experimental results similar to ones from RGB color space images, the method of contrast reduction can also be applied for video processing. Thus, these suggestions may be worth investigating to widen the use of contrast reduction for computer vision.

ACKNOWLEDGMENT

These research results were obtained from the commissioned research (No.05101) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010. DOI: 10.1109/TPAMI.2009.167.

[2] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and tracking system for video surveillance system," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3981–3996, 2021, ISSN: 1573-7721. DOI: 10.1007/s11042-020-09749-x.

[3] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998. DOI: 10.1109/34.655648.

[4] E. Arnold, O. Y. Al-Jarrah, M. Dianati, *et al.*, "A Survey on 3D Object Detection Methods for Autonomous Driving Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3782–3795, 2019. DOI: 10.1109/TITS.2019.2892405.

[5] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep Learning for Medical Anomaly Detection – A Survey," vol. 54, no. 7, 2021, ISSN: 0360-0300. DOI: 10.1145/3464423.

[6] B. Xu, W. Wang, G. Falzon, et al., "Livestock classification and counting in quadcopter aerial images using Mask R-CNN," *International Journal of Remote Sensing*, vol. 41, no. 21, pp. 8121–8142, 2020. DOI: 10.1080/01431161.2020.1734245.

[7] H. Choi and I. V. Baji´c, "Scalable Image Coding for Humans and Machines," *IEEE Transactions on Image Processing*, vol. 31, pp. 2739–2754, Mar. 2022. DOI: 10.1109/TIP.2022.3160602.

[8] G. Jocher, A. Chaurasia, A. Stoken, *et al.*, *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation*, version v7.0, Nov. 2022. DOI: 10.5281/zenodo.7347926.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016. DOI: 10.1109/CVPR.2016.91.

[10] J. Redmon and A. Farhadi, "Yolov3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[11] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft COCO: Common Objects in Context," *Computer Vision – ECCV 2014*, vol. 8693, pp. 740–755, Sep. 2014. DOI: 10.1007/978-3-319-10602-1_48.