A Study on Few-shot Object Detection for Warships Based on Data Generation Using Image Outpainting

SungWon Moon
Content Research Division
Electronics and Telecommunications Research Institute(ETRI)
Daejeon, Republic of Korea
moonstarry@etri.re.kr

Jiwon Lee
Content Research Division
Electronics and Telecommunications Research Institute(ETRI)
Daejeon, Republic of Korea
ez1005@etri.re.kr

Jungsoo Lee
Content Research Division
Electronics and Telecommunications Research Institute(ETRI)
Daejeon, Republic of Korea
jslee2365@etri.re.kr

Dowon Nam
Content Research Division
Electronics and Telecommunications Research Institute(ETRI)
Daejeon, Republic of Korea
dwnam@etri.re.kr

Wonyoung Yoo
Content Research Division
Electronics and Telecommunications Research Institute(ETRI)
Daejeon, Republic of Korea
zero2@etri.re.kr

***Abstract*—With the advent of hyperscale AI based on large amounts of data and supercomputing infrastructure to process them, AI has made remarkable progress in many areas. Among them, AI-based image generation technology has recently developed rapidly, and research on the use of AI training data is also active. Augmenting training data with synthetic image generation can be of great help in AI training for defense and medical applications, where data collection is difficult. In particular, in marine environments, where data collection is more difficult than on land, it is difficult to collect data on various weather conditions, so the application of AI-based image generation technology is very effective. In this paper, several images are generated by image outpainting based on objects in a real image, and the generated images are intended to be used as training data for an object detector.**

Keywords—object detection, data augmentation, synthetic image generation

I. INTRODUCTION

Attempts to develop universally applicable artificial general intelligence continue, and as part of this, there is fierce competition to develop hyperscale AI based on large amounts of data and the supercomputing infrastructure to process it. As a result, AI technology has grown rapidly in many fields, and image generation technology is no exception. Recently, AI-generated images are of such high quality that it is difficult to distinguish them from human-drawn images or photographs taken by cameras. Attempts to use AI-generated high-quality synthetic images as AI training data are ongoing, with some success. In the field of defense, where AI training is difficult due to the difficulty of collecting data and the high difficulty of taking photographs, if AI-generated synthetic images can be used as training data, it will have a great impact on improving the performance of defense AI and reducing training costs.

Generative Adversarial Networks (GANs), which have led the way in AI-based image generation in the past, can generate high quality images and control the generated images to some extent through calculations between latent spaces [1]. Exploiting these characteristics to use GANs for training data augmentation has achieved results in some areas [2]. However, the image generated by GANs is unsuitable for use in generating training data for object detection because the proportion of the whole image occupied by a given object is too large, and it is difficult to directly control the generation result. In particular, in order to use AI technology for defence surveillance images that need to detect a specific surveillance target, it is essential to use technology that can directly control objects in images.

Fig. 1. Example of a real warship image and cropped warships

Stable diffusion is suitable for data augmentation because it allows the control of objects, backgrounds, styles, etc. in generated images through multiple text prompts [3]. However, controlling generated images using only text prompts has limitations, making it unsuitable for use in data augmentation for special cases such as object recognition in medical and defense images. To solve this problem, data augmentation is required in a way that the background data is modified while the shape of the object to be detected is preserved as much as possible. In this paper, we aim to show that the performance of warship object detection can be improved by augmenting a real image by image outpainting, which preserves the shape of the object, and using it as AI training data for the object detector.

The rest of this paper is structured as follows. Section 2 introduces the existing synthetic image generation technology and object detection technology used in this paper, and Section 3 describes the proposed methodology. After presenting the experimental results in Section 4, we present the conclusions in Section 5.

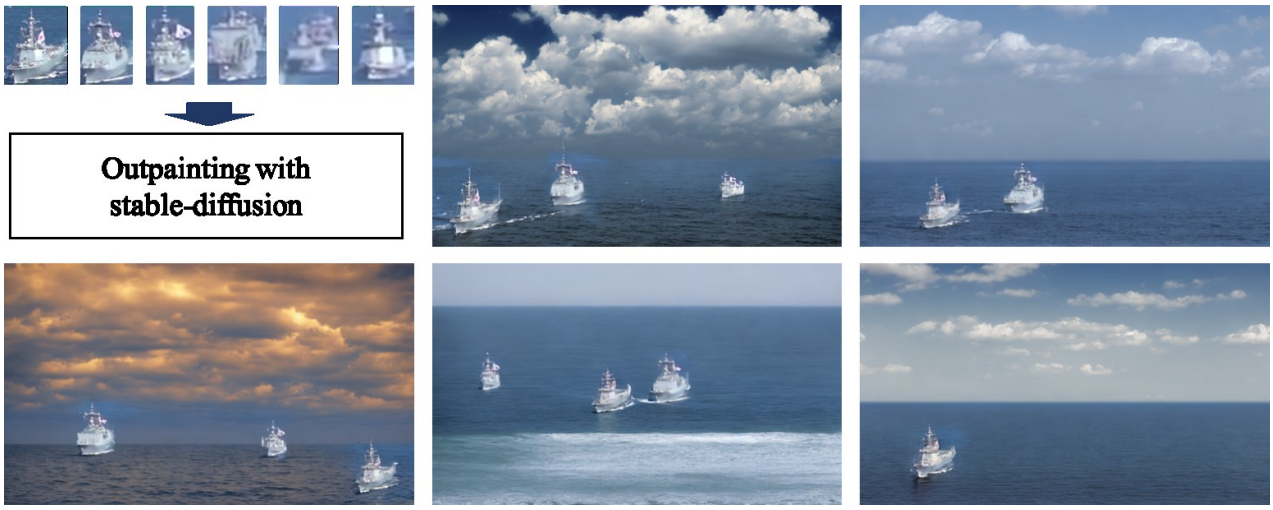

Fig. 2. Examples of high-resolution synthetic images

II. Related Works

A. Stable Diffusion

Stable diffusion is an artificial intelligence model, distributed under an open source licence by Stability AI, that generates images by taking text prompts as input [3]. Stable diffusion, a powerful text-to-image AI model, is based on latent diffusion and has been trained on the LAION-5B database, a large dataset. Due to the release of stable diffusion, computer vision research using image generation is rapidly developing, and it is used for application such as image generation using pose, sketch, etc. as input, inpainting and outpainting, in addition to text-based synthetic image generation through various plug-ins It is becoming. In this paper, stable diffusion is used for outpainting-based data augmentation, and the implementation details are presented in Section 3.

B. DAMO-YOLO

DAMO-YOLO, developed by the TinyML team of Alibaba DAMO Data Analytics and Intelligence Lab, has achieved higher performance than the state-of-the-art of the existing YOLO series [4]-[6]. DAMO-YOLO, a fast and accurate object detection method, uses new technologies such as Neural Archithecture Search (NAS) backbones, efficient Reparameterized Generalized-FPN (RepGFPN), a lightweight head with AlignedOTA label assignment, and distillation enhancement to improve performance. DAMO-YOLO provides models for several real-world scenarios, such as human detection, helmet detection, and cigarette detection, in addition to models pre-trained with the MS COCO dataset. In this paper, DAMO-YOLO was trained with a small amount of data and used as an object detector, as it was found to be appropriate for the scenario. Further implementation details are presented in Section 3.

III. Methodology

A. Data Augmentation

The method of augmenting warship data using the proposed outpainting is as follows. First, from a single 1920x1080 image including the warships to be detected, the bounding box areas including the warships are cropped. The target image and cropped warships are shown in Fig. 1. The size and location of the cropped warships are randomly changed and arranged, and synthetic data are generated by outpainting using stable diffusion. The actual outpainting was performed based on the open source stablediffusion-infinity, and the detailed settings are as follows [7].

- Model type : stablediffusion-2-inpainting
- Init mode : patchmatch
- Photometric correction mode : disabled
- Prompt : sea, sky
- Sample number : 5
- Scheduler : DPM
- Step : 25
- Guidance : 7.5

The synthetic data generated are of two types: low-resolution and high-resolution. First, five low-resolution synthetic images of 512x512 size were created to test the case where the shape of the actual data to be detected and the synthesized data are different. In addition, five high-resolution synthetic images of 1024x600 size with similar aspect ratio and warship size to the actual validation data were generated. The synthesized images can be seen in Fig. 2. and Fig. 3. Since the size and position of the warship used as input is known, an annotation file can be automatically created for each synthetic image.

B. Object Detector Training using Synthetic Data

We constructed three types of datasets for object detector training. First, dataset 1 consists of one real warship image, dataset 2 consists of one real image and five low-resolution synthetic images, and dataset 3 consists of one real image and five high-resolution synthetic images. The DAMO-YOLO model was trained using each dataset, and the detailed hyperparameter settings are as follows.

- Model name : TinyNAS_res
- Batch size for dataset 1 : 1
- Batch size for dataset 2, 3 : 6
- Base Learning rate : 0.00015625

Fig. 3. Examples of low-resolution synthetic images

The object detector trained with each dataset measures object detection performance using the warship object detection validation dataset, which consists of 20 real-world 1280x720 images. The objects to be detected were restricted to warships.

IV. Experimental Results

The experimental environment for verifying the change in object detector performance for the case of using only one real image for training and the case of using augmented synthesized images for training is as follows. The deep learning framework used was Pytorch 1.7.1 and the operating system was Ubuntu 20.04. An NVIDIA A100 GPU was used and CUDA 11.0 was used. The quantitative indicators used to evaluate the object detection performance were mAP, AP50, and mAR. A summary of the experimental results is shown in Table 1.

When only one real image was used for training, the mAP, AP50, and mAR were 17.9%, 42.4%, and 30.9%, respectively. The results of using the low-resolution synthetic images shown in Fig. 3. for training along with a real image are somewhat interesting. Despite the increase in training data, mAP decreased from 17.9% to 11.1%, and AP50 also decreased from 42.4% to 26.2%. The warships in the synthetic images were cropped from real data, and despite the increase in the total number of images used for training, it was confirmed that the shape of the data was different from the validation data, and the accuracy decreased significantly. This confirmed that even if the synthetic data is used to augment the data, if the shape is different from the actual image to be applied, simply using it to train the object detector may adversely affect the object detection performance.

When 5 high-resolution synthetic images similar to the validation data are used to train the object detector along with 1 real image, the object detection performance is as follows. With mAP 37.6%, AP50 64.9%, and mAR 52.1%, they increased by 29.7%, 22.5%, and 21.2%, respectively, compared to the object detector using only one real image for training. These experimental results confirmed that the detection performance of the object detector can be greatly improved if synthetic data is generated in an appropriate form using the outpainting technique. The results of the above experimental results are as follows. Data augmentation using outpainting can automatically generate annotations for images, and can help improve the object detection performance when used for object detector training because the object to be detected has little distortion. However, when generating synthetic images, the shape of the image must be controlled in a style similar to the image actually input to the object detector. Otherwise, the performance of the object detector may deteriorate. Based on these results, we propose to generate synthetic images for object detector training in a shape that is as similar as possible to the actual test data.

TABLE I. Object Detection Performance According to the Training Dataset

Training data type	mAP	AP50	mAR
1 real image	17.9%	42.4%	30.9%
1 real image + 5 low resolution synthetic images	11,1%	26.2%	31.5%
Proposed	**37.6%**	**64.9%**	**52.1%**

V. Conclusion

In this paper, we experimentally investigated the effect of data augmentation by outpainting-based synthetic data generation using stable diffusion on object detection performance. Through an experiment in which the object detector was trained on one real image and five synthetic images generated from it, it was confirmed that the use of synthetic data can have a great effect on the performance of the object detector. We expect that this study will help secure data through synthetic data generation when developing artificial intelligence for object detection in areas where data acquisition is difficult, such as national defense and medical care. In the future, we will study how to automatically generate large-scale synthetic images and use them for object detector training, and how to suppress image generation that negatively affects object detector performance by using negative prompts.

Acknowledgment

This work was supported by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. RS-2023-00223530, Development of the artificial intelligence technology to enhance individual soldier surveillance capabilities)

References

[1] I. J. Goodfellow *et al*., "Generative Adversarial Networks." *arXiv*, 2014.

[2] A. Antoniou, A. Storkey, and H. Edwards, "Data Augmentation Generative Adversarial Networks." *arXiv*, 2017.

[3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models." *arXiv*, 2021.

[4] X. Xu *et al*., "DAMO-YOLO : A Report on Real-Time Object Detection Design." *arXiv*, 2022.

[5] Z. Sun *et al*., "MAE-DET: Revisiting Maximum Entropy Principle in Zero-Shot NAS for Efficient Object Detection." *arXiv*, 2021.

[6] Y. Jiang *et al*., "GiraffeDet: A Heavy-Neck Paradigm for Object Detection." *arXiv*, 2022.

[7] lkwq007, "LKWQ007/stablediffusion-infinity: Outpainting with stable diffusion on an infinite canvas," *GitHub*, 2023. [Online]. Available: https://github.com/lkwq007/stablediffusion-infinity.