Drowning Detection based on YOLOv8 improved by GP-GAN Augmentation

Anonymous authors
Paper under double-blind review

Abstract

Drowning is a significant safety issue worldwide, and a robust computer vision-based alert system can easily prevent such tragedies in swimming pools. However, due to domain shift caused by the visual gap (potentially due to lighting, indoor scene change, pool floor color etc.) between the training swimming pool and the test swimming pool, the robustness of such algorithms has been questionable. To address this issue, we propose a domain-aware data augmentation pipeline based on Gaussian Poisson Generative Adversarial Network (GP-GAN). Combined with YOLOv8, we demonstrate that such a domain adaptation technique can significantly improve the model performance (from 0.24 mAP to 0.82 mAP) on new test scenes.

1 Introduction

Drowning is a significant global health problem, accounting for 7% of all injury-related deaths. Each year, an estimated 236,000 people die from drowning (1). To tackle this issue, we aim to utilize computer vision and machine learning techniques. Drowning detection can be simplified into a swimmer detection problem: When a swimmer disappeared within the pool area, there is a non-trivial chance of drowning.

Although previous works have made contributions to the swimmer detection problem (2) (3) (4) (5), they may not be reliable for new domains(due to train/test visual gap). Fortunately, recent advances in domain adaptation have shown promise in addressing the domain shift problem. For example, GANs have been proposed to bridge the domain gap in turbine detection applications (6). Following previous work, we proposes using GP-GAN (7) to address the domain shift issue. We conducted experiments on a swimmer detection training set (8) and a test set (9), which featured water scenes different from each other. We trained GP-GAN to augment the data and found that our domain adaptation approach outperformed YOLOv8 (10) in terms of predictions.
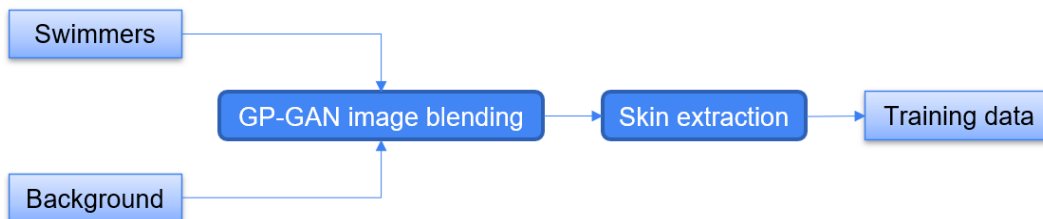
2 Methods

Figure 1: The generating process of images in the augmented training set. Background images of testing domain and swimmer snapshots from training domain are feed into the GP-GAN, followed by a skin extraction algorithm to maintain the original skin color

Most human detection datasets (11) typically feature pedestrians in street scenes, making them unsuitable for training a swimmer detection model. Unfortunately, there are very few available datasets

for swimmer detection (12) (13) (8), and only a handful of them meet our requirement, so we collected a set of 200 appropriate images from (8) to be the original training set. For the test set, we manually labeled 55 images from videos (9) captured by a surveillance camera installed in a swimming pool. They have a different water scene from the training set. We trained a YOLOv8 using the ImageNet pretrained weights as initialization.

In addition to the original training set, we generated an augmented training set by GP-GAN to compare the results. GP-GAN (7) is a powerful tool for blending two images realistically, which can help us blend swimmers into a swimming pool and generate new training data to close the domain gap between different water scenes.

The generating process is shown in Fig.1. We first extract swimmer images from the original dataset and collect several background images of swimming pools that are similar to the water scene of the test set. Next, we blend the swimmer images into the backgrounds using GP-GAN, with random location, rotation and resizing. However, the swimmers in the output images given by GP-GAN may be blurred, so we extracted their skin from the original images and put it back into the output images to ensure that all swimmers are clear enough. We use an established method (14) to extract human skin pixels from a picture: we convert the RGB color space into the YCbCr color space, and the Cb and Cr range distribution is used to determine the global skin color classifier, range as $(75 < Cb < 135)$ and $(130 < Cr < 185)$. We generated 200 images with above process, which constitute our new training set and have a similar domain to the test set.

3 RESULTS

Our main results are shown in Fig.2, the model's performance when trained on the original training set was even worse than that of a typical human detection dataset (11) (walker line), with a maximum mAP of just 0.24 and a minimum IOU of 0.5 on the test set. This poor score was mainly due to the significant domain differences between the water scene in the training set and the scene in the test set.
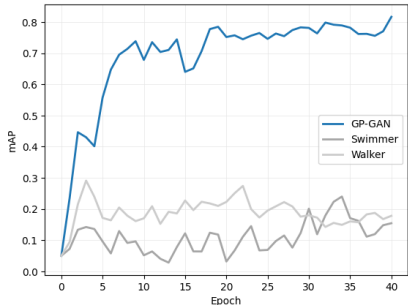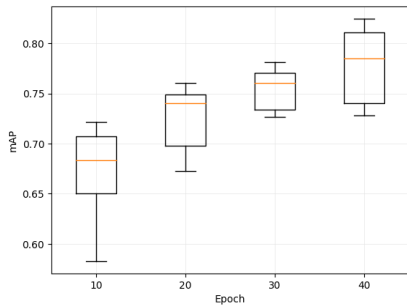

Figure 2: Results of different datasets.

Figure 3: Experiments on the new dataset.

With the help from the augmented training set, we were able to improve the performance of YOLOv8 on the swimmer detection task. After 40 epochs of training, the model achieved the best mAP of 0.82 shown in Fig.2 (GP-GAN line), which is a significant improvement over the original dataset. All training experiments were conducted using the same pretrained model and hyperparameters.

We trained the model six times on the new dataset, and the results with error bars are shown in Fig.3. The model demonstrated consistent and stable performance on the new training set, indicating that the generated data helped to close the domain gap.

4 CONCLUSION

We conclude that the GP-GAN is a promising technique that can battle the domain gap between training and testing in the swimming pool drowning detection problem. We believe that with more training data (like drowning ImageNet) to transfer knowledge from, such detection modules can be effectively saving thousands of lives worldwide.

5   URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

[1] WHO, "Drowning key facts," 2021, https://www.who.int/news-room/fact-sheets/detail/drowning.

[2] H. Xiao, Y. Li, Y. Xiu, and Q. Xia, "Development of outdoor swimmers detection system with small object detection method based on deep learning," *Multimedia systems*, pp. 1–10, 2022.

[3] N. Jacquelin, R. Vuillemot, and S. Duffner, "Detecting swimmers in unconstrained videos with few training data," in *Machine Learning and Data Mining for Sports Analytics: 8th International Workshop, MLSA 2021, Virtual Event, September 13, 2021, Revised Selected Papers.* Springer, 2022, pp. 145–154.

[4] D. Zecha, T. Greif, and R. Lienhart, "Swimmer detection and pose estimation for continuous stroke-rate determination," in *Multimedia on Mobile Devices 2012; and Multimedia Content Access: Algorithms and Systems VI*, vol. 8304.   SPIE, 2012, pp. 282–294.

[5] K. L. Chan, "Detection of swimmer using dense optical flow motion map and intensity information," *Machine vision and applications*, vol. 24, no. 1, pp. 75–101, 2013.

[6] C. Kornfein, F. Willard, C. Tang, Y. Long, S. Jain, J. Malof, S. Ren, and K. Bradbury, "Closing the domain gap – blended synthetic imagery for climate object detection," in *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022. [Online]. Available: https://www.climatechange.ai/papers/neurips2022/37

[7] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Gp-gan: Towards realistic high-resolution image blending," 2017. [Online]. Available: https://arxiv.org/abs/1703.07195

[8] drawnings, "drowning-finder dataset," https://universe.roboflow.com/drawnings/drowning-finder, jul 2022, visited on 2023-02-23. [Online]. Available: https://universe.roboflow.com/drawnings/drowning-finder

[9] M. Moon, "Review—surveillance+activity+2," 2019, https://www.youtube.com/watch?v=rWsbNclYgmE.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015. [Online]. Available: https://arxiv.org/abs/1506.02640

[11] L. Wang, J. Shi, G. Song, and I.-f. Shen, "Object detection combining recognition and segmentation," 11 2007, pp. 189–199.

[12] dan4zy@virginia.edu, "Swim clip 1 dataset," https://universe.roboflow.com/dan4zy-virginia-edu/swim-clip-1, apr 2022, visited on 2023-02-26. [Online]. Available: https://universe.roboflow.com/dan4zy-virginia-edu/swim-clip-1

[13] Swimmers, "Swimmers_tfg dataset," https://universe.roboflow.com/swimmers/swimmers_tfg, apr 2022, visited on 2023-02-26. [Online]. Available: https://universe.roboflow.com/swimmers/swimmers_tfg

[14] D. Joshi, A. Purwar, and M. S. Dasgupta, "Human detection based on explicit skin colour space thresholding and minor motion detection," *11th IEEE India Conference: Emerging Trends and Innovation in Technology, INDICON 2014*, 02 2015.

APPENDIX

Figure 4: The original training set, which contains 200 images collected from available datasets.

Figure 5: The test set, which contains 55 images extracted from videos caught by a surveillance camera near a swimming pool that has a different water scene from the training set.

Figure 6: The generating process of the new training set: images from left to right are the original image, the copy-paste image, the extracted skin and the final training image.

Figure 7: Predictions given by the model trained by the original dataset (left) and the model trained by the augmented dataset (right).
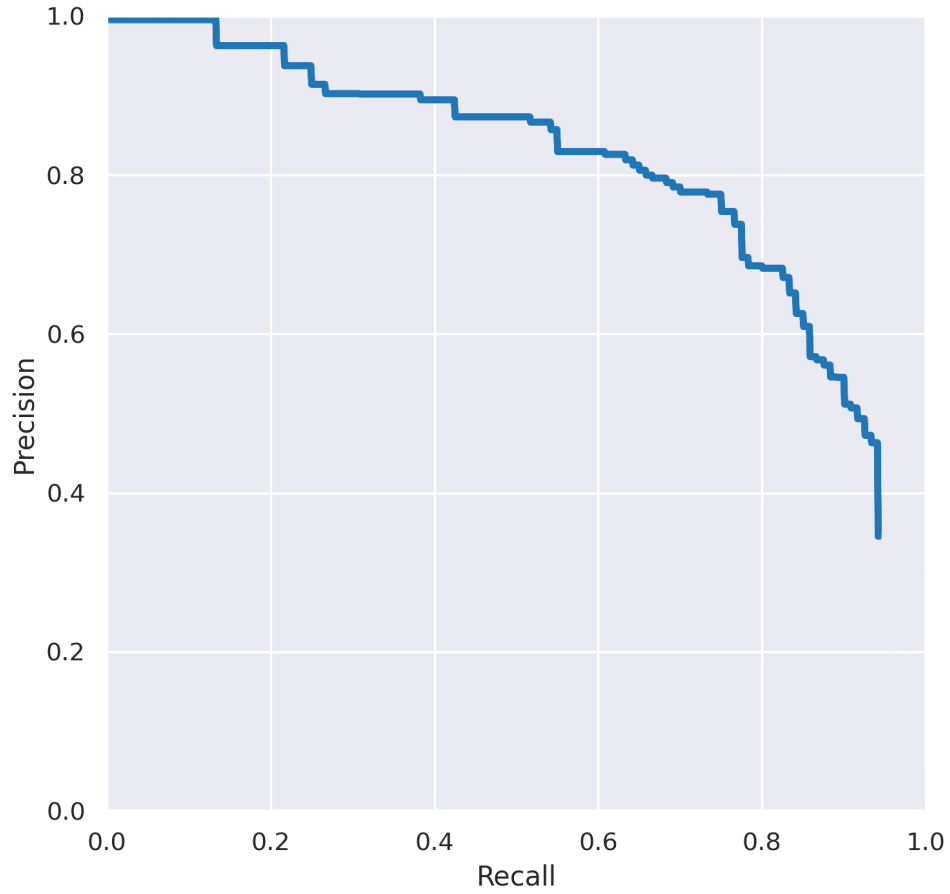
Figure 8: This is the Precision-Recall curve of the YOLOv8 trained by the augmented dataset. Precision measures the accuracy of positive predictions, whereas recall measures the model's completeness in identifying relevant instances. In the swimmer detection problem, recall is more important than precision to minimize the risk of missing any potential danger or drowning incidents.