Changing Background to Foreground: an Augmentation Method Based on Conditional Generative Network for Stingray Detection

Yi-Min Chou[1,2], Chien-Hung Chen[1,2], Keng-Hao Liu[3], Chu-Song Chen[1,2]

[1] Institute of Information Science, Academia Sinica, Taipei, Taiwan

[2] MOST Joint Research Center for AI Technology and All Vista Healthcare

[3] Department of Mechanical Electro-mechanical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

{chou, redsword26, song}@iis.sinica.edu.tw, keng3@mail.nsysu.edu.tw,

Introduction

Objective

Improving the performance of automatic stingray detection in the aerial images captured by an UAV to help ecological researchers in counting the number of stingray.

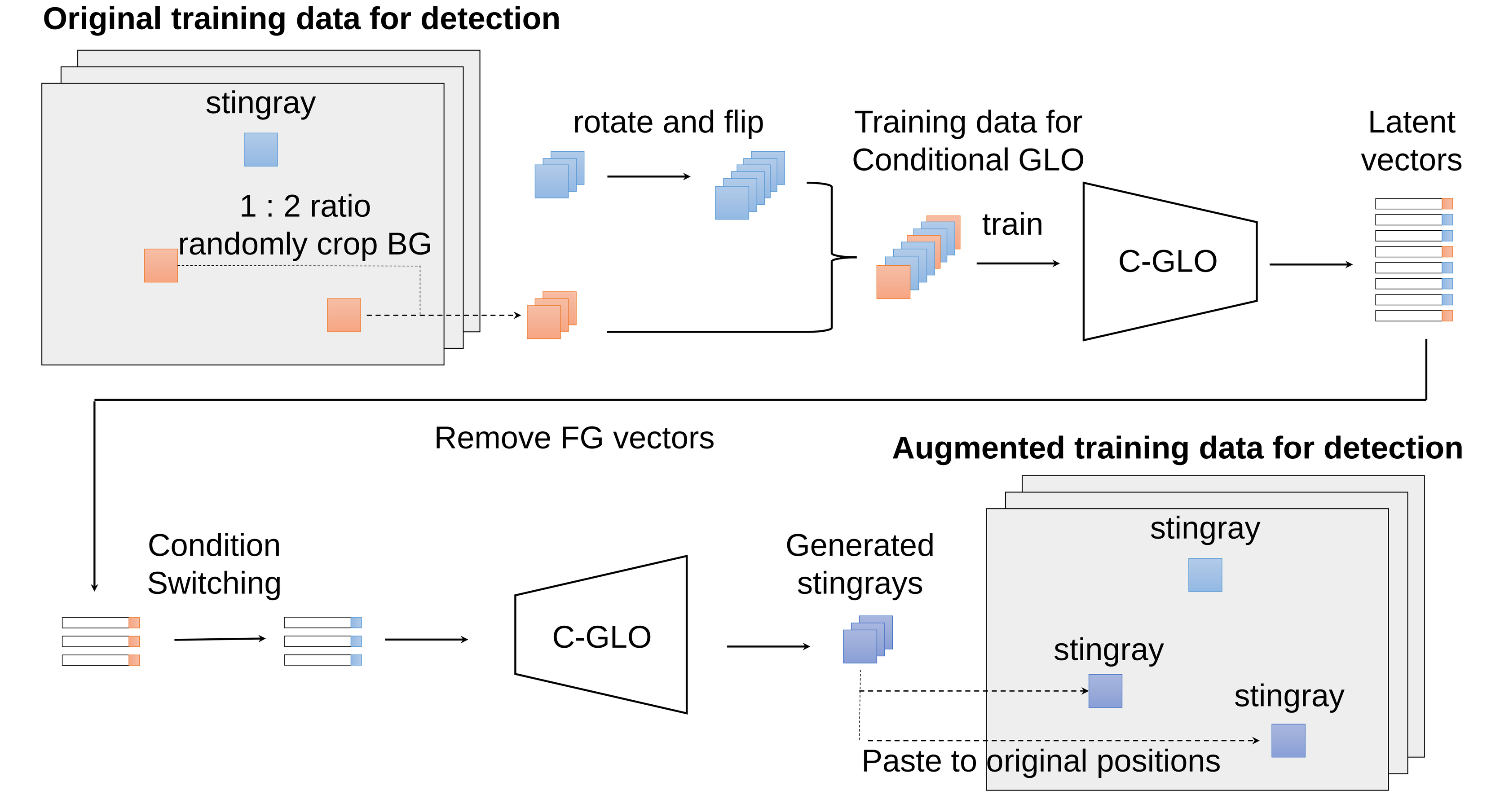
Problems

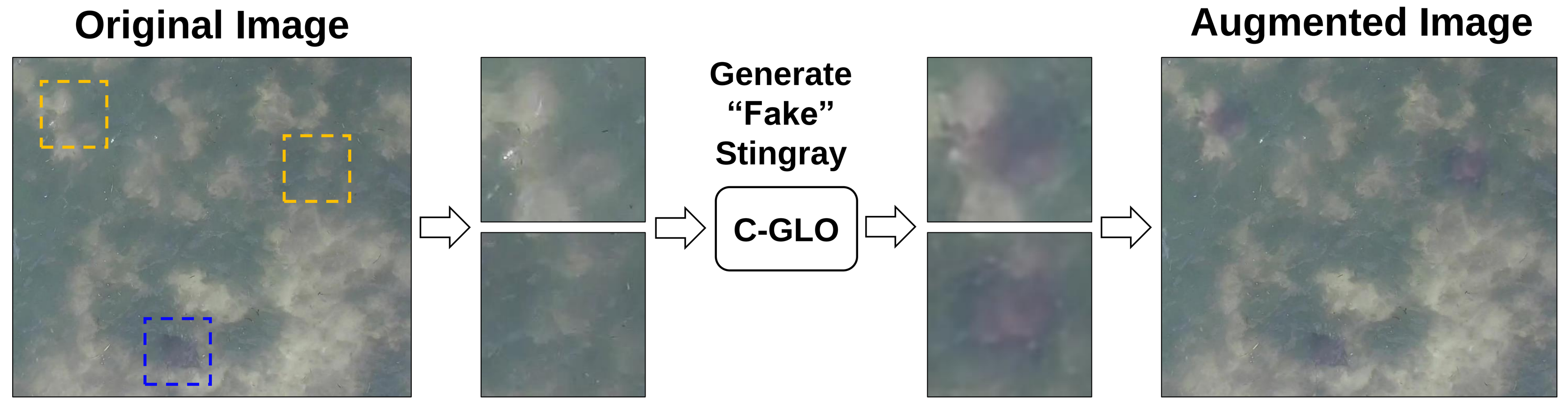
- A wide range aerial imagery (high resolution) limits the image batch number used for training in a deep learning object detection model (Faster R-CNN in this paper) due to memory issue.
- Stingray is always sparse in each image. It limits the batch number of the object for training.
- Using limited amount of training images results in worse detection results.

Solution

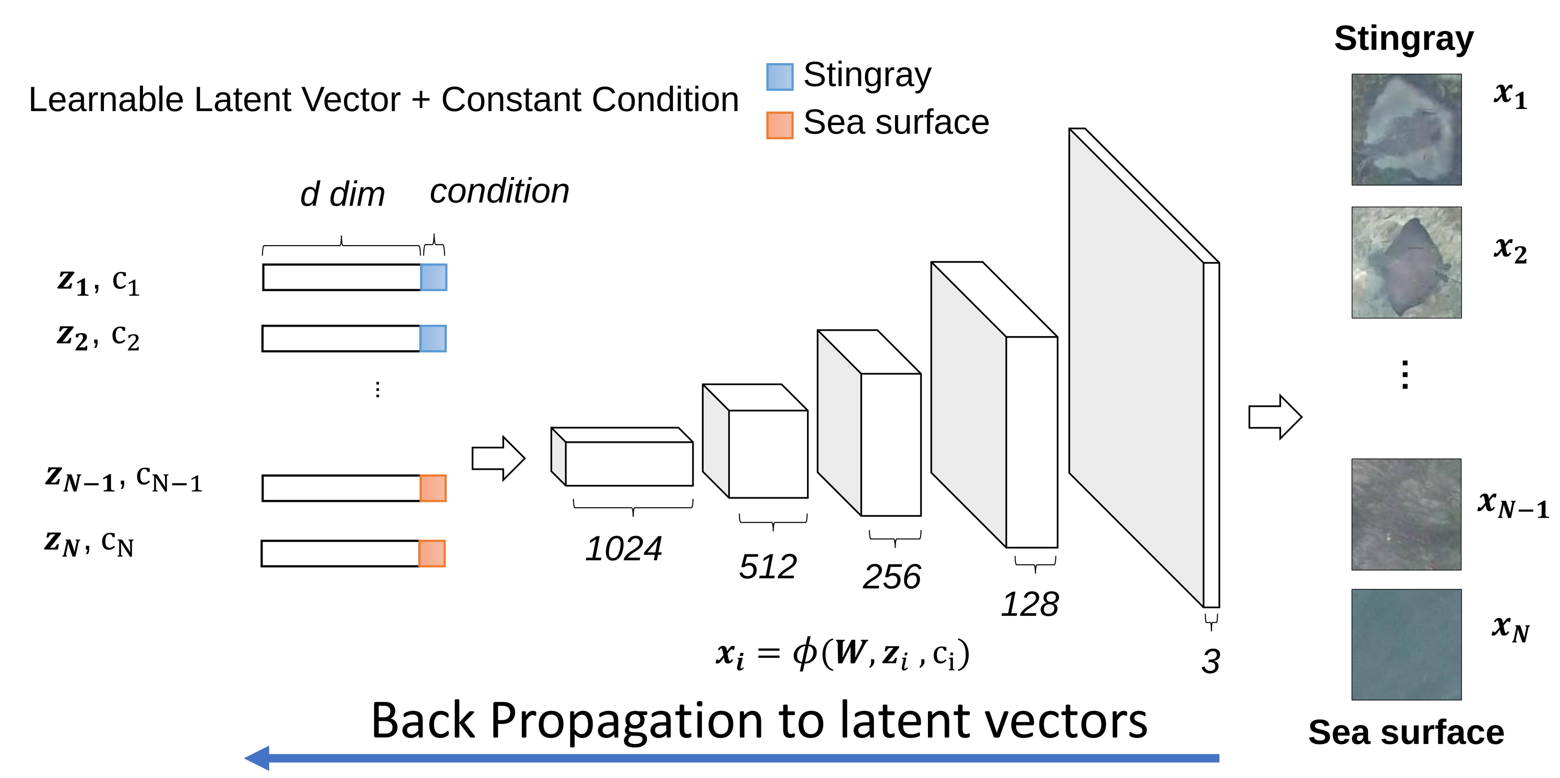
We propose a new data augmented method, called Conditional Generative Latent Optimization (C-GLO), to generate sufficient amount of stingray samples for reaching better object detection on marine aerial images.

Method Overview





The Architecture of C-GLO



Given unsupervised training images $I = \{I_1, \cdots, I_N\}$, C-GLO trains a generator ϕ(with the input z, c and network weights W), such that the following objective is minimized:

$$e(W, z, c) = \sum_{i=1}^{N} loss(\phi(W, z_i, c_i) - I_i)$$

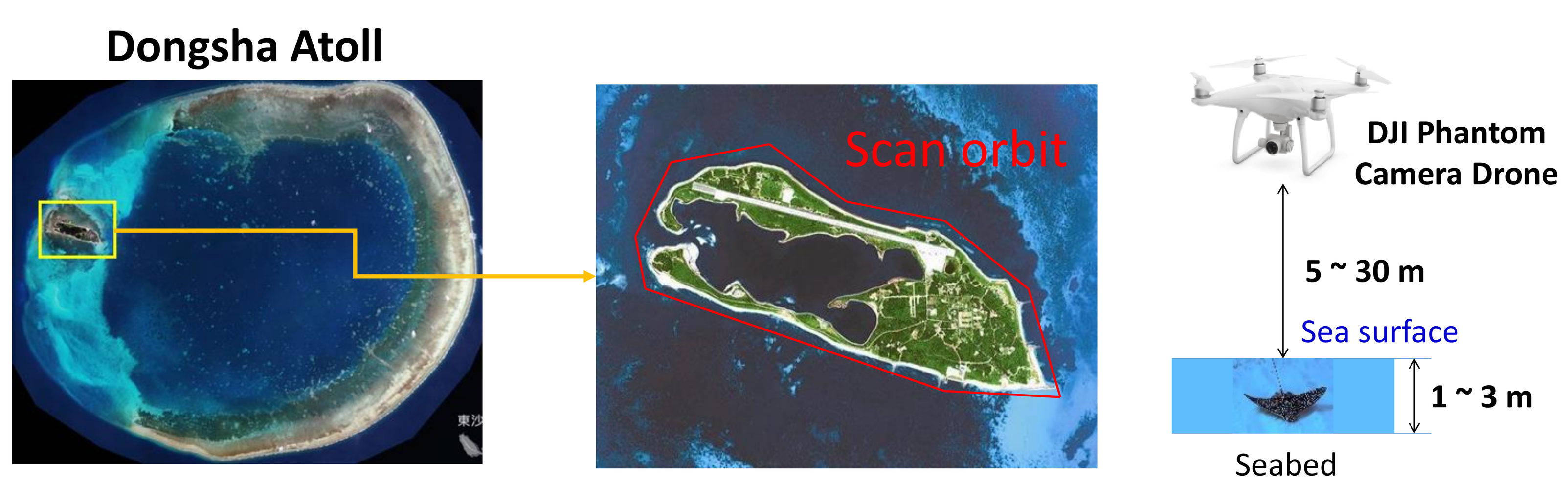
where $z = \{z_1, z_2, \cdots, z_N\}$. c = 0 indicates sea surface, and c = 1 indicates stingray.

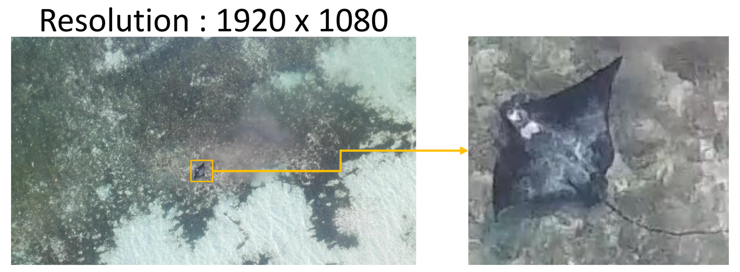
The training process of C-GLO is :

1. Given z, c, find W to reduce the total reconstruction loss of I .
2. Given W, c_i , find z_i to reduce the reconstruction loss of I_i , $\forall I$.

The above two steps are executed iteratively.

Data Description

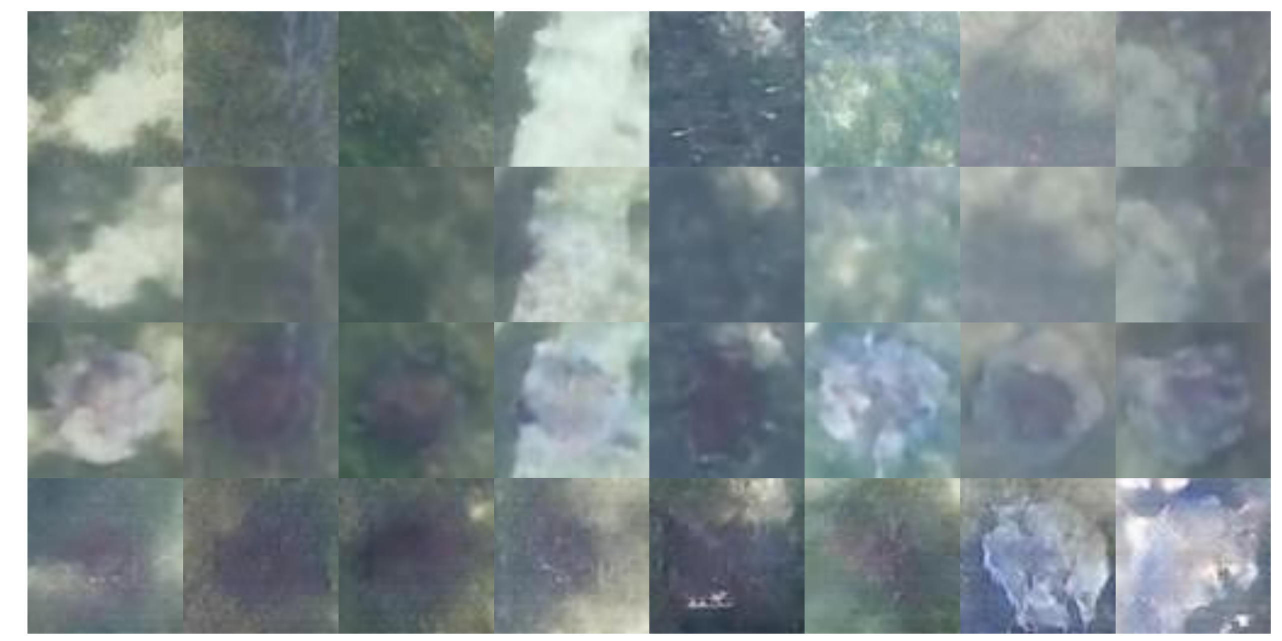




	Videos	Images
Training data	18	3245
Test Data	18	3147

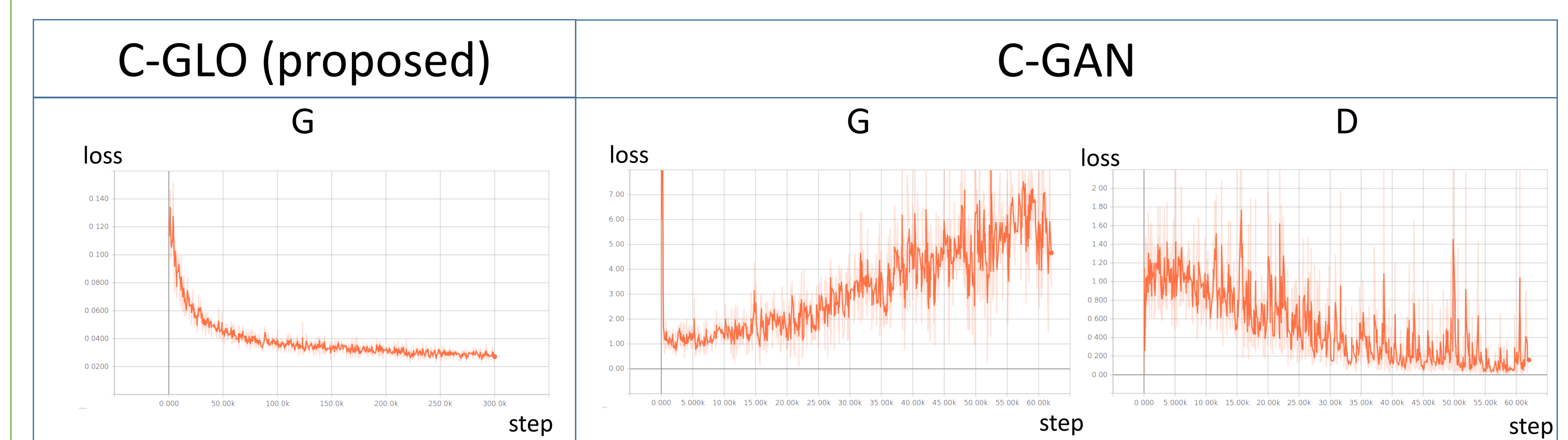
Results & Comparison

Generated Images



Comparison with Conditional GAN method

	Model Structure			Property	
	Encoder	Generator	Discriminator	Memory latent vectors	Invertible
C-GLO (proposed)	×	✓	×	✓	✓
C-GAN	×	✓	✓	×	×
ICGAN	✓	✓	✓	×	✓



Faster R-CNN Detection Performance (AP)

Network	Baseline	Ours-128	Ours-256	Ours-512
ZF	78.89	82.75	82.42	**83.04**
VGG-16	84.59	86.14	**86.61**	86.43

Conclusion

- We introduce a data augmentation method, C-GLO to fuse background patches and foreground objects.
- Experimental results reveal that the detection performance can be improved our approach.
- C-GLO performs more stable than GAN-based methods do in convergence of training.