ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

# Generative Data Augmentation for Vehicle Detection in Aerial Images

Hilmi Kumdakcı*, Cihan Öngün, Alptekin Temizel

January 11, 2021

- Aerial images are becoming very crucial in several applications

- Scarcity of training data is one of the prominent problems

- Data augmentation is a widely used method to increase the number of training samples and their variations

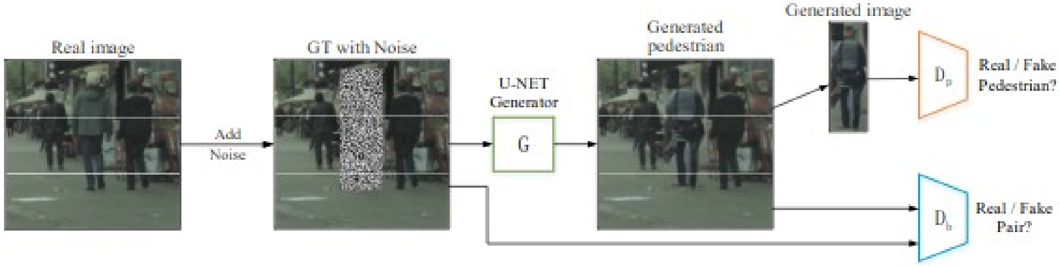
- We propose a generative augmentation method to improve vehicle detection performance in aerial images

- Well known traditional data augmentation methods (geometrical and color based transformations) are applied by practitioners and researchers

- It does not add any distribution by discovering deep features of the data and variety in terms of semantics

- Recently, generative neural networks come into play in data augmentation for object detection and classification
- Studies in the literature uses image-to-image translation methods to generate new object instances
- They are likely to need large scale datasets and sensitive to hyperparameters
- They generate worse outcomes from the noisy pure background, suffering from artifacts, and it is not guaranteed to generate good quality samples every time

- To insert objects into images in context aware manner, two discriminator approaches are proposed (PSGAN)




- In this approach, one discriminator is responsible from context whereas other one is the instance to be placed. Our framework can be integrated with this approach.

* Ouyang, X., Cheng, Y., Jiang, Y., Li, C. L., Zhou, P.: "Pedestrian-synthesis-gan: Generating pedestrian data in real scene and beyond." arXiv preprint arXiv:1804.02047 (2018).

- The proposed data augmentation method consists of a generator network and a detector network

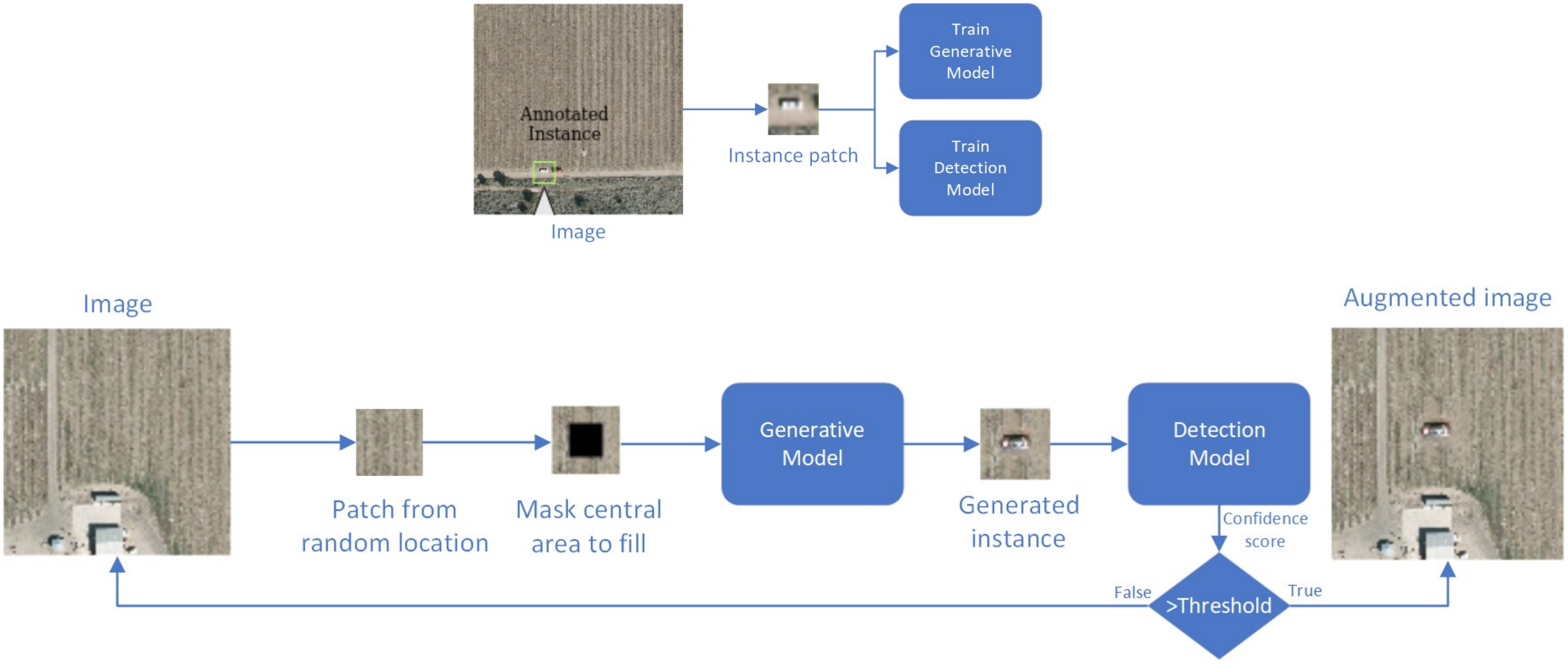
- The generator network is expected to generate new instances that fits the given background and the detector network must be able to detect corresponding instances with bounding boxes.

- For experiments, image inpainting network Pluralistic is mainly used as generator and Tiny-YoloV3 is used as detector for assessment and localization of objects

- Pluralistic uses two parallel paths, one is reconstructive and the other is generative, both are supported by GANs.
- When original image as $I_g$, the partially masked image as $I_m$,and the complement image as $I_c$. Pluralistic aims to sample from $p(I_c|I_m)$. The reconstructive path combines information from $I_m$ and $I_c$, which is used only for training.

- Uses anchor-based approach to regress bounding box coordinates
- Performs detection in two scales whereas YoloV3 does in three scales
- Preferred over other detectors due to its feasibility of usage and relatively low inference/training time

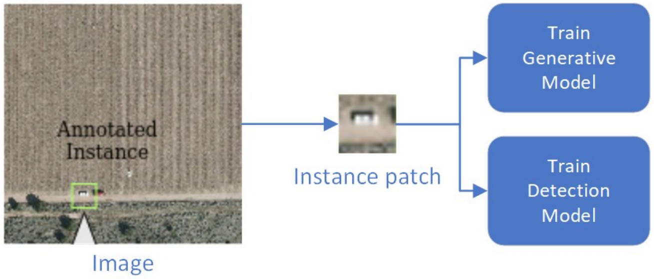
- The proposed method consists of 2 stages: training and augmentation
- Training stage involves independent training of a generative network and a detector network
- At the augmentation stage, the generative network is used to generate new samples and the detector is used to assess the feasibility of these samples for augmentation as well as localizing the synthetic objects
- The separate training procedure prevents the generator overfitting and provides diverse and realistic augmentations




Schematic of the proposed method. Top: Training stage, Bottom: Augmentation stage

# Proposed Method

- Due to the necessity of learning contextual background, both networks have been trained with same images, car instances and their surroundings
- Generator network learns to generate realistic instances on the given patches
- Detector is trained with bounding box annotations and the best network parameters which has the highest average precision is selected for the augmentation stage

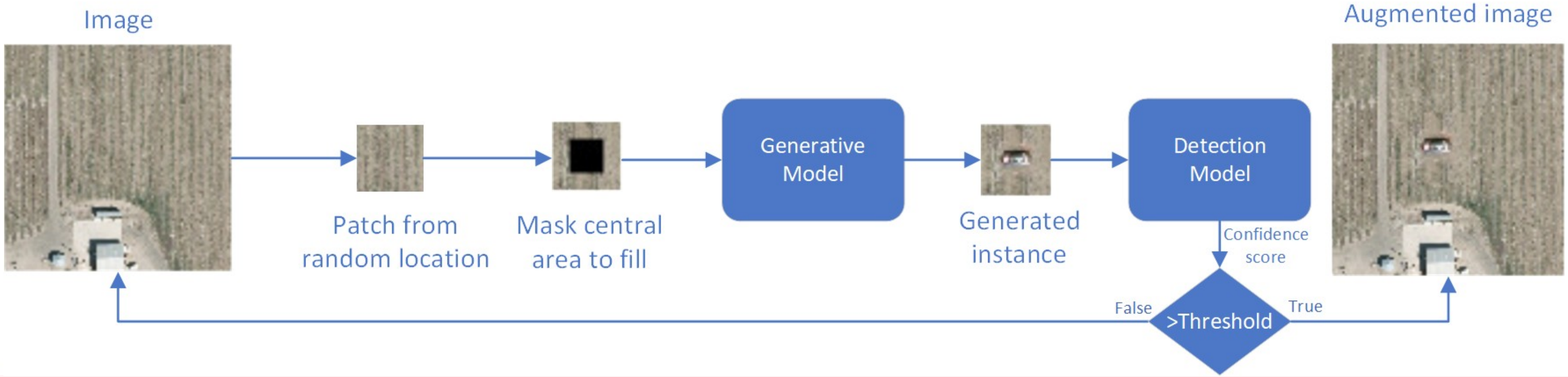



# Training Stage

- Based on the size distribution of car instances, 96x96 patches from random locations are extracted and their central 48x48 areas are masked

- The patches are fed into the trained generative model to generate synthetic object instances

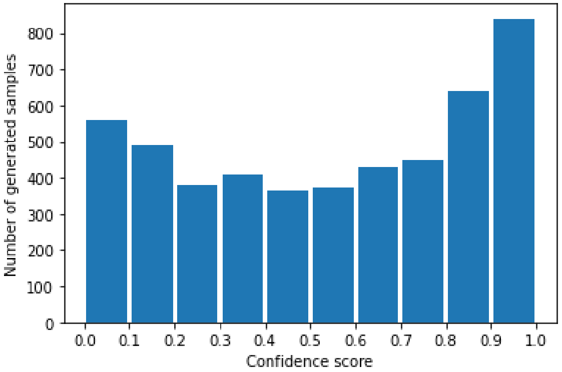
- The generated sample is accepted if the confidence score of detector network is higher than the predetermined threshold



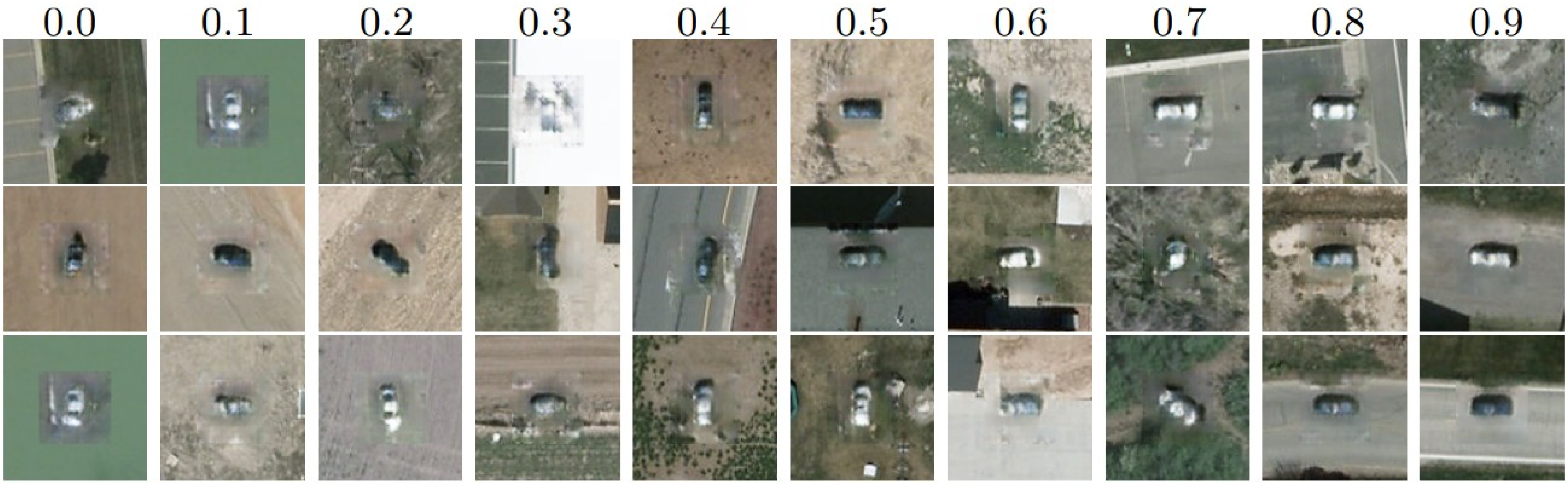

# Augmentation Stage

- We used Vehicle Detection in Aerial Imagery(VEDAI) which has 1272 RGB colored images with 1024x1024 resolution

- We divided the dataset as 500 and 772 for training and testing respectively




Histogram of confidence scores for 5000 generated samples which are fed into detector.




Examples from generated samples with corresponding confidence scores.

Raw and augmented images

# Average Precision results for different confidence thresholds

| Acceptance threshold | IOU 0.2 | 0.5 | 0.7 | Average | Augmentation Iterations |
|---|---|---|---|---|---|
| 0.0 | 56.20 | 36.11 | 7.26 | 33.19 | 1000 |
| 0.1 | 55.55 | 40.67 | 7.97 | 34.73 | 1172 |
| 0.2 | 55.42 | 39.51 | 9.23 | 34.72 | 1288 |
| 0.3 | 56.23 | 39.65 | 8.76 | 34.88 | 1423 |
| 0.4 | 56.38 | 44.16 | 10.31 | **36.95** | 1709 |
| 0.5 | 56.04 | 44.05 | 10.08 | 36.72 | 1818 |
| 0.6 | 56.16 | 42.29 | 10.81 | 36.42 | 2189 |
| 0.7 | 57.63 | 43.41 | 9.57 | 36.87 | 2667 |
| 0.8 | 57.64 | 43.15 | 9.90 | 36.90 | 3498 |
| 0.9 | 57.08 | 43.81 | 10.57 | **37.15** | 5902 |

# Detection performances (AP) with the selected parameters.

| Dataset Images | Augmented Instances | IOU 0.2 | 0.5 |
|---|---|---|---|
| 200 | - | 49.85 | 31.43 |
| 200 | 200 | 52.15 | 37.65 |
| 200 | 400 | 51.71 | 37.04 |
| 300 | - | 53.76 | 33.25 |
| 300 | 300 | 55.39 | 39.76 |
| 300 | 600 | 56.01 | 41.62 |
| 400 | - | 55.46 | 36.14 |
| 400 | 400 | 56.03 | 40.24 |
| 400 | 800 | 56.02 | 42.18 |

- The proposed method is generic and it can work with different generator and detector models

Detection performances (AP) when DeepFill is used as the generator model.

| Dataset Images | Augmented Instances | DeepFill |
|---|---|---|
| 200 | - | 31.43 |
| 200 | 200 | 35.96 |
| 200 | 400 | 39.31 |
| 300 | - | 33.25 |
| 300 | 300 | 38.93 |
| 300 | 600 | 41.81 |
| 400 | - | 36.14 |
| 400 | 400 | 38.73 |
| 400 | 800 | 45.40 |

- We proposed a generative augmentation framework for improving the performance of object detection in aerial images for small datasets.

- The proposed method
  - is generic and can work with different generator and detector models.
  - does not need any extra supervision than the bounding box annotations
  - increases the Average Precision performance of vehicle detection up to 25.7% by allowing detectors to be trained with higher number of instances

Thank you for listening.

*All references can be found in the paper.