Analysis of Utilization of Generative Models to Increase Image Quality for Object Detection in Satellite Imagery

Jason Zisheng Chang, Yifei Wang, Ruiqi Wang
Master in Interdisciplinary Data Science (MIDS), Duke University

Introduction

The goal of this project is to observe the performance gains of object detection using a generative adversarial networks (GAN) to increase the resolution of satellite imagery, and determine the difference in performance of object detection on 0.3 m/px and 1.2 m/px resolution data. Deep learning models built are a **You Only Look Once (YOLO)** system for object detection, and a **Super Resolution Generative Adversarial Networks (SRGAN)** system for super resolution (SR). Results showed that object detection performance of YOLO can be increased by SRGAN.

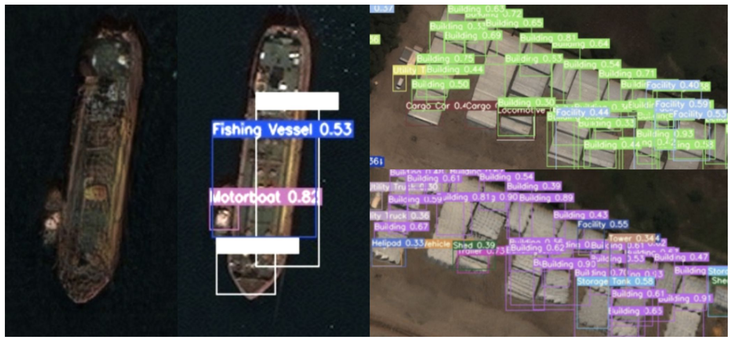


Figure 1. Images synthesized using SRGAN and original dataset.
Left column: left is SRGAN, right picture is original.
Right column: top is original, bottom is SRGAN.

Motivation

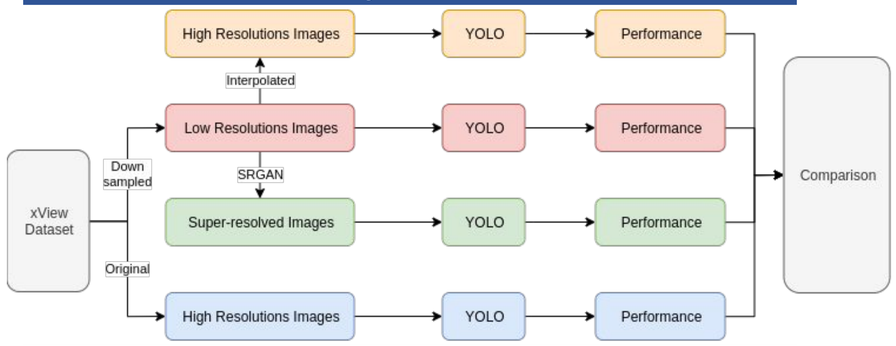
Object detection algorithms often struggle with low-resolution objects. SR seems to be an appropriate solution for this problem, but most approaches lack a quantitative measurement for SR images. In this project we quantitatively analyzed using a a SRGAN model to improve the performance of the baseline object detection model.

Data

Satellite imageries from the xView Dataset [1]. 50 training, 11 testing. Each images is around 10,000px x 10,000px in size. (3 km^2)

Class Label (train)	# of Ground Truth	Class Label (test)	# of Ground Truth
Building	59919	Building	7076
Small Car	1951	Small Car	840
Bus	1254	Truck	314
Cargo Truck	805	Bus	238
...	...	...	...

Project Flow



Models - YOLO

We utilize YOLO for unified object detection. YOLO detection system first resizes the input image into a grid, and thresholds the resulting detections by the models confidence. (Figure 2.)

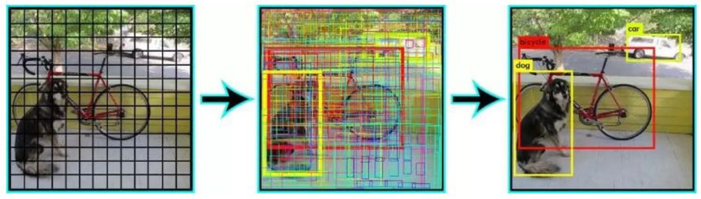


Figure 2. A demonstration of YOLO detection system [2]

Models - SRGAN

A generative adversarial network (GAN) is a model that contains two neural networks: a generator and discriminator). The GAN learns to generate new data with same statistics as the training set. A super resolution generative adversarial network (SRGAN) is a GAN that increases the resolution of low-resolution (LR) images by training on corresponding high-resolution (HR) images. By synthesizing sub pixel information in LR imagery a SRGAN generates super-resolved (SR) images.

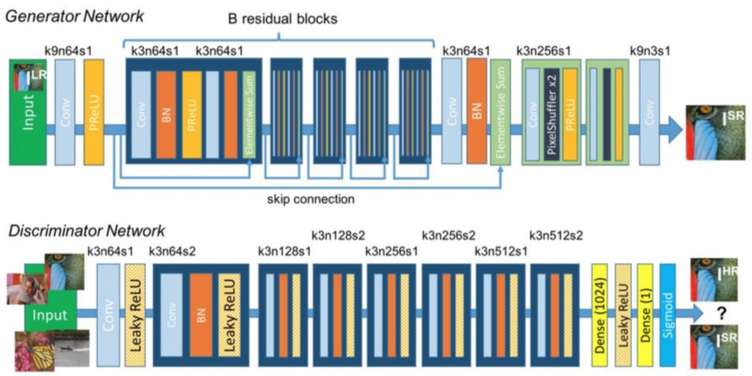


Figure 3. The network architecture of SRGAN system. [3]

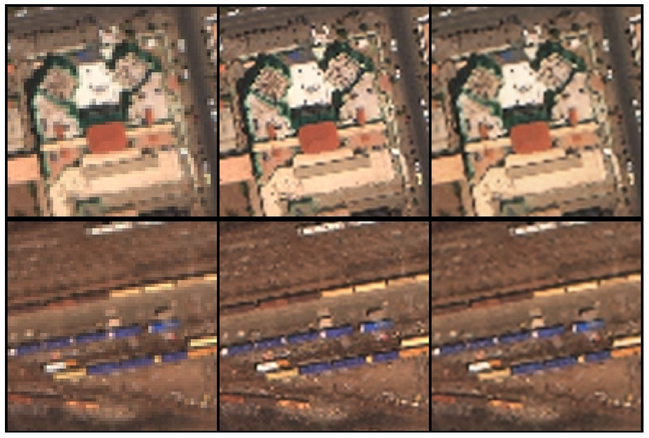


Figure 4. From left to right: the low resolution images, the original high resolution images, the super resolved images by the SRGAN.

Test Results

	#	Upsampled mAP	Original mAP	Generated mAP	Downsampled mAP
Plane	32	0.308709	0.590898	0.609714	**0.788502**
Building	7076	0.239417	0.190386	**0.328532**	0.18349
Yacht	86	0.0344961	**0.238569**	0.0307962	0.0523159
Car	840	0.0999877	**0.207925**	0.0285936	0

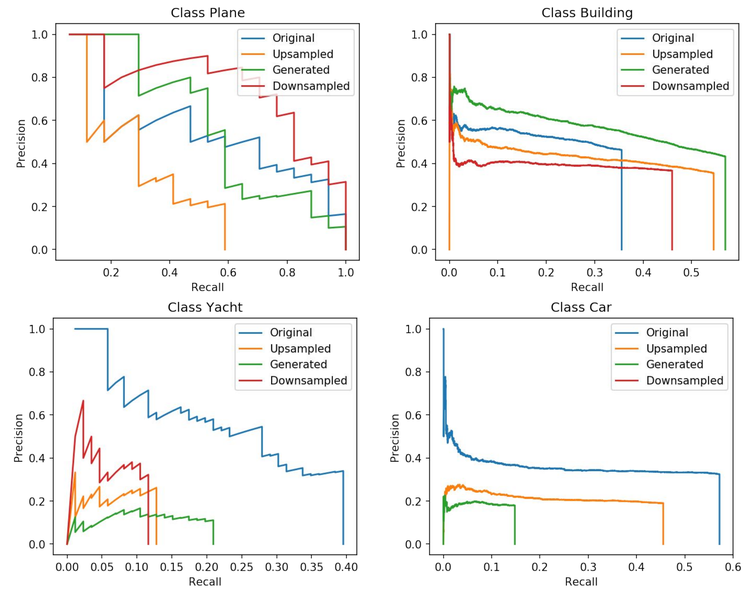


Figure 5. PR curves of YOLO detection for different classes of objects.

Contact

yifei.wang828@duke.edu
ruiqi.wang@duke.edu
zisheng.chang@duke.edu

References

[1] Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav Bulatov, and Brendan McCord. xview: Objects in context in overhead imagery. 02 2018.

[2] Image source: Efficient Implementation of MobileNet and YOLO Object Detection Algorithms for Image Annotation. https://hackernoon.com/efficient-implementation-of-mobilenet-and-yolo-object-detection-algorithms-for-image-annotation-717e867fa27d

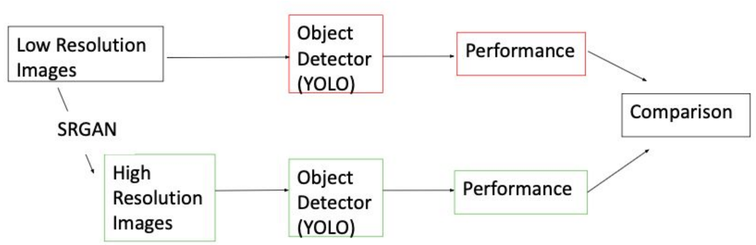
[3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo- realistic single image super-resolution using a generative adversarial network. pages 105–114, 07 2017.

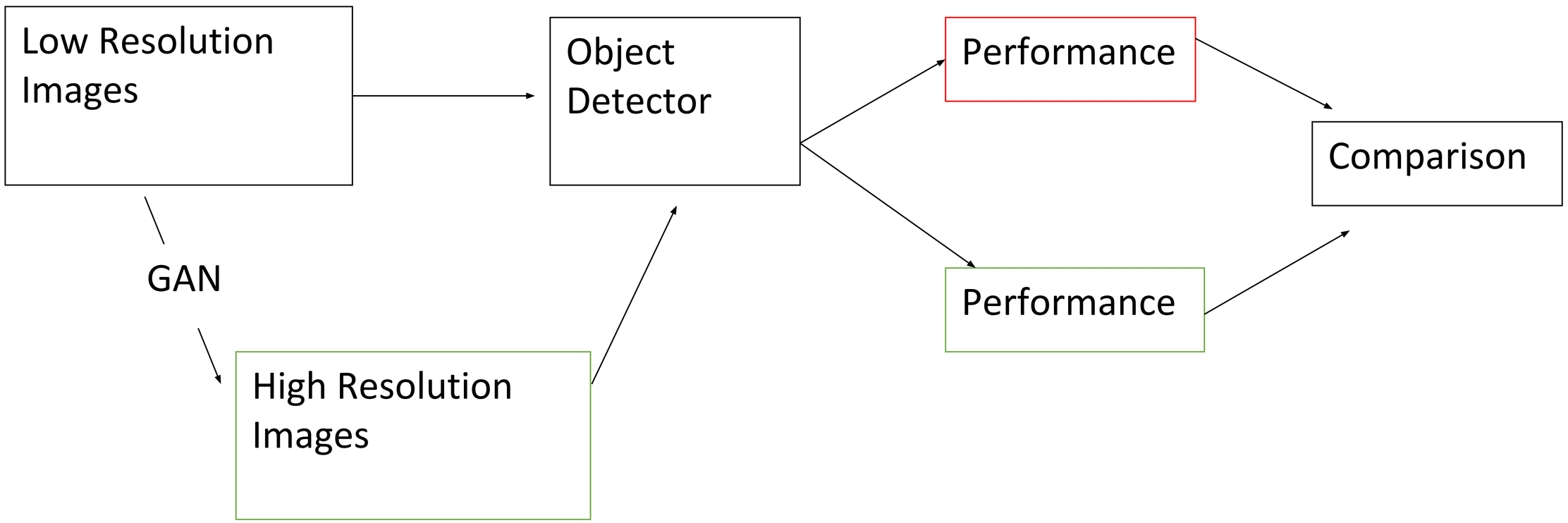
Analysis of Utilization of Generative Models to Increase Image Quality for Object Detection in Satellite Imagery

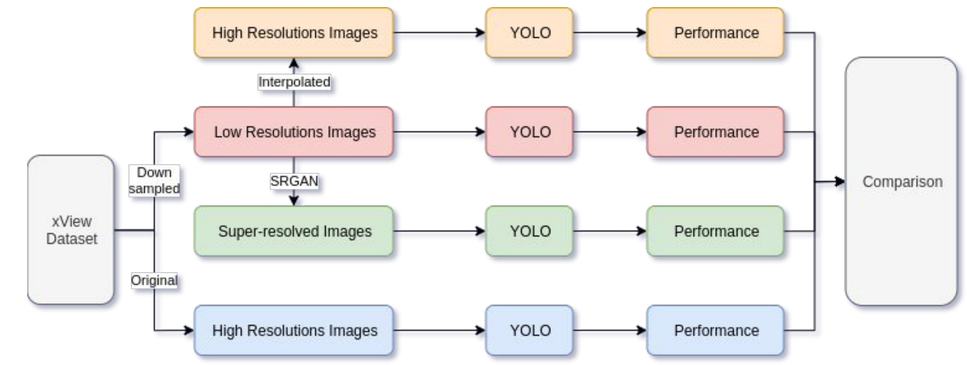
Jason Zisheng Chang, Yifei Wang, Ruiqi Wang

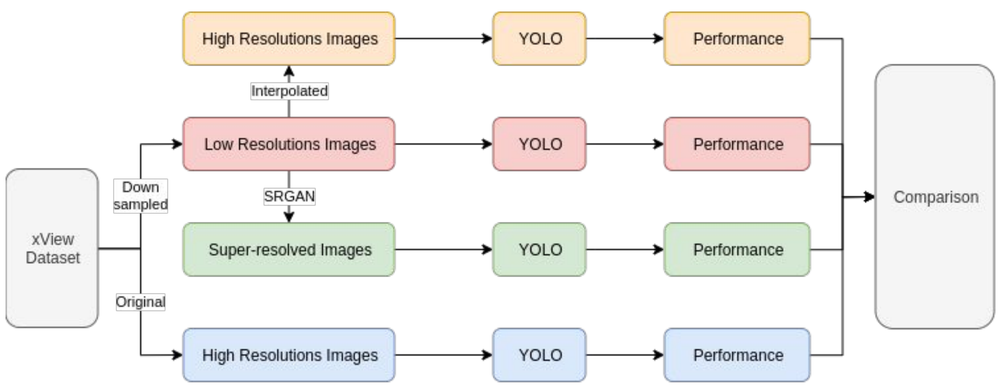
Project Goal

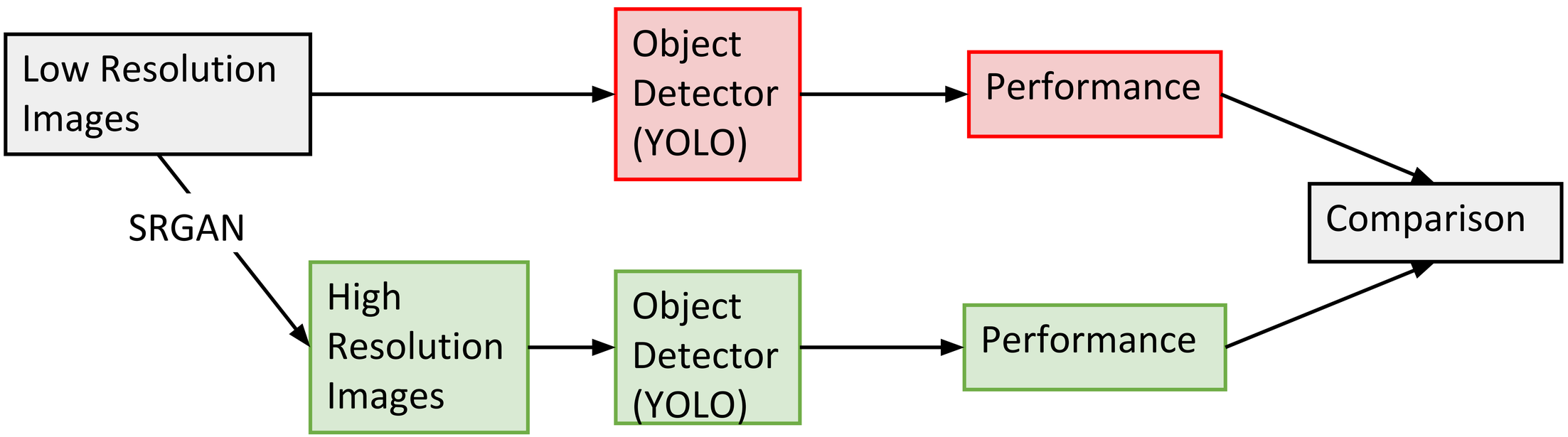
Observe the performance gains of using a generative adversarial networks (GAN) to increase the quality of satellite imagery for object detection






High Resolutions Images
YOLO
Performance
Interpolated
Low Resolutions Images
YOLO
Performance
Down sampled
SRGAN
xView Dataset
Super-resolved Images
YOLO
Performance
Comparison
Original
High Resolutions Images
YOLO
Performance


High Resolutions Images
YOLO
Performance
Interpolated
Low Resolutions Images
YOLO
Performance
Down sampled
SRGAN
xView Dataset
Super-resolved Images
YOLO
Performance
Comparison
Original
High Resolutions Images
YOLO
Performance


Low Resolution Images
Object Detector (YOLO)
Performance
SRGAN
Comparison
High Resolution Images
Object Detector (YOLO)
Performance

High Resolution Images

Object Detector (YOLO)

Performance

Low Resolution Images

Object Detector (YOLO)

Performance

SRGAN

Genrated high Resolution Images

Object Detector (YOLO)

Performance

High Resolution Images

Object Detector (YOLO)

Performance

Comparis

Data Overview

- Source
    - satellite imagery from the xView Dataset
    - xView: one of the largest publicly available datasets containing annotated images from complex scenes
- Image quality
    - 0.3 m/px
    - ~3200 pixels × ~3000 pixels per image
- Size
    - Train: 50 images
    - Test: 11 images

Data Overview

Image Contents in Training Set (Partial)

Class Label	Number of Ground Truth
Building	59919
Small Car	1951
Bus	1254
Cargo Truck	805
Utility Truck	463
Vehicle Lot	456
Trailer	423
Truck	401
Facility	240

Image Contents in Testing Set (Partial)

Class Label	Number of Ground Truth
Building	7076
Small Car	840
Truck	314
Bus	238
Trailer	139
Yacht	86
Cargo Car	67
Cargo Truck	54
Construction	53

Models Overview

- Object Detection
  - The model of You Only Look Once (YOLO) is used for object detection
  - YOLOv3 uses a single convolutional network to simultaneously predict multiple bounding boxes on and class probabilities for these boxes

- Super Resolution
  - The model of Super resolution generative adversarial networks (SRGAN) is used for super resolution
  - An SRGAN is a GAN that aims to increase the resolution of images, with its discriminator and generator being convolutional neural network (CNNs).
  - The super resolution process synthesizes sub pixel information in LR imagery to generate super-resolved images

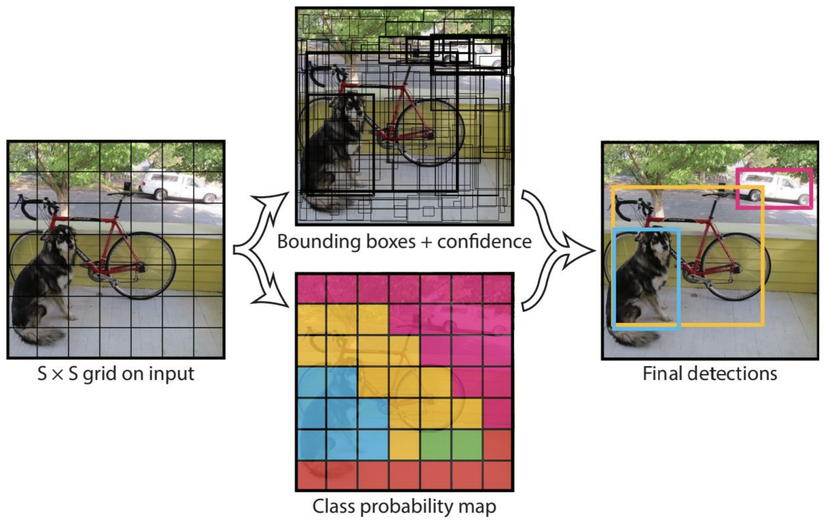
YOLO - Model Architecture

Resizes the input image into a grid

Runs a single convolutional neural network (CNN) on the inputs

Thresholds the resulting detections by the model confidence

in the detection



SRGAN - Method

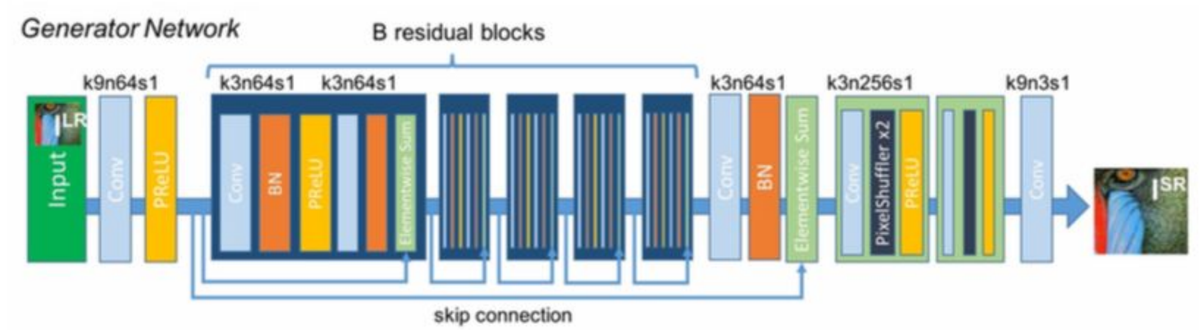
- A super resolution generative adversarial network (SRGAN) is a GAN that aims to increase the resolution of low-resolution (LR) images by learning the corresponding high-resolution (HR) images.

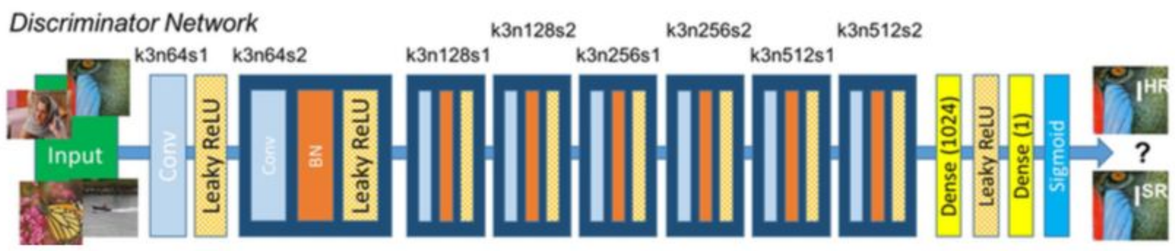
- The super resolution process synthesizes sub pixel information in LR imagery to generate super-resolved (SR) images.

- The LR images are obtained by downsampling operation with downsampling factor r = 4.

- For an image with C color channels, we describe a LR image by a real-valued tensor of size W × H × C, and HR/SR image by rW × rH × rC respectively.

SRGAN - Model Architecture





SRGAN - Loss Functions for generator

Generator loss consists of content loss, adversarial loss and total variation loss. The weights are chosen to make them comparable to each other.

$$l^{SR} = l^{SR}_{MSE} + 6 \times 10^{-3} l^{SR}_{VGG/i,j} + 10^{-3} l^{SR}_{Gen} + 2 \times 10^{-8} l^{SR}_{TV}$$

Where

$$l^{SR}_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y})^2$$

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

$$l^{SR}_{TV} = \frac{1}{rH(rW-1)} \sum_{x=1}^{rW-1} \sum_{y=1}^{rH} |G_{\theta_G}(I^{LR})_{x+1,y} - G_{\theta_G}(I^{LR})_{x,y}|^2 + \frac{1}{rW(rH-1)} \sum_{x=1}^{rW} \sum_{y=1}^{rH-1} |G_{\theta_G}(I^{LR})_{x,y+1} - G_{\theta_G}(I^{LR})_{x,y}|^2$$

$$l^{SR}_{Gen} = 1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))$$

SRGAN - Training

- Chip original images (~3200 × ~3000) to 6686 training and 1575 testing HR images (256 × 256)
- Downsample HR images to LR images (64 × 64) by a factor of r = 4
- i = 5 and j = 4 for the VGG-19 feature map content loss
- Adam optimizer with beta_1 = 0.9 and beta_2 = 0.999.
- Trained for 40 epochs with batch size 1 (GPU memory restriction)

SRGAN - Model Performance

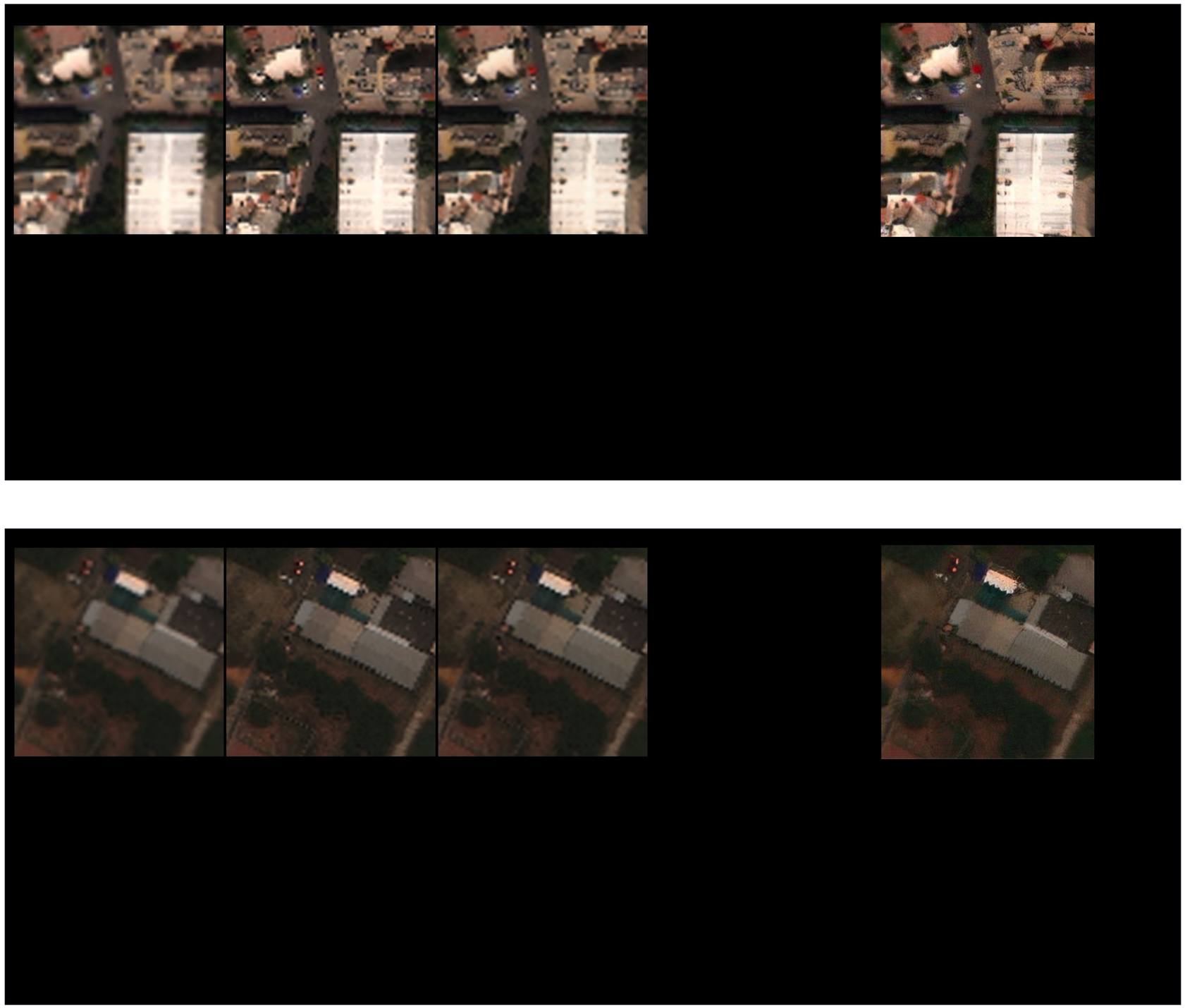
We use two following scores to evaluate the performance of the SRGAN.

- scores of structural similarity (SSIM)
- peak signal-to-noise ratio (PSNR)

	Train-SRGAN	Test-SRGAN	Train - API	Test - API
Number of Image Input	6686	1575	6686	1571
SSIM	0.9451	0.9541	0.7931	0.8065
PSNR	36.1508	38.1244	30.1537	31.6175

The results show that our SRGAN system is able to derive well super resolved images comparing to the super resolution API.

SRGAN - Model Performance



Two random examples of model output.
From left to right:
the low resolution images,
the original high resolution images,
the super resolved images by SRGAN,
the super resolved images by API.

Experimental Results

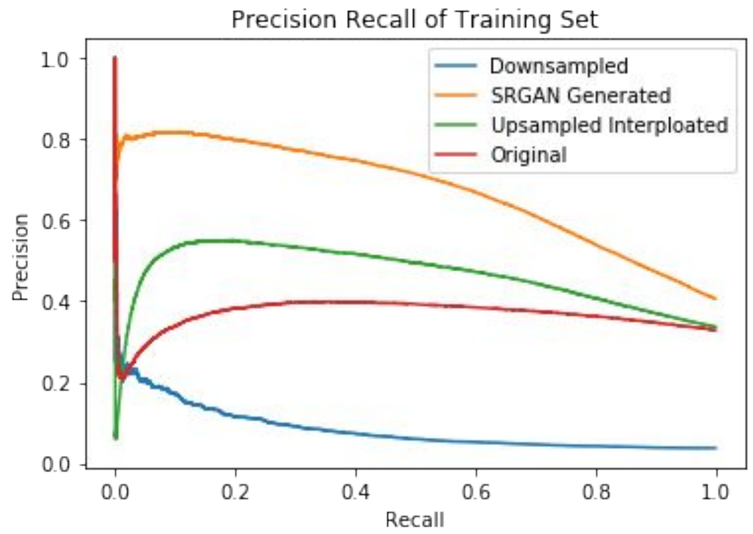
For the experiment we ran the YOLO model on five different input datasets:

1. Original 0.3 m/px
2. Downsampled 1.2 m/px
3. Interpolated Upscaled 0.3 m/px
4. SRGAN Generated 0.3 m/px
5. Super-Resolution API 0.3 m/px

In general, the object detection model trained on the generated dataset was able to outperform the models trained on the interpolated and original images

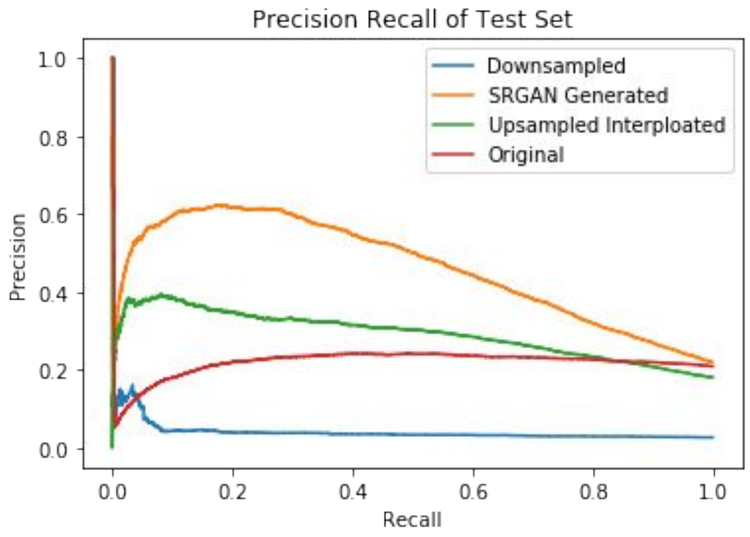
Experimental Results - Train

Train Dataset	**mAP**	**mAR**	**F1**
Original	0.1079	0.4162	0.2081
Downsampled	0.0507	0.1660	0.1789
Interpolated Upscaled	0.1198	0.3486	0.3094
SRGAN	0.0763	0.3377	0.2516
API	0.0753	0.3711	0.2493



Experimental Results - Test

Test Dataset	**mAP**	**mAR**	**F1**
Original	0.1146	0.4828	0.2087
Downsampled	0.0837	0.2821	0.1963
Interpolated Upscaled	0.1130	0.4133	0.2445
SRGAN	0.1202	0.5024	0.1771
API	0.0844	0.1497	0.2246



Example - Yacht Prediction Test Results

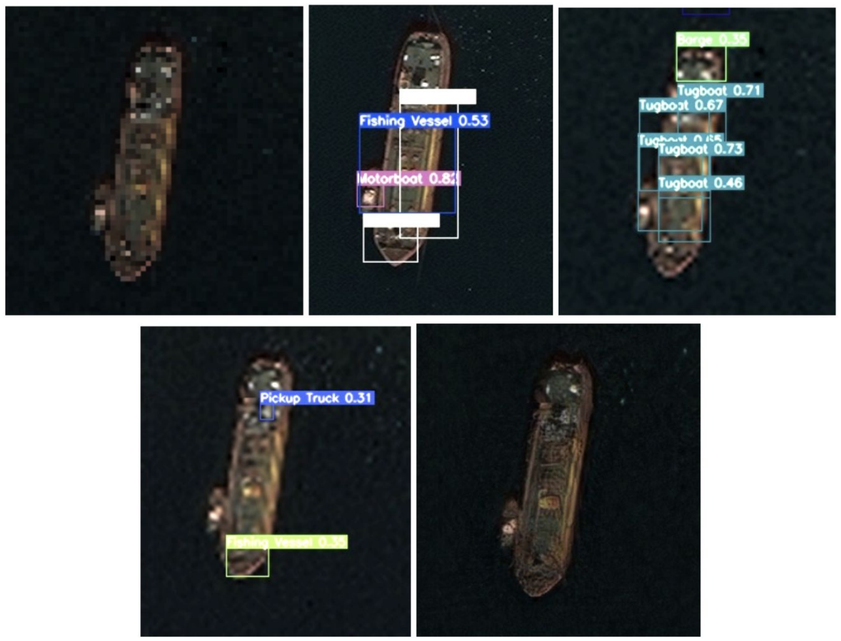
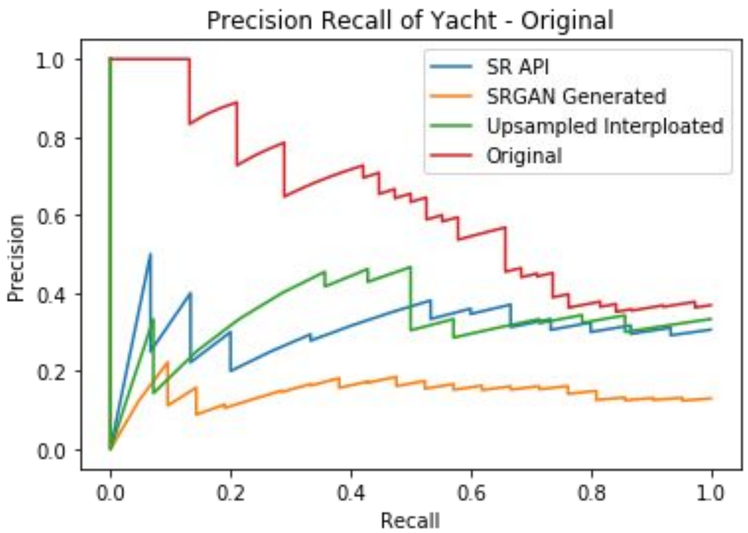


Figure 8: Samples of Yacht Predictions. Clockwise from top left: Downsampled,Original, SR API, SRGAN, Upsampled



Example - Building Prediction Test Results

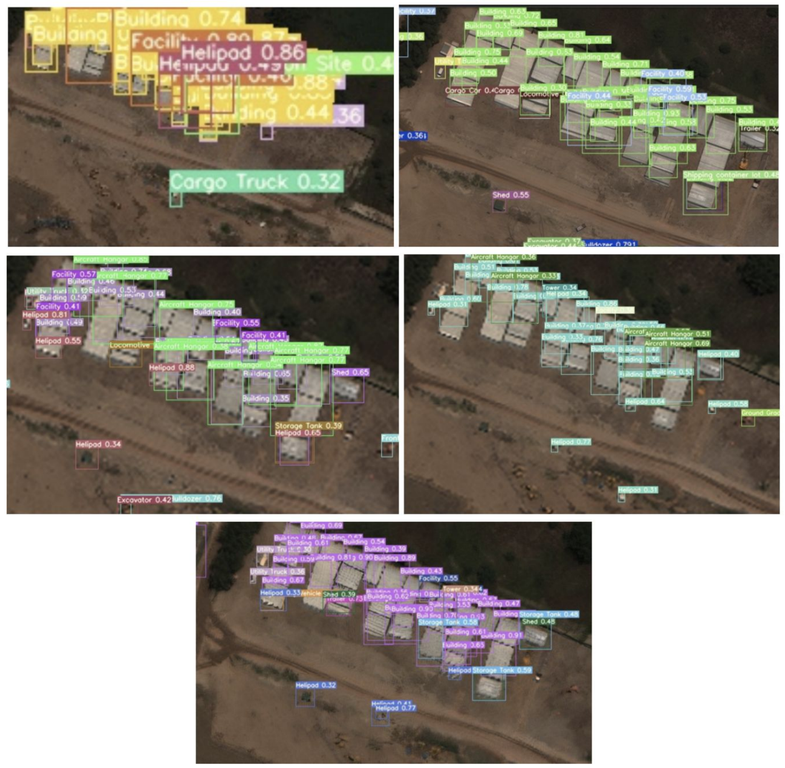
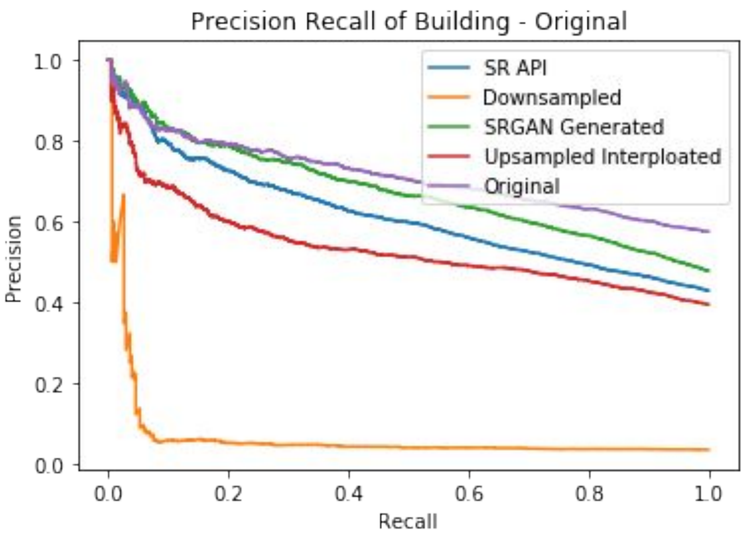


Figure 7: Predictions of buildings in xView dataset. Clockwise from top left: Downsampled,Original, SR API, SRGAN, Upsampled



Conclusion

- The generated images from the SRGAN was able to generate high resolution imagery for unseen low resolution images

- A YOLO object detection model trained on generated data was able to outperform a model trained on the original dataset

- Future work
  - Train the dataset on larger sample sizes
  - Compare the performance on even more subsampled datasets (10 m/px)

References

[1] Hussein Aly and Eric Dubois. Image up-sampling using total-variation regularization with a new observation model. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 14:1647–59, 11 2005.

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. pages 105–114, 07 2017.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. pages 779–788, 06 2016.

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556, 09 2014.

[5] Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav

Bulatov, and Brendan McCord. xview: Objects in context in overhead imagery. 02,2018.