Building footprint extraction based on RGBD satellite imagery

Andrey Syrov(asyrov@stanford.edu) Denis Ulanov (dulanov@stanford.edu)
Department of Computer Science, Stanford University

Presentation video link: https://www.youtube.com/watch?v=KzGsvpNMd5Q

Introduction

Motivation

- Buildings recognition is important for humanitarian and disaster response, commerce and defense applications
- Currently mostly done manually

Goal

- To train models that can recognize building footprints on satellite images
- **Input**: RBG satellite image with depth and terrain info
- **Output**: buildings boundary masks over the input image

Approach

- Mask-RCNN
- Conditional GAN

Dataset

Source

- USSOCOM Urban dataset, images of Tampa, FL and Jacksonville, FL
- 174 tiles in train set and 31 tiles in validation set, each 2048x2048 pixels
- Each tile contains RGB, surface height, terrain heights and ground truth with building labels
- Up to a few hundred buildings on each tile

Pre-processing

- Split into smaller images 256x256 pixels for faster training
- Rescaled and cleaned up terrain and surface inputs
- Data augmentation: vertical and horizontal flip, scale, rotate

Dataset size

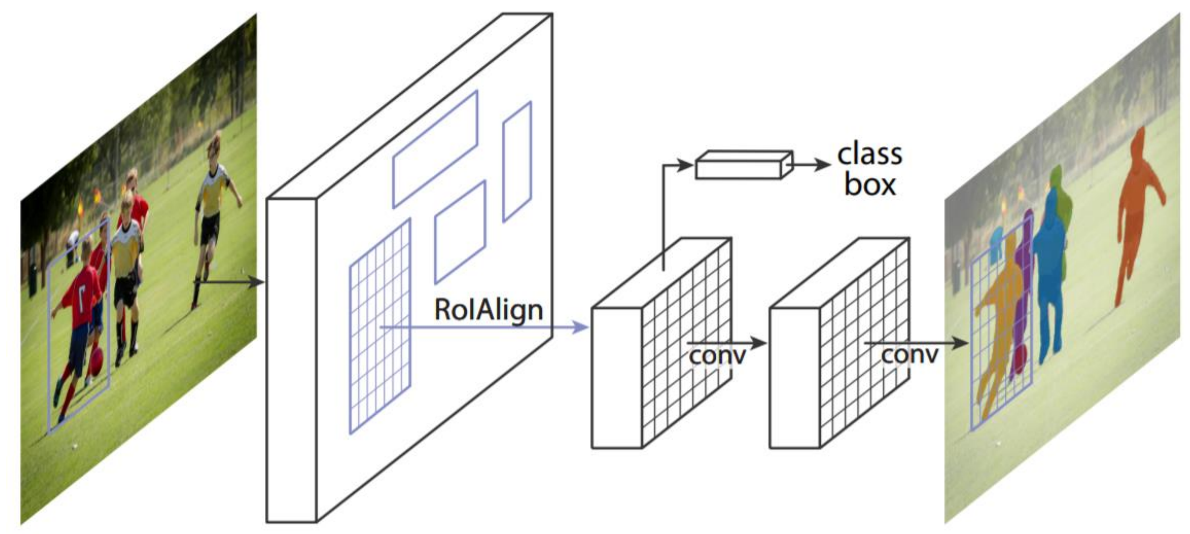
- Training: 11136 samples
- Dev: 1984 samples
- Test: 1984 samples

Methods

1. Mask-RCNN

State-of-the-art approach for instance segmentation problems. Model consists of the following parts:

- Backbone CNN for feature extraction (ResNet50) with Feature Pyramid Network on top
- Region Proposal Network to find anchor boxes
- ROI Classifier to propose regions-of-interest that may have classifying objects
- Segmentation Masks to identify objects masks



2. Conditional GAN

Discriminator task to recognize if the input is real building footprint or generated. Generator network task is to fool (maximize loss of) discriminator and also come close to ground truth image of building footprint.

Generator architecture: U-Net type encoder-decoder network with skip connections between corresponding layers in encoder and decode

Discriminator architecture: PatchGAN, multiple conv-relu-batchnorm layers. Models image as Markov random field.

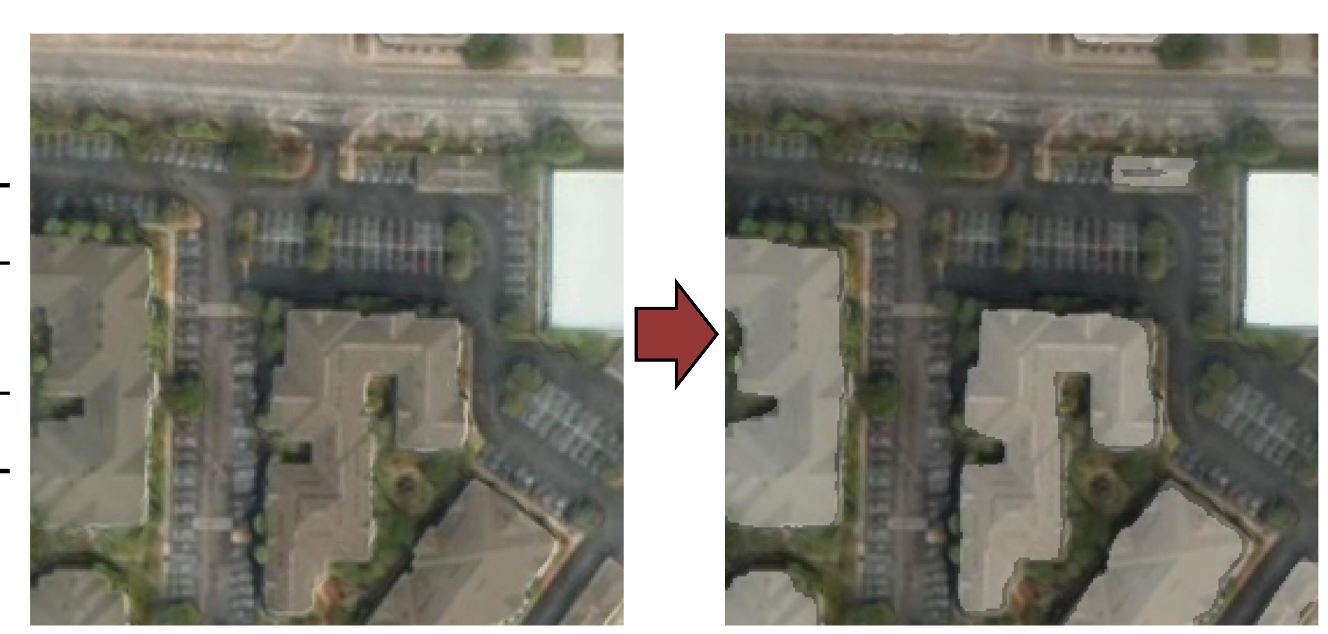
Results

1. Mask-RCNN

	Precision	Recall	F1	IoU
train	0.7734	0.7273	0.7496	0.6463
dev	0.7857	0.7127	0.7474	0.6506
test	0.7794	0.7278	0.7527	0.6411

2. Conditional GAN

	Precision	Recall	F1	IoU
train	0.8662	0.7826	0.8127	0.6982
dev	0.8518	0.7676	0.7956	0.6781
test	0.8437	0.7670	0.7899	0.6703



Evaluation

Metrics

- Define positive detection as IoU with ground truth higher than 0.5
- Compute Precision, Recall and F1 based on true positives, false positives and false negatives
- Compute average IoU with ground truth

Visualization

- Visualized results with saliency and occlusion maps

Conclusion

Conclusions

- Implemented 2 approaches to detect building footprints on satellite images
- Used variety of techniques to improve performance, such as transfer learning, augmentation, extra features and hyperparameters search
- Achieved performance comparable to human-level performance (F1=78.4%) for both models

Next Steps

- Train the model on bigger tiles to avoid misclassification of partially appearing buildings
- Further improvement is difficult cause we are close to human-level performance already

References

[1] He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn (2017). URL https://arxiv.org/abs/1703.06870.
[2] Isola, P., Zhu, J., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks (2016). URL http://arxiv.org/abs/1611.07004. 1611.07004.
[3] SpaceNet. Spacenet on amazon web services (aws) (2018). URL https://spacenetchallenge.github.io/datasets/datasetHomePage.html.
[4] Abdulla, W. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN (2017).
[5] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation (2015). URL http://arxiv.org/abs/1505.04597. 1505.04597.
[6] Jordan, J. Evaluating image segmentation models. (2018). URL https://www.jeremyjordan.me/evaluating-image-segmentation-models/.
[7] Pal, S. A guide to understanding convolutional neural networks (cnns) using visualization (2019). URL https://www.analyticsvidhya.com/blog/2019/05/understanding-visualizing-neural-networks/.