Building Detection from Satellite Imagery

Stanford University

Jasmine Hu, Manan Rai, Vivian Wong

CS230
Deep Learning
Spring 2018

Objective

To seek ways to improve the performance of models on building detection and segmentation on satellite images, our goals are to:

- Evaluate the performance of a Mask R-CNN with different backbones such as ResNet and DenseNet
- Evaluate the performance of various additions to a typical UNet architecture, and compare them with the Mask R-CNN
- Explore the effect of morphological pre and post-processing on the performance of a model

Data

Sources:

1. CrowdAI Mapping Challenge
2. Kaggle DSTL Competition

Details:

- RGB images with MS COCO or WKT format shape annotations

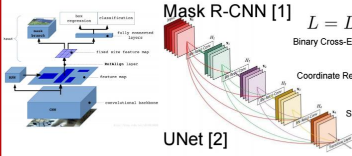
Train/Test/Val Distribution:

- 8366/910/910

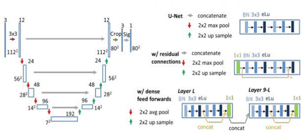
Model

Framework:

Mask R-CNN [1]



UNet [2]



Loss Func.

$$L = L_{cls} + L_{box} + L_{mask}$$

Binary Cross-Entropy

$$H = -\frac{1}{n}\sum_{i=1}^{n}[y\log(\hat{y}) + (1-y)\log(1-\hat{y})]$$

Coordinate Regression

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*).$$

Smooth L1 Loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Binary Cross-Entropy

$$H = -\frac{1}{n}\sum_{i=1}^{n}[y\log(\hat{y}) + (1-y)\log(1-\hat{y})]$$

Generalized Probability

$$J_m(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i \cdot \hat{y}_i}{y_i + \hat{y}_i - y_i \cdot \hat{y}_i}$$

Total Loss

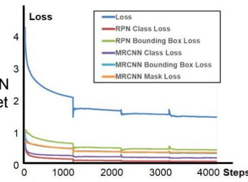
$$L = H - \log J_m$$

Results

I. Mask R-CNN vs. UNet

	Training IoU	Test IoU
Basic Mask R-CNN (trained over 40 epochs)	N/A	0.396
> w/ DenseNet	N/A	0.00023
Basic U-Net	0.792	0.765
> w/ res. con.	0.736	0.705
> w/ dense con.	0.803	0.788

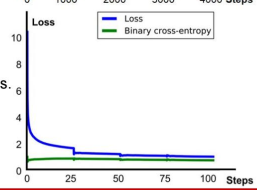
Mask R-CNN w/ DenseNet



Hyperparameter	Value(s)
Learning Rate	0.001
Weight Decay	0.0001
Learning Momentum	0.900
Mini-batch size	5
Steps per epoch	1000
Layers per Dense Block	[6, 12, 24, 16]

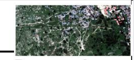
U-Net w/ res. con.

128 samples per step



Hyperparameter	Value(s)
Learning Rate	0.001
beta 1	0.9
beta 2	0.999
Mini-batch size	128
Steps per epoch	25
Contraction Layers	4
Expansion Layers	4
Initial Feature Width	12

2. Morphological Processes (Tested in UNet)

	Predicted Mask	# of Buildings Detected
No Morphology 		121
Dilation in Pre-Processing		101
Dilation in Pre-Processing + Erosion in Post-Processing		103
Watershed (building separation)		625

Mask R-CNN

U-Net w/ res. con.

Discussion

1. Mask R-CNN:

- Baseline with ResNet50 backbone, trained over 40 epochs, performs decently well
- DenseNet121 as backbone gives comparable relative performance after 4 epochs

2. UNet:

- Training performances of networks vary greatly on the features of different datasets
- Dense connections slightly improves the IoU for the CrowdAI dataset by 1%
- Residual connections improve performance on Kaggle dataset by 10%

3. Morphological Processes:

- Dilation in pre-processing improves detection but overlaps buildings
- Erosion in post-processing reduces overlap
- Watershed can separate overlapped masks

Future Work

- Using the U-Net as a backbone for the Mask R-CNN
- Train the Mask R-CNN for longer on the DenseNet and compare performance
- Train longer epochs on the U-Net model, and understand better how connections affect training
- Improve the watershed algorithm to separate overlapping building masks with less tolerance

References: [1] Ildoo Kim,"Deep Object Detectors," 2016. [2]Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation"