Satellite images segmentation for building detection using U-net

Guillaume Chhor, Ianis Bougdal-Lambert, Cristian Bartolome Aramburu
{gchhor, ianisbl, cbartolm}@stanford.edu

MOTIVATION

Building Detection

Monitor nomadic populations, refugee camps and urban development in remote areas

Real-time assessment of building destructions after natural disasters or in conflict zones

Estimate potential yield of installing solar panel on roofs in urban areas

Image segmentation

Classify each pixel

DATA

Satellite images & ground-truth labels

mapbox

OpenStreetMap

- Satellite images
- Crowd-sourced layers
- Dedicated APIs

Data Pipeline

docker

Data pipeline shared by Development Seed and run on a docker image for simplicity

developmentSEED

Data preprocessing

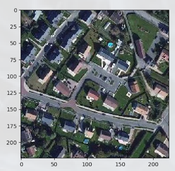
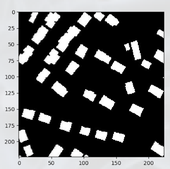
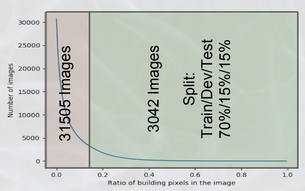


Image RGB
224 x 224



Binary mask
224 x 224

Normalized & Centered

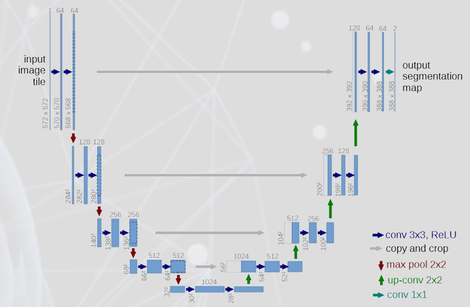


Threshold building pixels per image : 15%

MODEL

U-net

- Contracting path to capture context
- Symmetric expanding path for precise localization
- Combination of high-level features & upsampled output



Modification

- Use of Batch-normalization
- Use of “same” padding to conserve the output dimension
- Remove the 1024-deep convolutional layer to ease optimization (avoid vanishing gradient)

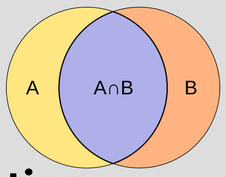
TRAINING

Metric & Loss

Dice coefficient:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} = 2 \times \frac{precision \times recall}{precision + recall}$$

$$Loss = -Dice$$



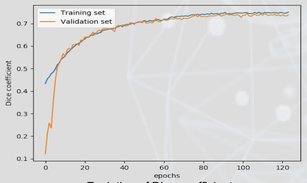
Real time data augmentation

- Cope with a small training set
- Avoid overfitting
- Shifting / Flipping / Rotation

Characteristics

- Adam optimizer
- Decrease learning rate on plateau (α_{init}= 0.001)
- Trained from scratch on GPU w/ batch size of 32 & on ~120 epochs

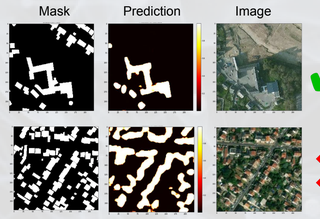
RESULTS



Evolution of Dice coefficient during training

	Training	Validation	Test
Dice	0.7413	0.7428	0.7506

IOU on Test set: **0.601**



Robust building detection (even on mislabeled data)

State of the art results obtained on high resolution satellite images:

	mean IoU	Acc. (Pixel)
Baseline FCN [2]	53.82%	92.79 %
Baseline FCN + MLP[2]	64.67%	94.42 %

Improve the boundaries detection:

We weight positively and negatively the inner and outer border of the building which impacts directly the loss and forces it to be more precise on the boundaries

We can show that a model trained on a Dice index will have this propensity to slightly encompass the positive pixels, resulting in unclear boundaries.

WHAT’S NEXT?

Address further the problem of preserving semantic segmentation boundaries. The approach can rely on the following ideas:

- Extraction of geometric properties (e.g: truncated distance of pixels to boundaries of buildings)
- Encoder-Decoder Network architecture
- Uncertainty Based Multi-Task Loss:

$$L_{total}(x; \theta, \sigma_{dist}, \sigma_{seg}) = L_{dist}(x; \theta, \sigma_{dist}) + L_{seg}(x; \theta, \sigma_{seg})$$