Aerial Image Labelling with Artificial Neural Networks

Microcosme consulting
2020

Agenda

Various Tasks in Computer Vision

Learning from Images : Convolutional Neural Networks

Aerial Image Labelling / Building detection

Various Tasks in Computer Vision

Object Detection

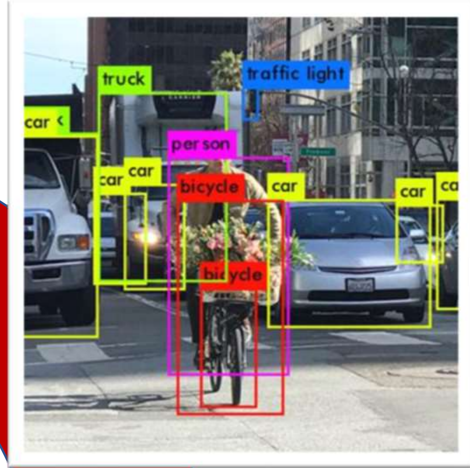
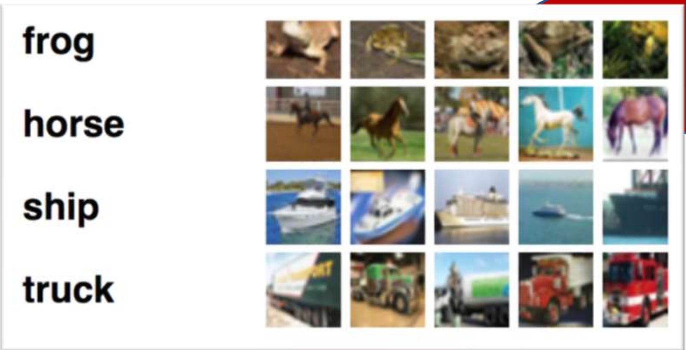
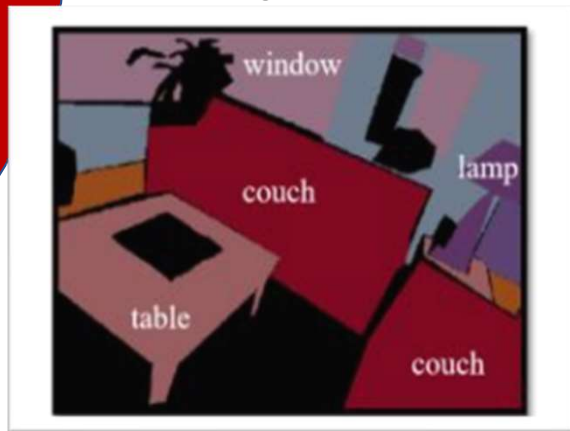


Image Classification



Keypoint Detection

Semantic Segmentation

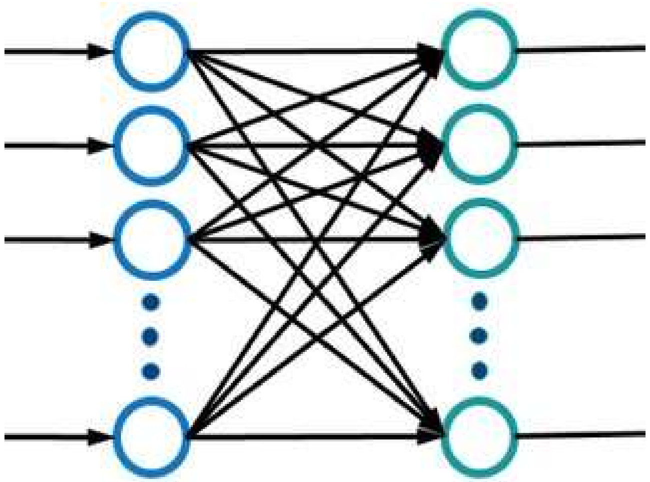


Face Recognition
Denoising, Captioning, etc.

Various Tasks in Computer Vision – Data Dimensions

How to encode these tasks into a neural network?

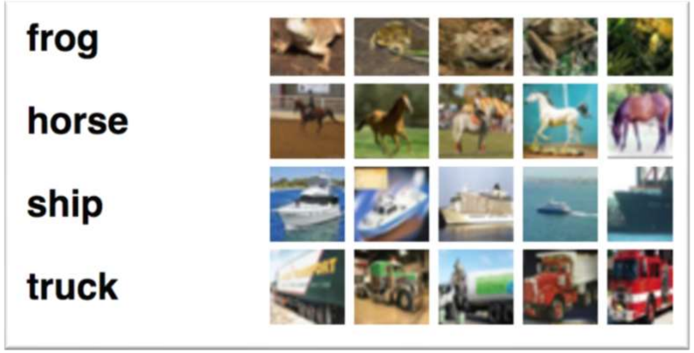
Input size?



Output size?

Various Tasks in Computer Vision – Data Dimensions

Image classification



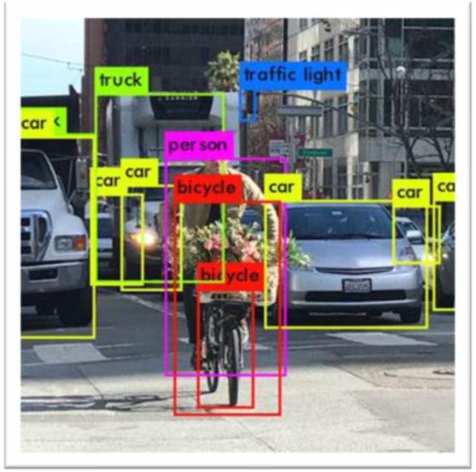
Input : Images size is *width* X *height* X *nbr channels* (channels: 3 colors if RGB).

Output : For each object category: *Yes/No* the image belongs to the category.

Yes/No statement can be inferred from a probability [0,1].

Various Tasks in Computer Vision – Data Dimensions

Object Detection



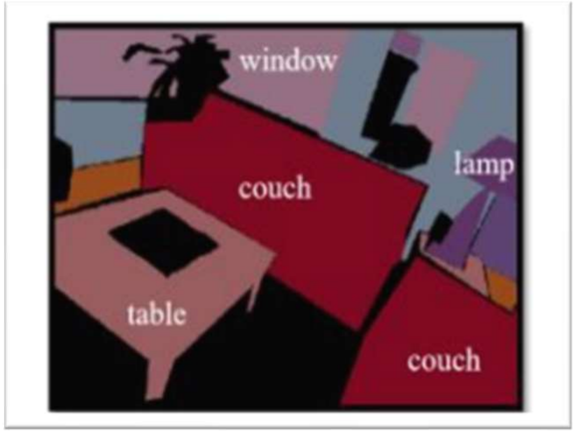
Input : Images size is *width* X *height* X *nbr channels* (channels: 3 colors if RGB).

Output : For each object category:

- *Yes/No* the object is present;
- Coordinates of center point *(x_0, y_0);*
- *Width* and *height* of bounding box.

Various Tasks in Computer Vision – Data Dimensions

Semantic Segmentation



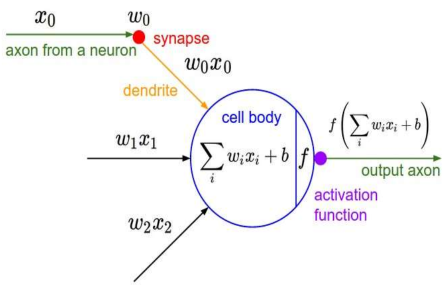
Input : Images size is *width* X *height* X *nbr channels* (channels: 3 colors if RGB).

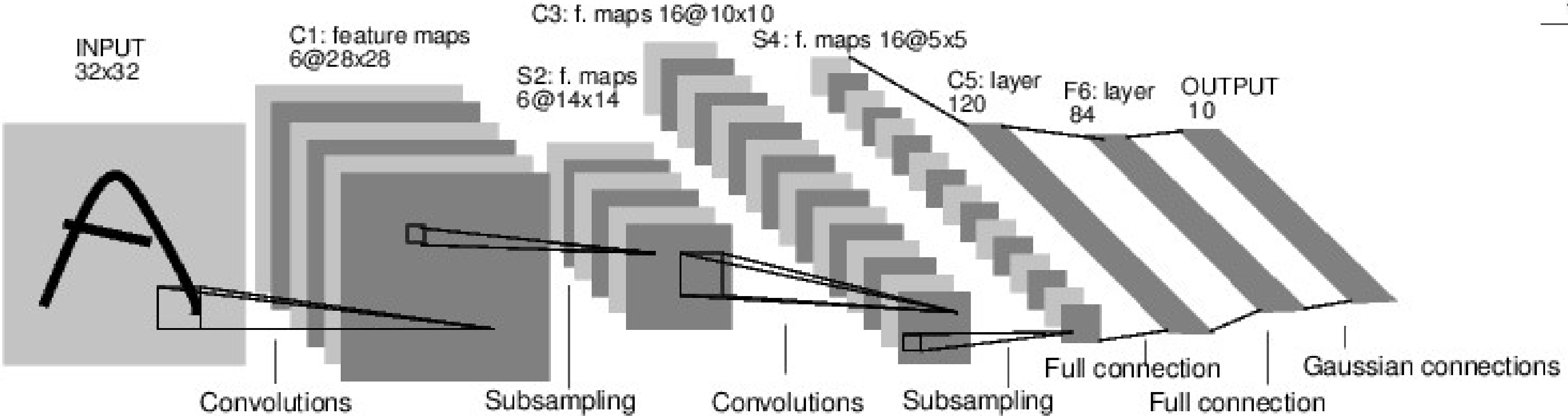
Output : For each pixel, for each category: *Yes/No* the pixel belongs to the category.

Learning from Images : Convolutional Neural Networks (CNN)

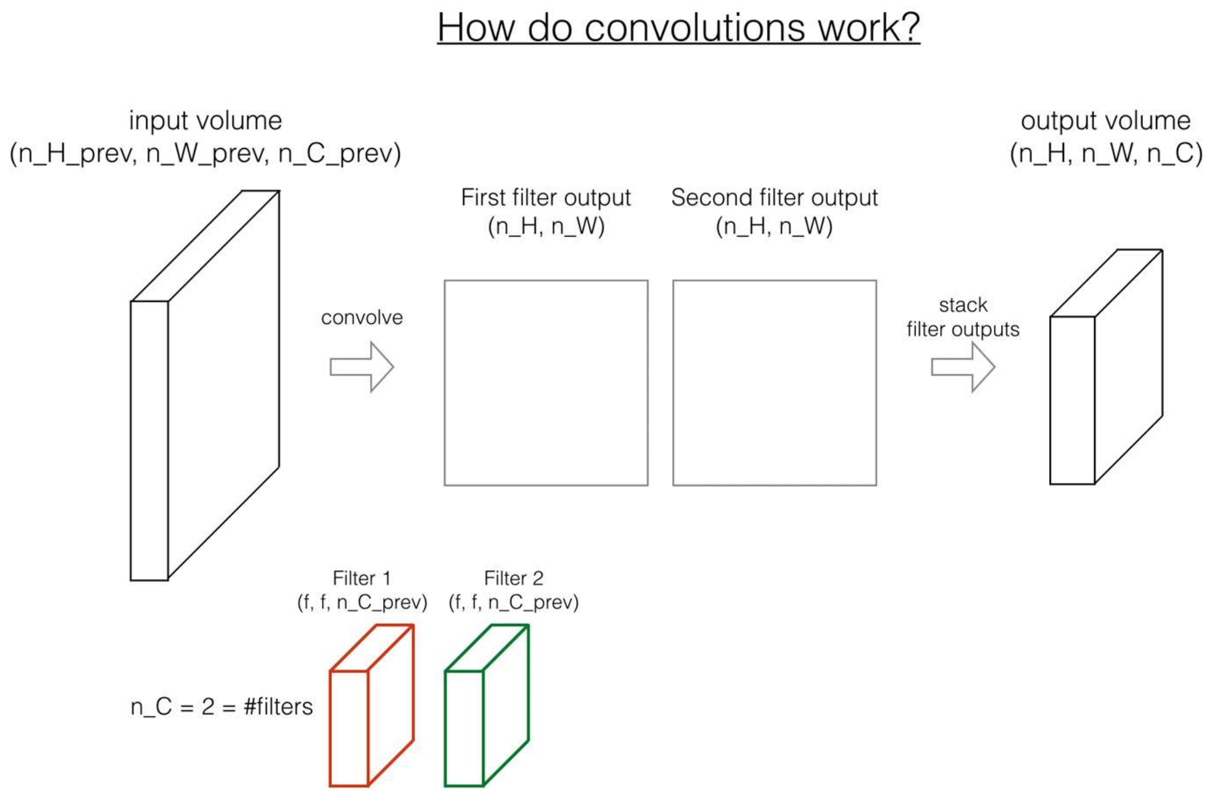
Convolution is spanning on a range of input data to produce the resulting output. The same convolutional filter is applied translationally to the full image.

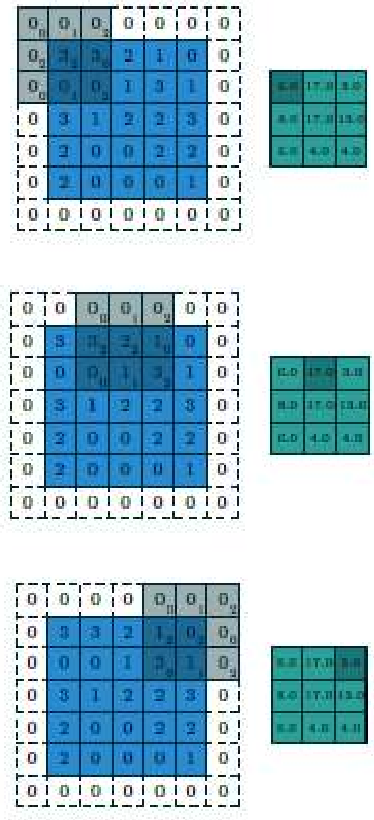
Standard inner neuron structure is kept.





Learning from Images : Convolutional Neural Networks (CNN)

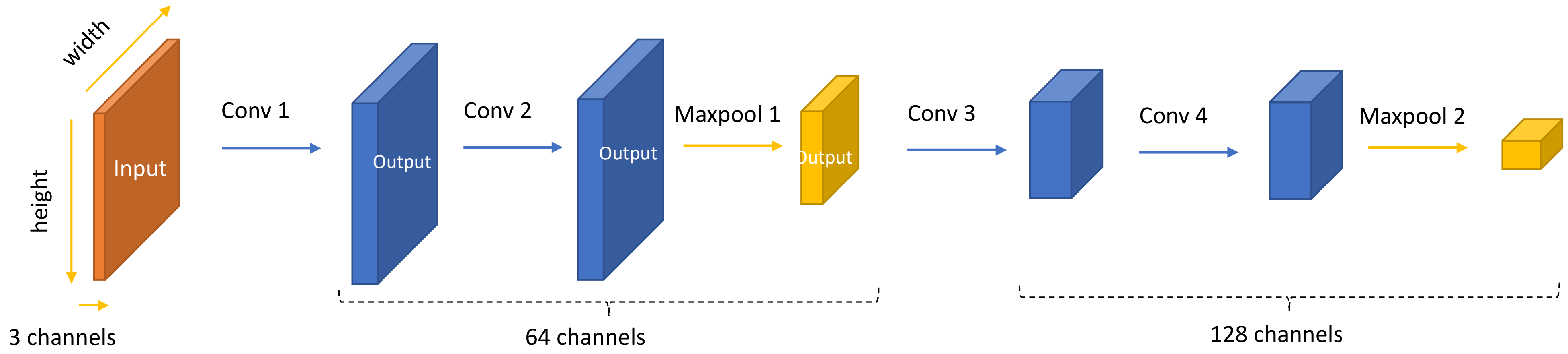




Learning from Images : Convolutional Neural Networks (CNN)

A typical building block architecture for CNN is involving

- repeated convolutional layers
  - conserving input width & height ;
  - but increasing number of channels/filters;
- followed by a width & height shrinking layer (« Max pooling »: max of *n* x *n* area).

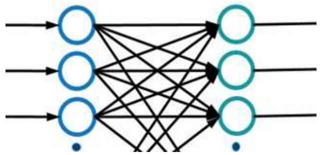


Learning from Images : Semantic Segmentation

Pixel to pixel identification uses

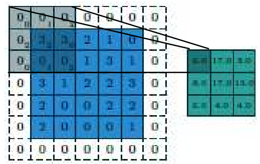
Fully Convolutional Neural Networks

No « dense » layers



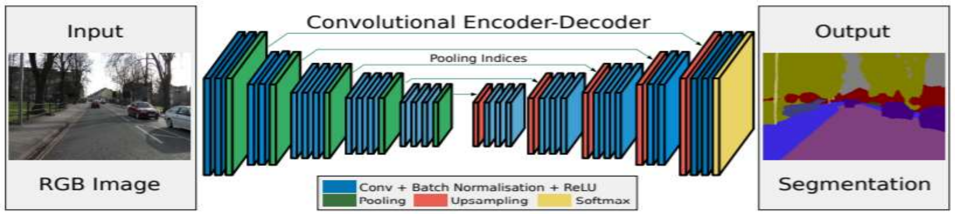
=

Only convolutional layers



=

Encoder-decoder architecture



Aerial Image Labelling / Building detection

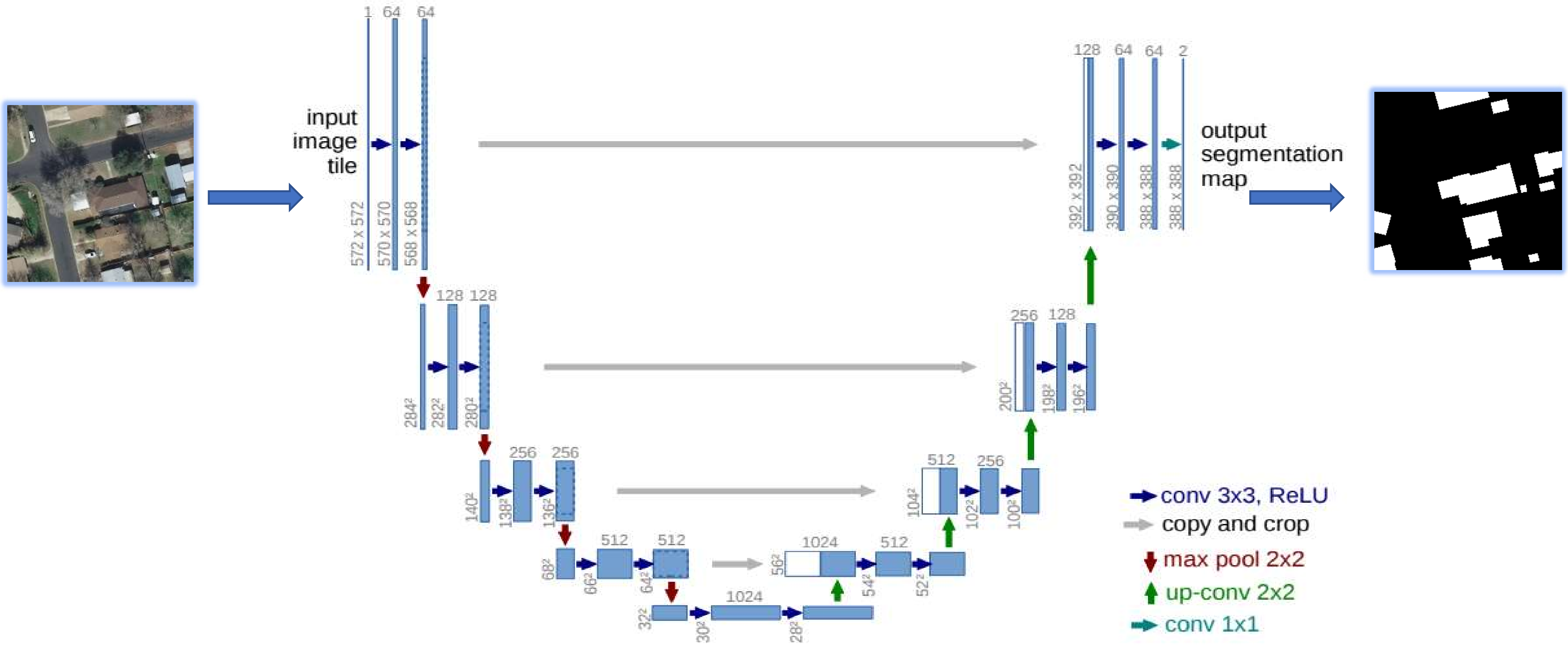
Building detection from aerial images is an example of semantic segmentation.
This application can be used for:

- rooftop sizing for solar panel potential estimate;
- electrical grid design;
- etc.

Bellingham Innsburck San Francisco Tyrol Chicago

[Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark;](#) Emmanuel Maggiori ; Yuliya Tarabalka ; Guillaume Charpiat ; Pierre Alliez; 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)

Aerial Image Labelling / U-Net Architecture

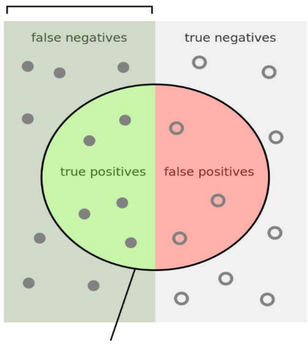


Ronneberger, Olaf; Fischer, Philipp; Brox, Thomas (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". arXiv:1505.04597 [cs.CV].

Aerial Image Labelling / Choose your metric!

Accuracy is a standard metric in classification problems.

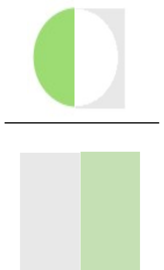
Pixels being **actually** from a building (ground truth)



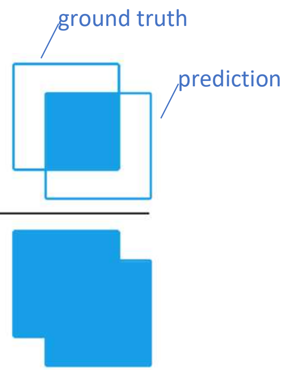
Pixels **predicted** as building

How many items are correctly classified ?

Accuracy =

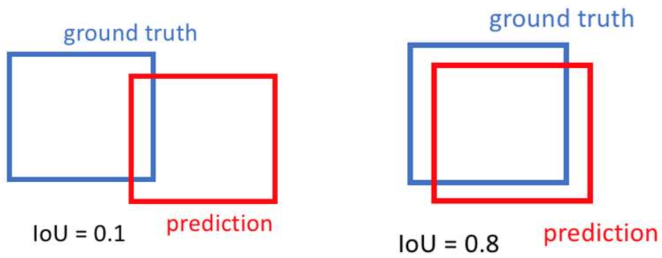


Intersection over Union (IoU) is a metric adapted for object detection.



$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Examples



Aerial Image Labelling / A Computing Challenge

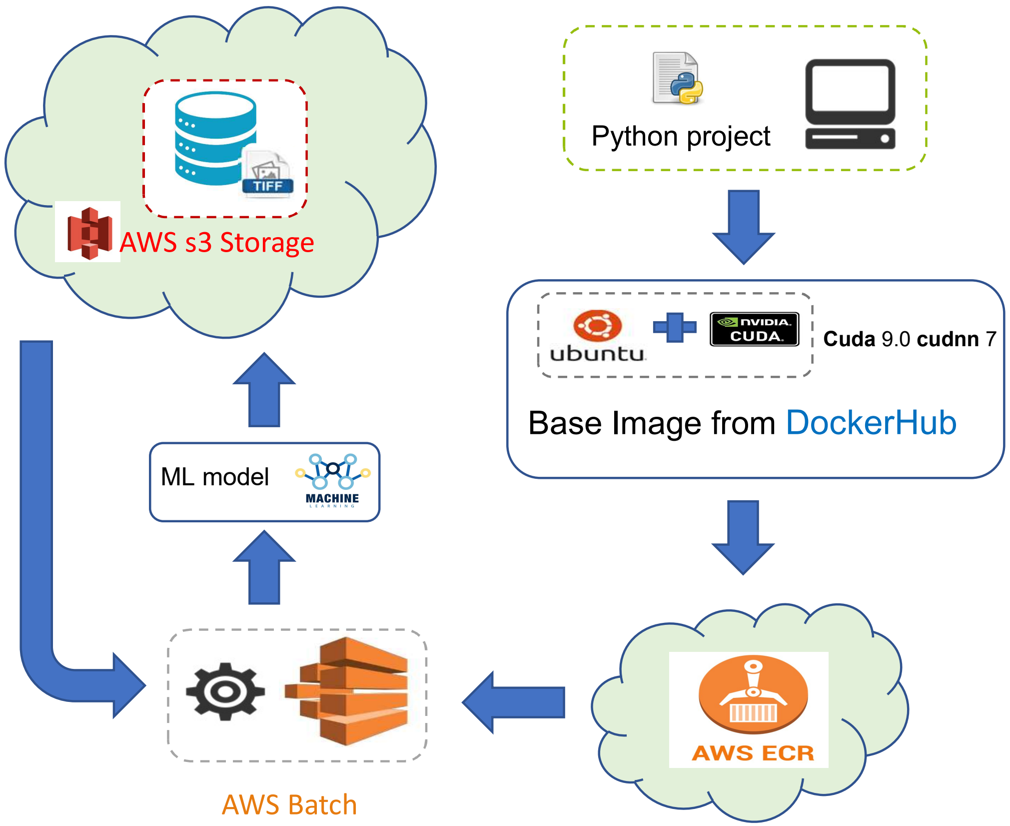
1. Training Data Set (INRIA):
   - 180 tiles with 0.3m resolution, 405 km^2;
   - Each tile $5000px^2$ (72Mb/image);
   - Divided in tiles of $256px^2$

→ ~65 000 images

2. Model Parameters:
   - 22 convolutional layers;
   - 18 batch normalization layers;
   - Total params: 3,842,241;
   - Trainable params: **3,839,297**;
   - Non-trainable params: 2,944;

Efficient computation requires vectorization of calculus for a set of sample (the larger the better).
Vectorization is best performed on GPU units.
Hardware is quickly a limiting factor (Performance / Out of memory error).

Aerial Image Labelling / When local resources are not enough



- Working outside your local resources needs to have a deployment strategy that is as simple as possible.
- **Automation** of the process is key! Production grade solutions are required.
- **Docker** is used for images/containers management;
- **AWS ECR**: Containers repository;
- **AWS Batch** to run heavy jobs (hardware virtualization: CPU/GPU, RAM/HDD);
- **AWS S3** for flexible storage: Train data set as input; Models as output.

Aerial Image Labelling / Results

Three data set split scenarios are considered to get:

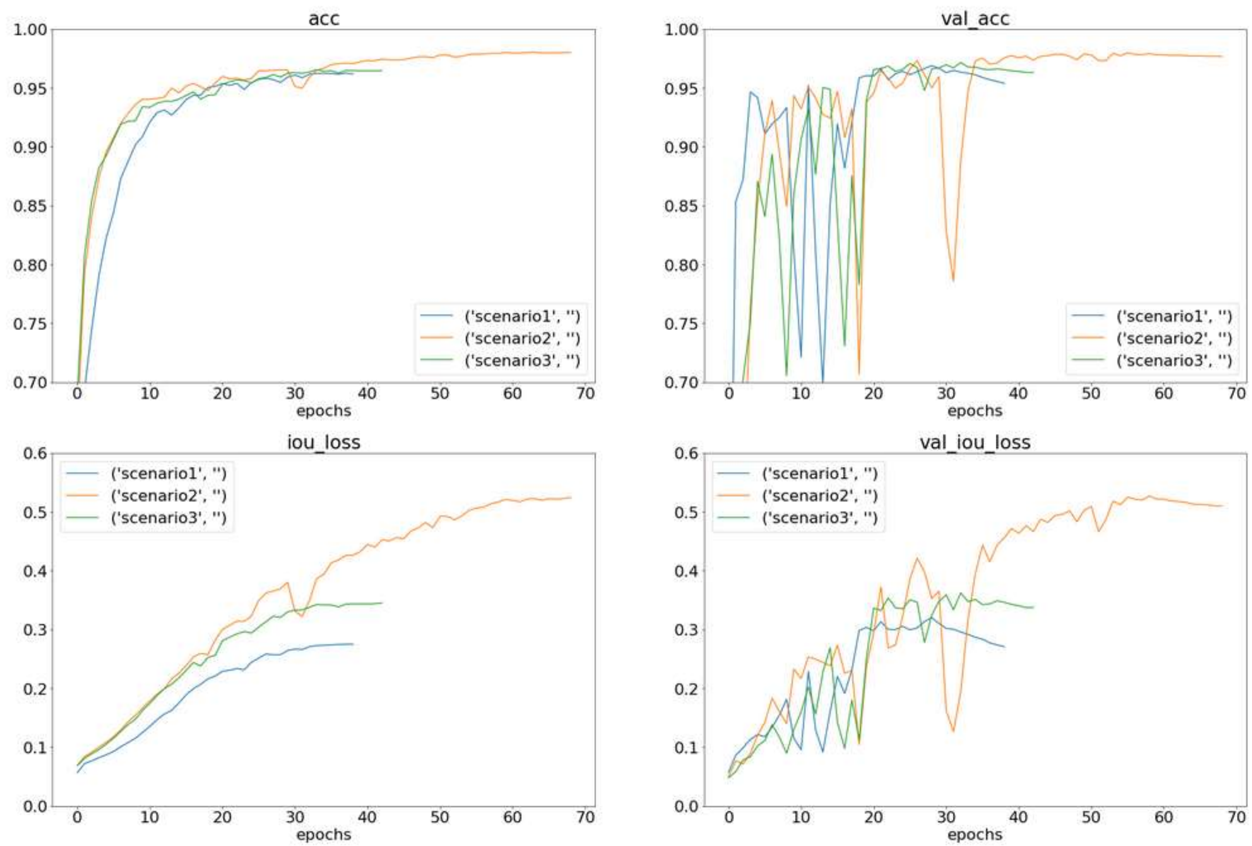
- 15% of data as **test set;**
- 85% of data as **train and validation sets** (80/20% train/val. split).

Figures on the right show the evolution of **accuracy** and **IoU** on the train and validation (val) sets along the training process.

Training is stopped when IoU on validation set is no longer progressing.

Training metrics

Validation metrics



Aerial Image Labelling / Results

Out of sample quality metrics are assessed on the reserved test set.

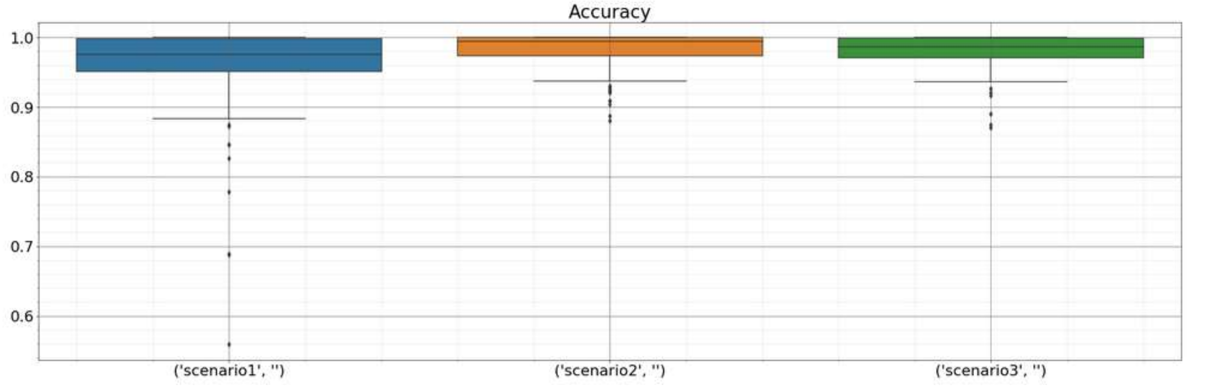
Binarization of the neural network prediction is performed before metrics computation.

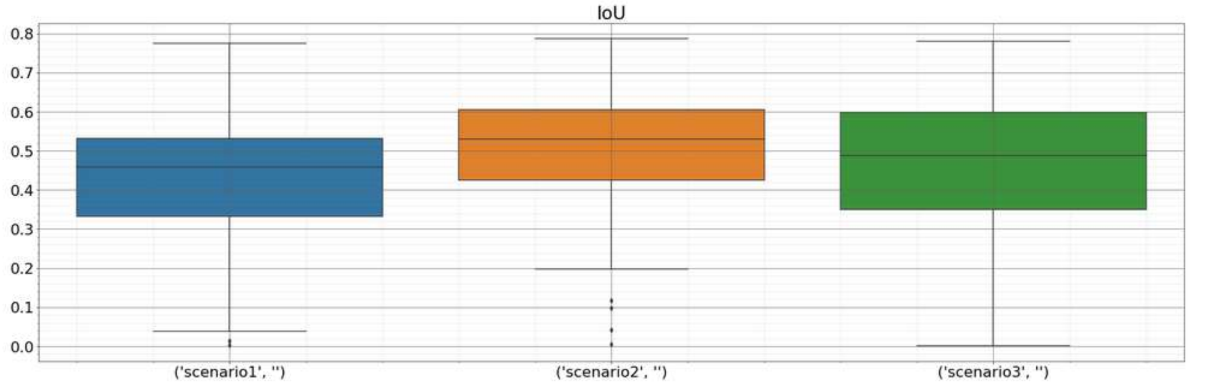
Median out of sample metrics are:

IoU (%)	Accuracy (%)
45,7 - 52,9 %	97,5 - 99,5 %

To be compared with SoTa metrics [2] on developed countries data set of:

	IoU (%)	Accuracy (%)
Baseline U-Net	~ 55 %	~ 93 %
SegNet	~ 70%	~ 95 %





[2] B. Bischke, P. Helber, J. Folz, D. Borth, A. Dengel "Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks". arXiv:1709.05932[cs.CV].

Aerial Image Labelling / REST Service

Once trained, the model is integrated in a service exposing a REST API for real life inference.