GROUP 10 - Dimitri Leandro de Oliveira Silva, Simon Leroy

Detecting various objects on satellite images

Introduction

In this project, we looked at a training dataset containing 3316 satellite images of size 256 x 256 pixels divided into 5 different classes: swimming pools, ponds, solar panels, trampolines, and background (none of the previous ones). The dataset was highly unbalanced: background reached 93% of total observations, trampoline 4%, solar 1%, pools, and ponds did not even reach 1% with respectively 20 and 9 observations.

Our goal was to detect pools, ponds, solar panels, and trampolines that were present on large satellite images (8000 x 8000 pixels) of mostly suburban areas. To do so, we trained different convolutional neural network models in order to find which one would give us the best performance. After that, apply a sliding window to determine the coordinates of the predicted objects.

Methods

Initially, we had to tackle the issue of the highly unbalanced training dataset we were given. Having a huge class imbalance and very few observations for certain classes (only 9 for ponds) would indeed lead to poor accuracy during the image classification. Trying to come over this problem, we manually labeled some of the unlabeled satellite images that we were given to us. To do so, we chipped those large images into 1444 smaller images (256 x 256 pixels) using a script with the sliding window approach. We didn't use more background images as we already had plenty of those. This fairly increased the number of samples we had for the classes *trampoline* and *solar*, but not as much for *pools* and *ponds*.

We still needed more data to train the model. Thus, we had to resort to data augmentation: using the *transforms* module of Torchvision, we managed to increase the dataset by rotating, flipping, blurring, and recoloring already existing images.

We believed that pre-trained models with satellite images would be the most suitable to develop this task. In this sense, we searched for pre-trained models on the internet and found the Satallighte library, which already had two pre-trained models with images from the EuroSat dataset, having 27000 samples divided into 10 classes: Annual Crop, Forest, Herbaceous Vegetation, Highway, Industrial, Pasture, Permanent Crop, Residential, River, SeaLake [1]. Furthermore, these images were presented in 240 x 240 pixels, very similar to the 256 x 256 format of the given images in the training dataset prepared for this task. The chosen model, named MobileNetV2, contains about 2.2 million parameters, with several convolutional, dropout, and batch normalization layers [2].

For the fine-tuning of the model, we carried out the following steps: definition of the transformations performed on the images; structure needed to load the provided images; model hyperparameters; checkpoints to save the best versions of the model; adaptation of the dense layers to fit the correct amount of classes; training of the dense layers; and, finally, all layers. It is worth mentioning that the hyperparameters were changed every round of 20 training epochs. For example, the value of the learning rate decreased each round, while batch size increased. These small changes helped the model to converge much faster.

Once the model was trained, we could finally carry out the last step of the project, which is detecting actual objects on large satellite images.

Results

In order to predict large images, we used a 50% overlap sliding window approach. In the example picture, you can see the overlapping windows of the predictions, the truth coordinates, and our final prediction, made after calculating the intersection of the windows predictions of a single object.

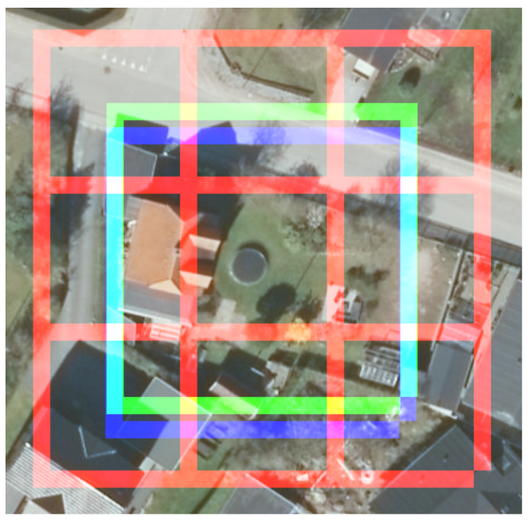

Image	TP	FN	FP	F1
DQIMQN.png	22	8	21	0.60
L7CT2I.png	5	3	16	0.34
UDPYYD.png	0	0	2	0.00

The results obtained with the validation set are shown in the table above. A true positive only happens when the *intersection over union* between the truth and predicted coordinates is greater than 50%. On the first image, the algorithm was able to detect 22 objects. On the second one, there were 16 false positives as the model interpreted part of the background as being trampolines. On the last image, that didn't contain any objects, only 2 false positives were raised.

Conclusions

In this work, we fine-tuned a pre-trained model to detect solar panels, pools, ponds and trampolines satellite images. The original model, MobileNetV2 network, was trained with the EuroSat dataset. After modifying the top layers and fine-tune the model, we applied a 50% sliding window overlap approach to predict large images. Using the intersection of predictions as the final prediction, the model reached a F1-score of 0.6 in one of the validation set images. In other one, there were only two false positives. We believe the model could have performed better with more training data, and also the usage of Non-max Suppression on new windows close to the intersection. But this would make the prediction time greater.

References

[1] Turan, C. Satellighte Documentation. https://satellighte.readthedocs.io

[2] Sandler, M.; *et. al.* MobileNetV2: Inverted Residuals and Linear Bottlenecks. https://arxiv.org/abs/1801.04381

Member contributions

Dimitri Silva

Worked on finding and setting the environment to modify and fine-tune a pre-trained model: dataset preprocessing, data transformations, model adaptation, training and checkpoint. Alongside, wrote the code to apply sliding windows and predict new satellite images using the intersection approach.

Simon Leroy

Experimented with pre-trained models, notably YOLOv3 but with poor results, and did an exploration phase to find which model to use. Also took care of the labeling of the unlabeled images by working with a script we had written.