

# Mine-RPN

Or how we were able to recognize pigs et. familia in Minecraft



**SAPIENZA**  
UNIVERSITÀ DI ROMA

Dario Loi, Davide Marincione,  
Benjamin Barda

Bachelor's degree in  
Applied Computer Science and Artificial  
Intelligence  
Sapienza, University of Rome

A. Y. 2021 - 2022

# Breaking the ice

FACCIAMO QUALCOSA DI SCEMO

## Faster RCNN

Developed in 2015 by Facebook's researches, Faster-RCNN is still today an industry standard thanks to it's accuracy and performance, getting a step closer to real time object detection

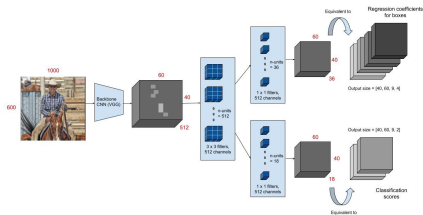


Figure: Faster-RCNN architecture.

# Why minecraft?

Minecraft has several desirable qualities:

- Simple graphics.
- Sandbox.
- Available to every team member.
- Distinguishable entity silhouettes.



**Figure:** A Minecraft promotional image.

# Behold, data!

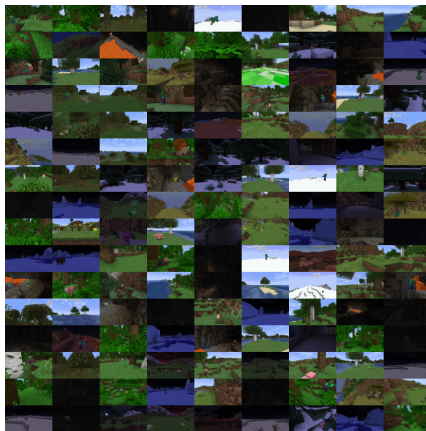


Figure: A representative chunk of our dataset

## Augmentation Techniques

In order to prevent overfitting and increase the amount of information available, we employed various data augmentation techniques, such as:

- Rotation and Reflections.
- Adjustments to Contrast, Brightness and Saturation.
- Sharpening and Blurring the image.



Figure: Our Dataset, Augmented.

# Tool

How to label 4000 images?

1. Load image
2. Create box / purge
3. Next

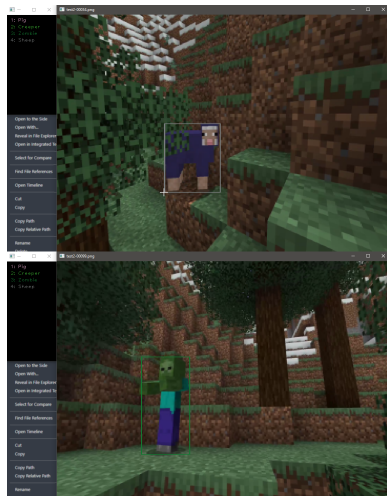


Figure: BBoxing in our tool

## Our Backbone

The backbone is the convolutional *heart* of our model, it is:

- Blazingly fast.
- Adaptable to any resolution.

While also offering:

- A 92% accuracy when used as a Classifier.
- A mean training time of  $\approx 2h$ .

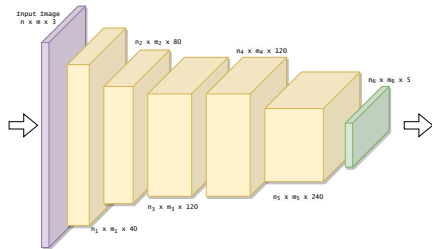


Figure: Our backbone.



## Our RPN

Our RPN network extends our Backbone and is composed mainly of two twin layers:

1. A Classification layer.
2. A Regression layer.

Before feeding data into those, it also performs some pre-processing:

- Anchor Splashing.
- Base convolution.
- Flattening (how do we get to fully connected otherwise?)

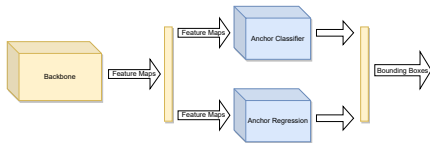
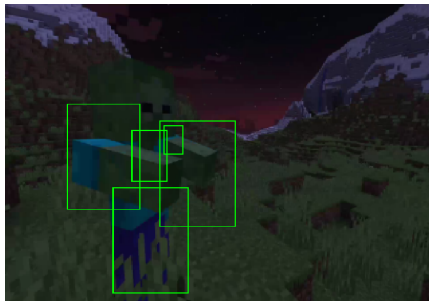


Figure: Our network's proposal layer.

## Examples

Even if many proposal are presented, the network realizes which is the objects to focus on, and which to discard.  
It is not always that easy ...



**Figure:** A zombie in it's natural enviroment

# Examples



Figure: A very confusing sunset



Figure: The network struggling

# The End.

Question Time! (it's an exam after all...).

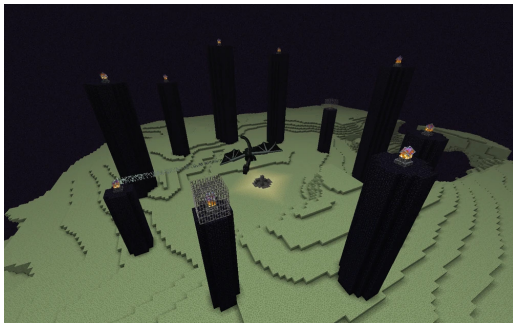


Figure: The End.