POLITECNICO
MILANO 1863

# Artificial Neural Networks and Deep Learning
# Homework 2 - Image Segmentation

Frantuma Elia - 10567359 - 945729,
Fucci Tiziano - 10524029 - 946638

A.Y. 2020/2021

# Table of contents

# Chapter 1

# Introduction

## 1.1 Description of the task

The homework consists in solving a visual question answering (VQA) problem on the proposed dataset. The dataset is composed by synthetic scenes (see example below), in which people and objects interact, and by corresponding questions, which are about the content of the images. Given an image and a question, the goal is to provide the correct answer.

(1) **Q**: Is the man's shirt blue? **A**: Yes

## 1.2 Dataset

The dataset is composed by 29333 total images, 58832 training questions and 6372 test questions.

### 1.2.1 Images

The images' properties are:

- Color space: RGB;
- image size: 400x700 pixels;
- file Format: png;

### 1.2.2 Answers

The set of the possible answers is static and made of 58 possible answers belonging to 3 possible categories: 'yes/no' answers, 'counting' answers (from 0 to 5) and 'other' (e.g., colors, objects, ecc.). In the following the labels associated to each answer:

### 1.2.3 Data augmentation

We have not performed data augmentation. The dataset dimension was big enough for the complexity of our model and performing augmentation would have required a lot of time to adapt all the question/answer couples. Furthermore, we could do it just on the images and not on the answers set.

## 1.3 Validation set

No automatic validation set is provided. This means that a subset of the training set must be used to perform validation.

In our case, we parametrized the number of training images to be moved into the validation set, with a 10% probability.

## 1.4 Test set

The test set is provided as a set of 6372 (image-question) couples, without the attached answers. Participants are required to provide the answers for the test images by submitting the solution with the correct submission format.

## 1.5 Evaluation

Submissions are evaluated on Multiclass Accuracy, which is simply the average number of observations with the correct label.

# Chapter 2

# Neural network architecture

For all the networks, we have modified the function `read_mask_example`. The modified function can be found into the `starting_kit` folder. The notebook expects that folder to be in `\content\drive\MyDrive`.

## 2.1 U-Net

Our first try was with U-Net, but the results turned out to be very poor. In particular, the high number of parameters made the tuning very difficult to perform. The training time was simply too much to experience some progress, since it took up to one hour and a half for one epoch. After some epoch, the result was the one shown below.

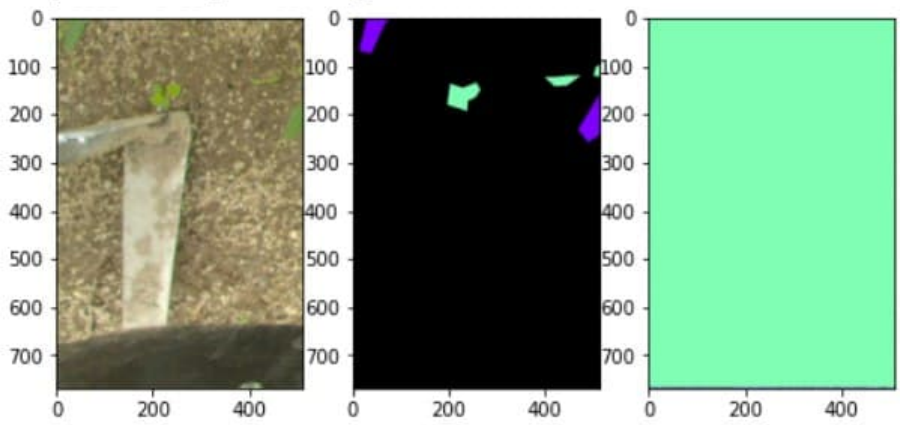


Figure 2.1: dataset image, target segmentation, actual segmentation

We tried to change the learning rate, the weights initialization and the

loss function, using a weighted one, but the situation remained the same.

## 2.2 Transfer learning with MobileNetV2

Then, we tried to use MobileNetV2, as it was suggested by the TensorFlow tutorial. After a while, we noticed that the trained model could not distinguish the maize plants from weeds. The result was a well defined foreground, but with a faded violet, since the two colours were mixed in correspondence of the plant. This was probably caused by the incompatibility of the weights with non-squared images, as shown by Jupyter in a warning.

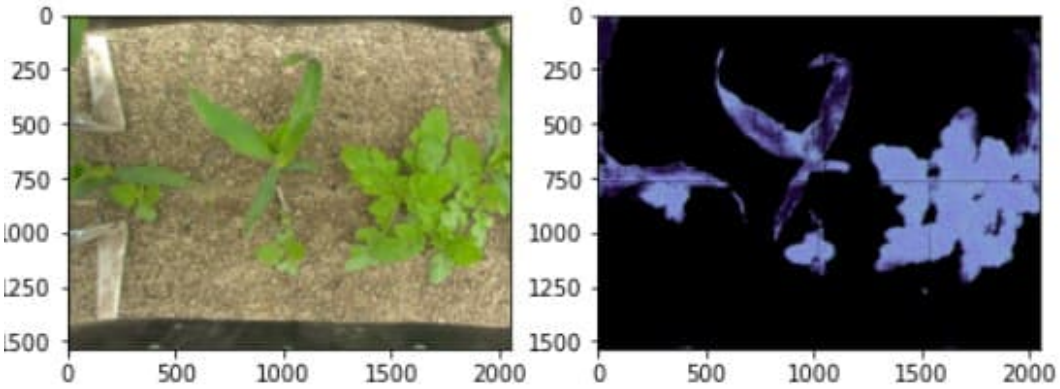


Figure 2.2: image segmentation

## 2.3 U-Net, with VGG as a backbone

### 2.3.1 Without skip connections

This architecture is taken from the notebooks seen during the exercise sessions of the course, adapted to this competition. After some epochs, the results were similar to the one shown below.

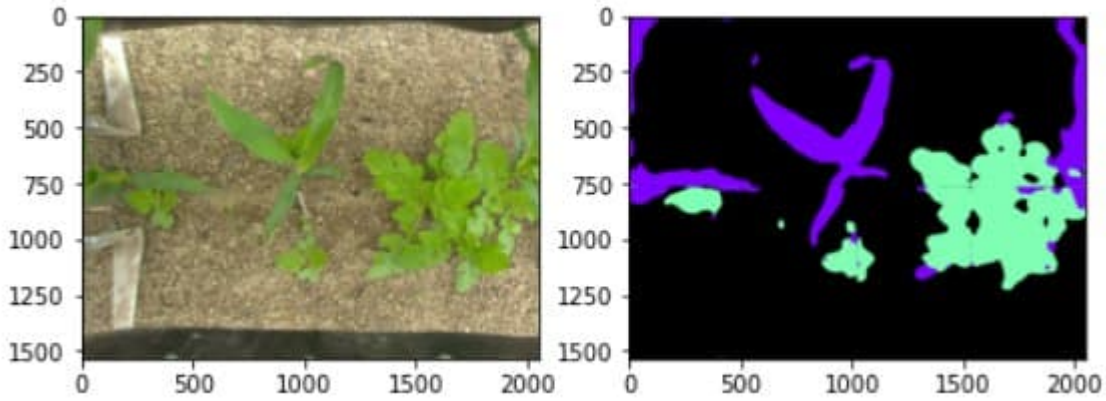


Figure 2.3: image segmentation

### 2.3.2 With skip connections

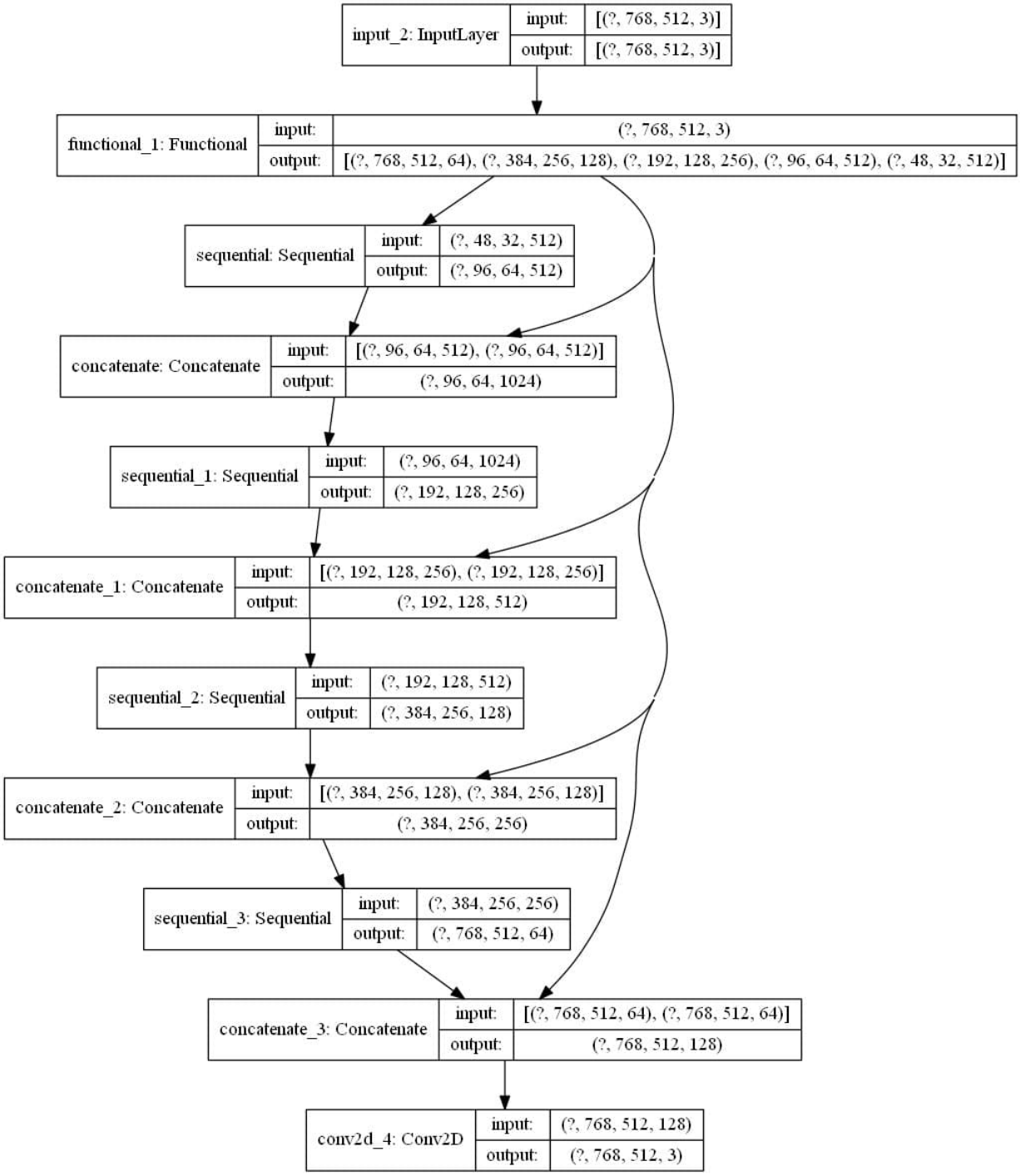


Figure 2.4: network obtained with VGG and skipped connections

After considering the former architecture, we decided to add skip connections in order to improve the performance of the upsampling layers. This led to a better definition of shapes and leaves and thus to a higher score.

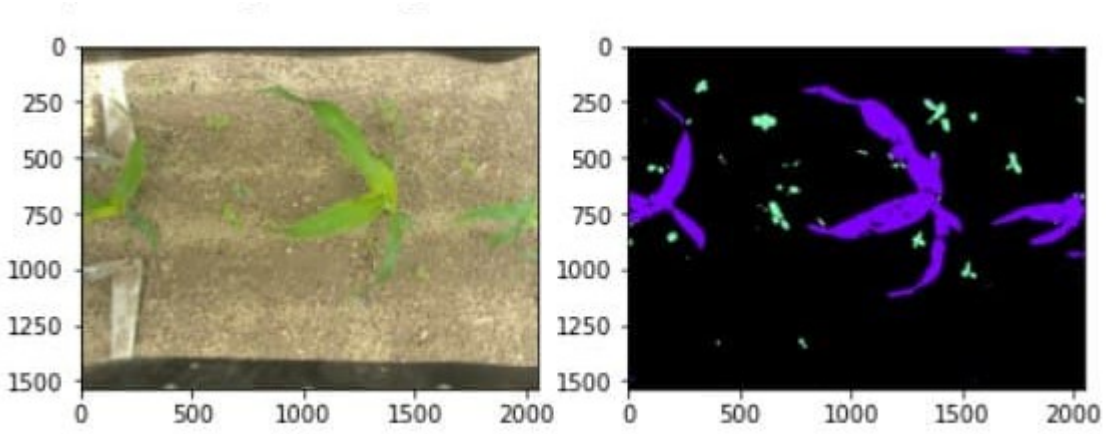


Figure 2.5: image segmentation

### 2.3.3 Score

The best score of the network was 0.7207, obtained with the skip connections.

# Chapter 3

# Main issues

During this competition we had to face problems that didn't occur in the first one: despite we could solve memory capacity issues by tiling the images, the main problem remained time. The classic network tuning procedure has been hardened by the great training time, up to one hour per epoch (obtained with U-Net). Even with the other network architectures, the training time was so long that it took hours, or days, to understand if the network was properly learning the model. This was the main reason because we used the Maize dataset only.

The tiling procedure took so long (8-10 hours) that we could not make many attempts on how many tiles were better. Furthemore, we had to manually fix the folder structure in the .zip file, since Colab didn't allow us to recreate the original dataset structure.

Other factors that limited our score were the absence of the last softmax layer and the TensorFlow preprocessing of predictions for VGG, that drastically lowered the performance of the network. We managed to correct these two mistakes just in the last days.

# Chapter 4

# References

## 4.1 Links

- GitHub repository of the project: `https://github.com/tizianofucci/A2NDLSegmentation`
- Competition web page: `https://competitions.codalab.org/competitions/27176`