Introduction

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

LORIENT AGGLOMERATION

Table of Contents

Context and actors

The project was created on Lorient Agglomération's request, and it's goal is to test if Deep Learning methods could be used to improve the work of counting and identifying birds. More precisely, these researches aim to help the study of birds in a mudflat during dredging operations. These researches could of course be used in other actions of avifauna protection.

The different organizations that take part in the study, on the effect of dredging on birds, are Lorient Agglomération, Bretagne Vivante, Master EGEL and our group from IMT Atlantique. The purpose of this document is to explain the reasons that lead us to use Deep Learning, its advantages and inconvenients.

Our work was made for our 2nd year Project at IMT Atlantique. The team is composed of 6 students of 2nd year:

- Antoine LAVRARD-MEYER
- Julien AUBIN
- Xueyao JI
- Xinyan HE
- Théo TEMBOU
- Pierre GAILLARD

Tutor : Frédéric Maussang.

Technologies

Classical Algorithms

The way we've been conceiving algorithms was kind of similar to a recipe. Programs were just a succession of simple actions that someone conceived theoretically. These simple actions put together could achieve difficult and complex tasks, such as regulating temperature in a building, calculating trajectories etc.

Before Machine Learning, the detection and classification of bird species was done by looking at characteristics of an animal. These characteristics had to be sufficiently different from each other, in order to be able to differentiate birds from other similar objects and then between them.

Thanks to algorithms that look for the colours **[1]**, shape **[2]**, movement **[3]**, the authors manage to get satisfying results. These results, however, sometimes require extensive knowledge of the techniques used. In particular, the setting of certain parameters also depends very much on the situation. The results reported by these algorithms, being less good than those proposed by neural networks, and being less generalizable, we decided not to develop these algorithms. For further application, **[4]** offers a good overview of the state of art, at a time when deep learning was not yet as developed as it is today.

The different techniques can however be complementary with machine learning, as in **[5]**, where the author uses edge detection techniques and a neural network, or in **[6]**, where the authors use Machine Learning and background subtraction in order to automatically detect humans presence in a field.

Machine Learning

As we've explained before, classical algorithms are like recipes to follow. But the more complicated the task becomes, the more heavy such an algorithm becomes. For example, to have an algorithm that recognizes a bird, one would have to write down, in the code, all the things that the program will have to look for (colours, shape etc), and these things change with the orientation of the bird, the luminosity etc. It would require thousands of lines of code to have an algorithm that is efficient in all conditions, and we would probably still miss some things that are useful for a machine to recognize a bird. On top of that, birds can be very different from each other, therefore it's difficult to define the common points that would be used. These problems led the famous mathematician, Alan Turing, to write about a technology that would allow machines to learn **[7]**.

Algorithms that use Machine Learning are very different from classical algorithms, because it doesn't need "rules" to do what it has to. In our case, no need to tell the algorithm that he has to look for wings, feathers etc.

In fact, the goal, with Machine Learning, is to find these rules based on Data, and then apply them to new Data. This way, the program will "learn" what a bird looks like, and then will identify birds on new pictures.

One could think that it doesn't have to work better than a classical approach. But in fact, the way to use Data, in order to infer the rules, is often more efficient. Indeed, it allows the algorithm to find many methods that we wouldn't have thought of to achieve its task. These methods (that work for algorithms but not humans) are often better suited to a machine than our methods are, and often more efficient than ours.

Deep Learning is a particular way of doing Machine Learning. It began working very well around 2012 and then never stopped achieving new things in many fields. In many cases, it's Deep Learning we're talking about when we say Machine Learning.

Deep Learning

Description of a Neural Network

The core of a Deep Learning algorithm is its neural network. As its name suggests, it is a technology inspired by our brain. A neural Network is composed of neurons that are linked with each other by synapses. The object that we give to the neural network is represented with numbers. For image recognition, it is 3 numbers by pixel, one for blue, one for red, and one for green, that give the intensity of each colour.

These numbers will excite the first neurons in a way that depends on their value. The signal will be transmitted to the next layer of neurons, threw the synapses, and on and on until the last layer of neurons, at the end of the network.

When we do species recognition for exemple, the number of neurons on the last layer is equal to the number of species we are trying to identify. To simplify, the species that will be detected is the one which corresponding neuron on the last layer is the most excited.

So a neural network is defined by the number of layers, the number of neurons on each layer, and how each layer is connected to the next one. For example, we could create a neural network composed of 3 layers of 8 neurons, with every neuron connected to all neurons of the next layer (we talk about Fully Connected Layers).

But this is not enough to define a neural network. To do this, we need two more things:

- Each neuron and each synapse must have a weight. These weights describe how the neuron or the synapse will transmit the signal. Concretely, a neuron with a small weight that receives a signal will send a weaker signal.
- We must define a way to evaluate the results of the neural network. It is for example how far from the truth the prediction of a species was. This is called a Loss Function, and we try to minimize it. Finding a good way to evaluate the final result is often very complicated and it depends a lot on what we are working on.

Training of a neural network

In order to train a neural network, we will use labelled data. For species recognition, each image contains an animal, and the label is the species of this animal.

So we “show” labelled data to the neural network, it gets excited, and makes a prediction. If it’s right, we want to reinforce the “reasons” that lead it to be right. If it’s not, we want to weaken these “reasons”.

Concretely, this process is made thanks to a Gradient Descent :

- We start at the last Layer.
- If neurons are excited while they shouldn’t (for species recognition it would be neurons that don’t correspond to the species of the animal that is being shown), then we reduce the weights of the neurons and synapses that lead to this excitation.
- On the opposite, for the neurons that should be excited, we will increase the weights of the neurons and synapses that lead to their excitation.

We repeat these steps for all the Dataset, and change the weights very slightly every time. At the end of the training, we have a neural network capable of doing the task we taught him.

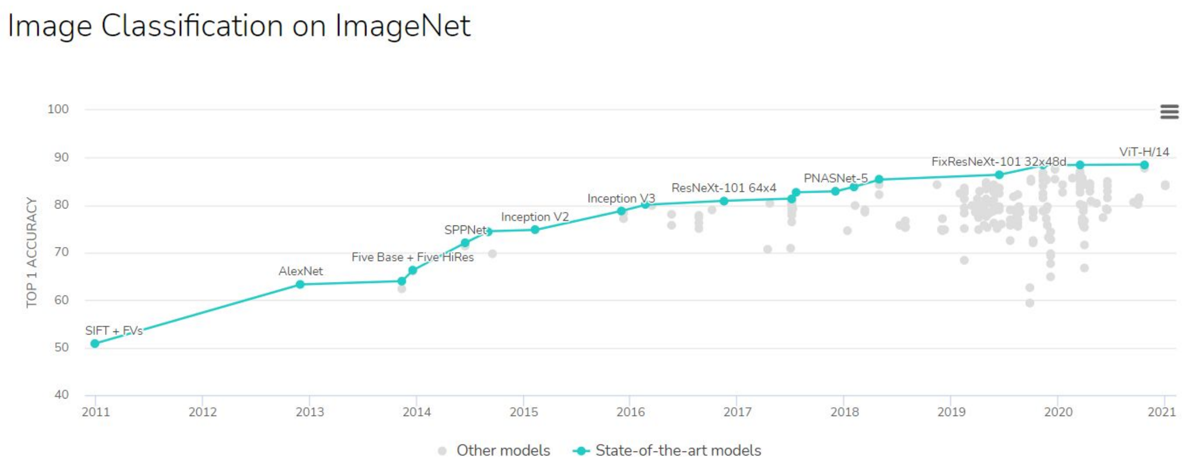
Importance of Data

The data are the key component of training : a good Dataset must contain lots of data, as varied as possible, and if possible close to real conditions. For example, with species recognition, we are looking for images in various luminosity conditions, on the beach or in the water, etc, so that the neural network doesn't associate a type of background or a level of brightness to a species.
These Data must be labelled, and this is what costs time, if you want to do it from scratch, or money, if you find an existing dataset.

State of art in image classification

Image recognition did progress a lot thanks to 2 things: the improvement of Deep Learning algorithms, and the huge augmentation of Data available, thanks to the internet.

The progress of Deep Learning is really visible on this graph that represents the scores of the best algorithms of image recognition over time (we can notice that AlexNet in 2013 was really a breakthrough).



https://paperswithcode.com/sota/image-classification-on-imagenet

Today, Deep Learning is used for example in order to analyze images for automated cars, surveillance cameras, deepfake, etc (you can go check about the algorithm Yolo **[8]** and the database Coco**[9]**, or ImageNet **[10]** for example). For more information about classification for image based species identification, several studies based on classification algorithms are reviewed, as well as automated species application **[11]**.

State of the art in object detection:

Choice of detection method

In order to count the birds, we need to know where they are in the image. This task is referred to as object detection. Some papers have implemented algorithms with good accuracy by using different techniques:

- Using UAV to take photos from above is a solution, and it doesn't seem to disturb the birds, as explained in this paper **[12]**. In this case, we don't have to separate birds from each other, and we can use classification algorithms that we apply to several places of our image like our tutor, Frédéric Maussang, in **[13]**. However, we didn't use this technique because of its cost and its difficulty of using it regularly, and the classification is harder for birds on the ground, if the image is taken from far above.
- Using Fix Cameras is another possibility, because it allows us to use the motion between two images, but the zone that we want to cover is too large to use this. Indeed, photos of birds that are too far couldn't be used. To address this issue, we could use automatic zooms, but it would be too expansive **[14]**.

These causes have led us to consider ground detection.

From an autonom to an assistant algorithm

The purpose of our research was to identify the algorithms that we would use based on the state of the art. One of the first things we understood was that the difficulty of object detection is linked to the fact that the birds are sometimes close to each other.

Even if our first goal was to find a way to completely automate the process of taking pictures and analyzing them, in order to have a very close study of the birds' behaviour, we had to give up on it. We decided to see the product as an assistant that will help the ornithologist studying the birds. The images will be taken by a human and the AI will find the birds, and identify them with a degree of trust. Then, the ornithologist would only have to scroll among the birds' close up that he would want to check. Here is a study that showed the benefits of combining algorithms and human knowledge **[15]**.

An algorithm made specifically for a mudflat

Another thing we understood, by looking at the state of art in bird detection, is the fact that it would be really difficult to design an algorithm adaptable in all conditions : the number of photos required also greatly vary in function of the diversity of the background, where the neuronal network is trained and applied. For example, a paper **[16]**, seeking to detect birds in an uncontrolled environment (which means that the neural network isn't trained specifically for this environment), have used 10 000 images in order to detect birds in a camera trap. This is a lot of images, and it is only for a classification algorithm, so it would be even more for detection, which is more difficult to do. Therefore, we decided that it would be better to design an algorithm and a Dataset only for the mudflat.

Algorithm choice

As we've seen, separating elements can be tricky, and getting such algorithms to work remains a challenge. Even in state of art, it is difficult to obtain a level of recognition similar to human vision. We will explain in these last paragraphs some neural networks used in object detection.

In our research, we found two kinds of object detection algorithms : one stage detector and two stages detector. The difference between both algorithms is that the first one directly predicts bounding boxes, corresponding to a given object, whereas the second firstly predicts possible object location, and then performs classification and regression of the bounding box. The first kind algorithm is often faster, and can work on video. The second algorithm is more accurate, but takes longer (a few seconds) to predict the location of the birds.

In the beginning of our research, we found the Yolo algorithm (one stage detector). It gives good results on some problems. However, this algorithm performs poorly when the objects that need to be detected are overlapping, which is our case. In order to implement a neuronal network and prove that we can detect overlapping birds, we use another one stage detector called the single shot detector (ssd). It gives better results than the Yolo algorithm, but still not perfect. According to our bibliography research, a two stage detector (faster rcnn), could have satisfying results. For example, faster-rcnn was used in **[15]** with great results. The authors manage to prove that their algorithm performed better than the specialist in some cases for bird recognition on an image. The greatest gain, however, was obtained when both the specialist and the algorithm were used on the same image. This article leads us to think that implementing a solution that will detect the birds in a picture is made possible by new advances in machine learning.

Problems

Lack of Data

The first problem we had to face during our project was the lack of data. There is in fact no data for the species we want to recognize, and especially no data that were acquired in the conditions we will face (but we will talk about this later). We also had no possibility to test our algorithm in true conditions.
Therefore we did split our algorithm in two :

- The first algorithm is used to detect the birds on an image, and set up a box as small as possible, that contains this bird. This step is called object detection.
- The second algorithm is used to identify the species of a bird on an image, which is called classification.

Then we picked up images from the internet for the classification part, and already existing Datasets for the detection part. We artificially added new versions of these images to the Dataset, by changing their brightness, orientation, etc. This process is called Data Augmentation.

Conditions of photography

How the images were taken really matters, and because of that, using “perfect” images (good resolution, zoom on the bird which is in the center of the image, good luminosity etc), from the internet was not a really good way to test if our algorithms could be efficient in real conditions.

In order to tackle this issue, we planned on modifying our images to reduce their quality, and we also planned on working with pictures taken in real conditions, but because of the pandemic and the lack of time, we didn’t get to do these tests. However, this could be easily done if the project goes on.

Focus on Resolution:

By looking at some research papers, in order to detect and classify the birds, a minimal resolution seems necessary. Although some papers focus on the detection of small pixel size (image around 100 pixels per bird) **[17]**, the result seems limited. Furthermore, we want to make some classification afterwards, that means that our work requires much better resolutions. In **[15]**, the author finds that the accuracy of the algorithm might be lower when the algorithm is applied on birds that occupy less than 1500px, but other issues also arise for birds filling too much space. In our experiments, we have struggled as well with this kind of issue.

Bibliography

N°	Subject	Link
1	Supporting the annual international black-faced spoonbill census with a low-cost unmanned aerial vehicle	https://www.sciencedirect.com/science/article/pii/S1574954115001697?casa_token=co2VVQ-SnncAAAAA:fbXYJrivcSXJjlaImC334XOI5ZkDhhHV_Cev2i-
2	Solving bird image overlapping for automatic population counts of birds using image processing	https://ieeexplore.ieee.org/abstract/document/8025241?casa_token=41vIU7Re2nwAAAAA:kmLkFGC9SmxcjSaebWtM2dlyLw-uz5tBJRst14ICXaM2mByRu8MK6Dq4NzoQwQesEcZCtz[FH
3	Automatic detection, tracking and counting of birds in marine video content	https://www.researchgate.net/publication/312560325_Automatic_detection_tracking_and_counting_of_birds_in_marine_video_content

4	A computer vision for animal ecology	https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2656.12780
5	Deep Learning Bird Detector (see counting part)	https://github.com/MichaelTeti/BirdDetectorCNN?fbclid=IwAR3p8pNJj6tnNT2yc371RtXm7NAmP4-SHc89HdYF3qjGmgr_hkyCz5hLrc8
6	DeepAnomaly: Combining Background Subtraction and Deep Learning for Detecting Obstacles and Anomalies in an Agricultural Field	https://www.mdpi.com/1424-8220/16/11/1904
7	Learning Machines by Alan Turing (go to the “Learning Machine” section)	https://academic.oup.com/mind/article/LIX/236/433/986238#164226651
8	YOLO	https://towardsdatascience.com/yolo-you-only-look-once-17f9280a47b0
9	COCO dataset	https://cocodataset.org/#home
10	ImageNet dataset	http://www.image-net.org/
11	Machine learning for image based species identification	https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13075
12	Can drones count gulls? Minimal disturbance and semiautomated image processing with an unmanned aerial vehicle for colony-nesting seabirds	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6308878/
13	Marine Bird Detection Based on Deep Learning using High-Resolution Aerial Images	https://ieeexplore.ieee.org/abstract/document/8867242
14	Deep learning–based automatic bird identification system for offshore wind farms	https://doi.org/10.1002/we.2492.
15	Automated Bird Counting with Deep Learning for Regional Bird Distribution Mapping	https://www.mdpi.com/2076-2615/10/7/1207
16	Detecting wildlife in uncontrolled outdoor video using convolutional neural networks	Detecting wildlife in uncontrolled outdoor video using convolutional neural networks - IEEE Conference Publication
17	Bird detection and species classification with time-lapse images around a wind farm: Dataset construction and evaluation	https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2135