UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

TREBALL FINAL DE GRAU EN ENGINYERIA FÍSICA

METEORITE IDENTIFICATION AND RECOVERY FROM DRONE-OBTAINED IMAGES

GERARD PONS RECASENS

Barcelona, June 2020

Supervisor

Jordi Llorca Piqué

INDEX

ABSTRACT

Meteorites are rocks that come from outer space. Studying them is a significant task, as they provide a better understanding about the past of the universe and also provide valuable information about planet compositions and the extraterrestrial resources available for future use. However, recovering fallen meteorites is a crucial step in studying them, but nowadays it is a slow and tedious task. This project seeks to find a way to fasten this process by identifying meteorites from aerial footage taken with drones. The data obtained go through a recognition algorithm, which is created using Machine Learning architectures. The learning process and the accuracy of these algorithms are determined by their architectural parameters and the characteristic properties of the meteorites chosen for the process. In this project, Machine Learning architectures are studied to obtain a model capable of correctly predicting the presence of meteorites in drone-made images.

1. INTRODUCTION

1.1 Meteorites

Meteorites are pieces of rock that arrive at the surface of our planet coming from the outer space. The vast majority of them come originated from asteroid collisions from the asteroid belt of the Solar System, located between Jupiter and Mars, although others come from the Moon or Mars. These asteroids can be from a few meters (usually called meteoroids if smaller) to hundreds of kilometers, such as Ceres, which has a diameter of 940 kilometers. Once these extraterrestrial pieces enter the atmosphere, a luminous effect called meteor occurs, as they collide with the air molecules of the upper part of the atmosphere, heating them to incandescence. This phenomenon is popularly known as shooting star. If the fragments make their way into the ground are then called meteorites.

The vast majority of meteorites were formed 4.5 billion years ago, and have been in the outer space for that long, thus the study of them is quite relevant. For instance, by studying the star dust trapped inside a meteorite, theories about stars and their nuclear process can be confirmed, information about the origins of Solar System, its formation rates and the estimated dates of certain events can be obtained. Also, by comparing the reflection spectra of the meteorites with the reflection spectra of the asteroids, one can identify the parent body, and so far the meteorites found are known to came from at least 135 different asteroids [1]. Moreover, it not only gives humans information about the past but it also provides useful knowledge about the resources available on the space that could be useful for future space travel, colonization or in the event of some resource shortages on Earth.

Meteorites fall all over the Earth, including oceans and zones with dense vegetation. The standard nomenclature uses the word falls for the seen and recovered meteorites, and finds for the not seen but recovered ones. Currently, around 62.000 meteorites [2] have been found, with the vast majority of them being found in deserts or plains with low vegetation due to the increase in visibility. The most important finding site is the Antarctic desert, whose finds correspond to the 63.2% of the total ones, primarily due to the expeditions promoted by the Japanese, American or Chinese governments, among others. In this case, the ice where the meteorites fall helps preventing its deterioration. Other important finding sites are the deserts of northern Africa, representing 20% of the total recoveries.

There, the heat and the dryness of the desert help on the preservation of the meteorites. Hence, the conditions and characteristics of the environment where the meteorites fall have an impact on the rate of their recovery.

1.1.1 Meteorite Classification

Meteorites can be organized following multiple classification criteria [3]. On the higher level, they are divided into differentiated or non-differentiated:

- Differentiated: said of the meteorites that came from parent bodies which experienced the process of differentiation, which occurs when the parent body is large enough to be heated by impacts, which melt parts of the interior resulting in the heavier metal sinking to the center and lighter rocks placing in the surface, forming the well-known crust, mantle and core structure.

- Non-differentiated or primitive: are meteorites which come from parent bodies which have not undergone a total processing event (but may have experienced mildly heating), so that they have been relatively unaltered, being remnants of the birth of the solar system.

More specifically, they are mainly classified by their nickel-iron content:

- Stony: are the most abundant meteorites, accounting for 97% of the findings. They are made of a mixture of olivine, nickel-iron minerals and pyroxenes. They can be classified in:

    A. Chondrites: are primitive meteorites, which contain spherical inclusions of crystallized minerals, called chondrules, which are among the oldest unchanged materials in the solar system and similar mineral compositions are not found in Earth. They are classified by their density or by the chemical and mineral composition of the chondrules.

    B. Achondrites: are differentiated meteorites which do not contain chondrules. They can be primitive achondrites, which come from a melted chondrite, or steroidal achondrites originated from the mantle or the crust of a differentiated parent body.

- Stony-Iron: are differentiated meteorites formed by a mixture of nickel-iron and silicate mineral. They can be pallasites, originated from mantle-core boundary of asteroids and are a mixture of crystalline olivine surrounded by iron-nickel, or mesosiderites, which are thought to be formed by the melting from impacts and contain an equal amount by volume of olivine and nickel.

- Iron: are differentiated meteorites coming from the core of asteroids. They have different structures depending on the iron-nickel rate and also the cooling rate. The more important groups are the Hexahedrite (4-6% Ni), Octahedrites (6-12% Ni) and Ataxites (12% Ni).

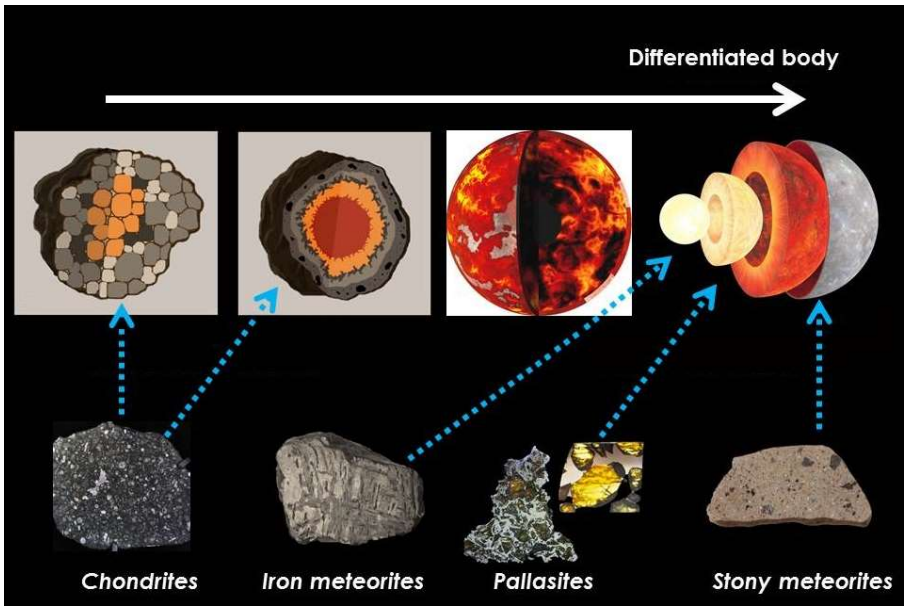

Figure 1.1 Image which contains examples of the most common classification types of meteorites and a representation of the different stages of differentiation of the bodies they originate from as well as the region of the body in the case of the differentiated ones.

1.1.2 Meteorite Recovery

As previously stated, some of the best places to look for meteorites are deserts, principally Antarctica, Sahara and Atacama. The typical modus operandi during the research of meteorites is covering large areas of those deserts using snowmobiles in cold desert of Antarctica and trucks or four wheelers in the hot ones. From them, rocks seaming to be meteorites are spotted and approached to determine whether or it was correctly identified or

it was a false positive. Another method consists in using a metal detector to inspect an area, which is a slower method but effective when it comes to finding meteorites that have similar features as the terrestrial rocks, in particular for iron meteorites. All the methods have in common that are time consuming and even physically exhausting if we take into consideration the climate of the zones in which the research is done. These circumstances were the ones that motivated the research to obtain a way to recover meteorites more easily, which is the one proposed and the objective of this work, and it is based on the use of a drone to cover the chosen surface.

1.2 Machine Learning for Image Recognition

In the past years, Machine Learning has been making its way into nearly every field which contains some data to work with. It is not strange that it has been found to be useful with a wide range of image recognition tasks, from distinguishing shapes, numbers or identifying faces, to more complex ones such as recognizing emotions or predicting medical conditions. The way it works is by exploring the data it is provided and learning from it by extracting patterns which are going to be used in the future to predict the content of new unseen data. The learning process occurs with the creation of an Artificial Neural Network (ANN), which is influenced by how our brains neural networks works. This artificial network contains different layers of neurons, or nodes, which are connected with the neurons of the following layer, to which they will transmit an input if they are activated. The strength of the connections and some activation mechanisms are parameters that the model will decide on its own during the learning stage. Depending on the input information and the data one is working with, different types of ANN are used. Precisely, image recognitions models work by seeing the images as numbers and extracting information from them through convolution operations, forming convolution neural networks, which are explained in deeper detail in Section 2.2.

The main objective of this project is to create a Machine Learning model which is able to accurately determine if any meteorite is present in the area studied, by learning from previously taken images, some of them containing meteorites. A research with nearly the same objective was performed in [4], but it did not reach a mature state and, as stated in the paper, its performance was far from optimal when the terrain contained other rocks or

terrain imperfections, in which their software got a lot of false positives. Hence, the aim of the model to be created is to be robust when it comes to identifying meteorites in ambiguous backgrounds. Moreover, it will also be researched if other types of data could be used in order to distinguish better the meteorites.

2. THEORETICAL FRAMEWORK AND METHODS

The overall view of the proposed meteorite recovery method has the following steps: a chosen surface is covered by a drone which will map the whole extension by collecting some data which is analyzed by an algorithm determining whether or not a meteorite is there. Thus, in this section the different candidate properties for meteorite recognition are presented and studied as well as the basics of the Machine Learning architecture creation process that have been followed.

Some considerations must be mentioned beforehand, as they have been made in order to delimitate the field of study and taken into account in further reasonings. First of all, it must be pointed out that meteorite identification method is created to be used in hot deserts, such as Sahara and Atacama. The cold desert of Antarctica has been ruled out as the extremely low temperatures drastically decrease the drone functionality, the background terrain differs widely from the one from other regions of the Earth, it is in a remote location with difficult accessibility, mobility and life conditions and also it is already widely scouted by governmental founded expeditions. Moreover, the methods and tests are focused on chondrite stony meteorites. This last limitation is based on the data from all the meteorites finds up to date, which shows that 94.9% of the finds have been from stony meteorites, and from those stony meteorites 93.5% are chondrites [2].

2.1 Meteorite's distinctive features

The study of the meteorite's distinguishable attributes is done to shed some light on which is the best data to be obtained from the area and thus which hardware the drone should be equipped with. It must be kept in mind that meteorites suffer from a wide range of processes once they enter the Earth's atmosphere and during the period of time they spend exposed to it, which is very variable.

Some meteorites are found within a few weeks or even days after they have reached the ground, mainly because its fall has been spotted and an active research for them has been conducted. These freshly fallen meteorites present one of the most characteristic modifications done by the Earth's atmosphere, which is the **fusion crust**. It appears when the meteorite enters the atmosphere at high speeds and the exterior starts to melt due to the hot wave of air created on its path. When the meteorite losses speed, it starts cooling down

and part of the melted rock solidifies around it creating a brownish glassy coating. In stony meteorites, the crust is dark brown to black, formed mainly by olivine, glass and iron oxides. In iron meteorites, it is composed of iron oxide magnetite, and its color is blue-black.

Other meteorites are found a long time after falling, even thousands of years later, so it is out of the question not to assume that these meteorites will present significant changes by the time they are found. **Meteorite weathering** [5][6] accounts for an important part of these changes and it is understood as how a meteorite composition is altered to be more stable on the Earth's surface. It is influenced by many factors, such as temperatures, pH, microbial activity or exposure to water or air and these factors make the different Earth's climates preserve the meteorites on diverse conditions. The Antarctic climate helps preserving the meteorites to up millions of years after they fall, and other dry or arid places up to tens of thousands. However, tropical climate or temperate zones present a very accelerated weathering process. The important weathering effects for the project are those that can be seen with unaided eye, and they can be classified into two types:

- Mechanical weathering that occurs due to wind or water erosion, extreme temperatures and the action of nearby living creatures, which results in ablation, fractures and shape changes.
- Chemical weathering that occurs due to the exposure to the Earth's atmosphere for long periods of time. This includes oxidation, rust, corrosion, hydration or carbonation, that also originate some visual changes to the meteorites. For instance, the curst can lighten due to the oxidation of the iron the meteorite contains.

Figure 2.1 On the left side, an image of a freshly fallen meteorite can be seen, presenting the black coating corresponding to the fusion crust. On the right-hand side, a heavily weathered meteorite can be observed, which has lost its characteristic black color. [5]

Knowing now how the appearance of the meteorites to be found is, three different characterization methods are explored.

2.1.1 Visual Recognition

When trying to locate an object using a drone, visual recognition is the first that comes into someone's mind as nowadays the vast majority of commercial drones include cameras already integrated into their structure. Both the intuitive aspect and the convenience of the already existing hardware for the data obtention process, made it is the first recognition method to be considered.

Different parameters of the image of a meteorite have to be studied, the most important ones being shape and color, to find out if meteorites present relevant characteristics which make them different from ordinary rocks. As meteorites are fragments of celestial bodies which can present any kind of form, and the rocks already found in the deserts also do not have a specific contour, the shape of the rock cannot be a primary distinctive characteristic. However, differentiating by color is a process that cannot be ruled out that easily, as meteorites present unique color characteristics.

First of all, as already explained, the **fusion crust** is created surrounding the meteorite when it enters the atmosphere, and has a black dark-brown color, which is usually different from the other rocks found in the deserts. However, when fragmentation occurs the different specimens are often not completely coated with crust or only display a modest coverage. Also, as stated previously, the weathering must be considered as it can induce a color change to it. The corrasion is one of the first visible effects on meteorites, which can occur just in a few months after they fall as the fusion crust becomes polished by the sand or other particles carried by the wind. The temperature gradient between the desert day and night, and the differences in expansion/contraction coefficients of the meteorite and its fusion crust can induce its exfoliation. Due to oxidation, the crust can lighten up but the characteristic dark color is still fairly preserved, and finally the rusting of the iron the meteorites contain can result in differences in color as a reddish-brown coating can appear.

Other important coloring effect is the **desert varnish,** also called desert patina, which appears in super arid deserts and it is a shiny black dark-brown patina composed of clay

and oxides of iron and manganese. Finally, **caliche coating,** which is formed of calcium carbonate and other minerals such as sand or clay, making a light grey or brown coating on the surface in contact with the soil.

Furthermore, as the identification process is meant to be done in the deserts of Sahara or Atacama, the respective autochthonous rocks or soil contents have to be taken into account when deciding if it could be differentiable from a meteorite.

Sahara

Although the first image that comes to our mind when thinking about the Sahara Desert is one of endless dunes and sand, these characteristics would not be propense for meteorite hunting, as when meteorites fall in sand they became embedded in the soil, which shields them from wind and other kinds of erosion. It is not only until deflation occurs that these old meteorites are brought to the desert deep surface, the desert pavement, which is an idoneal surface for the identification of these well-preserved meteorites. The surface is compound of relatively small angular/sub-rounded packed rocks which are usually red, brown or light brown (pebble and cobble size).

Atacama

Atacama Desert is a South American desert plateau, one of the most arid and driest places in the world. It has lower precipitation rates than the polar deserts and a very low erosion rate, a very important characteristic for the preservation of the meteorites. The pavement there has a reddish tone, and is usually compared to the Mars surface. The rocks are usually scattered, with high silicon concentration accounting for their pale tone.

Figure 2.2 The image on left shows the typical desert pavement from the Sahara Desert, whereas on the right the usual arid terrain of the Atacama Desert can be observed.

In conclusion, color identification of the meteorites could be perfectly valid for those which present a black dark-brown color because they have fallen recently before the search expedition or because they have not experienced weathering effects which changed their color to one similar to the rocks around them, which can be the case of corrasion, caliche coating or a combination of various of them. It must be noted that these weathered meteorites would also be unnoticed by a human research team on normal extensive expeditions, only molt detailed searches or the use of metal detectors could help with their identification.

2.2.2 Thermal Recognition

The second method of identification studied has been the infrared radiation (IR) that meteorites emit as a black-body, which is a method that was proposed but not evaluated in the Frontier Development Lab meteorite identification study [7]. In our approach, the drone should have to be equipped with an infrared thermal imaging camera and map the whole area of interest. In order for this method to succeed, it has to be studied if the meteorites present notably different temperatures than the objects surrounding them.

First of all, the **black color** expected in the vast majority of the recently fallen meteorites makes them absorb more light than the surrounding rocks, increasing their temperature. Other thermal properties, such as the **thermal conductivity** or the **thermal diffusivity** should be studied beforehand. The thermal conductivity of meteorites is significantly lower, by a factor of three to ten, than that of the pure minerals from which they are made of [8]. Hence, an estimation of how this property compares to other rocks found in the deserts cannot be made taking into account the meteorites composition. It has been shown that for chondrites, the thermal conductivity typically ranges from 0.5 $\frac{W}{m \cdot K}$ to 3 $\frac{W}{m \cdot K}$ at 300K. For typical desert rocks, the parameters are quite similar, as shown in Table 2.1. Regarding the thermal diffusivity, it is evaluated taking into account thermal conductivity, heat capacity and density. The only parameter on it which

	Thermal Conductivity $\frac{W}{m \cdot K}$
Sandstone	1.44
Siltstone	2.04
Limestone	2.29
Chromite	2.5

Table 2.1 Thermal conductivity values [9] for usual rocks found on the Sahara [10] and the Atacama [11] deserts.

differs between terrestrial and extraterrestrial rocks is the density, to which it is inversely proportional, as heat capacity values have been seen to be similar [12]. As density is nearly two times lower in terrestrial rocks, it can be expected to encounter lower values of thermal diffusivity in meteorites [13].

These terrestrial and extraterrestrial rocks behavior differences open a wide range of research possibilities. On the one hand, meteorites are expected to absorb more heat due to their darker color, but in the other they have a slower reaction to temperature changes. All these factors and behaviors have to be considered in a desert climate where there is a lot of direct and intense sun exposure, which makes them more interesting. In order to get a deeper insight on these parameters, temperature profiles from deserts have been studied. The data used to construct them was provided by a study carried between 1994 and 1998 which recorded every 30 minutes the temperatures from both air and rock in Yungay, an extremely arid part of the Atacama Desert [14]. From the database some parameters were obtained: the heating and cooling rate and the maximum and minimum temperatures from both air and rocks. The treated data, which is presented in Annex 1, shows that there is an important temperature gradient thorough the day, as the maximum temperatures go up to 30ºC, whereas the minimum ones even reach negative values. Moreover, it has also been seen that the maximum temperature of desert rocks was a lot higher than the air temperature, sometimes close to +20ºC, whereas the rocks minimum, while it was still higher than the air's one, was much closer to it. These daytime to nighttime temperature variations suggest that if there are any differences between the heating and cooling rates of the deserts rocks and meteorites, IR could be useful. In that scenario, the best time or times of the day to perform the meteorite search should be considered. Using the data from the same study, the usual temperature behavior of an ordinary day has been represented in Figure 2.3. Two well defined regimes or regions can be easily spotted, one when the temperature increases at a usually constant rate and another one when it decreases in a slower and not as constant one. These differences suggest that the meteorites heating and cooling behavior due to sun exposure could be investigated, in order to see if there are substantial differences with the desert rocks patterns. Different hypothetical scenarios could yield to proper time slots for meteorite identification with IR:

- If after the night both rocks and meteorites reach a similar temperature, the first hours of the day could be used to identify them depending on the differences in heating rates.

- If when the temperature starts to decay they both have reached a similar temperature during midday and present sufficiently different cooling rates, meteorites could be identified during the afternoon hours.
- Also, they could have a different base temperature, allowing the detection all day round.

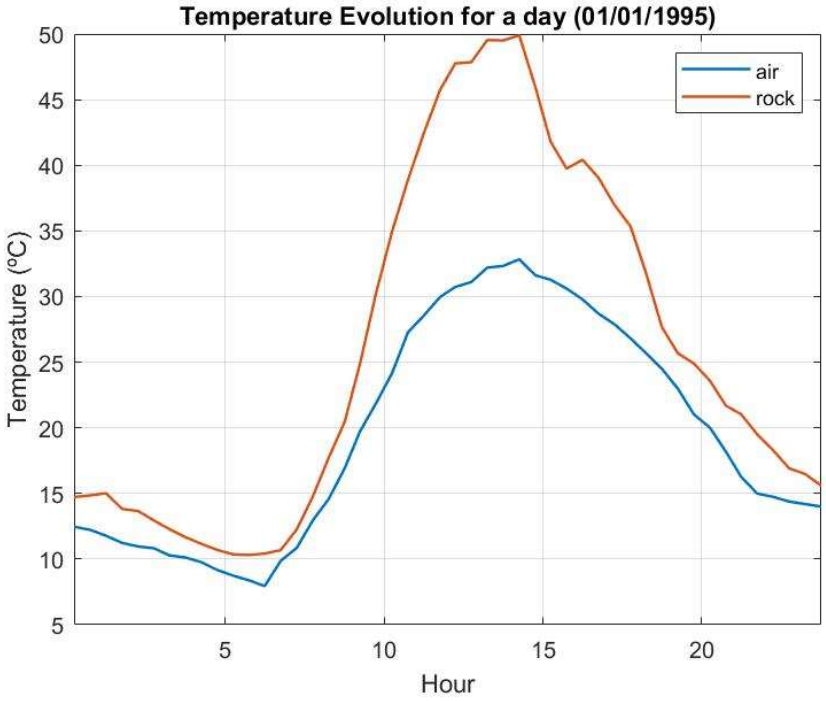

Figure 2.3 The temperature evolution during a day for both air and rocks of the Atacama Desert is shown. It can be easily spotted the heating and cooling regimes and also the huge difference of the temperature during midday.

2.2.3 Reflectance Recognition

Reflectance measures the light that a sample reflects once shined with a light source. It is used in meteorites to identify the asteroid they came from when comparing both spectra. Some of the components of the meteorites, such as olivine or pyroxene, have a distinctive spectrum, which helps to its identification. Although olivine is a fairly present mineral on Earth's surface and pyroxene is found in many metamorphic and igneous rocks, the contrast in proportions could lead to significant differences. However, the reflectance measurement has two major drawbacks: the hardware could not be carried easily by a drone and also the measurement will be a mix of all the elements present in the rocks and soil of the analyzed surface, which would yield to inconclusive results. Hence, the reflectance recognition method can be dismissed without further investigation.

2.2 Machine Learning architectures

Whether the recognition system is based on visual inspection or IR, the data the drone will obtain is composed of 2D images. The Machine Learning algorithms and architectures are programmed in Python, using the free open-source library Keras, backended by TensorFlow, another open-source software library. The first one is a modular neural-network library that offers a user-friendly possibility to experiment with deep neural networks by providing the building blocks, functions and optimizers. The latter, is a Google created symbolic math library used for numerical computation by using dataflow and differentiable programming, widely used for machine learning purposes. By making use of these two resources and a GPU equipped computer to speed up the learning process, the models have been created and tested.

Before getting into more detail about the models and architectures, some definitions of terms frequently used to describe the processes, structures and performance decisions are presented:

Weights: Value that determines the strength of a connection, in other words, the contribution of a feature to the next layer of features or to the final output. The weights are some of the parameters that will be tuned while training.
Epoch: is a measure of the number of times the training model is presented the data completely. Different number of epochs will be needed depending on the created models.
Activation function: is a function that decides the output of a neuron by transforming the value of the weighted sum of the inputs.
Hidden Layer: name that receives every layer that is placed between the input layer and the output layer. Each contains one or more activation functions.
Batch size: the number of elements our model will see before updating the weights, which is called an iteration. The number of iterations in an epoch will be the result of dividing the total number of elements in our dataset by the batch size.
Generalization: the ability a model has to perform accurate predictions from unseen data.
Overfitting: a model is said to be overfitted when it learns unimportant parameters form the training set, thus predicting it closely while failing to generalize.
Loss function: function that evaluates how good are the predictions made by the crated model. The higher the value is, the worse the model is predicting.

The first step to determine which models to use is to decide which kind of learning fits better the project demands. On the one hand, supervised learning is performed from labeled data, and it learns features from it to try to produce a correct output. On the other hand, unsupervised learning is an exploratory analysis that helps finding unknown patters and features from unlabeled data that would help with its classification. As the dataset is going to be created from scratch and the presence of meteorites can be easily spotted when preparing it, the dataset can be labeled without being a major inconvenience. Moreover, the visual or IR data has been thought to provide superficial information, so hidden patterns are not expected. Hence, to analyze the image, supervised learning algorithms are used, more precisely the ones based on Convolutional Neural Networks (CNN). The CNN models are a class of Machine Learning models widely used for image recognition and classification. To use them, a neural network has to be trained with an own dataset, the convolution filter values and neuron weights have to be obtained and then applied to the drone obtained images to detect meteorites. Different architectures have been used and studied to obtain the most accurate and reliable model, always aiming for the smallest model possible, to leave room to explore the possibility of incorporating the recognition software directly into the drone. Keeping in mind that often the architectural design of CNNs is referred to be more of an art than a science, the changes and explorations on the design have been made in a combination of understanding the network behavior and exploration purposes.

2.2.1 CNN structures and parameters

A wide range of CNN architectures have been proposed and studied recently for very different kind of image recognition tasks, but they are all mode of different combinations of the same common structures, which are the ones used in this project as well:

Input Layer:

In the input layer the RGB or grayscale images are fed to the architecture. The only difference between them is the information is stored in 3 values for a single pixel in the RGB case, whereas for the grayscale it only contains one channel. The pixel color values range from 0 to 255 on each channel, but they are normalized before entering the system to range from 0 to 1, as it is a common rule of thumb in CNNs input data because it helps with performance optimization.

Convolution Layers:

In each of these layers, N rectangular convolutional filters are applied to the image or the preceding convolutional results if they are not the first one. Each filter is useful to detect different parameters, from shapes and colors to gradient and shadows. The filter goes all around the image, generating another image of the same size known as activation map, so it ends up with N activation maps. The more convolutional layers the model has, the deeper it looks into the picture, so more sophisticated features are taken into account while classifying. The number of convolution filters each layer has (N) is chosen to be of powers of 2 because it presents a great advantage to the computational efficiency: it facilitates memory allocation on the GPU and also it experiences build in optimization efficiencies [15].

When moving the convolution filter, the image size of the output is designed to be the same as the one in the input, as the dimensions are planned to be reduced with the pooling layer, which is explained below. Moreover, the meteorite could be placed anywhere in the image, so no parts of the it can be left out. Hence, when moving the convolution filter centered on the corners and edges, the outer zone has been filled with zeros, action which is called 'same padding'. The size of this rectangular convolution filters is called kernel size and is a parameter which also has to be chosen according to the specific characteristics of the studied problem. The larger the convolution filter, the bigger the area of the features expected to extracted from the image are. Also, the size of the kernel is chosen to be an odd number because if an even-sized kernel was chosen, the output image will be asymmetrical with respect to the input one, producing aliasing errors.

Pooling:

After a convolution layer, a lot of new data is generated, and it is a common practice to reduce it by using pooling layers. The pooling layer that is used in the proposed architectures is known as Max Pooling Layer. This layer helps by reducing the activation maps size by keeping the maximum value in an 2x2 area which moves around the image every time it is applied without overlapping. With this procedure, the resulting layer will have its size reduced to a half of the original, as shown in Figure 2.4.

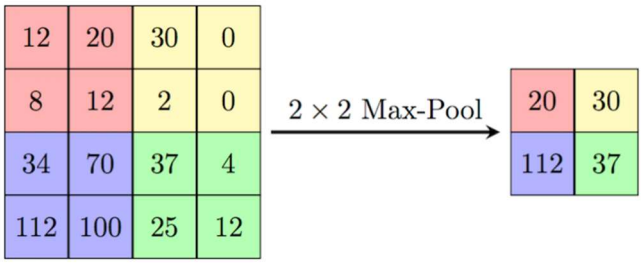

Figure 2.4 The procedure done in a Max Pooling Layer can be seen, where the maximum value of the 2x2 regions on the original map is kept at the output.

Dropout Layer

In this layer, the model randomly ignores some neuron outputs or activation map information each time it is applied, forcing itself not to rely on the same parameters, which could have a detrimental effect with regards to generalization.

Flatten Layer

After all the convolution, pooling and dropout layers, the output is represented as a single vector, so that it can be used as a standard Machine Learning neuron layer.

Fully Connected Layer (FCL)

This layer has N neurons connected to all the activations or neurons of the previous layer, enabling it to learn the different applied connection weights. Each neuron connection can be viewed as a learning parameter.

Output Layer

The last layer is an FCL with one neuron, and the value of it represents the probability of being a meteorite or a not.

This presented structures are the ones used and combined to create the models, but the way the neuron values are outputted before being transmitted to the following layer has also to be considered, which represents the election of the activation functions to be used. Its selection can have an effect on the model's speed, accuracy and convergence. In the studied case, the **Rectified Linear Unit** (ReLU) is used for the convolution layers, as it speeds up the process when the result falls in the 0-output region and has linear behavior which makes optimization easier [16]. On the other hand, as a binary classifier is going

to be created, the **sigmoid** activation is used for the final layer so the output is a value between to 0 or 1 depending on the model prediction. Both activation functions can be seen in Figure 2.5.

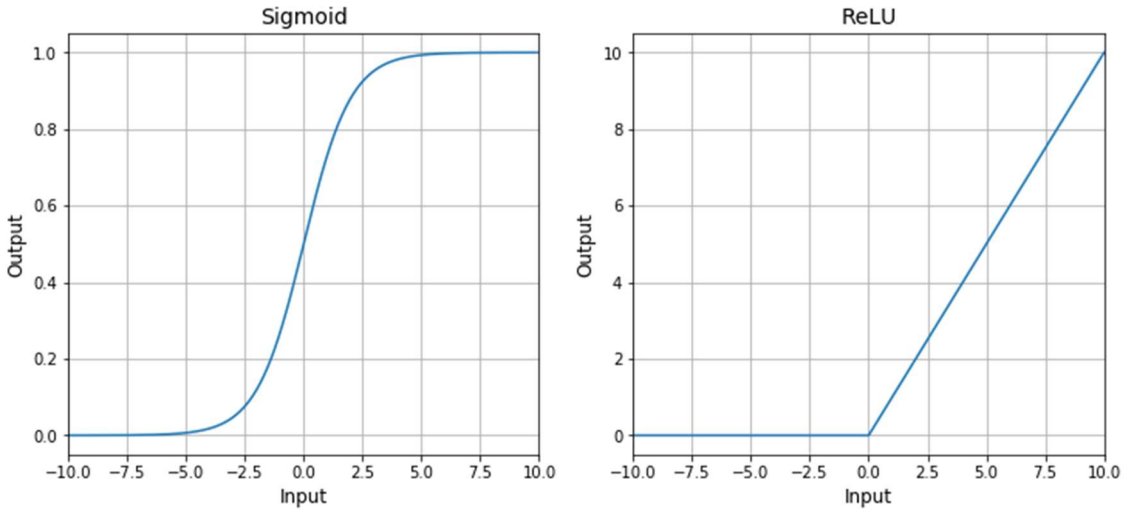

Figure 2.5 The different used activation functions are represented. The 0 to 1 range of the sigmoid function can be appreciated as well as the semi linearity of the ReLU function.

2.2.2 Machine Learning Steps

The steps followed to create and evaluate the models always stick to the same sequence: train, validate and test.

The first step is **training**, which has the objective of obtaining a set of weights and filter values which applied to new images would predict accurately what is on the them based on the knowledge obtained from the fed database. To do so, it changes the parameter values in order to obtain optimum performance. Usually the method used is the loss gradient descend, where the values of the weights are adjusted as follows:

$$\boldsymbol{w = w' + \alpha \Delta(w')}$$

Equation 2.1 The loss gradient equation is represented, where w' is the previous weight value, α is the hyper-parameter known as learning rate and Δ(w') is the gradient of the loss function.

In this mechanism, the minimum of the loss function is hopped to be approached moving according to the gradient. One of the concerns faced here is the election of the learning rate: a small one would for sure lead to the minimum of the function, but it could take a lot of iterations, whereas a very high one could fail to reach equilibrium if it kept oscillating back and forth. Moreover, it is recommended to schedule a decrease of the learning

rate as more epochs are performed to descend to deeper areas of the loss function. All the tuning and the change scheduling of the loss function parameter would require a lot of testing and computer power, so another optimization procedure is used. In the project, the adaptive learning rate stochastic gradient descent parameter known as Adam, which adjusts the learning rate making use of the first and second moments of the gradient has been chosen, as it has been shown to outperform other gradient descent algorithms [17]. The election of the Adam optimizer is also beneficial for the studied case, as it has little memory requirements, it is adequate for problems with a large number of parameters, is computationally efficient and has been widely studied for image recognition [18].

In parallel with training, the **validation** process also occurs. In this step, the obtained weights from training are applied to unseen images at every epoch, making predictions of its content. This is a very important step to make while the training process is being conducted, as it gives the ability to see how well the model does with data it has not been trained with. With these two processes the model parameters will be tweaked until desirable results are obtained. However, a last inspection has to be carried out, the **test** step. In this procedure, that only takes place once a model seems to be a promising candidate based on the training and validation results, images that were not used on the previous steps are predicted and the results are evaluated. This process is important to be sure that the model changes have been made in a general way, and not to satisfy the validation dataset characteristics.

2.2.3 Metrics

When evaluating the results in the training, validation or test steps, it is important to take into consideration how important are different metrics, such us precision, accuracy or recall to our specific problem. To define those parameters in Machine Learning terms, the following concepts are used:

- **True positives (TP)**: the model correctly predicts the presence of a meteorite on the image.
- **False positives (FP)**: the model incorrectly predicts the presence of a meteorite on the image.

- **True negatives (TN):** the model correctly predicts the absence of a meteorite on the image.
- **False negatives (FN)**: the model incorrectly predicts the absence of a meteorite on the image.

With these definitions, the metric used can be constructed:

Accuracy: is a parameter that evaluates how well the model predicts the presence or absence of a meteorite.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: is the ratio of times that actually a meteorite is present when the model predicts so.

$$Precision = \frac{TP}{TP + FP}$$

Recall: this metric evaluates the fraction of spotted meteorites by our Machine Learning model.

$$Recall = \frac{TP}{TP + FN}$$

All these metrics should be evaluated in the analysis in a different scale depending on the approach to try to get the best model possible. First of all, as done in the vast majority of Machine Learning models, the accuracy is analyzed as a general metric during the training and validation process, to get a quick glance of how well the model is doing. Moreover, once the training process is completed, one could look into the less curated data, such as the FP and FN, which also could shed some light on how to improve the model regarding the data it is fed. Additionally, a new metric called f-score is used in the evaluation and it combines precision and recall. It is a measurement which also takes values between 0 and 1, in which the relative importance of these two metrics is modeled by the parameter β.

$$f - score = (1 + \beta^2)\frac{precision \cdot recall}{\beta^2 precision + recall}$$

Equation 2.2 The f-score equation is presented. Values of β parameter >1 give more importance to recall, where values <1 give more weight to precision.

3. ARCHITECTURE DESIGN

In normal conditions, the next step performed would have been the realization of experiments and tests to determine which of the two remaining differentiation methods is the most adequate one. As it has not been possible to perform those tests or to obtain IR data from meteorites, the project has been focused on the visual identification of them.

3.1 Dataset creation

Once the drone-made photographs are taken, the process to investigate whether or not there is a meteorite in them starts. To perform so, the photographs are not analyzed as a whole but cropped into 256x256 pixel images and each of them is analyzed. Hence, the needed dataset consists of close-range images of meteorites and of non meteorites. In Machine Learning architectures for binary classification, the minimum number of photos available in the dataset is recommended to be of the order of 2000 for each class. For the background class, no problem is encountered to construct the dataset as for a single background photograph multiple images can be extracted. However, this is one of the first problems encountered when designing the meteorite dataset, as there is not an already created dataset for this specific project and not many images available online that match with our desired image characteristics, as they need to be taken from a zenithal point of view. In consequence, the original meteorite dataset was created as follows:

- Using a private collection of meteorites, images of them were taken in different background and environments. For instance, some of them were taken at the beach and in soils resembling the studied deserts' characteristics.
- Images in situ of meteorites that were available online and taken from an aerial perspective were downloaded and added to the dataset.
- To increase even further the dataset, meteorites cut from not usable images were adequately inserted into background images using an image treatment software, as done in [3].

However, these procedures resulted in initially less than a 7% of the desired database size. In order to proceed with the architectural design, different approaches were made to augment the data, each of them with different models and architectures.

3.2 Model creation

The models studied are created following the model that the Visual Geometry Group (VGG) of Oxford University created, with which they achieved top performance in various image classification competitions [19]. Apart from its excellent results in other applications, the main reason this has been the chosen reference model is that it works with small layers, reducing the number of parameters, which translate into smaller models, an important characteristic for the project purposes as previously stated. In general terms, the models are created by N pairs of convolutional and max pooling layers, followed by a flatten layer and ended with a group of FCLs, as in Figure 3.1. The parameters of this reference model have been studied and tweaked in order to obtain the best results possible.

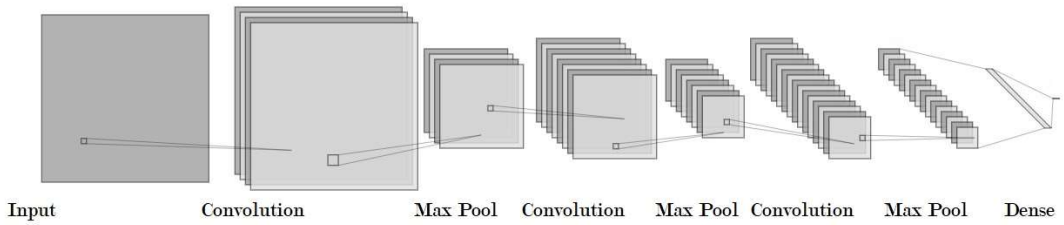

Figure 3.1 Schematic view the models created, which corresponds to one with three pairs of convolutional and max pooling layers.

3.2.1 Approach 1

In the first approach, the data was attempted to be augmented at the start of the algorithm, generating a dataset with a sufficient size, as performed in [3]. In order to do so, to each image a different number of rotations, vertical or horizontal flips and different color filters were implemented randomly, generating new images from the original one. With that, a new dataset of around 4.000 total images was created, with which started the model creation, evaluation and problem facing. This first approach is used as a reference and as a test to start evaluating different models, so when more complex and computer-power requiring algorithms are used, a generic line of work is already established. Moreover, the procedures and results are explained in more detail so that in further approaches they can be left out.

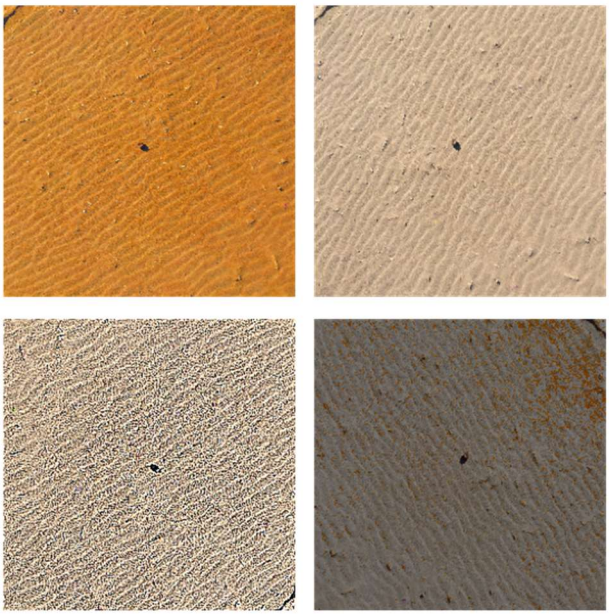

Figure 3.2 The image shows different examples of how the images are altered to increase the dataset on Approach 1. Rotations, translations and changes on brightness and sharpness can be appreciated.

Model 1: VGG with 3 Convolutional Layers

The first model that was evaluated was one with 3 convolutional layers:

input layer
convolutional layer (32, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (128, 3x3 kernel, ReLU)
max pooling layer (2x2)
FCL (128, ReLU)
FCL (1, sigmoid)

The architecture can be easily explained as follows: an input image is fed into the model, and 32 2D convolutions are applied to it, creating 32 new image activation maps. After that, the 32 activation maps are reduced to half of their size by the pooling function. The procedure is repeated with two more convolution layers with 64 and 128 different filters each. After that, the information of the activation maps is converted into a 1D vector, which is fully connected to 128 neurons. These neurons are again fully connected to a last neuron, which will produce the final output. As previously stated, the number of convolution filters is chosen to be of powers of two, and the activation functions are ReLU and sigmoid. Regarding the convolution filters, common kernel sizes (3x3,5x5,7x7) have been explored to see which provides the best results, and it has been observed that the

model is able to learn better with 3x3 sized filters. Furthermore, this result is also significant, as the number of parameters and mathematical operations performed will be the minimum possible, as 1x1 filters are discarded because it would mean that the features are so localized that are of the extend of a single pixel, which is not the studied case.

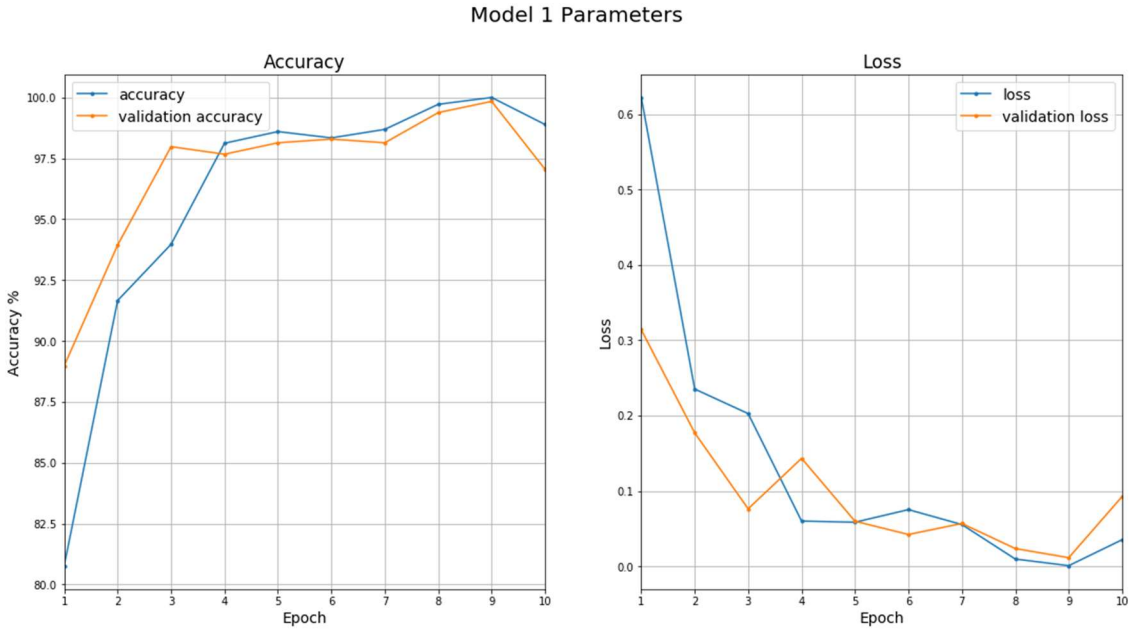

Figure 3.3 The accuracy and loss metrics from training and validation of Model 1 are shown.

When evaluating this first model (Figure 3.3), one can take a look into how many correct guesses it made, the accuracy, and how far these predictions are from the actual result, the loss function value, as the predictions are represented as probabilities. These metrics are studied both for the training and the validation processes.

At a first glance, without much CNN behavior knowledge, one should be impressed with the results, as the model with less than 10 epochs is able to predict almost always right the content of the image before starting failing with further training. Moreover, the loss is more than acceptable, finding a minimum at epoch 9. However, looking at the parameters while being a little more cautious, one can observe that at that epoch, the accuracy is of 100% and the loss is of the order of 10^{-4}. This points out that in fact, there is a very strong possibility that the model was able to memorize the inputs, so the prediction was not generalized. This idea can also be supported by the fact that the validation loss starts to significantly increase at the end, suggesting that the model drifts away from unseen data. One of the reasons this phenomenon may occur is the lack of data and how it was augmented, as some of the photos may be similar, inducing less generalized results. In order to address so, different possible solutions are presented to be studied:

- Reduce the number of filters each convolution applies, so that less information is obtained from each image.
- Introduce random dropout, so the model learns every time from a random fraction of features, complicating the overfitting of the parameters.
- Reduce the number of convolution layers, so our model retains more superficial information.
- Change the data augmentation approach so that the it does not train with similar images.

The first modification of the initial model was made by reducing the number of convolutions on each layer and also the number of neurons in the final FCL. In the studied case, the trainable parameters were reduced up to 2.116.673 from 16.870.721, which results in a reduction of 8 times in size. However, it is also observed (Annex 2) that adequate results are not achieved until a presumable overfitting starts occurring, observed as the accuracy improves while the validation loss starts increasing. This result reinforces the idea that the problem may reside on the generation of the dataset, yet the other solutions are to be explored to obtain a more solid conclusion.

Model 2: VGG with 2 Convolutional Layers

The second inspected model has the following structure:

input layer
convolutional layer (32, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
FCL (128, ReLU)
FCL (1, sigmoid)

Exploring this second model, it can be observed in Figure 3.4 that also extreme overfitting is encountered before reaching a desirable prediction score. Also, after three models evaluated so far, a pattern in the loss representations can be observed, as after the initial decay it stays in a plateau for 2 or 3 epochs before decreasing considerably, time when the overfitting starts to be noticeable. In this case, the plateau is notably higher than in model 1, meaning that the obtained predictions are worse or in other words, weaker, so model 2 will be discarded for future model architecture proposals.

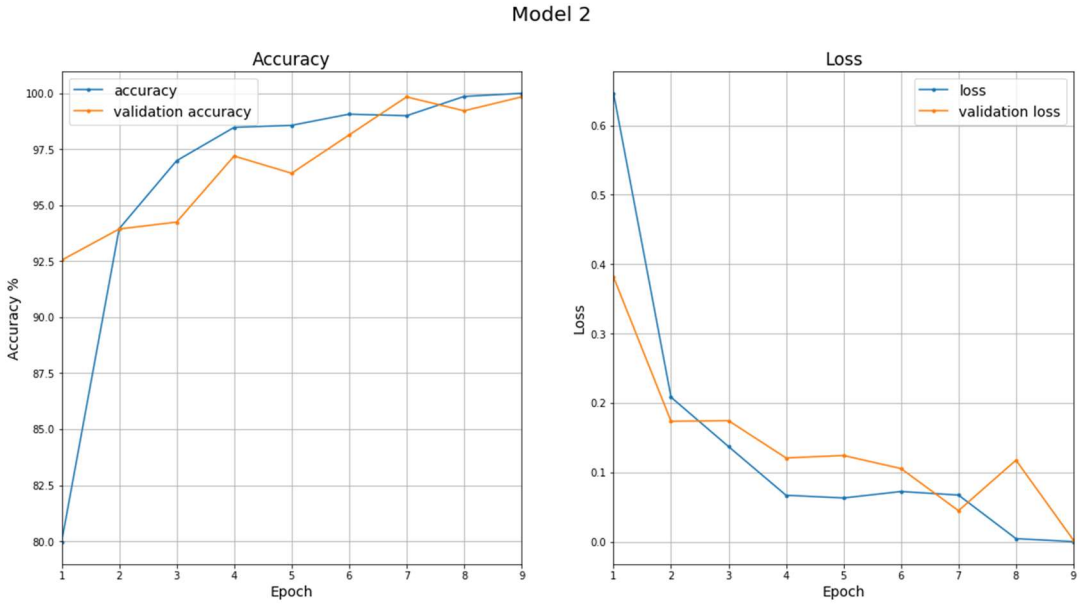

Figure 3.4 The accuracy and loss metrics from training and validation of Model 2 are shown.

Model 3: VGG with 3 Convolutional Layers and Dropout

Now, the dropout modification is explored in order to see if randomly excluding parameters help to prevent overfitting. The dropout is applied after each convolutional layer, so less information is passed to the next one, and before the final FCL.

input layer
convolutional layer (32, 3x3 kernel, ReLU)
max pooling layer (2x2)
dropout 20%
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
dropout 20%
convolutional layer (128, 3x3 kernel, ReLU)
max pooling layer (2x2)
dropout 20%
FCL (128, ReLU)
Dropout x%
FCL (1, sigmoid)

The phenomena that this new attempt aims to obtain is improving the validation results by making it difficult to memorize the images during the training process. If the model is forced to learn and predict with different extracted features every time, it is not able to rely on the same information, hence to retention of specific image characteristics is prevented. All the attempts used a 20% dropout in the convolution layers and the dropout on the FCL is changed from less aggressive rates (20%) to more aggressive ones (50%). The reasoning why only the dropout on the FCL layer is increased is because when aiming for

generalization, the penalization of the feature information is considered to be more relevant than the feature extraction.

First of all, the models with 20, 30 and 40% dropout on the FCL were studied. As seen in the accuracy plots of Figure 3.5, the learning rate decreases as the dropout increases, an expected result, but all the models are able to learn and predict correctly without reaching the incredibly high results the models were getting without dropout. It can also be observed that the validation loss decay is slower than the training loss one in the vast majority of the epochs. This is a good result, as it shows that the dropout of the parameters complicates the predictions during the learning and creates a more robust model to predict with, obtaining better results when the full information is used.

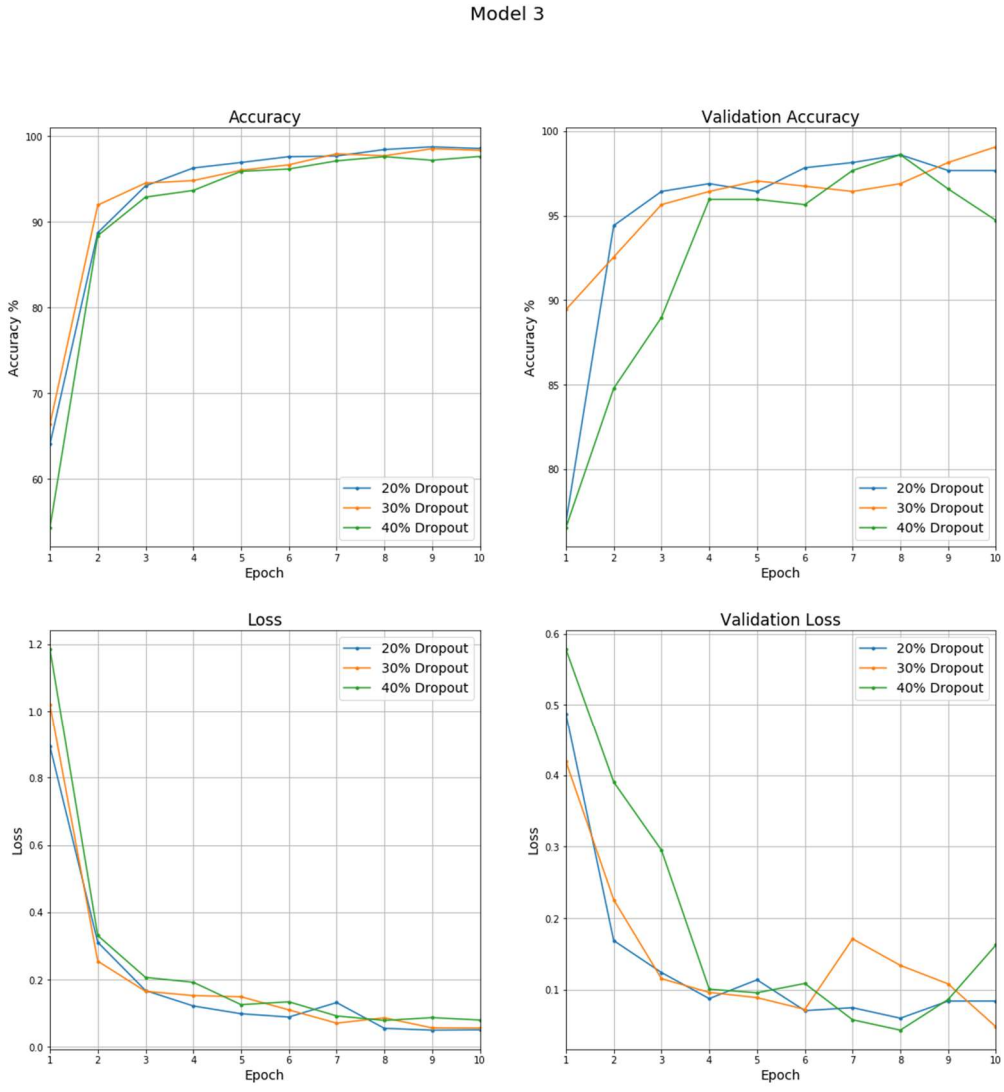

Figure 3.5 The figure contains 4 different plots of the accuracy and loss for both validation and training processes for the models with low dropout rate on the FCL.

Finally, the model with a 50% dropout layer at the end was tested. In the results it is observed (Figure 3.6) that the model struggles to learn during a high number of epochs,

but then the learning process starts and some good results are seen. It can also be seen that although 50% of the final parameters were randomly discarded, the training accuracy is high, which suggests that this is one of the most strong and robust models created so far, and also it shows great final validation loss values. However, the effect of the data augmentation process could still be a crucial factor to understand the high values of the accuracy.

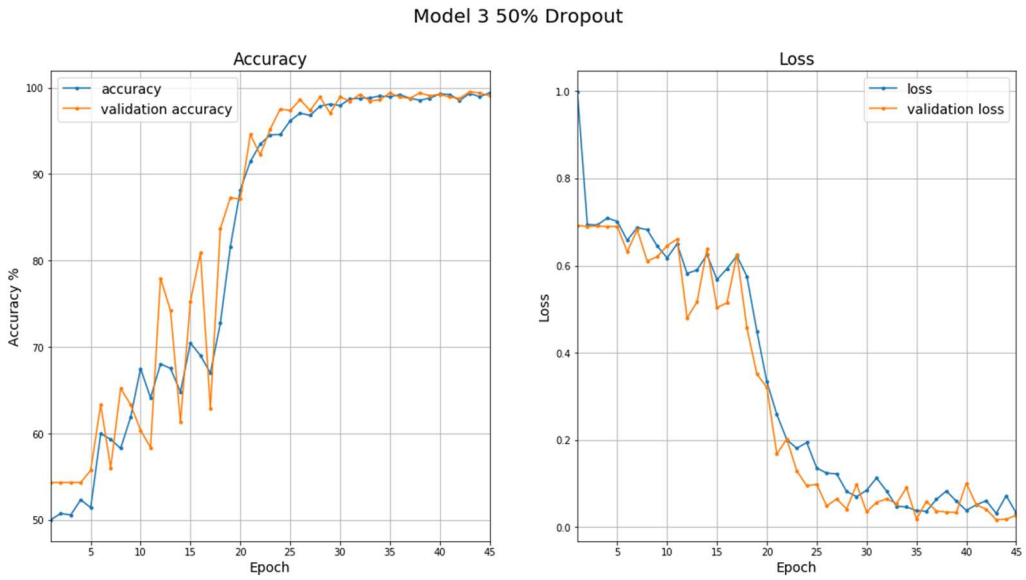

Figure 3.6 Metrics of the model with a 50% dropout on the FCL are shown. The different behaviors can be observed: first of all, a stationary process where the model is not able to learn followed by a learning period which ends stabilizing both for loss and accuracy.

Initial tests have been done with some unseen aerial images taken at the beach. It can be seen in Figure 3.7 that the model is able to correctly spot the meteorites, but there are some prediction errors which suggest that color and contrast took an extremely important weight on the Machine Learning algorithm, meaning that the model sometimes fails to differentiate important shadows in the background. In order to correct so, more data corresponding to darker sites of the background should be applied to the database.

Figure 3.7 *Results with model 3.5 are shown, were the meteorite has been correctly spotted.*

3.2.2 Approach 2

As it has been seen in the first approach, overfitting was encountered very quickly in the models. As the data was seen to be the main problem, a second approach was studied which is based on loading the data in batches and performing random changes to it each time an epoch is run, always guaranteeing that the output image is coherent with a real scenario, meaning that during training the model does not see the same exact image twice. However, the number of epochs needs to drastically increase as the volume of input data is reduced. The random changes introduced are:

- Rotation (from 0 to 30 degrees)
- Horizontal Flip
- Zoom (from 0 to 20%)
- Shearing (from 0 to 20%)
- Brightness (from -50 to +50%)

The first 4 changes are implemented to simulate different orientations and sizes of the meteorite and the brightness is applied to explore the different sunlight levels representing different times of the day.

Figure 3.8 The image shows 9 different modifications a single photograph of the dataset experienced. It can be observed that the images are clearly different, so the model will not observe exactly the same image more than once.

First of all, it was checked if with this new type of data generation, the model could potentially be able to work properly and obtain interesting results. To do so, starting with the general model used in Approach 1 and with a very limited dataset, it was trained to

see if it could be overfitted. The results were promising as overfitting was achieved quite quickly, so the model was suitable for getting a dataset size increase.

With this new approach, fine tuning the parameters of the architecture to obtain the best predictions possible was the first starting point. The first step was tuning the number of neurons of the last dense layer, which represents the number of features extracted from the last group of convolutions. To evaluate the results both accuracies were tracked once the plateau of the training curve was reached.

FCL Neurons	Accuracy	Validation Accuracy
64	0.503	0.681
88	0.575	0.500
128	0.841	0.732
150	0.583	0.559
180	0.810	0.680
200	0.808	0.74
210	0.833	0.733

Table 3.1 The different accuracies obtained for the various dense layer's neurons are shown.

It can be observed that for a low number of neurons in the last FCL, once the model learning process reaches a stagnation point it has not been able to learn properly how to differentiate. For a higher number of neurons, the accuracy increases to similar levels in the different tests, except for 150 neurons, where the model underperforms. However, it was also observed that with 128 neurons, the loss value is significantly lower, meaning that the model predicts correctly with more confidence. Moreover, as there is an interest in having a simple architecture with low hardware requirements, the optimal number of neurons for the FCL is chosen to be set at 128.

The next step was to explore the different performances for different number of convolutions on each of the three layers. The relation applied between the layers follows this basic principle: the deeper the layer, the more abstract the features it extracts. Hence, guided by that statement, the proposed architectures have convolutional layers with equal or higher number of convolution filters than the previous one. The common number of con-

volution filters used in a convolution layer of a typical CNN is the powers of two contained in the 32 to 512 range. To continue in the low requirements line, only the 3 lowest number of filters have been explored. The results are shown in Table 3.2.

L1	L2	L3	Accuracy	Validation accuracy
32	32	64	0.503	0.545
32	64	64	0.878	0.733
32	64	128	0.864	0.727
32	128	128	0.533	0.610
64	64	128	0.900	0.772
64	128	128	0.541	0.636

Table 3.2 Accuracy and validation accuracy for architectures with different number of convolution filters on the convolution layers are shown.

The obtained accuracy values show that a sequence of convolutions layers of 64, 64 and 128 convolution filters presents the best prediction results. However, the accuracy values are far from the desired ones. In order to address so, other important parameters of the models, the false positives and false negatives, are studied. The data on them shows (Figure 3.9) that although the model struggles at first at identifying the meteorites, as the false negatives values are high on the first epochs, soon enough it starts gaining traction and the vast majority of meteorites are detected. However, this increase on detections is due to a higher tendency of the model to identify doubtful images as meteorites, resulting in an increase on false

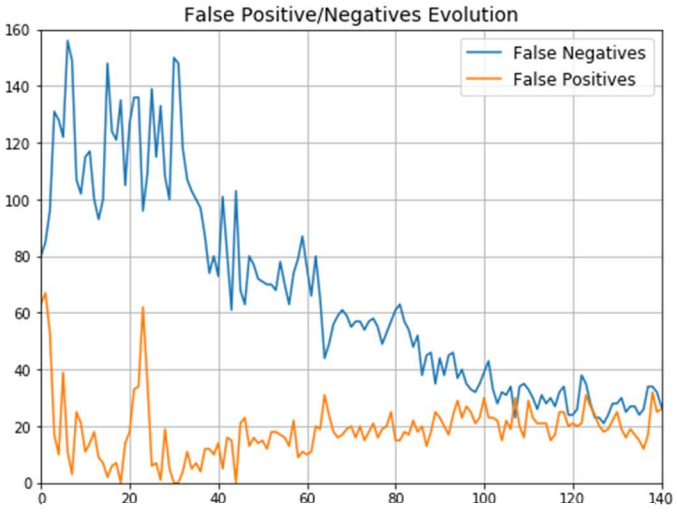

Figure 3.9 Evolution of the False Positives and False Negatives. A decrease on FN can be seen, but conditioned by a slightly increase on FP.

positives, mainly corresponding to the identification of rocks.
To face this problem, two different solutions are proposed:

- Increase further the database with more images of rocks that are not meteorites.
- Create two CNN models connected in series (Approach 3), so that the first one detects both rocks and meteorites, and the second one distinguishes between them.

After increasing the database, the first proposed solution was tested but no remarkable results were obtained. Thus, the third approach is explored.

3.2.3 Approach 3

This third approach has been proposed to reduce the amount of details the model has to learn at once, facilitating the meteorite identification. The reasoning behind the new proposed architecture is that the model could be struggling when trying to learn from non-meteorite images that contain rocks, as the entirety of the meteorite dataset is composed of elements with shapes similar or equal to normal earth rocks. With this new structure in mind, two new datasets have been created, reusing the existing images but also getting new ones to even the number of images in both classes of the dataset. This last step is a very important one to guarantee that the model weights are not biased to fit better the class with more images. Hence, the final structure is as follows: a first database, with one class for rocks and meteorites and another one just for background, and a second database, with one class for rocks and one for meteorites. Each database has a CNN architecture associated to it, which is created following the same steps as in Approach 2: first of all is using a very small database the model is overfitted to be sure that in some extend it is going to be able to learn and differentiate, and then different layer sizes, structures and parameters are tested to get the best model.

Architecture 1

The first architecture differentiates background images from images containing rocks or meteorites. In order to make this third approach work, very high accuracy was expected to be obtained on the first architecture, as it has been given a more general task. In this case, the model is expected to identify rocks or meteorites from more unspecific and general characteristics, such as shape or depth. This assumption, backed up by the fact that

rocks can be found presenting a wide range of colors, opened the possibility to work with grayscale images, which introduce two improvements to the model. First of all, it reduces the size of the model, as now the images only have one channel instead of the three RGB it had previously, which reduces the dimension of the convolution filters to 2D. Secondly, as the irrelevant color information is manually removed, the model is hopefully driven to focus the learning process on the desired features. With grayscale images, the filters and the images formed on the different convolutional layers are easier to observe and understand. As seen in Figure 3.10, on the two first convolution layers the filters tend to obtain activation maps which highlight the rock or meteorite as a whole, standing out from the background, whereas the third one focuses more on the contour lines.

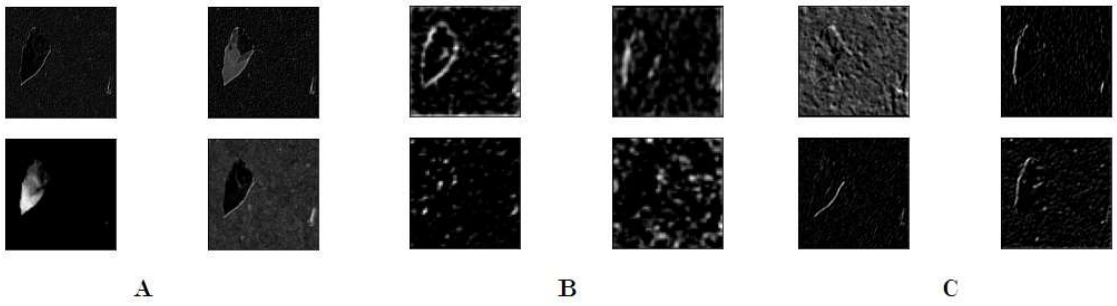

Figure 3.10 From the same input meteorite image 4 outputs from the first (A), second (B) and third (C) convolutional layers are shown.

After studying the different color of the input images with the usual parameter changes, interesting results were obtained. First of all, in both RGB and grayscale images, the models worked better with a lower number of convolution layer filters than the obtained in the previously designed architectures. This result can be understood to be due to the fact that in this architecture the model needs to obtain fewer features to classify, so if the model has less room left for feature selection, the less generalized characteristics have more tendency to be left out. When comparing the different colorings of the images (Figure 3.11), it is observed that similar values for the validation accuracy were obtained, topping at 98%, but the validation loss was found to be twice lower when working with grayscale images, suggesting the model predicts with more confidence. After further testing them to obtain better results, applying only a dropout layer at the FCL proved to encourage the model to achieve better results, as overfitting started occurring latter. This last model, whose overview is shown below, is the definitive one that is going to be used in the first architecture.

Architecture 1: Model Overview

input layer
convolutional layer (32, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
FCL (128, ReLU)
dropout 10%
FCL (1, sigmoid)

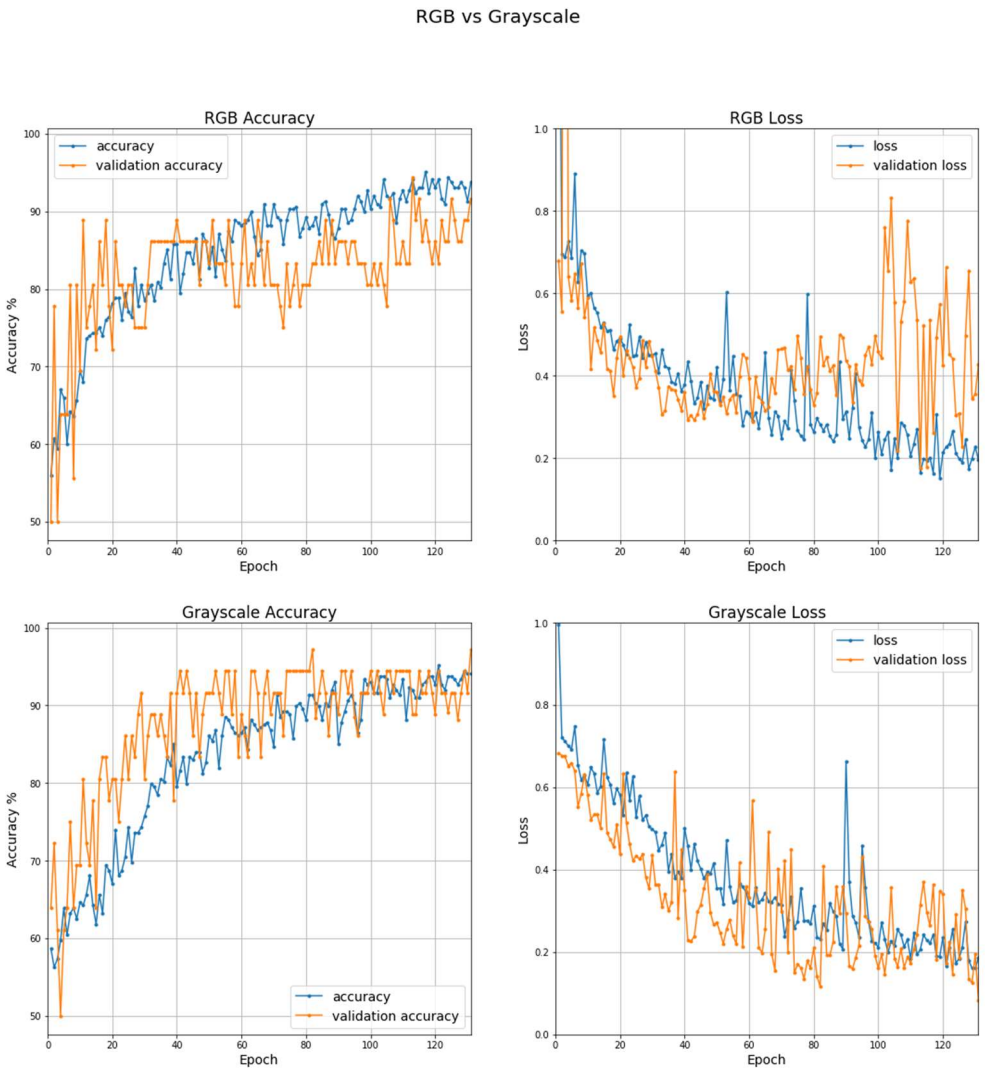

Figure 3.11 The figure shows the differences between the RGB and grayscale models. Regarding the accuracy, it can be seen that the grayscale model reaches an accuracy plateau higher than the one obtained on the RGB. With respect to the loss values, a clear overfitting can be seen in the RGB plot, as while the training loss keeps decreasing, the validation loss breaks its downward trend to start with a substantial increase. However, the grayscale losses decrease simultaneously allowing the obtention of more usable models.

Architecture 2

In this second architecture the model should differentiate between the two classes by inspecting some more in-depth characteristics than the ones explored in the first architecture, such as color or textures. The creation started with assumptions made following the same reasoning as in Architecture 1. First of all, due to the bigger number of parameters and features the model is expected to need and use for the identification, the models tested have a larger number of trainable parameters in comparison to the Architecture 1. Moreover, in this approach, the RGB is the chosen color scale, as the color is expected to be a crucial factor when classifying.

In this specific architecture, the precision and recall metrics are given more importance, always thinking accordingly to the problem with the analogy of the human search of meteorites. During the meteorite inspection, the encounter of some FP results is expected, as it also a common event in usual meteorite hunting when a detailed inspection is needed to determine if the rock is extraterrestrial. Hence, an unsuccessful check is not a huge inconvenience, as the frequency of meteorite identification is not expected to be very high. For this same reason it is certainly a problem encountering an FN, as leaving a meteorite behind is not desirable. So, more importance is given to the recall than to the precision of the method, making use of the f-score metric with the β parameter set to be 1.5.

The tests obtained the expected results as bigger models than in Architecture 1 showed better metric values, being a the 64x64x128 architecture the one with an overall better performance. Also, the number of neurons of the last FCL was tuned according to the f-score results, as seen in Figure 3.12. The results show that a higher number of parameters does not strictly imply better results, and the better ones were obtained in the range of 160 neurons on the FCL layer, which led to up to f-score values of 96%, a desirable value. Hence, the model of the Architecture 2 that will go through the testing phase will be made of that resulting model:

input layer
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)
convolutional layer (64, 3x3 kernel, ReLU)
max pooling layer (2x2)

convolutional layer (128, 3x3 kernel, ReLU)
max pooling layer (2x2)
FCL (160, ReLU)
FCL (1, sigmoid)

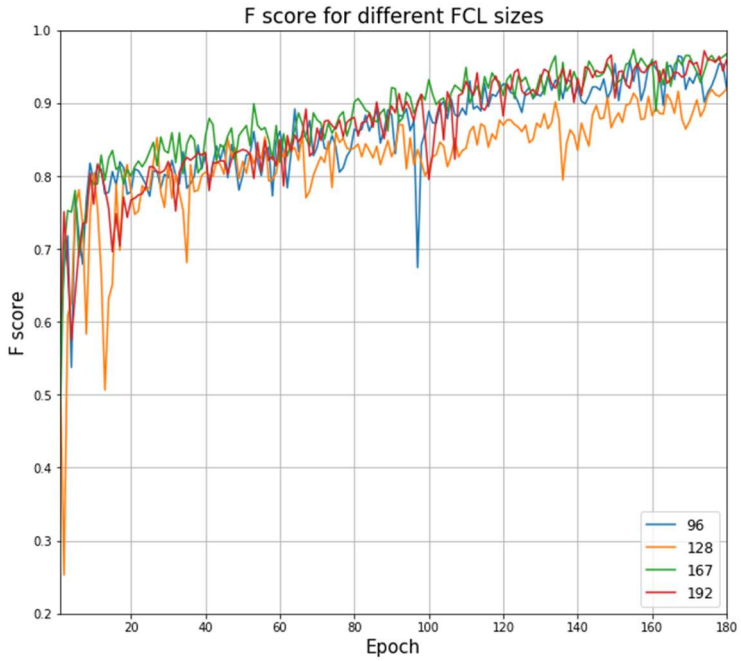

Figure 3.12 The validation f-score value evolutions for different neurons on the last FCL are show.

With both architectures created and with the completely acceptable results obtained, Approach 3 is selected to be the definitive one. Therefore, the whole model is tested to perform predictions made to some new images which now resemble the ones the drone would take. The photographs are cropped into 256x256 pixel images, and each of them is inspected. These image partitions are made assuring that the meteorite is fully visible in at least one image, as the database does not contain partial images of meteorites or rocks because their shape was designed to be a characteristic parameter. Therefore, the crops consist of two partially overlapped partition grids, the main one is as shown in Figure 3.13, and the second one is the same grid but shifted to start in the middle of the first partition square. Once the different smaller images are obtained, the model resulting from Architecture 1 is applied to all of them and their content is predicted, being the images containing rocks or meteorites spotted. This selected images are then fed into the model from Architecture 2, which predicts whether or not a meteorite is in them. As the model prediction results are outputted as probabilities, the confidence threshold to accept the prediction as valid can be chosen. Taking into account that the first architecture yield

nearly perfect results, and that some objects found on the surface could slightly confuse the model, its threshold is set higher than the one for Architecture 2. Finally, all the positive results are merged and reduced if there are repeated ones originated by the different partition grids. The test code is shown in Annex 3.

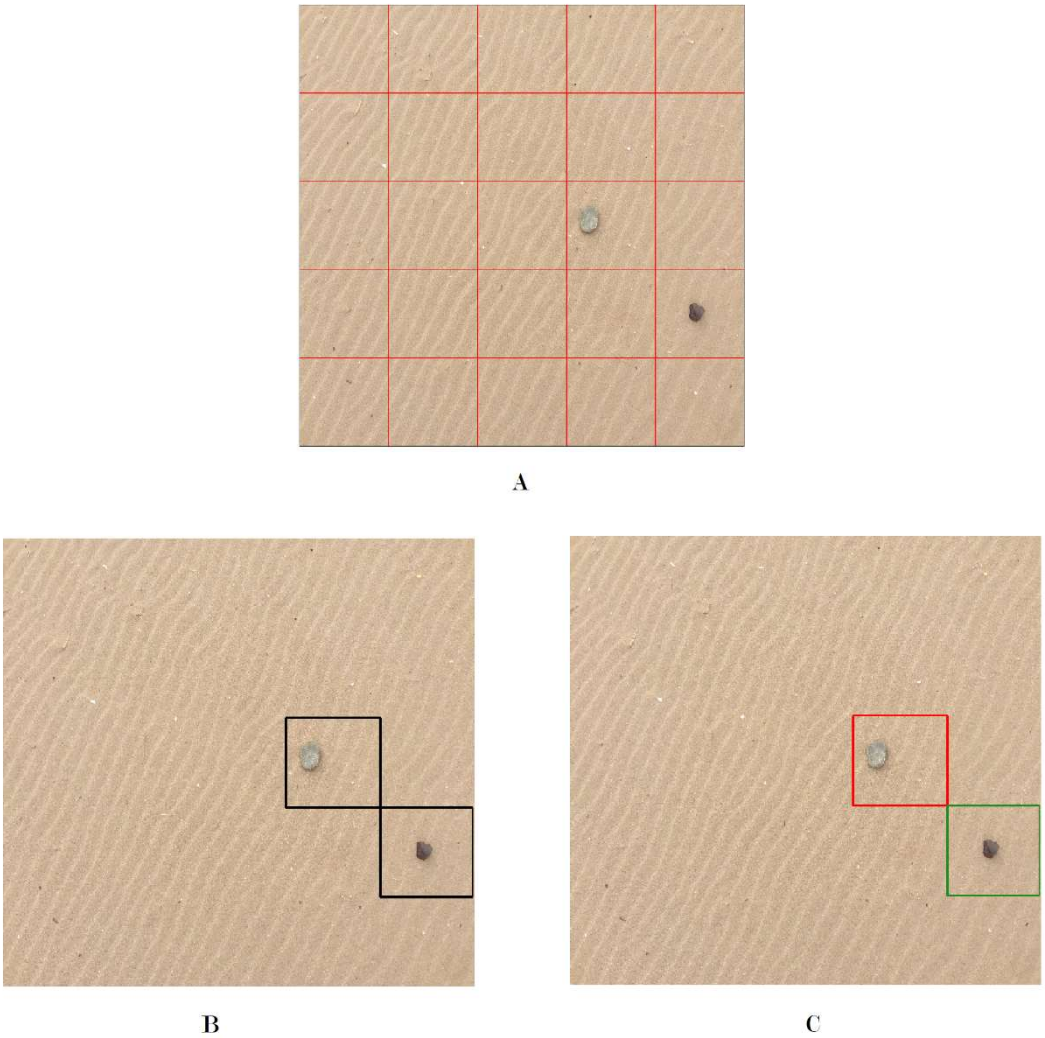

Figure 3.13 Image A shows one of the grids in which the drone taken photograph is cropped to perform the predictions. The different working stages of Approach 3 are also shown. First of all, the first CNN model of Architecture 1 spots the meteorite and rocks among the background images (B). Then, the second model classifies them into those two groups (C).

The test processes have given positive results, which do not differ from the ones obtained with the validation dataset. Therefore, Approach 3 is definitively accepted as the recognition model to use.

4. DRONE

The project is based on using a drone to perform the area coverage system. Multiple drones exist with different specifications, and a suitable one has to be chosen taking into account the desired functionalities. In this section, these specifications are explored.

The most important parameter that has to be taken into account in the election of the drone for the project is the autonomy versus area coverage, which is influenced by:

- Batteries: The type batteries used and its behavior with the desert temperatures can influence the autonomy.
- Fly path: An efficient path must be designed, considering or not double check routes.
- Height: Depending on the height of the flight of the drone more space could be searched in less time.
- Camera specifications: The field of view of the camera and its resolution have an important effect in the height parameter, thus the covered area.
- Payload: The more extra equipment attached to the drone and the heavier it is, the less autonomy the drone has.

4.1 Batteries

Long flights are of interest to cover as much area as possible with every battery change, so having large capacity batteries is a must. Generally, the larger the capacity, the heavier the battery is, so this added weight must be taken into account when exploring the payload capacity of our drone. Another important characteristic is the cell count or the battery voltage. A higher battery voltage results in more power provided to the drone motor, hence a higher speed, but those batteries are usually heavier as they contain more cells. However, for data collection purposes, covering the surfaces efficiently is more important than doing it quickly, so batteries with 4 or less cells are the more suitable for the task. Regarding the battery type, different types are available: Lithium Polymer (LiPO), Nickel Metal Hydride (NiMH) and Nickel Cadmium (NiCd) batteries. Considering the specifications [20], the wiser election are the 4 cell LiPO batteries, as they are the ones with

higher capacity and higher discharge rate for the same weight, and the only major drawback is a shorter lifespan.

4.2 Flight path, height and camera specifications

The flight path the drone will follow during the meteorite detection will be preprogrammed in order not to fail to scan any possible position. As the meteorite search will take place in the open flat extensions of the desert, the path will be designed with straight lines that turn to fill a rectangle, as shown in Figure 4.1. Currently, the path design can be easily programmed with a wide range of free user-friendly software apps [21].

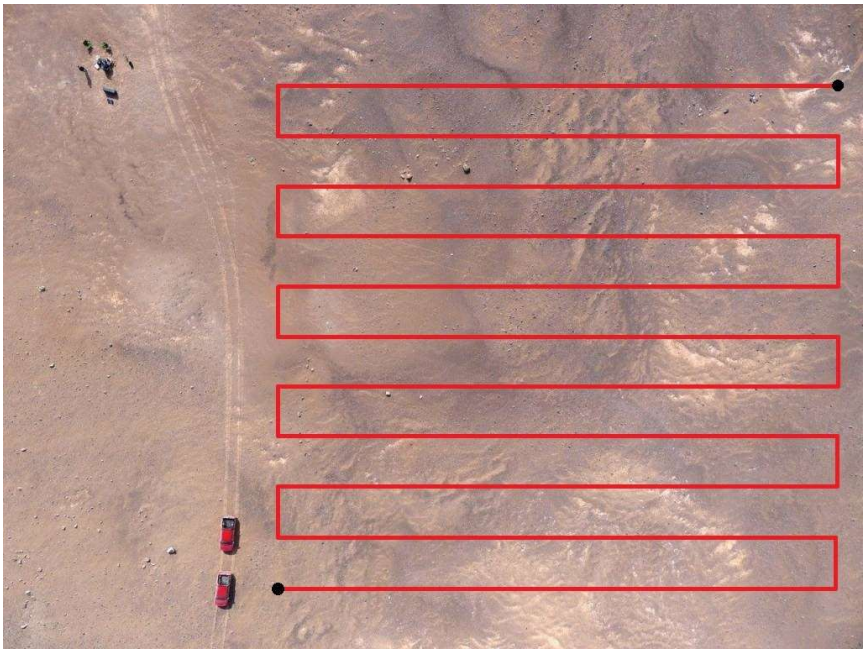

Figure 4.1 The image shows an example of a simple flight path of the drone over an area of the Atacama Desert. It could be extended with another return flight path perpendicular to the one shown on the image to deal with double checks.

One of the most important parameters in the flight path is the flight height. This parameter is influenced by different factors, some of which are related to one another:

- Separation between lines
- Camera Field of View (FOV)
- Image Resolution
- Size of the cropped images

The size of the cropped images was fixed during the creation of our dataset, so to be coherent and help the identification model with the detection, it will be kept to squares of 35 centimeters. This parameter, when transferred to pixels, was fixed during the Machine Learning architecture construction to 256x256 pixels. These parameters give us the cm/pixel value, which is of 0.136. This value also fixes the separation between lines depending on the chosen resolution. As having to rescale those images needs to be avoided unless is strictly necessary because it will lead to losses in quality causing a decrease in the identification model performance, the correspondence of real size and pixel values will be kept the same that is captured with the camera. This restriction, along with the image resolution and the FOV, which is the angular extend that can be observed with the drone camera at a particular time, fix the desired height of our drone flight. So, a wide range of FOV and different usual image resolutions in drone-suitable cameras have been studied to observe the flight height.

$$\boldsymbol{height} = \frac{(\boldsymbol{real\ image\ separation}) \cdot \frac{\boldsymbol{width\ pixels}}{\mathbf{256}}}{\mathbf{2} \cdot \boldsymbol{tan}(\frac{\boldsymbol{FOV}}{\mathbf{2}})}$$

Equation 4.1 The optimal height of the flight is estimated. The FOV value, which is provided in the camera specifications, is the one corresponding to the horizontal direction, which is the widest.

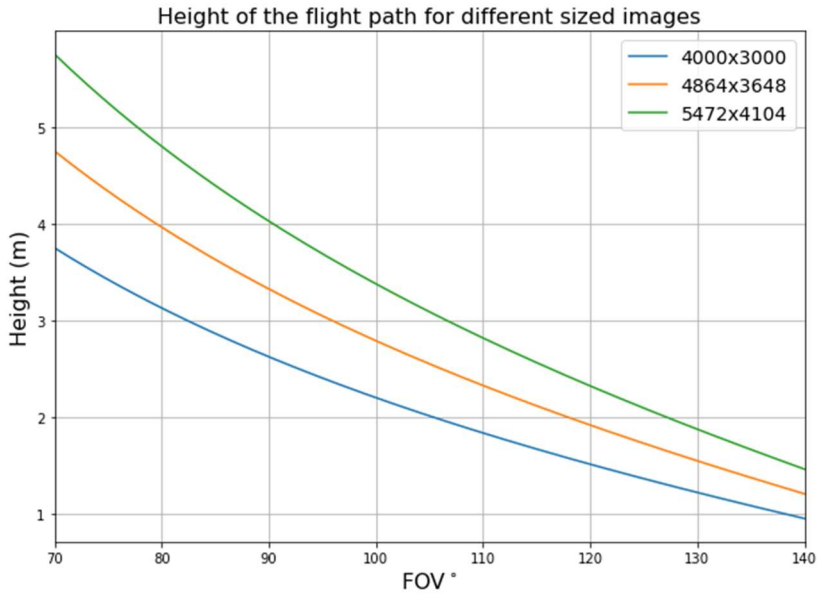

Figure 4.2 The plot contains the best altitude the drone may flight depending on the different resolutions of the image.

As seen below, a wider FOV allows us to flight nearer to the ground whereas higher resolution allows us to flight higher. To avoid the possible obstacles that could be found during the flight, a safety height of 2 meters has been arbitrary decided, so the election of specifications should lead to heights higher than it.

Finally, in order to be able to operate the drone for an extended period of time, the data, which includes the photographs along with the GPS coordinates of every image, will not be transmitted simultaneously with the flights. Doing it concurrently would apply unnecessary stress to the batteries, as the usual methods data transmission processes are battery consuming, reducing the autonomy of the drone. Instead, it will be performed after the flight has stopped which does not represent a problem. First of all, the meteorites are inanimate heavy objects, and a sudden significant change in position respect to the one determined by the drone images would only be possible if heavy rain, wind or other weather conditions were present, which are incompatible with a drone flight. Moreover, the drone flight path may cover some parts of the area more than once in order to be sure that no meteorite is left out by being on the edge of image or covered by the drone's shadow, for instance. Hence, moving to the meteorite find site would not be recommended while the drone is still flying. This process forces the drone to retain all the images on its hardware, and the volume of them could be enormous. So, the area covered by each photo for the different resolutions and the storage size occupied by the images is studied:

	Area covered (m^2)	**Storage Size (Mb)**
4000x3000	20.67	18
4864x3648	33.16	26.6
5472x4104	40.51	34

Table 4.1 The relation of covered area and the storage size of the images for the different resolutions is shown.

As it can be observed, a higher resolution will be able to cover a bigger area in each photo while it will not induce a memory storage problem as it could be thought. This means that with higher resolutions the drone will have to take less photographs and presumably less velocity changes, which translate in a more battery-efficient path.

5. CONCLUSIONS

As it has been proven with the results exposed in this work, recognizing meteorites from aerial images is a doable task that can be performed with high levels of accuracy using Convolutional Neural Networks. The method that showed better results consists in using two different CNNs connected in series, one separating rock-shaped objects and the other spotting the meteorites out of them, because when the rocks were included in the background dataset the model tended to wrongly label them. Moreover, the idoneal way to increase the dataset has been found to be by applying random changes at every epoch while working with a small dataset, since working with a full augmented dataset yield to generalization problems. Furthermore, no drawbacks or major problems have been found regarding the data obtention by means of a drone.

Additionally, some room has been left for improvement and further research regarding the identification of meteorites through IR as the theoretical study suggested that it could be a possibility. If tests were performed to try verify the IR differences between meteorites and rocks and they yield positive results, new classification algorithms could be proposed. On the one hand, independent CNNs models could be tested with the new IR data, but on the other hand, methods which worked with visual and IR data in parallel could be proposed and studied to improve the ones obtained in this study. In this scenario, the drone's autonomy would be reduced, as the IR camera placed on it would increase its payload.

Finally, finishing the project with a full test of the whole process would have been the icing of cake. Unfortunately, due to the strange circumstances of these last months, it has not been possible to advance on that line of work. Therefore, the project concludes leaving unexplored the possibility of using new data from the meteorites to identify them and without an in-situ full scale test.

6. REFERENCES

[1] C.M. Pieters and L. Mcfadden, Meteorite and asteroid reflectance spectroscopy, Annual Reviews of Earth and Planetary Sciences. 22 (2003)

[2] Meteoritical Society, Meteoritical Bulletin Database. https://www.lpi.usra.edu/meteor/metbull.php (2019)

[3] O. Norton, L. Chitwood. Field Guide to Meteors and Meteorites. Springer pp 75-175 (2008)

[4] R. I. Citron et al., Meteorite recovery using an autonomous drone and Machine Learning, Lunar and Planetary Science XLVIII. 2528 (2017)

[5] Svend Buhl. Notes on meteorite fusion crust. https://www.meteorite-recon.com/home/meteorite-documentaries/meteorite-fusion-crust (2018)

[6] P.Bland, M. Zolensky et al. Weathering of Chondritic Meteorites. University of Arizona Press, pp.853-867 (2006)

[7] S.Sinah et al., Meteorite discovery using unmanned Aerial Vehicles, NASA Frontier Development Lab Proceedings Pack, pp 52-57 (2017)

[8] C.Opeil, D. Safarik et al. Stony meteorite thermal properties and their relationship with meteorite chemical and physical states, Meteoritics and Planetary Science. 47, pp. 321,327-328 (2011)

[9] V.Cermak and L.Rybach. Thermal properties: Thermal conductivity and specific heat of minerals and rocks. Springer Verlag-Berlin. pp. 306-309 (1982)

[10] W. Verheye. Observations on desert pavements and patinas. Pedologie. 36, pp. 304-306 (1986)

[11] J. Tàpia, R.González et al. Geology and geochemistry of the Atacama desert. Antonie van Leeuwenhoek, 111, pp. 1274-1281 (2018)

[12] M. Szurgot On the Specific Heat Capacity and Thermal Capacity of Meteorites. Lunar and Planetary Science LVII . 1150 (2011)

[13] G.J. Consolmagno. Stony meteorite thermal properties and their relationship with meteorite chemical and physical states. Lunar and Planetary Science XXXV. 2108 (2004)

[14] C. McKay, E. Gomez-Silva et al. Temperature and Moisture Conditions for Life in the Extreme Arid Region of the Atacama Desert: Four Years of Observations Astrobiology. 3, pp. 393-406 (2003)

[15] Nvidia, Optimizing Convolutional Layers. User Guide. https://docs.nvidia.com/deeplearning/performance/ (2020)

[16] I. Godfellow, Y. Bengio et al. Deep Learning. MIT Press. pp. 96-120, 137-152, 224-255 (2016)

[17] D.Kingma and J. Lei Ba. ADAM: a method for stochastic optimization. International Conference on Learning Representations. Conference Paper (2015)

[18] S.Ruder. An overview of gradient descent optimization algorithms. CoRR pp. 4-8(2017)

[19] K. Simonai and A. Zisserman. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations. Conference Paper (2015)

[20] Drone Nodes. LiPo Drone Batteries https://dronenodes.com/lipo-drone-battery-explained/ (2017)

[21] Heliguy. Drone Apps https://www.heliguy.com/blog/2019/03/12/15-drone-apps-to-help-you (2019)