UNIVERSITY OF PIRAEUS & NCSR "DEMOKRITOS"
MSC PROGRAMME IN ARTIFICIAL INTELLIGENCE

Photography Style Analysis using Convolutional Neural Networks

by

Michael Zouros

Submitted
in partial fulfilment of the requirements for the degree of
Master of Artificial Intelligence
at the
UNIVERSITY OF PIRAEUS

Supervisor: Theodoros Giannakopoulos
Researcher B

Athens, 3 2022

Photography Style Analysis using Convolutional Neural Networks
Michael Zouros
MSc. Thesis, MSc. Programme in Artificial Intelligence
University of Piraeus & NCSR "Demokritos", 3 2022

UNIVERSITY OF PIRAEUS & NCSR "DEMOKRITOS"
MSC PROGRAMME IN ARTIFICIAL INTELLIGENCE

Photography Style Analysis using Convolutional Neural Networks

by

Michael Zouros

Submitted
in partial fulfilment of the requirements for the degree of
Master of Artificial Intelligence
at the
UNIVERSITY OF PIRAEUS

Supervisor: Theodoros Giannakopoulos
Researcher B

Approved by the examination committee on 3, 2022.

(Signature)	(Signature)	(Signature)
.....	.....	.....
Theodoros Giannakopoulos Researcher B	Ilias Magklogiannis Professor	George Giannakopoulos Researcher B

Athens, 3 2022

UNIVERSITY OF PIRAEUS & NCSR "DEMOKRITOS"
MSC PROGRAMME IN ARTIFICIAL INTELLIGENCE

Declaration of Authorship

(1) I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

(2) I confirm that this thesis presented for the degree of Master of Science in Informatics and Telecommunications, has

(i) been composed entirely by myself

(ii) been solely the result of my own work

(iii) not been submitted for any other degree or professional qualification

(3) I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or processional qualification except as specified.

(Signature)

.....

Michael Zouros

Athens, 3 2022

Acknowledgments

This thesis was prepared and submitted in the context of the Inter-institutional Postgraduate Program “Artificial Intelligence”, organized by the University of Piraeus and NCSR “Demokritos”. I would like to thank my supervisor, Principal Researcher Theodoros Giannakopoulos for his continuous and impactful guidance, as well as for the excellent communication we established throughout this first life cycle of this project. The professors and researchers of the inter-institutional program for offering their valuable knowledge in a plethora of different artificial intelligence related topics. My friends and family for their continuous support during these years. “Calderone” Theatrical and Photography Workshop for guiding and advising me in my first steps on the vast world of photography. “ATHENA” Research and Innovation Center for providing the infrastructure necessary for hosting our servers. Finally, Maria Stella Nikolaou, Cassandra Rose Heatley, Konstantino Tzivako, Catherine Dima, Evangelia Baou and Panagiota Tasiopoulou who helped with the cross-validation of the annotation process.

Abstract

This thesis studies the artistic nature of photography and tries to construct a framework for the definition of the term "photography style". It goes deep into the history of photography and analyzes a plethora of aesthetics that have been carved throughout the ages. Through this journey it collects the most important rules of aesthetics and groups them in specific categories. Then, with the help of deep learning and computer vision, it is able to train and predict on those specific categories.

Firstly, the reader is introduced to the world of photography. We present its historical background and then focus on its huge rise in the era of the social media. We then analyze some basics of photography, as well as some of the most known rules of aesthetics. We finally emphasize on the difficulty to bind those rules into a specific problem with specific tasks due to the subjectivity of photography and arts in general.

We then present a novel dataset of photographs annotated in terms of the respective image aesthetics. We also examine the ability of Convolutional Neural Networks (CNNs) to distinguish between the adopted photography style classes. In particular, we have defined five photography style classification tasks, related to the following aesthetic attributes: Color, Depth of Field (DoF), Palette, Composition and Type. We then followed an annotation procedure using on a set of 1832 photos selected from the Unsplash Full dataset. Multiple annotators have also been used, in order to measure inter-annotator agreement.

As soon as the dataset was compiled, we trained and evaluated a Residual Neural Network (ResNet50). The experimental results prove that, despite the imbalanced dataset, our model was able to achieve acceptable classification results. The dataset is openly provided, along with the trained models and Python code to use them. [1].

Keywords: photography · image aesthetics · deep learning · CNN · ResNet50

[1] github.com/magcil/deep_photo_aesthetics

Contents

List of Tables

List of Figures

List of Abbreviations

AAI	Authentication and Authorization Infrastructure
AI	Artificial Intelligence
ANN	Artificial Neural Networks
API	Application Programming Interface
AVA	Analysis of Visual Aesthetics
BCE	Binary Cross Entropy
DCNN	Deep Convolutional Neural Networks
DL	Deep Learning
DoF	Depth of Field
FPS	Frames Per Second
GAN	General Adversarial Network
GPU	Graphics Processing Unit
HDR	High Dynamic Range
IAQA	Image Aesthetic Quality Assesment
IT	Information Technology
MAE	Mean Average Error
ML	Machine Learning
MSE	Mean Square Error
NN	Neural Networks
RAM	Random Access Memory

ReLU	Rectified Linear Unit
REST	REpresentational State Transfer
RGB	Red Green Blue
RL	Reinforcement Learning
SGD	Stochastic Gradient Descent
SS	Shutter Speed
UI	User Interface

Chapter 1

Introduction

Photography (Greek: φως+γραφή = "writing with light") is the art of "capturing a moment" by recording light, either electronically by means of an image sensor, or chemically by means of a light-sensitive material such as photographic film. Advances in electronics have enabled the creation of high quality image sensors even in smartphones during the last decade, while professional-level digital cameras have been made of reasonable cost due to the same reason. This, combined with the existence of social networks, has led to a major growth in digital photography, with thousands of photographs produced in a daily basis. Either through a camera or through a smartphone, behind each shot there is the personal expression of its photographer, usually imprinted as an aesthetic touch in the photograph.

1.1 Problem description

The huge amount of visual content that is being continuously produced has led to the need for accompanying metadata, to enable effective indexing, search and recommendation in the respective platforms. Such information is usually provided by the content creators as content-tags. However, this is rarely meaningful and, most importantly for our interest, non-related to the aesthetics of the photography itself. At the same time, one of the major challenges a hobbyist photographer faces, is the struggle to express herself through selecting among different combinations of styles and concepts. Therefore, a method that automatically recognizes tags related to the aesthetics of a photograph can be used to effectively index and retrieve photographs based on deeper, artistic attributes.

This project is an effort to identify some of the most basic rules of aesthetics in the art of photography and create a framework in which we are going to train an algorithm to be able to recognize those rules. The study follows all the necessary

steps, from data collection and annotation to model training and tuning. We are also expanding beyond that, with the creation of a basic infrastructure (REST API), aiming for a fully fledged experience.

1.2 Related Work

The task of automatically assessing the aesthetics of images has gained a great interest in the last decade, aiming to use computer vision and machine learning techniques, in order to simulate the human perception related to the aesthetics of an image. The domain [1] forms a novel interdisciplinary field of computer vision, computational aesthetics, psychology and neuroscience and has several applications on various fields raging from photography and cinema to fashion and graphic design. Appropriate lighting [2], contrast [3], color palette [4] and image composition [5] are just a few of the attributes adopted in related studies. In combination with the development and great image classification success of Deep Convolutional Neural Networks [6], [7], [8], as well as the extensive use of pretrained models through transfer learning [9], image aesthetics quality assessment has gained a tremendous growth. Due to the many aspects involved in characterizing the aesthetics of a photograph, most papers resort in experimenting with highly aesthetic structured datasets containing a rich variety of metadata [10] and usually one very well defined aesthetic [11]. Due to their huge size and the fact that a big part of metadata is user-generated, there is a high change that those datasets are biased. Most studies also tend to create frameworks of discrimination between high and low quality images [12], [13], [14], where ground truth is strongly associated with generative user-provided aesthetic quality scores, depending on their intuition and beliefs [13].

1.3 Motivation

As a hobbyist photographer, I know the struggles of trying to learn the basics and improve your photography, while trying to find your own unique style through the lens. As a potential AI researcher, I wanted to experiment with alternative, more contemporary and fun ways of learning those basics and tracking your style progress as it evolves. Through this study I am taking the first step of achieving this goal.

1.4 Contribution

In this thesis, we propose a more detailed analysis in the most important aspects that define the aesthetics of a photograph as a whole. Specifically, the main

contributions of this work are summarized as follow:

- A novel and meaningful taxonomy of image aesthetic tasks is defined: 5 image aesthetic tasks that are considered basic aesthetic dimensions when composing a photograph are used. Those tasks are carefully selected to express a broader range of images, meaning they represent aesthetic rules that can be found in almost any image. In addition, unlike most studies, this work does not aim to assess the quality of each image and distinguish between "good" and "bad" aesthetics: it aims to identify specific aesthetic aspects that can be found in almost every image.

- Based on the definition of these 5 tasks, we compiled and annotated a dataset of almost two thousand images, from the Unsplash Full Dataset. The annotation process for this dataset, was completely manual, meaning no scripts, programs or image metadata (eg. exif data, as in previous works) have been used to decide on the classes for each task, since such information has been proven to be quite noisy in many cases. Validation on the annotation process was also performed by experts and inter-annotator agreement metrics were also used for the same purpose. This detailed process has led to a high-quality dataset in terms of annotations on the selected image aesthetics tasks.

- We demonstrate the ability of Convolutional Neural Networks transferred from typical image classification tasks to distinguish between classes of the defined aesthetic tasks, and we provide access to the respective trained models in an open-source repository.

1.5 Thesis outline

The remainder of this thesis is organized as follows. In Chapter 2 we introduce the world of photography, referring to its history and its basic concepts. We also explore its relations with other forms of art and get to know about fundamental image aesthetics. Chapter 3 is dedicated to artificial intelligence. From machine to deep learning, computer vision and convolutional neural networks. Chapter 4 describes the different stages of dataset collection and annotation process, as well as a detailed description for each task. It also presents the inter-annotator agreement and the strategy we followed for validating the annotation process. Chapter 5 describes the classification process from start to finish. Project structure, logic, models and settings can be found there. On Chapter 6 we present the results of our final models through confusion matrices, classification reports and training/validation loss diagrams. We also present prediction examples on unseen data. Finally, on

Chapter 7 we discuss about the results and the ways we, as well as others can expand and improve the specific project.

Chapter 2

Photography

2.1 History of Photography

Photography has a long and very interesting history, which spans several centuries. It all starts in the 4th century BC from Greek philosopher and polymath Aristotle. With the help of Euclidean geometry, employed by the Greek mathematician Euclid, he described and made use of the principles of the camera obscura, in which an image is projected inverted (upside-down) and reversed (left to right) through a small hole. During the Industrial Revolution many researchers achieved the creation of the first photographic equipment, based on the already known principles of camera obscura. One of the very first cameras was presented in 1847.

At a span of twelve years (1827-1839), many researchers from various fields contributed on photography. Specifically, Louis Daguerre alongside Nicéphore Niépce, William Henry Fox Talbot, Hippolyte Bayard and John Herschel achieved the stabilization of an image into a photosensitive material. The climax of their study comes in 1939, when Louis Daguerre officially presents his method at the French Academy of Sciences, thus photography is officially born.

The term "Photography" (from the Greek words φως + γραφή = "writing with light") was first used in 1839 from Sir John Herschel, son of the distinguished British astronomer Sir William Herschel. In his extensive and very detailed study "Note on the art of photography, or the application of the chemical rays of light to the purposes of pictorial representation", Herschel inferences on the effect of light on photosensitive materials.

Photography owes its existence in the principles of theories of Optics and Chemistry. Without the understanding of the properties of light and the function of vision, but also without the knowledge of the mechanism of redox and photochemical reac-

tions the pioneers of the photographic process would not have achieved the desired result: an image of the world created by the sunlight alone.

Finally, another major factor for the evolution and progress of photography were the sociopolitical conditions and the economical situations of 18th and 19th centuries. he insatiable desire of the bourgeoisie for easy and quick profits combined with the process of change from an agrarian and handicraft economy to one dominated by industry and machine manufacturing, shaped the history of photography as we know it today.

Below we present some nodal chronological points in this history. We have divided those keypoints into three sub-lists, emphasizing the three distinct sciences which helped on the evolution of photography; Optics at first, Chemistry later and Informatics in the last years:

- 350 BC: Aristotle describes the way a camera obscura works.

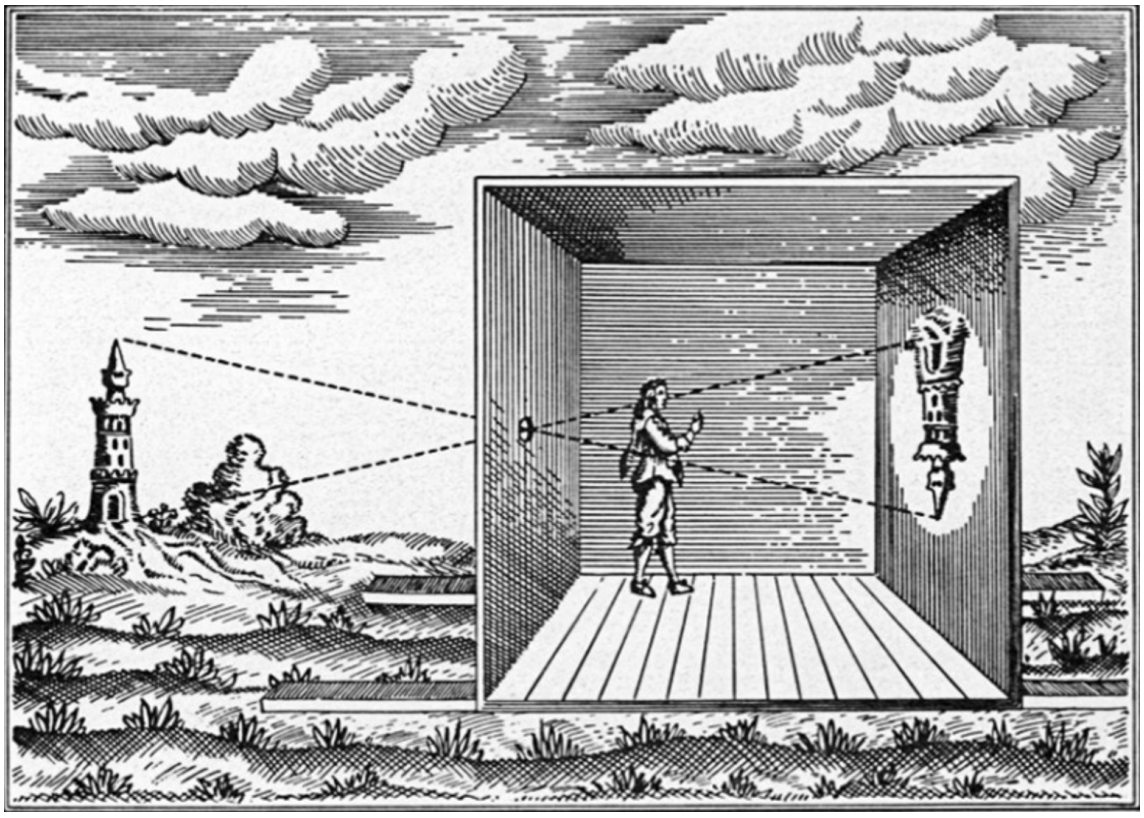


Figure 2.1: Camera Obscura, source

- 1530: Italian cleric and diplomat Daniele Barbaro mounts the first lens on a camera obscura for better results.
- 1550: Italian physician Gerolamo Cardano configures Daniele Barbaro's lens to also accept an apperture mechanism for better sharpness.

- 1604: Italian physicist Angelo Sala observes that some compounds of silver changed color under direct sunlight.
- 1605-1620: Austrian astronomer Johannes Kepler uses the first "portable" camera, which operates simillar to a camera obscura.
- 1676: German mathematician Johann Sturm creates the first camera with variable focal length and a mirror able to inverse the image.

At this point mathematics and optics have little to offer. The need for the contribution of chemistry is imperative. A contribution which will come 50 years later:

- 1725: German researcher Johann Heinrich Schulze achives to take a ephemeral photo by exposing silver salts in the sunlight.
- 1800: Distinguished British astronomer Sir William Herschel invents infrared radiation.
- 1821: Sir John Herschel uses sodium thiosulphate and achieves the stabilization of the idol.
- 1826: French inventor Nicéphore Niépce creates the first photograph in the history, which took approximately 8 hours (from sunrise to sunset). For this reason, his method was called "heliography". At the same year, French artist Louis Daguerre, later partner of Niépce, creates his own method known as daguerreotypes (a photograph taken by an early photographic process employing an iodine-sensitized silvered plate and mercury vapor).
- 1835: William Henry Fox Talbot creates the first negative on paper.
- 1837: Louis Daguerre uses sea salt as a way to stabilize his daguerreotypes.
- 1839: Invention of the photography is announced on France. At the same year, French researcher and pioneer Hippolyte Bayard creates the first positives on paper and presents the first ever photography exhibition with his personal portofolio.
- 1840: American Alexander S. Wolcott alongside John Johnson creates the world's first commercial photography portrait studio and patented the first U.S. camera, manifactured by Voigtländer.
- 1841: William Henry Fox Talbot achieves exposure times of 30 seconds.

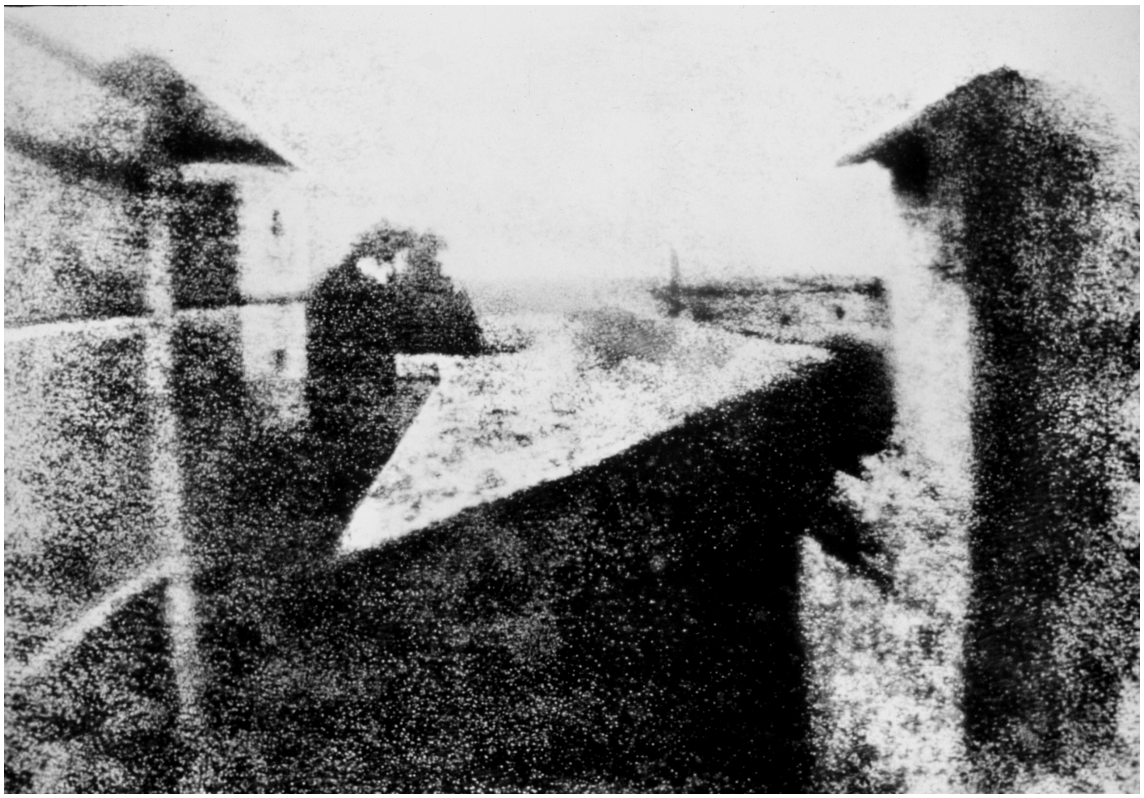


Figure 2.2: View from the Window at Le Gras, source

- 1843: Scottish painter, photographer and art activist, David Octavius Hill creates photographs of incredible beauty that are considered unsurpassed even today. It is the first time elements of aesthetics are distinguished in a personal work.

Figure 2.3: (a) Miss Matilda Rigby, (b) Group at Bonaly Tower source

- 1844: William Henry Fox Talbot publishes the first ever anthology of photographs.

- 1847: Calotype, a photographic process in which negatives were made using paper coated with silver iodide, gets perfected.
- 1853: First photography studio is created in Athens, Greece, by Filippos Margaris. The first calotypes by a Greek photographer are presented.
- 1855: War press photographers Roger Fenton and James Robertson take the first ever war photographs from the war in Crimea.
- 1856: French photographe Gaspard-Félix Tournachon, known by the pseudonym Nadar, takes the first ever series of aerial photographs from an aerostat.

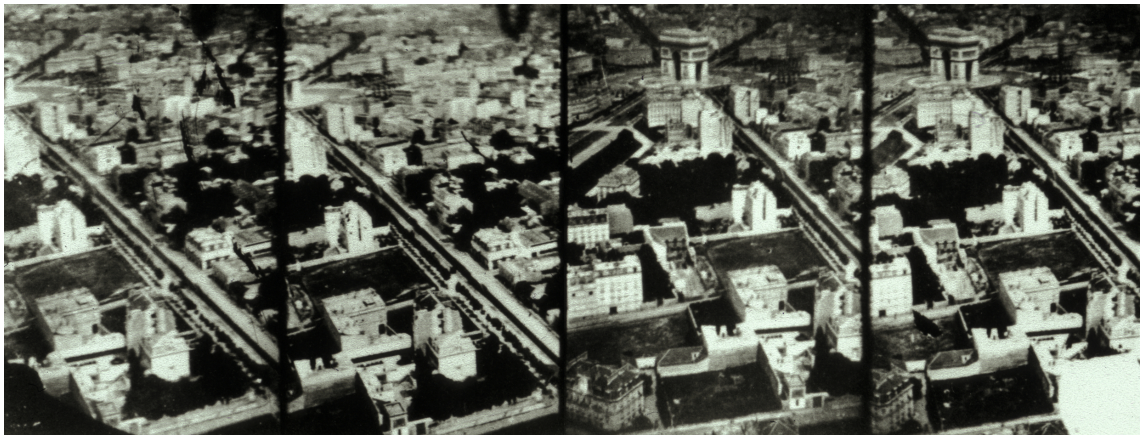


Figure 2.4: Nadar, Aerial view of Paris, 1868, source

- 1861: Scottish mathematician James Clerk Maxwell presents the electromagnetic theory. This consists the dawn of color photography.
- 1865: Sir Humphry Davy uses magnesium powder on the first artificial light source (the first ever flash).
- 1880: A photograph is been published for the first time in a newspaper, using the method of photozincography.
- 1888: American entrepreneur George Eastman launches a photographic film and presents the first affordable camera labeled Kodak in USA. In the same year, the first issue of the famous photo-centric National Geographic magazine is launched too.
- 1890: Photographic scientists Ferdinand Hurter and Vero Charles Driffield invent photometry.
- 1895: First ever cinematic projection in Paris.

- 1904: French manufacturer Auguste Lumière takes the first complete colorful photograph.
- 1908: With the help of a zoom lens, the first ever telephoto is taken.
- 1911: Leica presents its first test camera model.
- 1916: First ever colorful film from Agfa is launched, called Agfachrome
- 1926: The first Leica model publicly launches. Its small size combined with its many features were groundbreaking for the customers.
- 1928: Another famous brand launches its first camera, Rolleiflex.
- 1940: Photography officially enters the museum of modern art of New York.
- 1948: The first ever polaroid camera made its appearance. In the same year the most famous photo news agency, Magnum is established.

Subsequently, the advancement of technology in computer science had its own positive influence in the world of photography:

- 1959: The first photos of the earth from a satellite are published.

Figure 2.5: Explorer 6 satellite crude TV images of Earth from space, 1959, source

- 1990: The first ever digital camera is presented.

- 1997: The first photographs from the planet Mars are published.
- 2000: Mobile manufacturers add a camera on their cell phone products.

We notice that despite its young age, photography has a very long and interesting history. There are many more things that happened, too many to even mention in our list. Throughout those years, techniques and methods have evolved alongside photography. The basic ideas and principles of photography though have remained the same. On the next section we present those basics, from a camera's functions and various settings to the concepts behind taking a right photo.

2.2 Basics of Photography

In this section we present the basics of photography in a simplistic and easy to understand way. We skip the hard math and give optical examples for each concept we come across.

2.2.1 Camera's main function

The core of photography is about capturing the light. Nowadays, this is very easily achieved due to the massive technological progress of the last years. The process of capturing the light and taking a snapshot of the world is done through cameras. There are many camera types, depending on their format (slr, rangefinder, mirorrless, medium format, twins lens reflex, etc), their size, their use and their capabilities. Despite their variety, the function is the same:

a The light goes into the camera through the lens which has an aperture inside - it opens and lets light go inside.

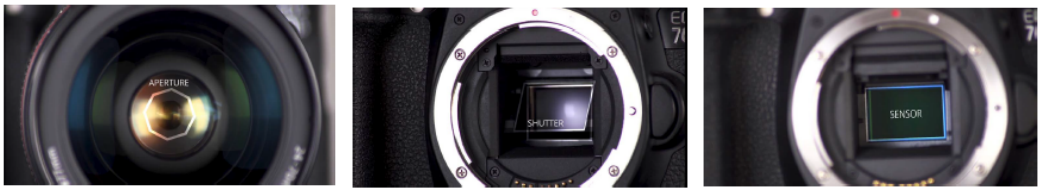
b It passes through the camera shutter

c It hits the camera's image sensor (analog or digital)

d The snapshot is captured and saved as an image

2.2.2 Exposure

Exposure is the amount of light which reaches the camera sensor or film. It is a crucial part of how bright or dark our pictures appear. Achieving a proper exposure requires a mixture of different settings in the corresponding amount of light our



(a) (b) (c)

Figure 2.6: (a) Aperture, (b) Shutter, (c) Sensor

camera sensor accepts each time. A properly exposed photograph looks natural. On the other hand, sometimes by accident or deliberately (for artistic reasons) our final photo seems different and unnatural. That happens when we either overexpose or underexpose in the respective lighting conditions. Overexposing usually makes our photographs too bright, meaning the whites prevail on the final image. Underexposing is the opposite, where the blacks prevail and our photograph gets too dark. An overexposed image has reduced details in the whites, while an underexposed image has reduced details in the blacks. On Figure 2.7 we present the three different types of exposure:

(a) (b) (c)

Figure 2.7: (a)Under Exposure, (b) Proper Exposure, (c) Over Exposure

2.2.3 Exposure Triangle

The Exposure Triangle or Photography Trinity is a common way of associating the three aspects that determine the exposure of a photograph: aperture, shutter speed, and ISO. One must balance all three of these to achieve a desired result, as an adjustment on one aspect usually requires adjustments of at least one of the other aspects.

- Aperture (also known as an "iris") is a mechanical component that provides a variable-sized aperture in the optical path that can be used to control the amount of light passing through the lens. The F-stop or F-number is the scale that represents the size of the aperture, with smaller f-numbers representing

larger apertures and larger f-numbers representing smaller apertures. Aperture also affects the depth of field, one of the basic aesthetic tools (more on that later). Finally, aperture alongside shutter speed are the two main means of controlling exposure.

- Shutter Speed (SS) or exposure time is the speed of the shutter. Shutter is a mechanism in our cameras with metal curtains that opens and closes as fast as we define them to (Figure 2.6 (b)). In other words, shutter speed is the time that our photosensitive material, film or digital sensor will be exposed to light. Shutter speed's scale is represented by fractions of a second, with values varying from 1/4000 (very fast shutter speed) of the second to 30 seconds (very slow shutter speed) and even more. The faster the shutter speed the less light reaches the sensor and vice-versa. Shutter speed is widely used as an aesthetic tool, even defining a unique type of photography, called "Long Exposure".
- Another mean of controlling the exposure, ISO is a unit of measurement of the sensitivity of a digital sensor or film to light. When the ISO value is low it has less sensitivity to light, while when the ISO value increases it has a higher sensitivity to light. In other words, increasing the ISO increases the exposure, while decreasing it decreases the exposure. ISO starts from the number 100 and doubles for each next value (200, 400, 800, ..). Depending on the camera, very high values of ISO will lead to digital noise/grain on the final image. A disadvantage which under certain circumstances can turn into a very strong aesthetic tool.

2.2.4 Basic Concepts, Settings and Techniques

In this section we are going to briefly mention some of the main concepts, settings and techniques in the world of photography. Basic concepts include:

- Sensor. An image sensor is an electronic device that converts an optical image into an electronic signal. Each camera has a different size and type of sensor. A bigger sensor captures more into the frame. Usually, the biggest the sensor the most expensive the camera is. Most famous ones are:
  - Full frame sensor. Full frame sensor is a large sensor equivalent to a 35mm film camera.
  - Cropped (APS-C) sensor. It consists the most used sensor. It is a bit smaller than a full frame in size, but it is also much cheaper. It usually has less megapixels than a full frame sensor.

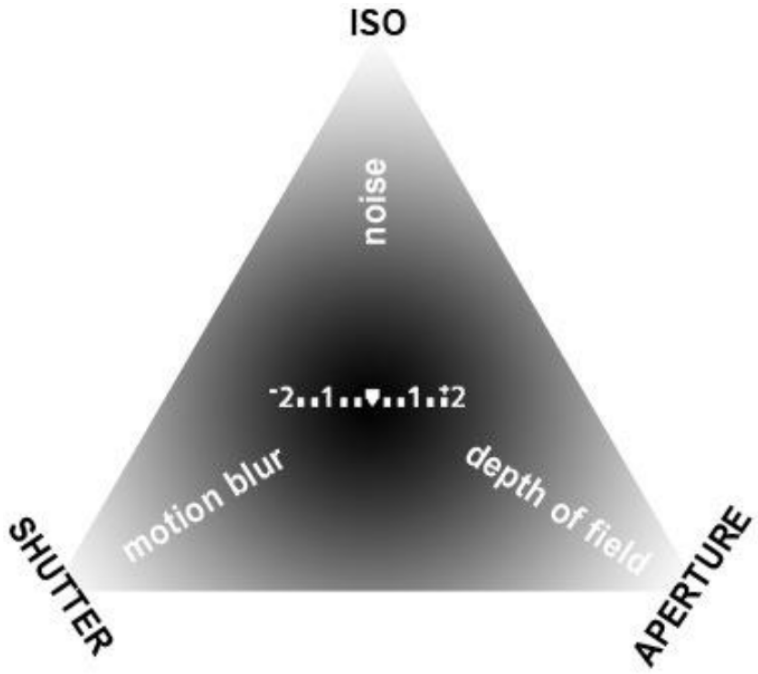


Figure 2.8: Exposure Triangle

  - Micro four thirds. An even smaller sensor, common in Olympus and Panasonic cameras.

- Resolution. Resolution refers to the number of pixels in an image. Resolution is sometimes identified by the width and height of the image as well as the total number of pixels in the image. A pixel is the smallest unit of information that makes up a picture. The more pixels a sensor has the more detailed is the image it produces.

- Frame rate. Frame rate (expressed in frames per second or FPS) is the frequency at which consecutive images called frames appear on a display.

- File types. There are 2 different categories for each image we produce. The uncompressed (RAW or NEF or CR2 depending the camera brand) and the compressed which comes in many extensions (.jpg, .jpeg, .png, etc). There are many notable differences between those 2 categories. The uncompressed image files are always of better quality, carrying the maximum amount of information. For those reasons they are also larger in size and consists perfect files for post-production. On the other hand, the compressed image files have lower quality, because they carry a smaller amount of information. That makes them smaller in size too, perfect for instant reviewing and sharing online.

- Stops. Stops is the maths behind photography. A stop is a doubling or halving

of the amount of light we let in when we take a photo. Increasing by 1 stop means we get x2 the light, while decreasing by 1 stop means we get half the light. Stops are used to measure all of the three exposure aspects we saw in section 2.2.3. It is important noting that ISO's predefined values have only full stops (from 100 to 200 is 1 full stop), while aperture and shutter speed predefined values gives us 1/3 of a stop each time.

- White balance. White balance defines what the color white looks like in specific lighting conditions, which also affects the hue of all the other colors. White balance is another creative tool, where the user can affect the final result of his/her photograph. White balanced is measured in Kelvin, a unit known for the measurement of temperature. Lower Kelvin values will make our images look cooler, where higher Kelvin values will make them look warmer. White balance is a very important concept, and knowing how and when to adjust it can add up in ones creativity.

- Lighting. Maybe the most important of all the photography related concepts. In the end it's all about lighting. Knowing how to read the lighting sources (natural or artificial) and how to position yourself, your camera, your objects, your background, is the quintessence of photography and the most hard to master thing. In short, when taking a photo, we need to consider the following things:

  - Position of the sun and any other light that we don't add on our frame intentionally. Is the sun behind a cloud or mountain or it stares directly on our frame? Note that many things may be considered as lightning sources, from the perspective that almost any object bounces back light.
  - Hour of the day. Sunrises and sunsets (golden hour) are considered the best hours to shoot, because the sunlight is not as harsh.
  - Position of our objects. Should the light fall on their back, front or side? Should we use a diffuser or not?
  - Flash numbers, positions and power.

Basic camera settings include:

- Light Meter. A light meter is a device used to measure the amount of light. A light meter is used to determine the proper exposure for a photograph.

- Metering Mode. The metering mode refers to the way in which a camera determines exposure. There are 3 different metering modes:

  - Matrix mode, where the camera reads exposure from all areas of the frame.
  - Spot metering, where the camera reads exposure from a single focus point.
  - Center-weighted metering, where the camera read exposure from the center of the frame and its surrounding area.

- Exposure compensation. Exposure compensation is used to alter exposure from the value selected by the camera, making photographs brighter or darker.

- Priority modes. There are many different priority modes, where the user can set a specific value on a specific setting of the exposure triangle (aperture, shutter speed) plus the ISO and the camera software will decide for the other value. For example there is the aperture priority mode, where the user can set the aperture and the ISO and the camera will decide for the value on the shutter speed, depending on the lighting and maintaining the values for the other two settings. There is also a manual mode where the user decided on all of the values.

- Scene modes. Each camera comes with predefined scene modes which have different settings regarding to what type of photography we are shooting (what our main subject is). Some of the most notable are portrait mode for portraits, macro modes for macro photography, landscape and even night modes. Depending on the scene mode the camera adjusts its settings differently.

- Histogram. The histogram is a graph that shows the exposure of every part of the frame (from darkest to lightest tones). Each color channel (R)(G)(B) has also its own histogram.

- Burst mode. Burst mode, is a shooting mode where several photographs are captured in quick succession by either pressing the shutter button or holding it down.

Some of the basic techniques include:

- Dynamic range. Dynamic range in photography describes the ratio between the maximum and minimum measurable light intensities (white and black, respectively).

- HDR photography. HDR photography is a photo technique combining multiple exposures into one image.

- Exposure bracketing. Exposure bracketing is taking multiple shots of the same image, in order to find the optimum single shot for the exposure.

Finally, we present some basic concepts about lenses. Apart of the camera's body, a camera's lens is also a very important aspect, which gives a plethora of additional settings to experiment with:

- Lens's type. There are 2 major lenses, the prime lenses and the zoom lenses. Their difference lies in their focal length. Prime lenses have a stable focal length, while zoom lenses tend to have a variable focal length.
- Focal length. The focal length of the lens is the distance between the lens and the image sensor when the subject is in focus, usually stated in millimetres (e.g., 50 mm). In the case of zoom lenses, both the minimum and maximum focal lengths are stated, for example 18–55 mm.
- Field of view. Field of view is the area of the inspection captured on the camera's sensor. In other words, it states how much we can see through the specific lens. The smaller the focal length, the wider the field of view.
- Depth of field. Depth of field (DoF) is the distance between the nearest and the farthest objects that are in acceptably sharp focus in an image. We state depth of field here, because it's a multi-factor concept, which between everything else is also affected by the lens.

2.3 Aesthetics

Aesthetics (or esthetics) is the section of philosophy that deals with the theory of beauty. Is the study of sensory and emotional values, sometimes called judgments of sentiment and taste. The word aesthetic is derived from the Greek "αισθάνομαι", meaning I perceive/feel/sense, related to "αίσθησση" (sensation).

Aesthetics is closely associated with the philosophy of art, covering both natural and artificial sources of aesthetic experience and judgment. The philosophy of art specifically studies how artists imagine, create, and perform works of art, as well as how people use, enjoy, and criticize art [1]. Beauty is very subjective, thus making aesthetics and art subjectives too. That's why people have specific tastes and like specific artworks that other people don't.

Art is a very powerful form of expression. It can affect our moods or even our beliefs. Photography is no exception. The nature of photography relies on its ability

[1] https://en.wikipedia.org/wiki/Aesthetics

to replicate a scene in a believable manner. It can also distort the reality giving the sense of illusion to the viewer. As a close relative to painting, photography has developed similar aesthetic tools and rules which can make a photograph much more appealing to the audience. In the following sections we present several of those aesthetic rules, some of which we used in this study. Those rules are not mandatory to follow, but are widely accepted as aesthetic touches that beautify our photographs.

Finally, we want to emphasize two very important things. Firstly, we believe that in art there is no parthenogenesis, meaning that specific aesthetic techniques have been curved and tested through the ages. And secondly, an artist must learn, train and test those methods, techniques and rules. Only then he will be able to merge them or even break them, creating his own unique style in the process.

2.4 Style

Style is another concept related to photography and arts in general. When we talk about style, we are referring to a photographer's personal aesthetic approach when shooting a photograph. This approach is what makes his/her photographs distinct and recognizable. A photographer's style is influenced by his/her personal experiences, character and interests. It as also influenced by the camera, lenses, and settings he/she chooses to work with. Style is often determined by a photographer finding and repeatedly using the tools that help them create the images they envision [15]. The whole idea behind this project is trying to identify the style of a photographer through his work.

2.5 Aesthetic tools and principles

In this section we present some of the most important aesthetic tools and principles. We discuss about the various patterns of visual elements a photographer may choose to create beautiful images.

2.5.1 Type

The most basic thing we need to decide when we photograph is what we want to photograph. Each person has his/her own tastes. Personally, I started by photographing landscapes and animals, because these were the subjects I was interested in. Along the way of my photographic journey I experimented with other different subjects, like humans or even stars. Starting with subjects you like and find interesting is crucial for investing in this art. There are many types of photography, as there

are many different objects you can photograph in the world. From landscape and animal, to urban and cityscape, it is important to understand the essentials for each genre. During our dataset creation and annotation, we came across many different perspectives and opinions about what each type express and deals with. As the list is quite big, we are going to mention some of the most controversial ones, so we have the opportunity to explain their differences:

2.5.1.1 Documentary vs Travel vs Street

These three types of photography are very much alike and most of the times puzzles people. There is one fundamental difference though, and that is the approach. To understand this approach we will give a short example of what a photographer of each genre does. A documentary photographer will focus on the subject itself as his/her goal is to give voice to the subject and its condition. A travel photographer will approach the subject seemingly as the documentary photographer does, but the point of reference is the cultural context that surrounds the subject and not the subject itself. The difference between the approach of a documentary and a travel photographer is pretty similar but still, there some significant differences. The street photographer, on the other hand, does something quite different. For the street photographer, the subject is just an occasion. What the street photographer wants is to present a very personal and often unique viewpoint upon social reality. To do that, the street photographer has to approach things differently, reforming their appearance and function through technique, and mostly through composition. Street photography, in other words, is a composition-driven type of representational imaging. The main goal of the street photographer is to offer a personal vision about urban life, focusing and uncovering contradictions or relations between unrelated elements in the urban environment, seen only by the photographer and afterward by the audience through the photographer's pictures. By putting them into some sort of visual relation or conflict, the street photographer gives new roles and meanings to things, telling a unique story about life in the city.

2.5.1.2 Pet vs Wildlife

It seems obvious, but there is a great controversy of what is considered a pet, and most of the times it's up to each person preferences and feelings. A snake may be considered as a pet for some and as a wild animal for others. On the other hand, we can't say that a cow falls under the category of "wild" life, can we? The problem lies in the fact that we are doing a scientific study and we need to well-define all

these little (in a very subjective domain) things, so we can proceed with structuring our problem. For that reason, for the purposes of our project, we are abiding by wikipedia's definition about what a pet means. We quote:

"A pet, or companion animal, is an animal kept primarily for a person's company or entertainment rather than as a working animal, livestock or a laboratory animal"

So yes, for the purposes of this study, cows, sheeps and horses, all fall under the wildlife category.

2.5.1.3 Architectural vs Cityscape

Both of those photography types consist of buildings inside a city. Their basic difference is that in architectural photography we are focusing specifically in a single architectural structure, while in cityscape photography we are usually getting a broader photograph of a city and its life. This distinction becomes sometimes hard, when you are focusing in a big architectural structure inside a city, so this structure is part of many other structures (buildings) inside your photograph. For example a bridge or a skyscraper inside a town. One possible way that helps with this distinction is where the main focus in the photograph is.

2.5.2 Color and Palette

> "One very important difference between color and monochromatic photography is this: in black and white you suggest; in color you state. Much can be implied by suggestion, but statement demands certainty... absolute certainty."
>
> – *Paul Outerbridge*

A debate of "ages" in the world of photography, with many fans on both sides. Colorful versus monochrome photographs. Each has its own aesthetic glamour and can be used in every aspect of photography. The monochrome format becomes more and more popular nowadays, especially with the reappearance of film cameras and the movements of vintage and retro increasing at a fast pace.

Colours have long been associated with certain emotions. Different colors evoke different emotions. For example red is passionate and aggressive while stating importance. Green is natural, stable and prosperous, while blue evokes serenity and trustworthiness. Colorful images tend to captivate viewers and have a notable im-

pact on their emotional world. Choosing your color "palette" is a very important task if you are going down the colorful path. Knowing which colors to use and how to combine them is essential for the story you are trying to tell and the emotions you want to evoke. For example, using complimentary or contrasting colours within the same frame allows the viewer's eye to take in the entire frame, while using only one or two colors bestows a sense of simplicity and minimalism on the image.

On the other side, monochrome photography eliminates the distraction of colour, which often makes us to observe the different elements that are present in the image, as well as their relations, more thoroughly. The monochrome format, adds an emotive feeling as the black-and-white photography is emotive by nature. Besides this emotional load, the monochrome format suggests a timeless perspective upon the subject while the color contemporizes it and brings it into the realm of Now.

A quick note here. Many consider white and black as colors. In fact, those colors does not exist in photography. White and black are not colors, they're shades. Shades that augment colors and can still evoke feelings. Black in particular is called "blank" or "vacuum" in photography, because pure black means the absence of any light. Unlike white and other hues, pure black can only exist in nature without any light at all.

(a) (b)

Figure 2.9: (a) Black and White, (b) Colorful

2.5.3 Depth of Field

Depth of Field or "DoF" in short is the distance between the nearest and the farthest objects that are in acceptably sharp focus in an image. We can characterize a photograph as having a Deep (more sharpness) or Shallow (less sharpness) depth of field. Depth of Field is affected by a mixture of different things. The most important are the aperture, the distance between the camera and the subject and the focal length of the lens. It is important to mention that in most cases, the area that appears acceptably sharp in the image corresponds to 1/3 in front of the subject in focus and 2/3 behind of the subject it focus. That is not always true though, as with different lenses, distances and settings this ratio changes.

Depth of Field is a very important aesthetic tool. Knowing when to blur or sharpen objects and backgrounds in your image may tremendously alter the story you are going to tell to your audience. Most of the times, depth of field goes hand to hand with the type/genre of photography you are shooting. For example, on an "en face" portrait of a person, we tend to reduce our distance from our subject, use prime lenses with big aperture values (<f1.8) and focus in the face (mainly in one of his/her eyes). This has as a result a very blur background with a very sharp face, emphasizing on the facial features of the person being photographed. On the other hand, when we want to take a picture of a beautiful landscape, we want our image to be as clear as possible, so every little element in it can be distinguished. For that reason, we tend to use small aperture values (>f/22.0), making our whole frame as sharp as possible.

Figure 2.10: Shallow

Figure 2.11: Deep

2.5.4 Composition

Photography is the art of taking a snapshot of the real world and presenting it inside a frame. This frame may consists of many different elements, like trees, buildings and people. The way you are going to present it, is another powerful aesthetics tool called composition or technique. By composing a photograph in a certain way, you can draw the viewer's eye to your main subject/s. Perspective plays a dominant factor during this composition. How you will position yourself and your camera, your subject/s and your background are all part of this great process.

As we have mentioned before, photography has very close relations with painting. Most (if not all) of the composition techniques are already known and used on the canvas. We are going to present some of the most famous techniques, the application of which can truly change the aesthetic of ones photograph.

2.5.4.1 Centering

Probably the most famous and widely used technique, is positioning your main subject in the center of the frame. Centering your main subject immediately draws the viewer's eye in the center of the image, and makes clear what you want him to see and study. Centering composition works very well with certain types of photography like portrait, where you usually want your subject to be the main element in your frame. Centering consists one of the easiest to perform techniques and applying it can easily "add up" in your image.

Figure 2.12: Center

2.5.4.2 Rule of Thirds

Another famous and very versatile composition technique, which encourages dynamism. It helps draw the viewer's eye into the image and places more emphasis on the subject. Ideally, the empty space that's left should be in the direction the subject is looking or heading into.

Rule of thirds divides the frame in nine equal rectangles, with the use of four imaginary lines. Their intersection creates 4 "power points", where one can place its subject, or part of it. There are many different uses of the rule of thirds. One can place his/her subject on top of one of the four imaginary lines. Another can place his/her subject in on of the power points mentioned above. Last but not least, one can also divide subjects, define thematics or even split the background by carefully placing them between the three and the remaining six of the total nine rectangles.

Figure 2.13: Rule of Thirds

2.5.4.3 Frame within Frame

Frame within frame is called the composition technique in which you place a secondary frame of any shape within your initial frame. It's a very powerful technique that adds context, intensity and depth in your image. It also helps to emphasize on your main subject. Frame within frame is not hard to apply, but there is sometimes a misconception about what a frame within a frame is. When you want to apply this technique in you images, is it substantial that your secondary frame has depth. For example a window which is blurred and you can't see inside it is not considered a frame. On the other hand, a mirror which reflects the figure of a person and his background (meaning that we have depth), is considered a frame within a frame.

Figure 2.14: Frame withing Frame

2.5.4.4 Leading Lines

Another very interesting technique is called leading lines. Maybe one of the best ways to engage your viewer's eye in a long visual journey, leading lines can often create a very strong and distinct visual experience, instantly catching the viewer's attention, even from a distance. Either by literal lines, like the lines in a road, or imaginary lines that are created by the way you framed your image, leading lines consist another powerful aesthetic technique you can use.

Figure 2.15: Leading Lines

2.5.4.5 Minimal

Minimal composition or minimalism photography in general is called the genre of photography (not to be confused with the types of photography on section 2.5.1) that is focused on simplicity. Minimalism photography attempts to explore how much information can be taken away from within a frame, before it loses its impact. It consist a very interesting approach and quite difficult to achieve. By eliminating or minimising as much as possible in the frame, you make the viewer concentrate more on the subject itself and less on the background. Minimalism also forces you to pay closer attention to color, contrast, shadows, textures, patterns and lines. Finally, due to this abstraction, minimal composition usually evokes a sentiment of calmness and serenity.

Figure 2.16: Minimal

Chapter 3

Artificial Intelligence

Artificial Intelligence (AI) is a the branch of computer science concerned with the design and the implementation of computer systems that are capable of performing tasks that typically require human intelligence or mimic elements of the human behavior. John McCarthy alongside Alan Turing, Marvin Minsky, Allen Newell, and Herbert A. Simon, are considered the fathers of artificial intelligence. Artificial intelligence has seen an exponential rise nowadays, mainly due to the rapid development of IT related technologies like computational resources (GPUs, RAMs), as well as the huge growth of data and information (big data) that circulate the web. Let's break down the two words and explain their definitions:

- **Artificial** is something that is made or produced by human beings rather than occurring naturally, especially as a copy of something natural.

- **Intelligence** is the ability to acquire and apply knowledge and skills, as well as *the ability to adapt to new circumstances*. This is very important, as even lots of animals don't have this ability (famous digger wasp "Sphex ichneumoneus" example [16]).

Human intelligence is a combination of many diverse abilities. Likewise, artificial intelligence is focused on a wide array of sectors, including learning, reasoning, problem solving and perception [17].

3.1 Machine Learning

Machine Learning (ML), or predictive analytics, or predictive modelling, constitutes a sub-field of artificial intelligence. There are many definitions of what machine learning is. Two of the most famous include:

- "Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959).
- "Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience (Tom Mitchell Hill, 1997)"

In other words, machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves [18]. Machine learning has many applications and is affecting our everyday life, sometimes even without knowing it. Some of the most important include:

- Automatic speech recognition
- Recommendation systems
- Virtual personal assistants
- Face recognition
- Self driving vehicles
- Searching (web, image)

3.1.1 Basic Steps

From unlocking our phones via facial recognition and getting a recommendation on a new movie to self driving cars, we notice the many different applications of machine learning. There is a very well defined framework describing the steps one should follow to provide a computer program or a system to learn and improve on experience. Those are:

a. Data collection. Through data collection, we search for potential open source datasets that can help on our specific task/s. We always need to take into account possible legal or privacy issues.

b. Data preparation. Data usually aren't in the desired form for our problem. We need to preprocess, curate and sometimes clean them. Most of the times, depending on the type of learning, we need to annotate them too. Annotation is the process where we (or a group of experts) decide on the values of the

classes we have define for our task/s. For example, a specific song (data) can be characterized as either of jazz (class1) or rock (class2) genre (task).

c. Data exploration. In this step, we try to find some kind of repetitive patterns, correlations or outliers (anomalies) inside our dataset. This can be done with simple plots and graphs. We also try to discover possible biases that may be hidden and eliminate them before proceeding with the training.

d. Model selection. During this step we create our model structure. We may need to revisit this step if our initial model doesn't seem to train well on our data. We can also choose to use an already well trained (on probably different task/s) model and retrain it on our task/s (transfer learning).

e. Model training, where we train the model on specific settings. Some of these settings include the epochs, the learning rate, the activation function and the batch size. More on those later.

f. Model evaluation, where we evaluate on how well our model was able to train on our data with those specific settings.

g. Model tuning, where we tune (change) the settings, aiming for better results.

h. Prediction/Inference, where we consider our model well-trained and proceed on the prediction of new, unseen data.

The list above constitute the basic steps of teaching a model to predict on one or more predefined specific tasks, via machine learning. We can also include one more optional, but sometimes necessary step, called data augmentation. Either through more data collection and preparation or through the use of specific algorithmic techniques, we can augment our initial dataset with the purpose of having more samples to train to.

There is also the question, why we even want a computer program to teach itself anything. This is the most fundamental question which usually fires up our desire and curiosity to start exploring ways of structuring a problem and trying to find a solution through machine learning. Why we decide to study the specific problem, how can we contribute in the existent bibliography and how will we structure our task/s are some important things we need to define in order to proceed with the basic steps mentioned above.

3.1.2 Types of Learning

There are four types of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement.

In supervised learning, the machine is taught by example. We provide our algorithm with a dataset that includes desired inputs (examples) and outputs (their labels). The algorithm accepts these inputs, tries to identify patterns and correlations, makes observations and concludes in some predictions (outputs). In short, given some data and the correct output, the algorithm tries to find a relationship between the data and the output. Supervised learning in its turn is divided into different sub-tasks. Those include:

- As per the number of outputs:
    a. Binary, where we have only 2 outputs, so our examples are assigned exactly one of two classes.
    b. Multi-class, where we have more than 2 outputs, so our examples are assigned exactly one of the many classes.
    c. Regression, where target output can take any value within a range.
- As per the use of the dataset:
    a. Model-based, where the algorithm uses the training data to create a model that has parameters learned from the training data.
    b. Instance-based, where the algorithm uses the entire dataset as the model.
- As per the number of classes an object (example) can belong to:
    a. Single-label, where the object can belong in only 1 class.
    b. Multi-label, where the object can belong in 2 or more classes.

In semi-supervised learning we use both labelled and unlabelled data. With labelled data we provide our algorithm the correct answers (output) on the input, just like in supervised learning. Unlabelled data lacks that information. With this combination, machine learning algorithms can learn to label unlabeled data.

Unsupervised learning is the exact opposite of supervised. In this type of learning, we do not provide our algorithm with the correct/desired outputs for each input. In other words, the algorithm tries to learn patters, correlations and relationships from the available unlabelled data. There are 2 main sub-types of unsupervised learning:

a. Clustering, where our algorithm tries to find similarities (or distances) on observations and group them together into distinct (generally non-overlapping) groups.

b. Dimensionality reduction, where our algorithm tries to summarise the data in a reduced number of dimensions (e.g. 2D). This usually happens by calculating new variables from the existing ones, which are reduced in number in comparison with their initial ones.

We can use them separately to solve different problems, or sometimes combine them.

Finally, Reinforcement Learning (RI) is concerned with learning how intelligent agents will take actions inside their environment that will maximize their reward. We define the rules and the algorithm explores different paths and possibilities. Through trial and error it teaches itself by learning from past experiences, while aiming to achieve the best possible result each time.

3.2 Deep Learning

Deep Learning (DL) is a sub-field of machine learning concerning techniques and methods inspired by the structure and function of the brain. More specifically, in deep learning we are using computing systems and algorithms known as Artificial Neural Networks (ANN), or in short Neural Networks (NN), that were inspired by the biological neural networks that constitute human and animal brains. Deep learning has seen a very rapid growth in the last years, mainly due to the huge amounts of data that is easily accessible for use, as well as the improvement of the various computational resources, mainly GPUs. On Figure 3.1, we present the basic structure of a simple neural network. The following network accepts 3 inputs, has 1 hidden layer and produces 2 outputs.

3.2.1 History of Deep Learning

Deep learning may have seen a great progression nowadays, although its roots are dated back 70 years ago, at 1950. Below we present the most important deep learning achievements chronologically [19], from 1950 onwards:

- 1950: English mathematician and computer scientist Alan Mathison Turing predicts that computers will achieve human-level intelligence by the year 2000.

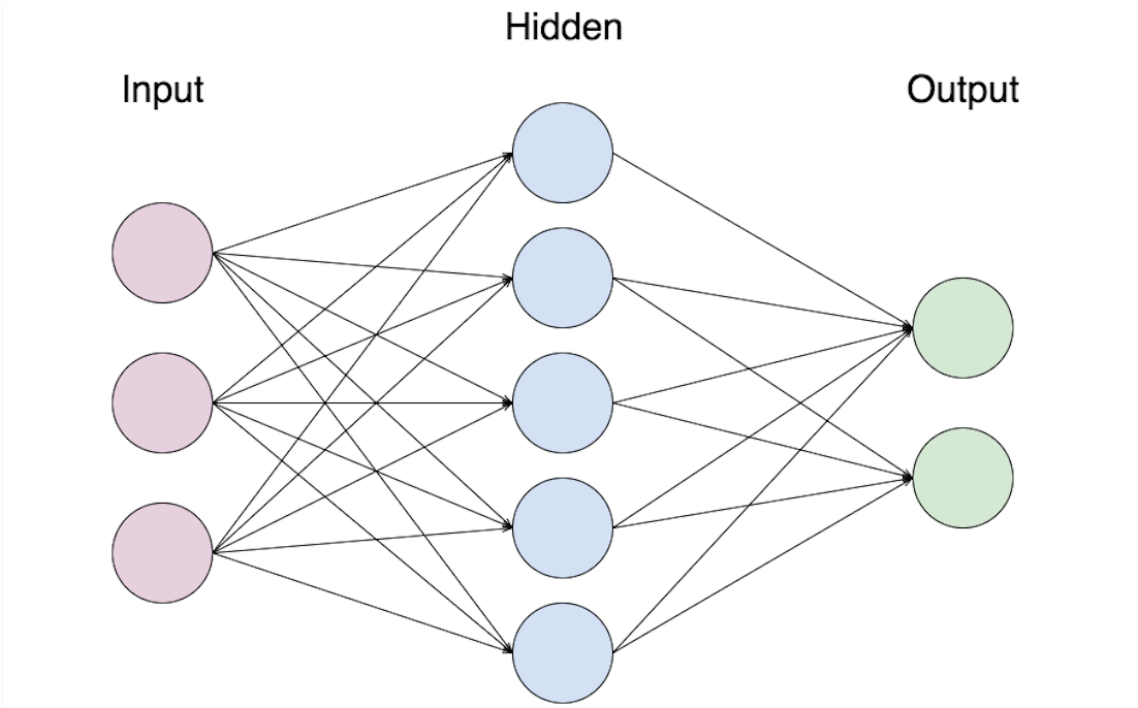


Figure 3.1: Basic Structure of a Neural Network, source

- 1965: Soviet mathematician Alexey Ivakhnenko develops the Group Method of Data Handling (GMDH), a method of inductive statistical learning.
- 1982: Neuroscientist John Hopfield's invents the first recurrent neural network knows as Hopfield network.
- 1985: Computational neuroscientist Terrence Joseph Sejnowski pioneers the application of learning algorithms to difficult problems in English speech (NETtalk).
- 1986: Computer scientist Geoffrey Hinton and several more researchers introduce a new technique called backpropagation for improved shape recognition and word prediction.
- 1989: Computer scientist Yann LeCun, the founding father of convolutional nets, invents a machine that can read handwritten digits.
- 1992: IBM researcher Gerald Tesauro creates an artificial neural network ("TD-Gammon") that can play backgammon.
- 1997: IBM's Deep Blue beats world No. 1 chess champion Garry Kasparov at his own game.
- 2006: "Deep Learning" finally gets its name, by computer scientist Geoffrey Hinton.

- 2011: IBM's Watson crushes its human opponents on the TV game show Jeopardy.
- 2012: AlexNet, a GPU implemented CNN model designed by Alex Krizhevsky, wins ImageNet Large Scale Visual Recognition Challenge with accuracy of 84% (up from 75% of previous models).
- 2014: Facebook develops DeepFace, which can identify faces in photos with human-like accuracy. On the same time, doctoral student Ian Goodfellow invents the General Adversarial Network (GAN).
- 2016: Google DeepMind's AlphaGo beats No. 1 Go world champion Lee Se-dol at his own game.
- 2017: Libratus, created by researchers at Carnegie Mellon University defeated four top players at No Limit Texas Hold 'em, after 20 days of play.
- 2019: Yoshua Bengio, Geoffrey Hinton, and Yann LeCun wins Turing Award for their immense contribution in DL and AI.

3.2.2 Basic Concepts

The world of deep learning introduces new concepts and functionalities. In this section, we introduce the most important ones. First of all we have the cornerstone of deep learning, the perceptron. Perceptron consists the structural building block of neural networks. In fact, a neural network consists of many perceptrons (neurons). A perceptron accepts features as inputs multiplied by their corresponding weights. The sum of this multiplication then passes through a node called activation function. Activation function, another basic concept in deep learning, is the tool which helps restrict the neuron output to a certain limit. This is important, because very high values on the output can cause a lot of computational problems. Another very important use of the activation function is that provides non-linearity on the network. This essentially means that our network can successfully approximate functions that do not follow linearity or it can successfully predict the class of a function that is divided by a decision boundary which is not linear. This is crucial, as there are very few phenomenons in the world that follows linearity. Some of the most used activation functions are:

- Sigmoid or Logistic activation function. It consists the most obvious alternative to the simple thresholding approach of the perceptron. Here, small

changes in the input results in small changes in the neuron's activation. The issue lies when there are very small or very large changes (close to 0 or 1). That leads on very small changes on weights (w) and bias (b), meaning no real change on the output. That is called neuron saturation. The main reason of using the sigmoid activation function is that it ranges between 0 and 1. Therefore, it is especially used for models where we have to predict the probability as an output, and mainly for binary classification tasks.

- Tahn or hyperbolic tangent activation function. Tahn is similar to the sigmoid activation function, but ranges between -1 and 1. Because it centers on zero (zero-centered), neuron saturation is less likely when used as input to next layers. It is usually prefered over the sigmoid. The advantage is that the negative inputs are mapped strongly negative and the zero inputs are mapped near zero.

- ReLU (Rectified Linear Unit) activation function. ReLU is the most used activation function to date. It is inspired by bio neurons, meaning it has only non-negative values. It ranges from zero to infinity. The issue with ReLU is that all the negative values become zero immediately which decreases the ability of the model to fit or train from the data properly. Despite this, ReLU learns must faster than sigmoid or tahn. It is most widely used on convolutional neural networks (CNNs).

- Leaky ReLU, an attempt to fix ReLU's issue by increasing its range below zero, usually around 0.01.

There are many other activation functions, each one with its own pros and cons. A honorable mention is the softmax activation function that consists a more generalized logistic activation function which is used for multiclass classification. On Figure 3.2 we present the graphs of the four activation functions:

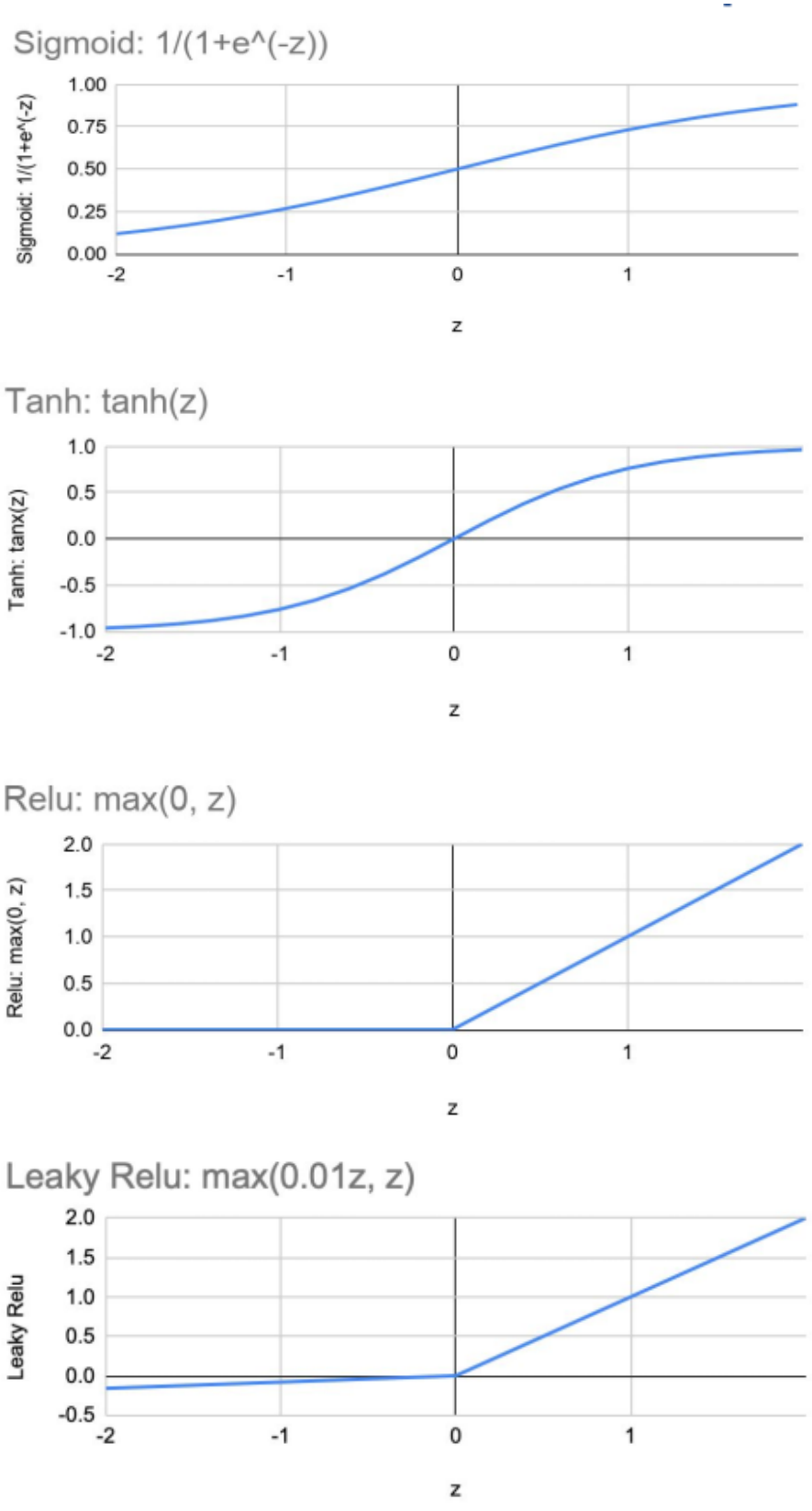


Figure 3.2: Most famous activation functions: Sigmoid, Tahn, ReLU, and Leaky ReLU

There are some more technical concepts related to the ways a neural network works, too many to extensively explore in this study. Indicatively we mention:

- Loss function. A loss or cost function is a method of evaluating how well your algorithm models your dataset. Specifically, it computes the distance between the current output of the algorithm and the expected output. There are many different loss fucntions, like Binary Cross Entropy loss (BCE) used on classification problems and Mean Average Error (MAE) or L1 loss, Mean Square Error (MSE) or L2 loss used on regression problems. Choosing the right for your task loss function is a very important step for the optimal training of your model.

- Backpropagation. Backpropagation or backward propagation of errors basically consists the essence of neural networks and the reason why they work so well. It is the method of fine-tuning the weights of a neural network based on the error rate obtained in the previous epoch (= 1 full forward propagation/iteration of the whole model). It helps calculate the gradient of a loss function with respect to all the weights in the network, aiming for the best possible generalization.

- Learning rate. Learning rate or step size is a configurable hyperparameter used in the training of neural networks that controls how much a model must change. In other words it represents the speed at which a machine learning model "learns". Choosing an optimal learning rate may prove challenging as there is a tradeoff between longer training times (small values) vs an unstable training process (larger values).

- Optimizer. An optimizer is a function or an algorithm that, in response to the output of the loss function, modifies the attributes of the neural network, such as weights and learning rate, in order to reduce the losses. There are many different optimizers like Gradient Descent (Simple, Stochastic, Mini-batch), AdaGrad, RMSProp, Adam and more. Each one with its own advantages and disadvantages.

3.3 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a deep learning algorithm specialized to work with image data. It accepts an image as an input and tries to find patterns and relations between its pixels. They are a powerful image processing tool able to

perform many different tasks, like facial recognition, document analysis and visual search. On Figure 3.3 we present a basic CNN architecture.

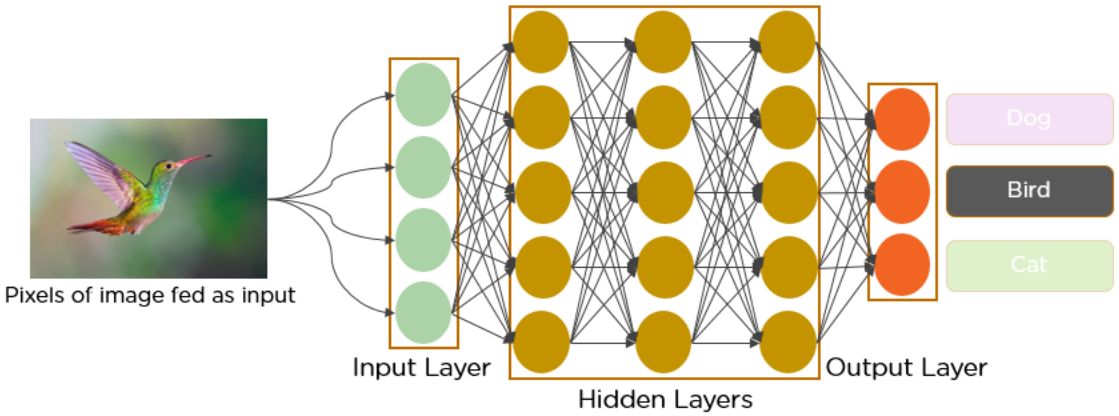


Figure 3.3: Convolutional Neural Network, source

The biggest advantage of a CNN over a traditional feed-forward neural net, that makes them work really well with images, is its ability to successfully capture the spatial and temporal dependencies in an image through the application of relevant filters. A CNN performs a better fitting on an image dataset, due to filters that reduce the size of each image (features) without losing the initial representation, patterns and relations between each image's pixels. On a traditional neural network, we would need to flatten the image, which means that we feed our model a vector with no real description about its features' relationships, plus this vector will be pretty big in size meaning extra computational complexity.

The word "convolutional" better describes what CNNs are all about. They perform convolutions (in fact cross-corellations, because the kernel does not rotate), between the image and a specific kernel (or mask). This kernel defines the filter we want to apply on the image. Different kernels serve different purposes. There are kernels that help us remove noise, locate features, find edges, reduce the size of the image or even transform it. A convolutional neural network uses a plethora of those filters in order to discover some meaningful information about the representation of an image. In Figure 3.4 we present a convolution example:

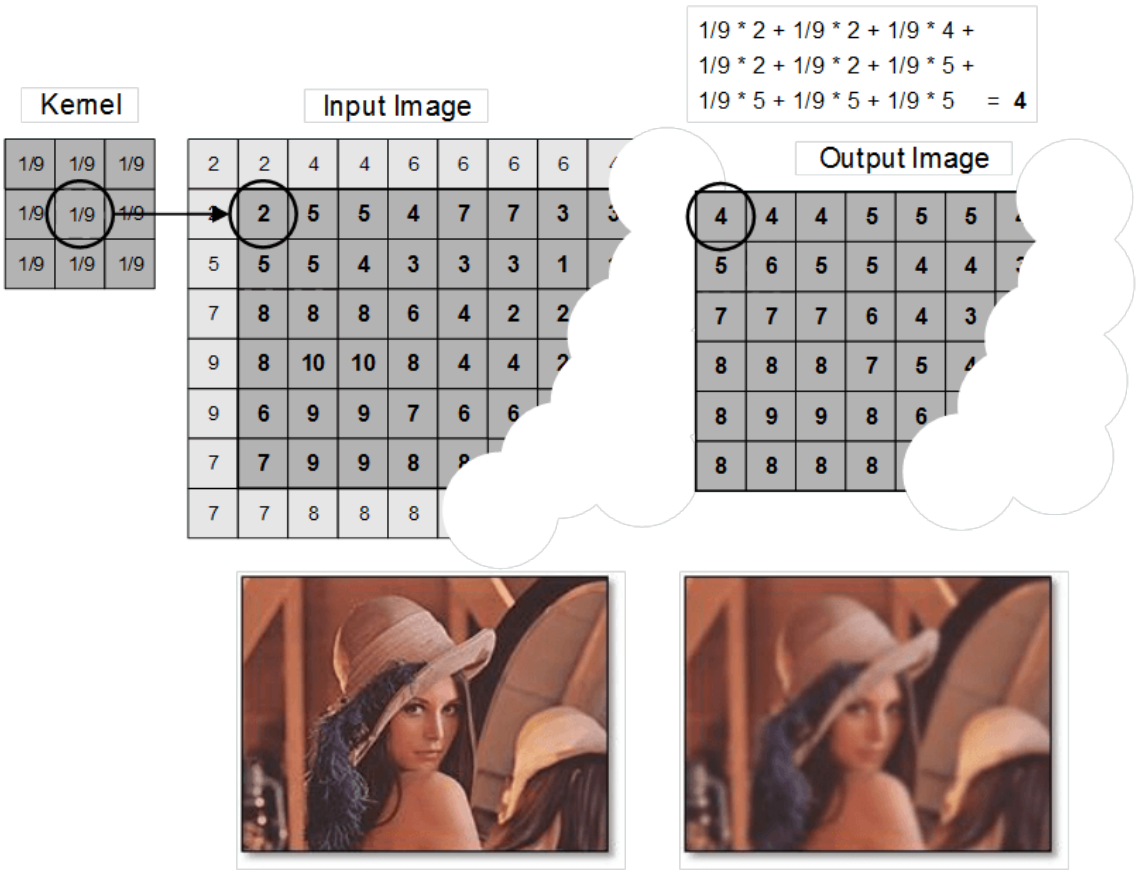


Figure 3.4: Cross corellation between an image and a kernel, source

There are many different filters, each with its own purpose during the convolution. Below we present the most important ones:

- Edge Detection. This filter tries to detect sharp edges/differences in the image color/brightness.
- Padding. When you need to apply filters of a fixed size, but you don't want to lose width and/or height dimensions in your image, that's when padding comes into play. Padding adds extra space (usually zeroes) around your initial image, so that you don't lose valuable information when applying other filters.
- Gaussian. This filter is usually used to remove random noise from the image. By doing that, it also blurs it too.

Chapter 4

Dataset

4.1 Images

Regarding the data collection, we requested and got granted access to the Unsplash Full Dataset. Unsplash Full Dataset consists of thousands of photographs from various amateur and professional photographers. These photographs are made available as URL links in a CSV form, along with other characteristics and metadata attributes for each photograph (such as photographer's name, submission date, exif info, etc). We followed an active and agile rationale for our data collection process, that took part in 3 stages:

- In the first stage, we extracted and annotated 200 images to use as an initial dataset to facilitate the task definition procedure and start training and evaluating baseline initial classifiers.
- A second, larger round of dataset gathering and annotation using 800 more images followed. This led to a more detailed and robust model evaluation using 1000 images in total.
- Based on the evaluation of the previous task, we then proceeded to a third, active selection of images focusing on underrepresented classes and on classes with insufficient performance metrics during the evaluation process. Towards this end, we used a combination of metadata-based search and manual curation on the retrieved images. This led to 832 more images. Some of the target classes we achieved to pinpoint and expand in size are "Black and White" for the "Color" task, "Astro", "Fashion", "Underwater" and "Wedding" for the "Type" task and many more. Our dataset size is now 1832 images in total.
- Since the three of the five classification tasks are multi-label, we needed to

follow a per-task data augmentation process to further handle the class imbalancing issues. For the particular case of the "Color" task, we performed downsampling in the most dominant class, leading to a decrease in the total number of samples. For all other cases, we performed typical image-based data augmentation using rotations and flips. For the "Depth of Field" task there was no need for augmentation, as it was very well balanced in the first place. Our aim was to have at least 30-50 images from each class. So, including our dataset augmentation and/or downsampling, our final, ready for training dataset becomes as follows:

Table 4.1: Number of samples per augmented task

Task	Dataset Size
Color	1384
Depth of Field	1832
Palette	4167
Composition	2365
Type	3492

On Figures 4.1-4.5, we present the Label Distribution for our binary and multi-label tasks. Blue indicates the initial, green the augmented and red (read from right to left) the downsampled size of a class. For example, on figure 4.1, for the "Black and White" class there are 139 images before and 684 images after augmentation. For the "Colorful" class, there are 1693 images before and 700 images after downsampling.

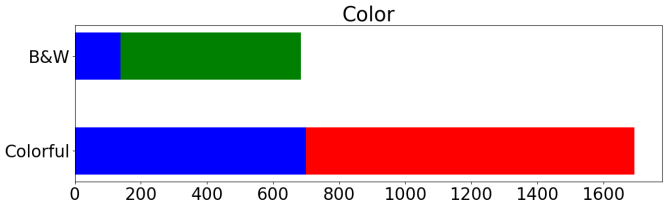


Figure 4.1: Color Label Distribution

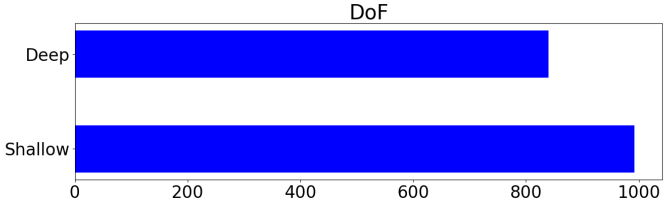


Figure 4.2: Depth of Field Label Distribution

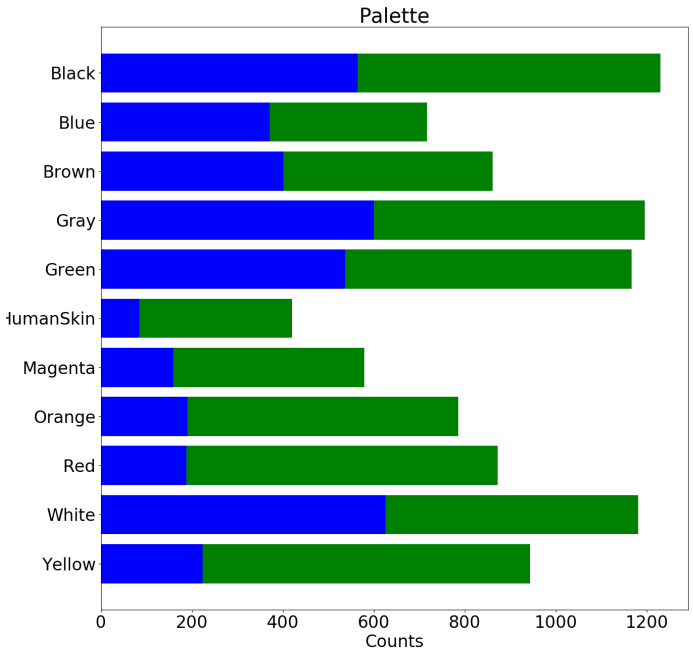


Figure 4.3: Palette Label Distribution

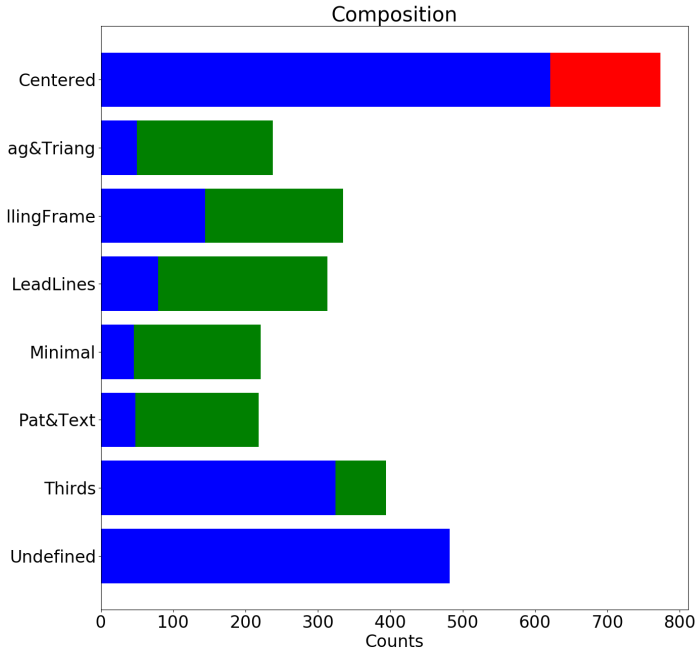


Figure 4.4: Composition Label Distribution

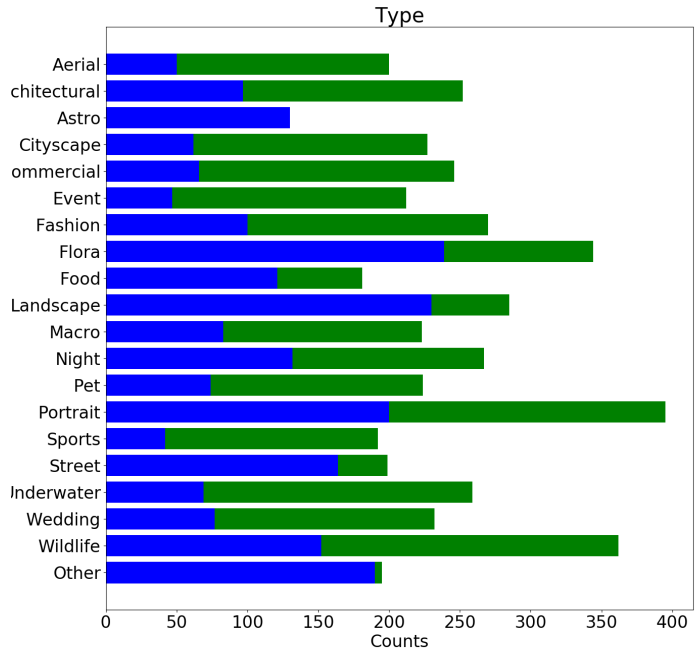


Figure 4.5: Type Label Distribution

4.2 Classification tasks

Style Analysis in a photograph can be a very vague term. What is style to begin with? What are we looking for in a photograph in order to decide its style?

Photography is an art closely related to painting. Photography (in its modern version) may be relatively new, but painting has a long history, with the oldest known paintings being approximately 40,000 years old. With this intro we want to state that in art (especially in its newer forms like photography) there is no parthenogenesis. Patterns, techniques, color matches, textures, framing and many more characteristics of photography have been carved through the years, giving us specific "rules" of aesthetics. Those rules are not mandatory to follow, thus the subjectivity of photography and any other form of art.

Style refers to a person's particular way of expressing themselves. This expression is affected by the person's continuous interaction with its environment. In photography, we can say that style refers to a photographer's favorite "rules" of aesthetics to follow. There are photographers who prefer to shoot in black and white only, and photographers who excel in a specific type of photography, like street photography.

That being said, we collected and narrowed down those rules of aesthetics and grouped them into 5 already well known categories/tasks. These include:

- Color. This simple binary task discriminates photographs to Colorful or Black and White. We consider as Black and White all the monochromatic (one tone and/or one color) images too (e.g. Sepia). The existence of Color in a photograph consists one of the fundamental aesthetics, especially nowadays where we have the option to choose between colorful and black and white shots. Colours have long been associated with certain emotions. For example, green is relaxing, while red is the color of danger and passion. Yellow brings joy and cheerfulness and white brings hope and simplicity. Colorful images tend to captivate viewers. On the other hand, black and white photography is a powerful aesthetics tool: black and white photography eliminates the distraction of colour. In this way the viewer explores more thoroughly the different elements that are present in the image, as well as their relations.

- Depth of Field. Depth of Field or DoF in short is the distance between the nearest and the farthest objects that are in acceptably sharp focus in an image. We can characterize a photograph as having a Deep (more sharpness) or Shallow (less sharpness) depth of field. DoF is one of the hardest classes to annotate with confidence, as it heavily depends on a number of camera settings, like aperture, focal length and distance from the subject. From the aesthetics perspective, DoF is a dominant factor that is related to the photographer's intent with regards to where she wants her viewers to focus on the photograph. Sharpening or blurring specific parts of an image, can change the whole story behind it. For example, one may have two photographs of the same landscape, one sharp (deep depth of field) where we can see every detail of this landscape and one blurred (shallow depth of field) where we focus on a specific small tree of the whole landscape. Despite those images capture the same landscape, the storytelling behind them is very different.

- Palette. With this (multi-label) task we refer to the dominant colors of the image. We cover only the basic colors which we apply on their derivatives too. For example, we consider Light Green, Dark Green, Forest Green, Lime, Emerald, etc as Green colors. The color palette is another powerful aesthetics tool and choosing the right colors for your frame is a very important task. For example, using complimentary or contrasting colours within the same frame allows the viewer's eye to take in the entire frame, while using only one or two colors bestows a sense of simplicity and minimalism on the image.

- Composition. This multi-label task is probably the most aesthetic-related. There are many different composition techniques, each of which attach a dif-

ferent aesthetic in the photograph. Placing the subject in the middle of the image (centered) is one of the most famous and easy to use techniques. Rule of thirds divides the frame in nine equal rectangles, with the use of four imaginary lines. Their intersection creates 4 "power points", where one can place the subject, or part of it. Rule of Thirds encourages dynamism, where the viewer sees a key element in the side, then takes a visual journey throughout the rest of the image. Another frequently used example of composition is the "Leading Lines" technique, which tends to create a very strong and distinct visual experience through the use of literal or imaginary lines inside the frame.

- Type. There are many types of photographs based on the content itself, such as: Portrait, Pet, Street and Astrophotography to name a few. The technical and creative skills required can be independent to photography genres: one can improve in one area and can learn valuable lessons and techniques that make her a better photographer across other types of photography as well. On the other hand, different types of photography sometimes require different handling. For example, sports photography requires quick reflexes and speed, while wildlife photography requires lots of patience. Type is also a multi-label task.

And their respective classes/labels:

Color	Depth of Field
Colorful	Deep
Black and White	Shallow

Table 4.2: Binary Tasks Classes

Type	Palette	Composition
Aerial	Black	Centered
Architectural	Blue	Diagonals and Triangles
Astro	Brown	Filling the Frame
Cityscape	Gray	Leading Lines
Commercial	Green	Minimal
Event	Human Skin	Patterns and Textures
Fashion	Magenta	Rule of Thirds
Flora	Orange	Undefined
Food	Red	
Landscape	White	
Macro	Yellow	
Night		
Pet		
Portrait		
Sports		
Street		
Underwater		
Wedding		
Wildlife		
Other		

Table 4.3: Multi-label Tasks Classes

A quick note here. Most of the photographs in the dataset (95%) have been taken by amateur photographers. That means that it is harder and sometimes impossible to locate a specific class from our predefined classes. That is why we also included labels like "Other" and "Undefined" for specific tasks.

4.3 Annotation Process, Aggregation and Inter-Annotator Agreement

The annotation process took place on Label Studio, a flexible, easy to install data annotation tool, which supports custom UI creation depending on the needs of each problem. The annotation was done in 3 stages, related to the dataset's collection stages mentioned above. On figures 4.6-4.8 we present some of the interfaces of Label Studio:

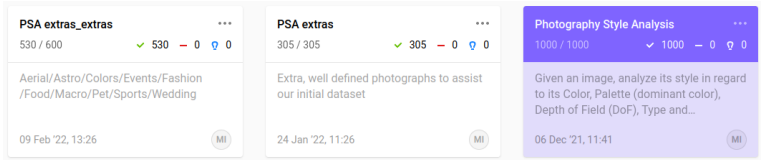


Figure 4.6: Initial dataset and extra images

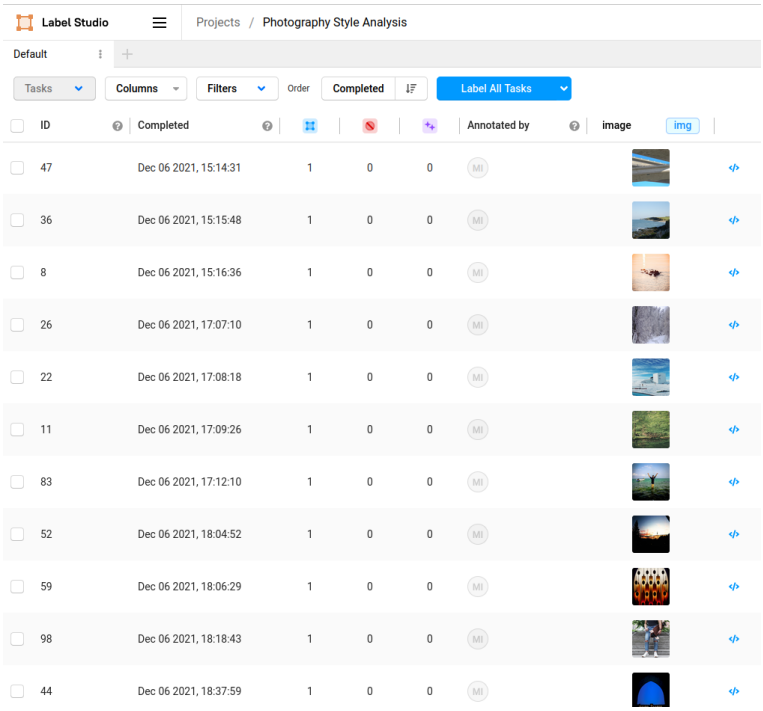


Figure 4.7: Annotation User Interface

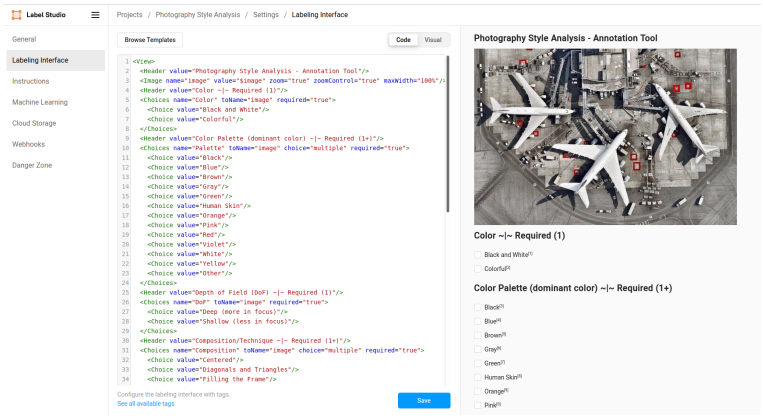


Figure 4.8: Customizable UI settings

The dataset has been annotated by the author who is a semi-professional photographer. However, since this is a subjective annotation task, as described above, there is a great need for validation of the whole annotation process. The free version of Label Studio does not support multiple annotators, so to deal with this we created an external questionnaire (Google Form) that consists of 20 carefully selected,

representative images. This questionnaire was shared to six different amateur and professional photographers and functioned as a secondary annotation tool. These six experts served as extra annotators for the purpose of calculating the inter-annotator agreement, that is the percentage of answers that agree on a label for each task. For our binary tasks, that was easy. We just had to calculate how many of the validators' answers were the same as our answer. We present the formula below. Note that we include ourselves and our answers in the whole process:

```
for each task:
    for each image:
        calculate all answers same to ours
        divide with no of validators
    add up all results
    divide with no of images
```

For the multi-label tasks the formula follows pretty much the same logic. Instead of calculating a single answer, we calculate multiple answers for each image. That means we also divide with the numbers of those answers before adding up to the results. Note that we don't take into account answers different than ours, except if those answers constitute the majority (more info on that later). We present the formula for the multi-label tasks:

```
for each task:
    for each image:
        for each answer:
            calculate all answers same to ours
            divide with no of validators
        add up all results
        divide with no of answers
    add up all results
    divide with no of images
```

On Table 4.4 we can see the results of the inter-annotator agreement (percentage):

In all cases, we notice that the inter-annotator agreement is high, considering the number of possible classes in each task. Composition proves to be the hardest classification task.

Another very important reason why we chose to have a standard 20 image dataset and not give each validator a random 20 image sample, was so as to be able to check if the majority of the validators were agreeing in an answer which was different from

Task	Percentage of Agreement
Color	0.91
Depth of Field	0.73
Palette	0.79
Composition	0.55
Type	0.72

Table 4.4: Inter-Annotator Agreement

ours. In the case of a tie, we kept our original answer as the final annotation for the specific image/task. If the majority was agreeing in an answer different than ours, then in the case of our binary tasks (Color, Depth of Field), we would change our original answer. In the case of our multi-label tasks (Palette, Composition, Technique), we would add the extra label in our original answer set. Some examples given below:

Author	Nick
colorful	black and white

final verdict: colorful

Table 4.5: Tie Case

Author	Nick	George
deep	shallow	shallow

final verdict: shallow

Table 4.6: Majority Disagrees - Binary Case

Author	Nick	George
red, pink	red, blue	blue, yellow

final verdict: red,pink,blue

Table 4.7: Majority Disagrees - Multi-Label Case

4.4 Potential biases

During the annotation process we noticed a lot of repetitive patterns between classes of the "Type" task. These repetitive patterns may lead in certain biases, especially because they appear on the same task. Below we mention some of these potential biases:

- The majority of "Wedding" images we annotated were expressing the western culture. We were able to find some few "Wedding" images from other cultures too.
- The majority of "Macro" images we annotated also belonged to the "Flora" class.
- The majority of "Underwater" images we annotated also belonged to the "Wildlife" class.
- The majority of "Astro" images are pictures of the moon.

One positive decision we took to avoid possible bias was to add the "Human Skin" as a color in the "Palette" task, so to be able to label all of the different skin tones we encountered into one class.

Chapter 5

Classification

Regarding the classification process, we used the same images as our core dataset for all tasks. Because we had 5 different tasks, 3 of which were multi label, it was not possible for our dataset to cover every single imbalancement of the classes. For that reason, we proceeded with different stategies in an effort to achieve the best possible balance each time. For the "Depth of Field" task we didn't need to do much, as it was pretty well balanced already. For the "Color" task, the "Colorful" images where outperforming our "Black and White" images 10 to 1. For this reason, and because we believed that the specific task would be the easiest one for our model to understand, we proceeded in a simultaneous augmentation on our "Black and White" images and at the same time downsampling for our "Colorful" images. Finally, for our three multi label tasks, we had several very underwhelming classes. For that reason, we proceeded with augmenting all of the those underwhelming classes. It is important to note that the whole augmentation process was very carefully designed, with augmented data existing only in the training set and random shuffling in the validation set. Each image has its own entry on a csv file, deriving from the Label Studio tool. Each entry consists of the image name/id plus the image's labels for each task. We modified the initial csv file, so to get a new csv for each task (5 in total), where each new csv has as its headers all of the possible classes for the specific task. Each image entry is now described by zeroes and ones, depending on their annotation value for each class (1 for True / 0 for False). Figures 5.1, 5.2 present the csv transformations:

Image	Id	Color	Palette	DoF	Composition	Type
photo-1624881254177-60066bed0802.jpg	962	Colorful	Yellow	Shallow	Centered	Landscape
photo-1473238588443-2db3ad66c1a4.jpg	135	Colorful	Blue	Deep	Undefined	Other
photo-1588972536098-7c1a904f5129.jpg	1322	Colorful	["Brown", "Green"]	Deep	Undefined	Aerial
photo-1614287151654-2c4b663670b0.jpg	1660	Colorful	["Gray", "Orange", "White"]	Deep	Centered	["Pet", "Portrait"]
photo-1622272542129-b076527e9c0c.jpg	918	Colorful	White	Deep	Rule of Thirds	Other
photo-1607849973876-c3def37d771b.jpg	1329	Colorful	["Blue", "Green"]	Deep	Centered	Aerial
photo-1591893280191-1ecbb3f67137.jpg	651	Colorful	["Gray", "Green", "White"]	Shallow	Undefined	Landscape
photo-1564981381462-e39978497c64.jpg	388	Colorful	Yellow	Shallow	Centered	Commercial
photo-1613332420690-97442f6953a8.jpg	836	Colorful	["Blue", "White"]	Deep	Undefined	Landscape
photo-1572526504444-f8f2e6cf76b6.jpg	466	Black and White	["Black", "Gray", "White"]	Deep	Leading Lines	Street
photo-1609090917257-d2e6a056fa45.jpg	780	Colorful	["Brown", "White"]	Shallow	Centered	Pet
photo-1589502998736-ae4ac06f2eef.jpg	601	Colorful	["Black", "Brown", "Gray"]	Shallow	Centered	Portrait
photo-1501703269644-eb49bfcc7393.jpg	211	Colorful	["Gray", "Red", "Yellow"]	Shallow	Rule of Thirds	Flora
photo-1620374643809-b69c702d0ed4.jpg	1570	Colorful	["Brown", "Green", "White"]	Deep	Centered	Food
photo-1566327011423-029eb055efb3.jpg	423	Colorful	["Blue", "Gray", "Human Skin"]	Shallow	Centered	Portrait
photo-1580805902282-ae40f58585fc.jpg	503	Colorful	Brown	Deep	Rule of Thirds	Landscape
photo-1461258098785-4570b760f16c.jpg	115	Colorful	Blue	Deep	Rule of Thirds	Landscape
photo-1598343079259-25a72ccccb9e.jpg	723	Colorful	["Brown", "White"]	Deep	Frame within Frame	Other
photo-1563483628704-ba28baa9df2b.jpg	364	Colorful	Brown	Shallow	["Centered", "Frame within Frame"]	Wildlife
photo-1612795298862-bd268ece1c7c.jpg	828	Black and White	["Black", "Gray", "White"]	Deep	Rule of Thirds	Wildlife
photo-1500852852497-118bcf47addd.jpg	208	Colorful	Blue	Shallow	Undefined	Landscape
photo-1519992950826-eb8791ac367d.jpg	1065	Black and White	Black	Deep	["Centered", "Minimal"]	Astro
photo-1522570980439-883c66249569.jpg	314	Colorful	["Green", "Orange", "Pink"]	Shallow	Centered	Flora
photo-1591671747743-84e2af6a7fbc.jpg	1490	Colorful	["Blue", "Green", "White", "Yellow"]	Shallow	Centered	["Fashion", "Portrait"]
photo-1610478920395-df32e4af2afb.jpg	790	Colorful	White	Shallow	Undefined	Street
photo-1530777780045-c11965d8f4c7.jpg	335	Colorful	["Black", "Blue", "Human Skin"]	Shallow	Centered	Portrait
photo-1616717392423-7468d9b91a99.jpg	1896	Colorful	["Black", "Gray", "Yellow"]	Deep	Rule of Thirds	["Commercial", "Food"]
photo-1540283551754-87a8aa9f9b61.jpg	1832	Colorful	["Brown", "Green", "Orange"]	Shallow	["Minimal", "Rule of Thirds"]	["Macro", "Wildlife"]
photo-1580506665033-b0ba8c947cd6.jpg	498	Black and White	["Black", "White"]	Deep	Undefined	["Cityscape", "Night"]
photo-1596688051781-67e73abfb77f.jpg	684	Colorful	["Blue", "Green"]	Deep	Rule of Thirds	Landscape
photo-1583525008095-89a9917c1d02.jpg	536	Colorful	Black	Shallow	Centered	Commercial
photo-1540108816274-ecb656218800d.jpg	1740	Colorful	["Black", "Gray", "White"]	Deep	Leading Lines	["Night", "Street"]
photo-1565837257446-1c73ba1a8348.jpg	414	Colorful	["Black", "Blue", "Brown"]	Shallow	Rule of Thirds	Landscape
photo-1560721397-4eb6096da3d0.jpg	66	Colorful	Green	Shallow	Centered	Sports
photo-1519713776281-1605db57fca1.jpg	303	Colorful	Other	Deep	Centered	Commercial
photo-1557745271-23ad4dd34d53.jpg	39	Colorful	Other	Deep	Undefined	Street
photo-1572322633904-ebbd90cb2f9b.jpg	1013	Colorful	White	Shallow	Centered	Wedding
photo-1560171128-f37b5cf8900d.jpg	56	Black and White	["Black", "White"]	Deep	Undefined	Architectural
photo-1620095286514-83898b1b73aa.jpg	1128	Colorful	Black	Deep	Centered	Astro
photo-1591078855982-0fc3afdbae24.jpg	1024	Colorful	White	Shallow	Rule of Thirds	Wedding
photo-1626817774457-7abc3194e32f.jpg	1466	Colorful	Orange	Deep	["Minimal", "Patterns and Textures"]	Macro
photo-1562525510-791587b11a1f.jpg	85	Colorful	["Black", "Red"]	Shallow	Undefined	Other
photo-1461601511666-d70469638a16.jpg	116	Colorful	["Blue", "Brown"]	Shallow	Undefined	Landscape
photo-1553244269-09ca8bff0cbf.jpg	1199	Colorful	Blue	Deep	Centered	["Underwater", "Wildlife"]

Figure 5.1: Initial csv from Label-Studio

Image	Composition	Centered	Undefined	Rule of Thirds	Leading Lines	Minimal	Patterns and Textures	Filling the Frame	Diagonals and Triangles
photo-1624881254177-60066bed0802.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1473238588443-2db3ad66c1a4.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1588972536098-7c1a904f5129.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1614287151654-2c4b663670b0.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1622272542129-b076527e9c0c.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1607849973876-c3def37d771b.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1591893280191-1ecbb3f67137.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1564981381462-e39978497c64.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1613332420690-97442f6953a8.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1572526504444-f8f2e6cf76b6.jpg	['Leading Lines']	0	0	0	1	0	0	0	0
photo-1609090917257-d2e6a056fa45.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1589502998736-ae4ac06f2eef.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1501703269644-eb49bfcc7393.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1620374643809-b69c702d0ed4.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1566327011423-029eb055efb3.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1580805902282-ae40f58585fc.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1461258098785-4570b760f16c.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1598343079259-25a72ccccb9e.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1563483628704-ba28baa9df2b.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1612795298862-bd268ece1c7c.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1500852852497-118bcf47addd.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1519992950826-eb8791ac367d.jpg	['Centered', 'Minimal']	1	0	0	0	1	0	0	0
photo-1522570980439-883c66249569.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1591671747743-84e2af6a7fbc.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1610478920395-df32e4af2afb.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1530777780045-c11965d8f4c7.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1616717392423-7468d9b91a99.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1540283551754-87a8aa9f9b61.jpg	['Minimal', 'Rule of Thirds']	0	0	1	0	1	0	0	0
photo-1580506665033-b0ba8c947cd6.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1596688051781-67e73abfb77f.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1583525008095-89a9917c1d02.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1540108816274-ecb656218800d.jpg	['Leading Lines']	0	0	0	1	0	0	0	0
photo-1565837257446-1c73ba1a8348.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1560721397-4eb6096da3d0.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1519713776281-1605db57fca1.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1557745271-23ad4dd34d53.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1572322633904-ebbd90cb2f9b.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1560171128-f37b5cf8900d.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1620095286514-83898b1b73aa.jpg	['Centered']	1	0	0	0	0	0	0	0
photo-1591078855982-0fc3afdbae24.jpg	['Rule of Thirds']	0	0	1	0	0	0	0	0
photo-1626817774457-7abc3194e32f.jpg	['Minimal', 'Patterns and Textures']	0	0	0	0	1	1	0	0
photo-1562525510-791587b11a1f.jpg	['Undefined']	0	1	0	0	0	0	0	0
photo-1461601511666-d70469638a16.jpg	['Undefined']	0	1	0	0	0	0	0	0

Figure 5.2: Reformed csv for "Composition" task

Throughout the life cycle of this study, we have evaluated several custom CNN architectures on some tasks, trained from scratch on some data, but in all cases the results were rather low. We then tried two pretrained CNNs, namely the DenseNet121 and ResNet50 models [20], [21]. We chose to use ResNet50 mainly for computational complexity reasons. Specifically, DenseNet121 is a deeper network (121 vs 50 layers) meaning it is more complex leading to bigger training times. It also means that for our relatively small dataset, it was more prone to overfitting [22]. ResNet50 has been trained on millions of images from the ImageNet database, into 1000 different

object categories. As a result, the network has learned rich feature representations for a wide range of images. We wanted to take advantage of this, especially while training on the "Type" task. Classes like flora (trees, flowers), pet (dog, cat) and so on, are some of the "objects" ResNet50 has already been trained on. Figure 5.3 presents the ResNet50 architecture:

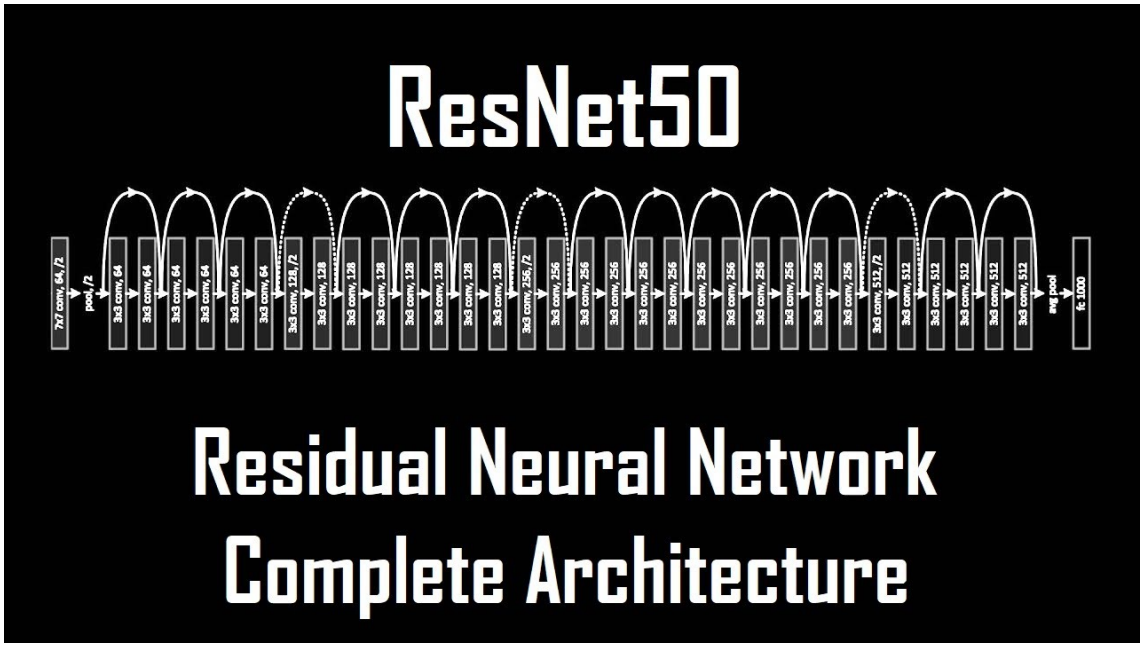


Figure 5.3: ResNet50 Architecture
source

ResNet50 consists of 48 convolutional and 2 pooling layers. Having no fully connected layers in between gave us the flexibility to use bigger input sizes, without worrying about retraining those layers. Specifically, the images we fed our model were of size 400x400. To achieve transfer learning to our five individual tasks, we only replaced the last fully connected layer depending on our tasks's needs.

The project was implemented in Python 3.8. Some of its main frameworks and libraries include the PyTorch deep learning framework, computer vision and image processing libraries like OpenCV and Matplotlib, and many more. On Figure 5.4 we present the whole project structure.

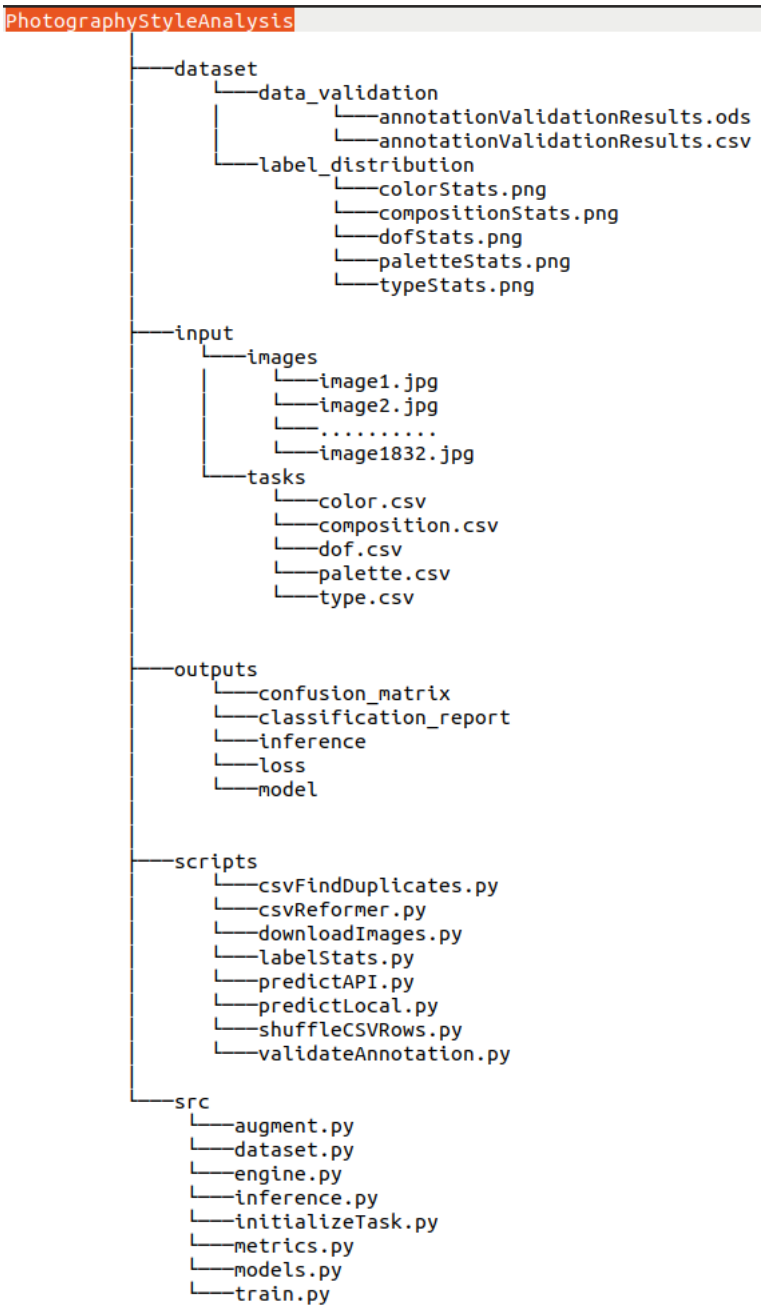


Figure 5.4: Project Structure

Last step before starting training our ResNet50 neural network on our various tasks was to decide the values of crucial variable settings needed for our algorithm to run. We experimented on a number of different epochs, learning rates and batch sizes in order to optimize our results. During the dataset split we also keep 10 images as a small Test set for inference. To achieve transfer learning to our five individual tasks, we took advantage of only the last layer, which we replaced to retrain on our tasks. On Table 5.1 we present the CNN training parameters that have been adopted in our experimental setup. These settings were the same for all

the tasks. To approach the multi-label tasks, binary cross entropy (BCE) has been used as a loss criterion and a Sigmoid activation function has been applied at the output. Note that we run those settings for each task separately. We could try and train successively one task after another (each time with the updated model), but this way there was a high probability of biasing our whole model. That is, because some tasks are closely related to one another. For example, a "Landscape" type photograph is closely related to "Green" and "Blue" palette colors, to "Deep" depth of field and to "Colorful" color.

Setting	Value
Dataset Split	85/15
Learning Rate	0.0001
Epochs	30
Batch Size	32
Optimizer	Adam
Loss Criterion	BCE Loss

Table 5.1: Transfer Learning Settings

At this point, it is important to explain how we approached our multi-label problems. We use as criterion the binary cross entropy loss (BCE), so to generate all outputs as a form of probability. Then we apply the Sigmoid activation function to get all outputs between 0 and 1. So far, for each class on a specific task, we got a prediction score. How are we going to make it a multi-label classification? The method we used is pretty simple and straightforward and works really well. After we get all the sigmoid outputs, we are going to choose the top 'X' scores. We can use the indices of those scores and map them to the specific class of the specific csv of the corresponding task.

Chapter 6

Experimental Results

6.1 Performance metrics

We used plenty of metrics to measure the performance of our model. The most important were validation loss, classification reports and confusion matrices. Note that we didn't calculate the accuracy of our model, because accuracy is not optimal for imbalanced data. When we use accuracy, we assign equal cost to false positives and false negatives, so accuracy becomes a poor measure of evaluation for our classification model. We also report the baseline F1 value for each task, as the performance of the random class estimation. Per-class and average F1 values are presented in Tables 6.1, 6.2, 6.3, 6.4, 6.5. For the binary classification tasks, macro F1 score is already 40-70% above a baseline, with respectively 0.85 score for Color and 0.71 score for Depth of Field tasks. Regarding the multi-label classification tasks, we implemented and tested on a random selection classifier, with resulting scores ranging between 0.15 - 0.20 for the F1 metric. Reviewing our model's results on the multi label tasks 6.3, 6.4, 6.5, the relative improvement is 70% up to 300% related to the random selection (baseline).

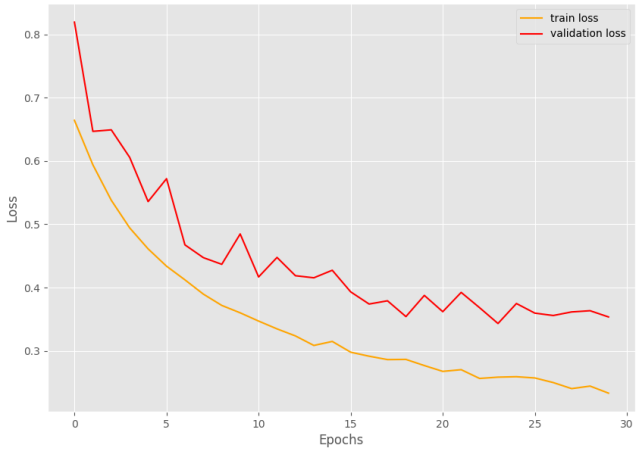


Figure 6.1: Color Validation Loss

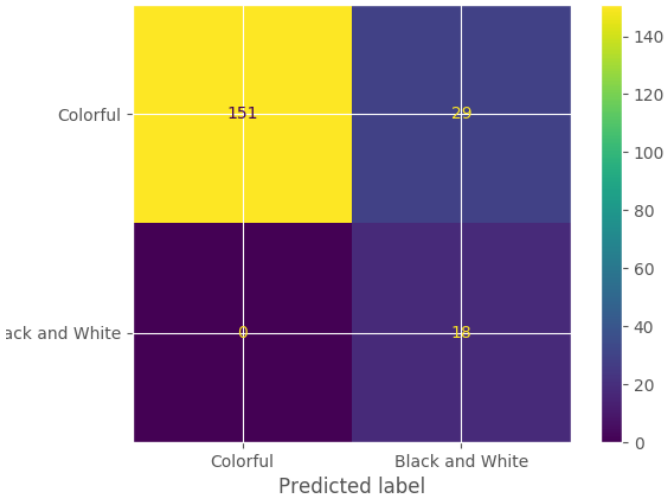


Figure 6.2: Color Confusion Matrix

Color Classification Report			
Color	Precision	Recall	F1-Score
Black and White	0.38	1.00	0.55
Colorful	1.00	0.84	0.91
macro avg	0.85	0.85	0.85 (0.4)

Table 6.1: Color Classification Report

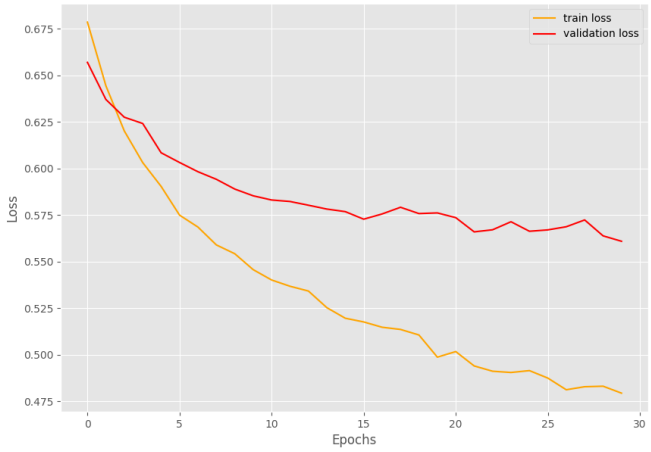


Figure 6.3: Depth of Field Validation Loss

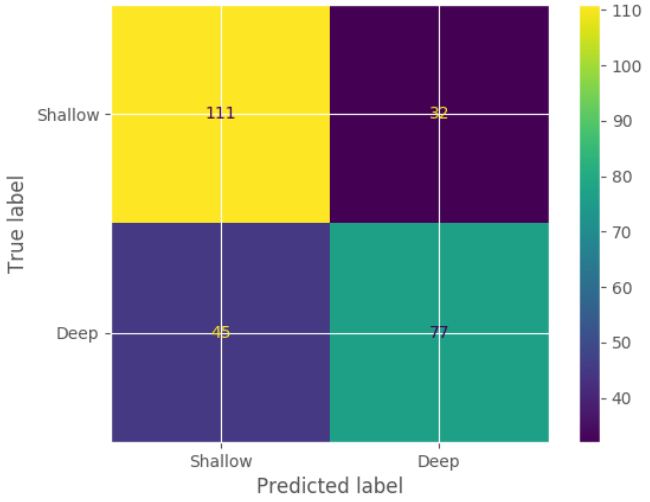


Figure 6.4: Depth of Field Confusion Matrix

Depth of Field Classification Report			
Depth of Field	Precision	Recall	F1-Score
Deep	0.71	0.63	0.67
Shallow	0.71	0.78	0.74
macro avg	0.71	0.71	0.71 (0.46)

Table 6.2: Depth of Field Classification Report

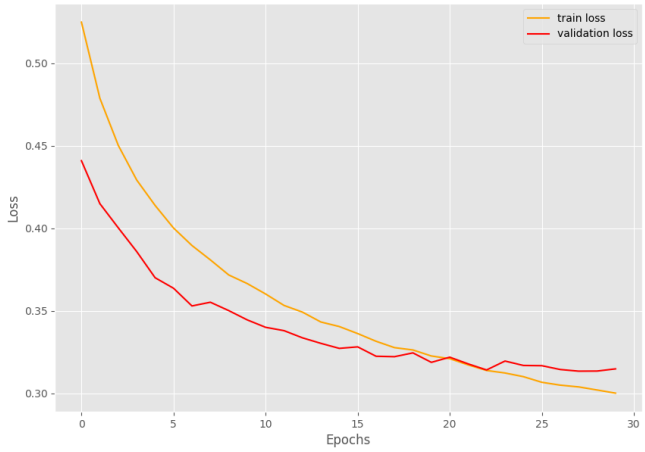


Figure 6.5: Palette Validation Loss

Figure 6.6: Palette Confusion Matrices

Palette Classification Report			
Palette	Precision	Recall	F1-Score
Black	0.89	0.57	0.69
Blue	0.73	0.44	0.55
Brown	0.71	0.26	0.38
Gray	0.70	0.41	0.52
Green	0.91	0.49	0.64
Human Skin	0.25	0.38	0.30
Magenta	0.88	0.39	0.54
Orange	0.50	0.16	0.24
Red	0.47	0.23	0.30
White	0.86	0.25	0.38
Yellow	0.47	0.18	0.26
macro avg	0.67	0.34	0.44 (0.19)

Table 6.3: Palette Classification Report

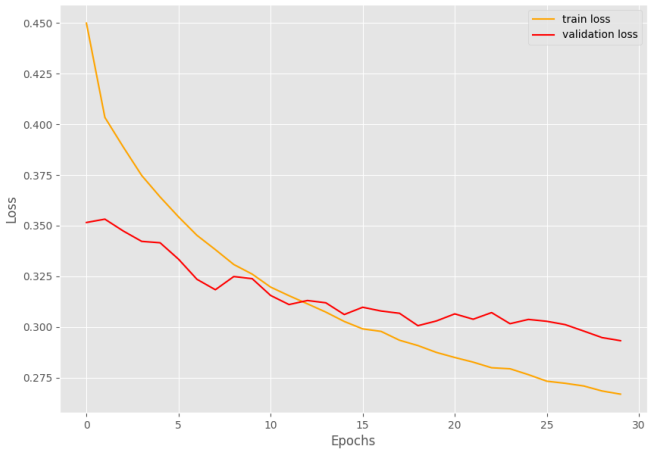


Figure 6.7: Composition Validation Loss

Figure 6.8: Composition Confusion Matrices

Composition Classification Report			
Composition	Precision	Recall	F1-Score
Centered	0.42	0.42	0.42
Diagonals and Triangles	0.10	0.20	0.13
Filling the Frame	0.33	0.44	0.38
Leading Lines	0.28	0.42	0.33
Minimal	0.35	0.73	0.47
Patterns and Textures	0.38	0.50	0.43
Rule of Thirds	0.46	0.07	0.13
Undefined	0.53	0.56	0.54
macro avg	0.36	0.42	0.35 (0.14)

Table 6.4: Composition Classification Report

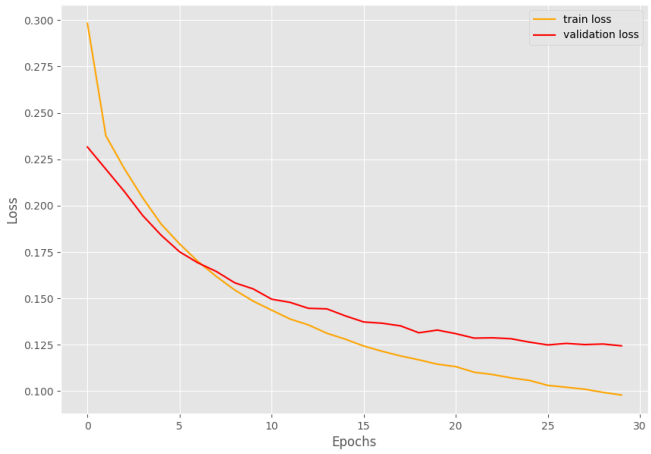


Figure 6.9: Type Validation Loss

Figure 6.10: Type Confusion Matrices

Type Classification Report			
Type	Precision	Recall	F1-Score
Aerial	0.33	0.30	0.32
Architectural	0.50	0.79	0.61
Astro	0.89	0.67	0.76
Cityscape	0.38	0.43	0.40
Commercial	0.12	0.43	0.19
Event	0.15	0.29	0.20
Fashion	0.54	0.28	0.37
Flora	0.87	0.79	0.83
Food	0.78	0.90	0.84
Landscape	0.71	0.84	0.77
Macro	0.50	0.26	0.34
Night	0.60	0.20	0.31
Pet	0.65	1.00	0.79
Portrait	0.67	0.82	0.74
Sports	0.17	0.50	0.25
Street	0.79	0.48	0.60
Underwater	0.70	0.64	0.67
Wedding	0.40	0.12	0.19
Wildlife	0.83	0.29	0.43
Other	0.56	0.50	0.48
macro avg	0.66	0.55	0.60 (0.06)

Table 6.5: Type Classification Report

6.2 Results

Throughout the experiments, there were also some classes that we failed to sufficiently populate, so we excluded them from our final experiments. Those include the "Documentary" from the "Type" task and "Frame within Frame", "Symmetrical" from the "Composition" task. We also proceeded in a merging between the "Pink" and "Violet" colors from the "Palette" task, under the "Magenta" name.

Apart from the aforementioned performance experimentations, we also demonstrate the ability of the models to infer on new unseen data. On figures 6.11, 6.12 we present some images with their "Actual" vs "Predicted" label, while we inference on our test set (10 unseen images for each task):

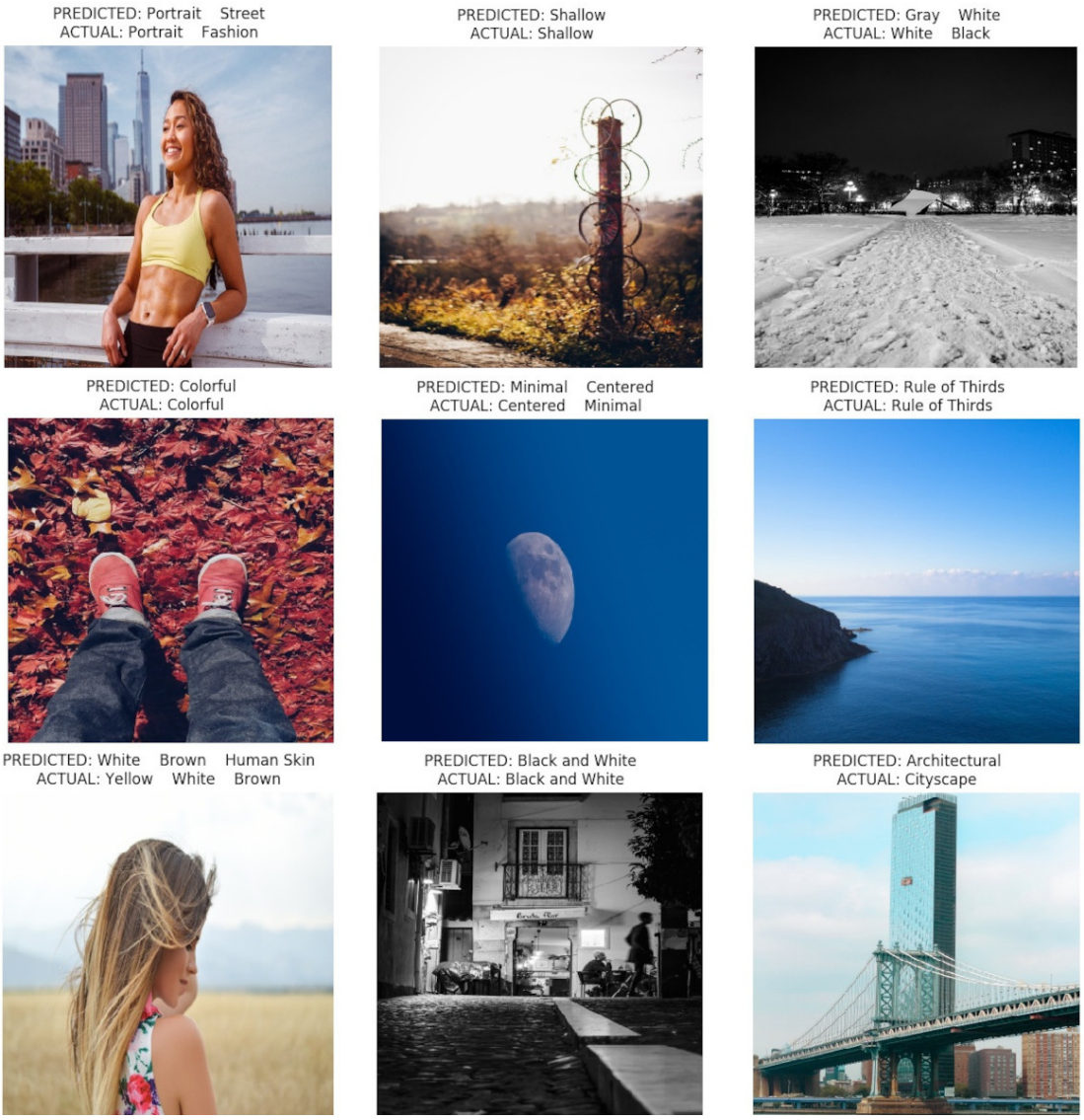


Figure 6.11: Inference - Predict on Test Set

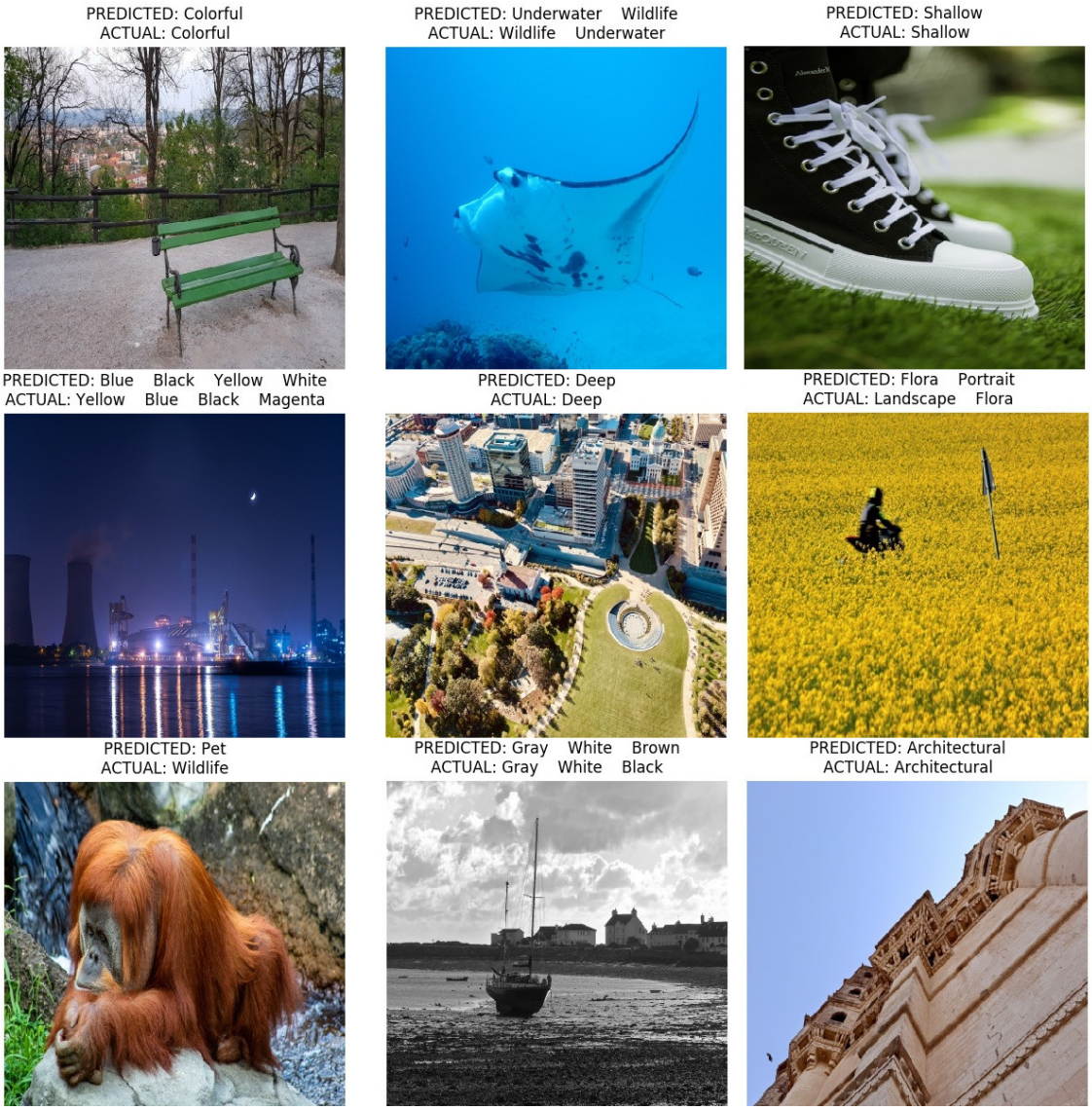


Figure 6.12: Inference - Predict on Test Set

We notice some good predictions, as well as some pretty close predictions on all tasks. For example, in the first image of 6.11, the photograph could easily be considered as of a "Street" type. And in the last image, "Architectural" and "Cityscape" are two pretty similar -under the right circumstances- concepts, which can easily puzzle our model. Similarly, on 6.12 our model has wrongly predicted an orangutan as a "Pet", closely related to "Wildlife" as both consists of animal images. Those results indicate that our model trained quite well on the classes we had on plethora. That also means that with an even more active targeting of classes to train we can achieve a very well rounded model, able to predict really well on all of our predefined classes for all our tasks.

Finally, we want to re-emphasize the importance of subjectivity of photography as an art. Even when following "specific" rules, our interpretation during the an-

notation process on some tasks may be not right at all. In the end, no one really knows what a photographer wanted to capture on his/her photograph when he/she was taking the picture. And that is the whole purpose and the beauty behind photography (and arts in general). The fact that the public sees (and feels) different things on a photograph than his/her creator.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this work, we have presented a novel dataset for classifying photographs with regards to their aesthetic attributes. Towards this end, two binary and three multi-label classification tasks have been defined and more than 1800 images have been annotated, and inter-annotator agreement has been estimated for each task on a a sub-set of images. Those tasks included the "Color", "Depth of Field", "Palette", "Composition" and "Type" of a photograph. In addition, CNNs have been evaluated on these tasks.

Specifically, we implemented a ResNet50 and results proved that their ability to discriminate between aesthetic classes was 3 times higher than the baseline performance. The dataset is openly provided, along with the trained models and Python code to test them [1].

7.2 Future Work

The proposed work can be extended by adding more aesthetic tasks, such as the "Exposure" of a photograph (over/under), or its "Key" (high/low). In addition, more difficult tasks include predicting the "White Balance" or the "Zoning System". In addition, a wider experimentation with CNN architectures and transfer strategies can be evaluated in a future work. Finally, in the context of a wider and production-level annotation procedure, an improved version of the ground truth could be generated if multiple annotators were used for the whole dataset (instead of just using multiple annotators to estimate the inter-annotator agreement on each task).

[1] github.com/magcil/deep_photo_aesthetics

Some of the most important things for immediate improvement on the performance of our model include:

- more data, especially for our underwhelming classes
- multiple annotators per image for a better structured inter-annotator agreement and therefore better dataset integrity
- periodical validation on the annotation process by experts
- transfer training on GPU and experiment with more RAM-intensive model tuning

We are also continuing to improve this project, through more data collection, annotation, cross-validation and model tuning. We have already set up a server and a basic RESTful API [2] for people to be able to get predictions on their own photographs. We will continue to scale the server, with updated API calls, better security through AAI integration and later a UI. We are aiming for a fully fledged experience through which individuals and groups interested in photography can learn some basics about this art, track their unique style progression and have lots of fun.

[2] API

References

[1] Hongtao Yang, Ping Shi, Saike He, Da Pan, Zefeng Ying, and Ling Lei. A comprehensive survey on image aesthetic quality assessment. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pages 294–299. IEEE, 2019.

[2] Michael Freeman. *The complete guide to digital photography*. Sterling Publishing Company, Inc., 2008.

[3] Daan Zwick. Contrast in photography. In *Glare and Contrast Sensitivity for Clinicians*, pages 139–144. Springer, 1990.

[4] Junho Cho, Sangdoo Yun, Kyoung Mu Lee, and Jin Young Choi. Palettenet: Image recolorization with given color palette. In *Proceedings of the ieee conference on computer vision and pattern recognition workshops*, pages 62–70, 2017.

[5] Pere Obrador, Ludwig Schmidt-Hackenberg, and Nuria Oliver. The role of image composition in image aesthetics. In *2010 IEEE International Conference on Image Processing*, pages 3185–3188. IEEE, 2010.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[7] Gautam Malu, Raju S Bapi, and Bipin Indurkhya. Learning photography aesthetics with deep cnns. *arXiv preprint arXiv:1707.03981*, 2017.

[8] Marco Leonardi, Paolo Napoletano, Alessandro Rozza, and Raimondo Schettini. Modeling image aesthetics through aesthetics-related attributes. In *London Imaging Meeting*, volume 2021, pages 11–15. Society for Imaging Science and Technology, 2021.

[9] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning*, pages 647–655. PMLR, 2014.

[10] Hiya Roy, Toshihiko Yamasaki, and Tatsuaki Hashimoto. Predicting image aesthetics using objects in the scene. In *Proceedings of the 2018 International Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia*, pages 14–19, 2018.

[11] Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. Aesthetic quality classification of photographs based on color harmony. In *CVPR 2011*, pages 33–40. IEEE, 2011.

[12] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006.

[13] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *European conference on computer vision*, pages 288–301. Springer, 2006.

[14] Dong Liu, Rohit Puri, Nagendra Kamath, and Subhabrata Bhattacharya. Composition-aware image aesthetics assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3569–3578, 2020.

[15] Kyle DeGuzman. Aesthetic photography — composition, lighting, and color. `https://www.studiobinder.com/blog/what-is-aesthetic-photography`, Studiobinder, 28 Feb. 2021.

[16] H Jane Brockmann. The control of nest depth in a digger wasp (sphex ichneumoneus l.). *Animal Behaviour*, 28(2):426–445, 1980.

[17] B.J. Copeland. artificial intelligence. `https://www.britannica.com/technology/artificial-intelligence`, Encyclopedia Britannica.

[18] Expert.ai Team. What is machine learning? a definition. `https://www.expert.ai/blog/machine-learning-definition`, expert.ai, 6 May 2020.

[19] MLK. Brief history of deep learning from 1943-2019 [timeline]. `https://machinelearningknowledge.ai/brief-history-of-deep-learning/`, machinelearningknowledge.ai, 24 Nov 2019.

[20] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[22] Keno K Bressem, Lisa C Adams, Christoph Erxleben, Bernd Hamm, Stefan M Niehues, and Janis L Vahldiek. Comparing different deep learning architectures for classification of chest radiographs. *Scientific reports*, 10(1):1–16, 2020.