INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Deep learning-based graffiti detection: a study using images from the streets of Lisbon

Joana Pereira Fogaça

Master's in Data Science,

Supervisor:
Ph.D. Tomás Gomes da Silva Serpa Brandão, Assistant Professor, ISCTE - University Institute of Lisbon

Co-Supervisor:
Ph.D. João Carlos Amaro Ferreira, Assistant Professor with Habilitation, ISCTE - University Institute of Lisbon

November, 2022

Department of Information Science and Technology

Deep learning-based graffiti detection: a study using images from the streets of Lisbon

Joana Pereira Fogaça

Master's in Data Science,

Supervisor:
Ph.D. Tomás Gomes da Silva Serpa Brandão, Assistant Professor, ISCTE - University Institute of Lisbon

Co- Supervisor:
Ph.D. João Carlos Amaro Ferreira, Assistant Professor with Habilitation, ISCTE - University Institute of Lisbon

November, 2022

I dedicate my dissertation to my loving parents.

Acknowledgements

I want to leave my most sincere thanks to all the people who helped, accompanied, and supported me throughout the course.

First, to my family, who supported me so much and always believed in me. Especially to my parents, thank you very much for all the love, joy, support, and assistance throughout these years.

To my friends, for understanding and supporting me even in the most stressful moments.

To the advisor and co-advisor for all the support, help, and knowledge-sharing throughout all the meetings and emails exchanged.

And finally, to the Lisbon City Council, for supporting this partnership and providing the necessary images for this work.

Resumo

A Câmara Municipal de Lisboa está interessada em desenvolver um sistema que detete automaticamente e em tempo real os graffitis ilegais na cidade de Lisboa, utilizando carros equipados com câmaras fotográficas. Este sistema permitiria uma identificação e limpeza mais rápida e eficiente dos graffitis ilegais que estão constantemente a ser produzidos. Uma resposta imediata a este tipo de vandalismo tornou-se cada vez mais pertinente não só para assegurar que a cidade continua a ser um local limpo e seguro onde os cidadãos se sentem confortáveis e felizes como também para desencorajar este tipo de atos.

Esta tese consiste numa prova de conceito da viabilidade do sistema, de forma a compreender se faz sentido dedicar mais esforços à criação do sistema. Foram fornecidas e recolhidas imagens de diferentes fontes que incluíam graffitis ilegais, imagens com graffitis consideradas arte de rua e imagens sem graffitis.

Foi desenvolvida uma pipeline que primeiro classifica a imagem com uma das seguintes etiquetas: graffiti ilegal, arte de rua ou sem graffiti. Caso seja um graffiti ilegal, é utilizado outro modelo que deteta as coordenadas do graffiti na imagem.

Foram utilizadas técnicas de pré-processamento, aumento de dados e transferência de aprendizagem para treinar os modelos.

Quanto ao modelo de classificação, foi obtida uma acurácia global de 81,4% e F1-score de 86%, 81% e 66% para as classes *street-art*, graffiti ilegal e imagem sem graffiti, respetivamente. Quanto ao modelo de deteção de graffiti, foi obtida uma Interceção sobre a União (IoU) de 70,3% para o conjunto de teste.

Abstract

The Lisbon City Council is interested in developing a system that automatically detects in real-time illegal graffiti present throughout the city of Lisbon by using cars equipped with cameras. This system would allow a more efficient and faster identification and clean-up of the illegal graffiti constantly being produced. More immediate response to this kind of vandalism has become more and more relevant to ensure that the city remains a clean and safe place where citizens feel comfortable and happy and to discourage this kind of act.

However, because there were only images available, this thesis became a proof of concept of the system's viability to understand if it makes sense to engage more effort in creating the system. Images were provided and collected from different sources that included illegal graffiti, images with graffiti considered street art, and images without graffiti.

A pipeline was then developed that first classifies the image with one of the following labels: illegal graffiti, street art, or no graffiti. And then, if it is illegal graffiti, another model was trained to detect the coordinates of graffiti on an image.

Pre-processing, data augmentation and transfer learning techniques were used to train the models.

Regarding the classification model, an overall accuracy of 81.4% and F1-scores of 86%, 81% and 66% were obtained for the classes street art, illegal graffiti and image without graffiti, respectively.

As for the graffiti detection model, an Intersection over Union (IoU) of 70.3% was obtained for the test set.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

Glossary of Abbreviations and Acronyms

- API - Application Programming Interface
- CNN - Convolutional Neural Network
- CRISP-DM - Cross Industry Standard Process for Data Mining
- FC - Fully Connected
- FPN - Feature Pyramid Network
- IoU -Intersection over Union
- OCR - Optical Character Recognition
- PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- R-CNN - Region Based Convolutional Neural Networks
- RoI - Region of Interest
- RPN - Region Proposal Network
- SIFT - Scale Invariant Feature Transform
- TOF - Time-Of-Flight
- UAV - Unmanned Aerial Vehicle
- XML - Extensible Markup Language

CHAPTER 1

Introduction

Ellis describes graffiti as “someone’s urge to say something— to comment, to inform, entertain, persuade, offend or simply to confirm his or her own existence here on earth” [1]. The identification of graffiti as art or crime has long been discussed from various social perspectives, such as culture, art, politics, and economics. There is still a significant disagreement in society, with some defending and supporting graffiti as a positive aspect and a form of artistic expression, while others consider it an act of vandalism [2], [3].

Ross and Wright define the term graffiti as “words, figures, pictures, caricatures, and images that have been written or drawn on surfaces where the owner of the property has not permitted this activity” and street art as “stencils, stickers, and wheat-pasted posters (e.g., non-commercial images) that are affixed to surfaces where the owner of the property has not permitted this activity” [4, p. 2]. Both approaches are typically associated with acts of vandalism since the property owner usually does not permit the action in question [4].

However, street art has shown great cultural importance in some cities, such as Lisbon. The Portuguese capital has also been slowly standing out in the world of urban art with the intense production of such works over the last year. Thus, Lisbon has been gradually positioning itself in this field worldwide and obtaining one additional motivation factor for tourism [5].

Unlike many other cities, the Lisbon City Council provides specific spaces and walls spread around the city that the artists can apply to create street art, encouraging the creation of more of these works by making them legal and publishing them in the Lisbon urban art gallery's website [6].

As Campos states, “In international terms, street art has gradually become a city asset while at the same time it has grown in prestige and value for the contemporary art market” [5, p. 1]. For this reason, some studies, such as the one conducted by Novack *et al.* [7] already begun to focus on identifying and analysing these works of art to support their mapping for tourism purposes.

On the contrary, illegal graffiti, which does not add any value to the place, has become increasingly financially prejudicial due to the costs associated with its prevention and cleaning [8], [9]. Besides that, illegal graffiti is known to cause a negative impact on the local economy: since general people associate it with dirtiness and insecurity, areas containing a wide presence of illegal graffiti are subject to a decline in consumer demand for products and services (such as restaurants, cafes, shops, houses, bus stops, etc.) [10], [11].

According to Capucho in [12], the same can be observed in Lisbon. A “Diário de Notícias” report states that the Lisbon City Council spends about half a million euros annually cleaning graffiti and tags throughout the city [12]. In addition to this, the Portuguese transporter "Comboios de Portugal" (CP) claims that the money spent on cleaning graffiti from trains between 2008 and 2019 would allow the company to buy a new train [13]. According to Ferreira [13], cleaning each square meter painted with graffiti costs, on average, costs 7.35€ to Portuguese taxpayers.

To try to control and minimise damage, surveillance systems are often used. However, it is costly and impractical for surveillance personnel to monitor and detect graffiti simultaneously on multiple images and cameras [8], [10]. For these reasons, more and more effort has been made to control and facilitate graffiti detection through automatic algorithms [11].

The Lisbon City Council is interested in developing a system to automatically detect graffiti using real-time videos captured by cars that will navigate the city. Therefore, the process of supervision, identification, planning, and communication with the team of Urban Hygiene of the city and the removal of graffiti will be faster and more effective. This way, it will become easier to mitigate a significant problem in the town of Lisbon related to vandalism and damage to public spaces through graffiti.

The work developed in the context of this Dissertation intends to be a proof of concept that aims to provide evidence of the feasibility of an automatic system for the identification and classification of graffiti using machine learning algorithms.

1.1. Motivation

One of the most challenging objectives of the City Council focuses on constantly maintaining a clean and safe city, free of vandalism. One of the recurrent forms of vandalism in Lisbon is the elaboration of graffiti on buildings, walls, and objects in the city. To tackle this problem and keep the city clean, the Lisbon City Council has Urban Hygiene teams in the city whose goal is to remove graffiti. This removal occurs when the team is notified of the presence of graffiti and its location. However, before this notification is sent to the team, much work is necessary from the Lisbon City Council employees in capturing, gathering information, and filtering the images to detect/identify the graffiti to be removed.

A classification and detection system for graffiti could be an excellent asset for the City Council workers as it makes the process of selecting the graffiti to be notified for cleaning much more effective and faster.

1.2. Objectives

The primary objective of this study is to improve the process of detection and georeferencing of graffiti in Lisbon through a system that automatically identifies and classifies an image as having illegal graffiti (Figure 1), street art (Figure 2) or no graffiti.

Figure 1. Examples of images with illegal graffiti

Figure 2. Examples of images with street art graffiti

In addition, in cases where the image is classified as illegal, the system also detects the region of the image associated with the graffiti to notify the Urban Hygiene teams which places need to be cleaned.

This system would allow the allocation of the work done by the Lisbon City Council members in selecting the places to be cleaned to other more critical tasks. Thus, a tedious and time-consuming process can become a simple, easy, and effective process that only requires uploading images to an application.

The grand ambition of the City Council is to develop a system that, through a camera system implemented in cars driving around the city, can detect walls that need cleaning and automatically notify the corresponding team of their location.

This Dissertation developed a contribution/proof of concept to understand the viability of the identification and classification of graffiti and, this way, obtain the necessary certainty and confidence to determine whether it makes sense to invest in a system of automatic identification and classification in real time of the existed graffiti in the city of Lisbon.

1.3. Research Questions

In addition to the objectives mentioned above, the system and its analysis for automatic graffiti detection and classification allow us to answer the following question:

1. Can a deep learning model successfully discriminate the differences between street art and illegal?
2. How accurate is the automatic identification and location of illegal graffiti on images acquired in the streets of Lisbon under loose controlled conditions?

1.4. Research Methodology

The system development methodology was based on the Cross Industry Standard Process Standard for Data Mining (CRISP-DM). CRISP-DM is a methodology that provides an overview of the life cycle of a data mining project, describing all phases of the project [14].

However, it was necessary to adjust the methodology to comply with the needs and characteristics of the problem and the data in question**.** The main difference to the problem at hand is the type of data. In this case, since it is a computer vision problem, the data consists of images, which require a different kind of processing and collection**.** These adjustments can be observed in Figure 3.

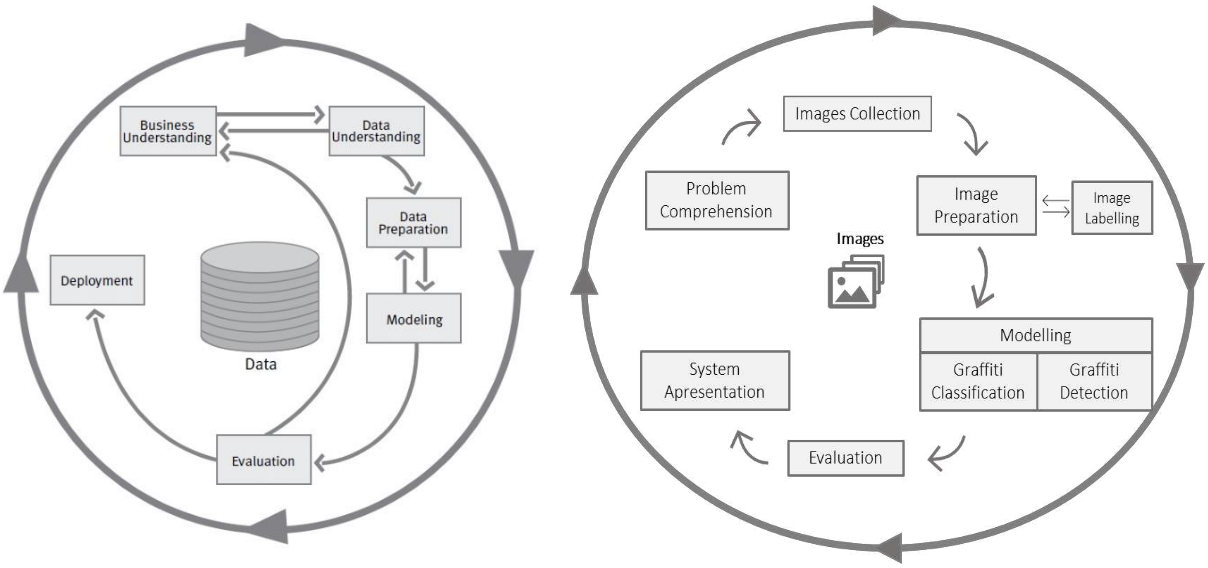

Figure 3. CRISP-DM and CRISP-DM proposed adjustments

Considering the diagram of the adjusted system, the different phases of the project's development are as follows:

1. **Problem Comprehension:** Through several meetings with Lisbon City Council members from various departments, several topics were discussed to retrieve the necessary understanding of the problem, the current procedure used, the objectives, and the data available for the implementation.
2. **Data collection:** The available data comes from various sources: 1) a shared folder provided by members of the Lisbon City Council containing images with diverse formats, resolutions, quality, illumination conditions, points of view and distance to the object of

interest; 2) images from the Lisbon urban art gallery website; and 3) images collected from various internet sources. Due to this image data diversity, an extensive and time-consuming process of collecting and organizing the different images was required.

3. **Image Preparation:** Duo to the different types, formats and backgrounds of the collected images, an exhaustive process of data preparation was necessary for them to be considered suitable for use in the models. This phase involved a considerable procedure of selection, elimination, extension conversion, and organization of the images.
4. **Data labelling:** Aligned with the preparation of the images, it was necessary to annotate a large number of images. This process was done using the *LabelImg* tool[1] that allows selecting the area corresponding to the graffiti in each image and saving its coordinates in files that are then used as inputs in the model.
5. **Modelling:** During the modelling stage, two models with different purposes were trained and saved: a classification model that allows the classification of an image into three classes (with illegal graffiti, with street art, or without graffiti) and a model for the detection of illegal graffiti in an image. This procedure involves an extensive process of testing several different neural network architectures and fine-tuning parameters.
6. **Evaluation:** To evaluate the models obtained in the previous phase, for the classification model, several metrics were used, such as accuracy, precision, recall, and F1-score. For the detection model, the Interception of Union metric was calculated as it allows us to measure how similar the predicted bounding boxes are to the true ones.
7. **System presentation:** Finally, for the presentation of the system and the work done, this dissertation was written and served as documentation, and the system and some results were presented at the 1st Meeting of the Urban Data Laboratory of Lisbon. Besides this, there is the ambition and possibility of creating a paper.

[1] https://github.com/heartexlabs/labelImg

1.5. Dissertation Structure

After the Introduction, this dissertation is organized according to five additional chapters distributed as follows:

- Chapter 2 presents the literature review, including the description of the systematic review process: systematic review methodology, discussion of the work done so far in this field, analysis of the selected papers, and final discussion about the importance of the theme, the different approaches already developed, and the gaps verified.
- Chapter 3 provides an overview of the developed system and the various procedures performed for the implementation and modelling of the two deep learning models: the graffiti classifier and the graffiti detector. Besides this, this chapter also presents the data used in this work and describes the procedures applied from its collection to its preparation for model training and evaluation.
- Chapter 4 describes the process of finding the best-fitting model for each objective, the experimental setups and the results and metrics calculated for the different trials tested. It also describes and demonstrates the limitations and gaps of each of the models.
- Chapter 5 concludes the dissertation and provides suggestions for future work.

CHAPTER 2

Related Work

This chapter will briefly introduce the literature review process, from the methodology used to the conclusions and algorithms used in similar projects related to identifying street art and illegal graffiti through video or images.

2.1. Systematic Review Methodology

The systematic literature review was based on the PRISMA methodology. The research methodology for the literature review started by gathering articles related to the theme through a joint search in the abstracts and citations from Scopus, web of science, and google scholar databases.

The query used to search the articles was the following:

(graffiti OR street art OR (painting AND (wall* OR facades OR building*)) AND ("deep learning" OR "computer vision" OR "image analysis" OR "machine learning" OR " data science" OR "neural network" OR detection OR classification)*

Different queries were considered and tested. However, this was the one that best suited the problem, resulting in articles covering the study's two main themes: graffiti (of any kind) and data science. Additionally, the retrieved articles describe methods and models for detecting/analysing paintings on walls using machine learning or image processing algorithms.

However, further filtering was necessary since, in addition to these articles, the results also present several related works from other fields, such as analysing the best ways to remove graffiti, to detect the material of a surface covered with graffiti, and even duplicate articles. For this selection, in some cases, a simple analysis of the titles was enough, while others required a closer look at the abstracts.

At the end of the process, additional articles were removed from the list by further analysis of their contents, on the other hand, by inspecting the reference lists, additional articles were added to the selection. The included articles address not only other attempts of automatic classification and detection of graffiti but also their cultural impact, where discussions and opinions on the subject were debated.

This Methodology generated a total of 20 articles, where 15 of which are related to machine learning systems or image processing algorithms.

2.2. Related Work

Regarding the detection of acts of vandalism, there are some systems that, instead of only detecting graffiti, also detect the act of graffiti. An example is the work published in [8], which implemented a system that aims to identify stationary visible changes based on the detection of modifications concerning a reference background that is stationary in space and time. However, since other objects, such as people standing still and parked vehicles, can also display stationary patterns in space and time, this system is prone to false positives.

Both [10] and [11] try to make the detection more effective by, in addition to analysing visible changes due to variations of the appearance, also analysing visible changes due to variations of the 3D geometry of the scene, i.e., the information relative to the depth. This way, this application can improve the results of [8] since new objects in the scene will change brightness and depth; thus, the algorithm will be less prone to false positives. The work in [10] was subject to the TOF (Time-Of-Flight) camera's limitations since the resolution of the camera used in the experiments was 64 x 64 pixels for both intensity and depth images, which did not allow a distance of more than 1.5/2 meters between the camera and the wall under test. However, the experimental results of [11] allowed to verify the robustness of the method used in different situations, such as crowded scenes, abandoned objects, static intrusions, and illumination changes.

Nahar *et al.* [15] proposed a system based on an autonomous Unmanned Aerial Vehicle (UAV) that can detect graffitied walls and cover them with spray paint if necessary. This was designed to clean hard-to-reach public places such as bridges and highways. The video stream of the UAV is sent to a machine-learning server containing a trained model developed from scratch for detecting graffiti images. This neural network model was built using the machine learning library TensorFlow. However, the article does not mention any results obtained, thus making it impossible to compare with other algorithms regarding their performance.

Similarly, Wang *et al.*[9] also proposed a semi-autonomous UAV graffiti detection and removal system, but this time based on the *ssd_mobilenet_v2_coco* transfer learning model pre-trained on the COCO data set from the tensor flow API (Application Programming Interface). However, in this case, two different models were developed, one for graffiti detection on traffic signs and another for graffiti detection on walls. The model for detecting graffiti on walls also recognises some graffiti styles. The graffiti are classified into: Throw up Graffiti, Wildstyle Graffiti, Cartoon Graffiti, Throw up Alphabet, Wildstyle Alphabet or Cartoon Eye. The tests performed on both models showed an accuracy of up to 99%. However, the authors also state that the system needs to be further tested in more complex environments.

In [16], the subject is also the detection of illegal graffiti. The paper proposes the creation of a graffiti map based on the amount of graffiti. Its purpose is to tackle vandalism in places with high concentrations of graffiti and discourage future acts. The model was trained using 632 images acquired in São Paulo City, using a Resnet 101-layers backbone model pre-trained on the Coco dataset. The results from this transfer learning method showed an average precision of about 0.57.

Studies focused on detecting other types of graffiti, such as murals or street art, can provide information and methods that help in the creation of maps with the exact locations of the artworks for their divulgation to the interested community, as in the study conducted by Tessio Novack *et al.* [7]. This study uses the VGG16 convolutional neural network (CNN) model pre-trained on the ImageNet dataset, with three fully connected layers and a dropout rate of 0.5. Using the binary cross-entropy loss function and the AdaGrad optimisation algorithm, an overall accuracy of 93% was achieved. This algorithm allowed the production of a density map containing the graffiti artworks found in the central part of London.

Munsberg *et al.* [17] try to investigate how a CNN model performs in the detection of art graffiti. The main contribution of this paper is to demonstrate that, when using transfer learning, instead of removing the last fully connected layer for a layer containing the desired number of final classes in the neural networks, it is more efficient to maintain it. Munsberg *et al.* argue that removing the last layer may result in a loss of relevant information for the new task, and, with this approach, they were able to achieve the results faster, with a smaller number of epochs.

Besides the previously mentioned works, there was one article whose goal was the detection of any type of graffiti, whether it is considered art or not, such as the approach applied to Medellín City [3] that uses the PyTorch library to implement an R-CNN, the Resnet-50 classifier, already pre-trained on the ImageNet dataset. This research allowed the construction of a visualisation tool through heat maps that, besides helping define measures to improve sectorial policies, also allows better control and definition of efforts to preserve the areas rich in art graffiti and restore those with a negative aspect. As future work, the paper mentions a possible improvement to a more in-depth graffiti classification based on their form or purpose.

In addition to graffiti detection, a new topic is being increasingly discussed with graffiti data: gang identification by segmenting the graffiti based on their similarities. The analysis and interpretation of gang graffiti can help law enforcement better understand their activities and where they need to operate to respond and have an idea of the gang's intentions [18]–[21].

The system implemented by Wei Tong *et al.* [17] assumes that two graffiti are more likely to be created by the same graffiti artist if they have high similarities in visual and contextual aspects. This system starts by extracting visual features (such as letters, numbers, and symbols) through OCR (Optical Character Recognition) and selecting the most similar images. Then the similarities between the images are calculated, and this way, it is possible to identify the most similar photos, which will correspond to those with a higher probability of having been drawn by the same individuals. The results obtained achieved an accuracy of about 64%.

Graffiti-ID [18] is a research project conducted at Michigan State University that aims to return similar graffiti from their database. For this, the Scale Invariant Feature Transform (SIFT) was used to extract the most relevant visual features (which are referred to as critical points). Then, the graffiti association is based on calculating the similarities via Euclidean distances between the critical points of the two images. Local geometric constraints are added to try to reduce false associations.

A CNN model was adopted by He Li *et al.* [18] to classify graffiti into different classes based on a set of graffiti components. The model was composed of five convolutional layers followed by three fully connected ones and a final softmax layer that achieved an overall accuracy of 87%.

Parra *et al.* [22] proposed a model-based system that, by analysing graffiti images, can present relevant information about the gangs associated with the graffiti. This system is composed of three methods: colour recognition (taking advantage of the capabilities of the modern mobile devices' touch screen and the user's help to trace the path along the colour region), segmentation of the colour image based on Gaussian thresholding, and content-based retrieval of the graffiti to detect the graffiti and identify its components as objects and shapes (such as stars, pitchforks, crowns, and arrows). From here, a hierarchical k-means clustering is used to create vocabulary trees. As the authors of the article state, "the main advantage of using a vocabulary tree for image retrieval is that its leaves define the quantisation, thus making the comparison dramatically less expensive than previous methods in the literature" [22, p. 3].

2.3. Research Outcome

The documents selected for this research came from various sources such as articles, newspaper reports, and books, but most came from conference papers. Furthermore, the areas associated with each article present a wide variety of results, with computer science presenting the highest percentage of articles, as shown in Table 1.

Table 1. Percentage of selected articles by field

Subject area	% of articles
Computer Science	27.5%
Social Sciences	17.5%
Engineering	12.5%
Decision Sciences	12.5%
Arts and Humanities (and News)	12.5%
Mathematics	10%
Physics and Astronomy	7,5%

Figure 4 displays a bar plot with the number of articles per country. It is possible to observe that the United States is the country that has contributed the most to the advancement of systems in this field. It is mainly related to the segmentation of graffiti for a better understanding of some gangs' behaviours, locations, roots, and ambitions since the number of gang-related crimes have increased in the US [23]. None of the selected articles was written by Portuguese authors.

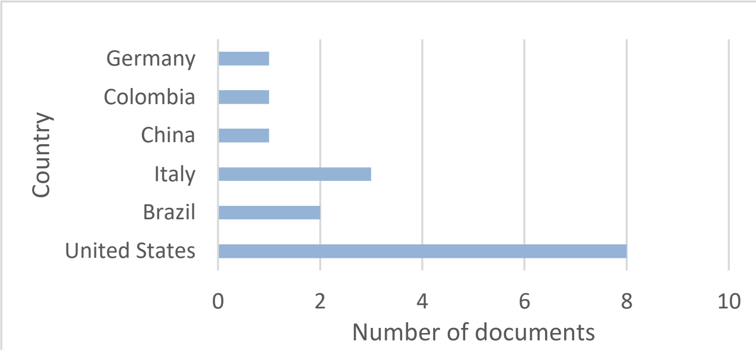

Figure 4. Number of documents by country

Although the detection and classification of graffiti is a topic that can help not only to minimise the damage related to illegal graffiti but also to spread the benefits of urban art, the literature review allowed to realise that this subject has not yet been sufficiently explored in terms of technology. Of the selected articles, only 15 articles are related to graffiti identification or classification methods.

From the reviewed articles, the most frequent target is the detection of illegal graffiti, as can be observed in Table 2. This can be justified by the importance of the negative connotation associated to with legal graffiti, which brings discontent from the population and associated costs [8]–[11]. As stated by Angiati *et al.*, "for many people, graffiti's presence suggests the government's failure to protect citizens and control lawbreakers" [8, p. 1] and, for this reason, the need to find alternatives to control and minimise the costs related to the issue has become more and more certain.

In addition, the segmentation of graffiti based on their similarities for the purpose of detecting gangs has also shown to be very valuable in this area because it helps law enforcement agencies understand the activities and territories of each gang [21], [23].

Of the selected articles, only two focus on classifying street art graffiti (only classifying as street art or no graffiti), and only one report [3] covers the detection of both types of graffiti without distinguishing them.

Table 2 shows the number of articles based on each detection type.

Table 2. Number of articles by type of detection

Type of detection	Number of articles
All graffiti types	1
Street art	2
Gangs	5
Illegal graffiti	7
Total	**15**

Regarding graffiti detection (illegal or art), the two most studied and developed methods were: image processing, where detection was based on videos from surveillance cameras, and neural networks. The CNN (convolutional neural network) architecture appears to be the most thorough in this area since it has been proven to be the best algorithm for image understanding and to provide very successful results in segmentation, classification, tagging, detection, and retrieval tasks [24], [25].

83% of the cases that used neural networks for detection were through transfer learning. Transfer learning is used to enhance the machine learning process of a domain by transferring information from a related problem instead of starting and learning from scratch [26].

2.4. Research Discussion

As mentioned above, graffiti can be damaging or beneficial to the location in question. In the case of illegal graffiti, it is essential to control and act quickly on it to avoid giving the author notoriety, thus discouraging this act. In addition, the graffiti components and their details can also provide important information about how and where some graffiti artists or gangs operate. In the case of urban art, publicising it can help attract people to the area, thus improving the local economy.

The 20 articles selected for the elaboration of this literature review allowed not only to understand the concerns and disagreements within the theme but also to know and comprehend some of the algorithms implemented both for the detection of graffiti, for the segmentation and classification of the creators of the graffiti.

The study made it possible to understand that, although interest and importance of this field have been growing, there are still few implementations for these purposes. Moreover, no artificial intelligence system in Portugal has yet been developed to address this issue. Another verified gap was the lack of algorithms that distinguish between street art and illegal graffiti. There are already some systems for detecting illegal graffiti and others for detecting street art. However, a system that integrates both concepts was not found in the literature. Only one study [3] addresses this type of classification, and only as future work.

CHAPTER 3

Graffiti Identifier

3.1. General Overview

The proposed system, the graffiti identifier, uses two deep learning models to automatically classify the type of graffiti on an image (between street art, illegal graffiti or no graffiti at all) and to localise it for the illegal graffiti case.

Initially, the images go through an image classification model that tries to identify the type of graffiti on the pictures. Through this classification, the City Council department is allowed to plan how to proceed. If the output class is street art, the image and its geographic location can be used for marketing purposes since there is also a particular target niche of tourists interested in this type of art. The image and its geographic location can, for instance, be included in a map for its disclosure or be added to the website of the urban art gallery where several urban arts of the city of Lisbon are featured, thus keeping the site constantly updated.

For the case where the graffiti is classified as illegal, the image will go through a second deep learning model, but in this case, the objective is the automatic detection of the coordinates of the graffiti on the image. Once an illegal graffiti is detected, an alert can be sent to the corresponding cleaning and sanitation team to proceed with its cleaning.

This pipeline allows us to automate and facilitate a process currently done by members of the Lisbon City Council as soon as they receive images that report the presence of new graffiti in Lisbon.

Figure 5 shows a diagram that summarises the implemented system.

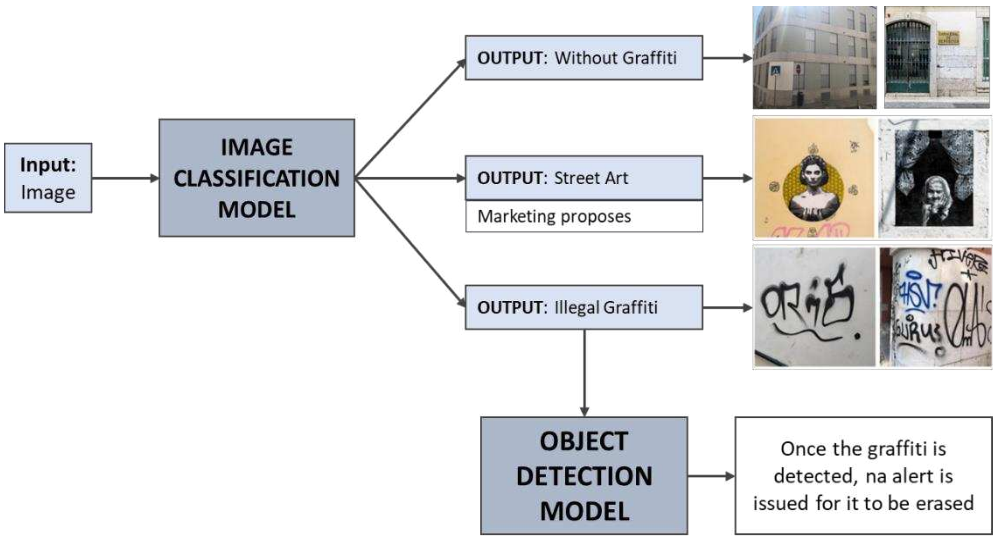

Figure 5. Summary of the implemented system

3.2. Deep Learning Models

Two deep learning-based models were developed to respond to the two different objectives. First, a classification model that identifies the type of graffiti (or absence of it) in an image. And secondly, a detector of illegal graffiti that locates it in the image.

Both models were developed using pre-trained machine learning models through transfer learning. Using a pre-trained model for a larger-scale image classification problem, we can take advantage of some learned feature maps that allow us to start at a more advanced point of learning, already with some generally valuable features that will enable a faster and more advanced model creation.

3.2.1 Graffiti Classifier

As mentioned, the graffiti classifier aims to classify an image into one of three classes: image with illegal graffiti, image with street art, or image without graffiti.

Since transfer learning involves the use of a pre-trained neural network, several architectures previously trained for image classification problems were tested, adding only four new training layers at the end**:**

- A Flatten layer to transform the multi-dimensional output from the Keras application model into a single-dimension tensor.
- A Dense layer. Three different activation functions were tested ('linear', 'relu', and 'tanh').
  A Dense layer has a deep connection. In other words, all neurons in this layer are connected with all neurons from the previous layer, allowing to learn information from all combinations of features from the previous layer.
- A Dropout layer to prevent overfitting.
- A Dense layer with a SoftMax activation function that allows changing the dimensionality of the output to be accordingly to the three different classes: illegal, street art or without graffiti.

The weights used in the tested architectures were obtained using the ImageNet dataset[2], a large dataset organised according to the WordNet hierarchy, comprising over 14 million images categorised into about 22 thousand different object categories. Although the images included here exhibit considerable differences relative to the images used in this dissertation (regarding graffiti), pre-trained networks with weights optimised for this large dataset can be useful as feature extractors. For

[2] https://image-net.org/index.php

example, a network that can already identify walls correctly can be a valuable contribution to the problem at hand since graffiti is usually present on them.

The pre-trained architectures tested on the scope of this Dissertation were *Resnet*, *EfficientNet*, *VGG*, *DenseNet*, *Xception* and *InceptionResNet*.

The Resnet architectures use residual blocks (or "skip connections") to solve a problem often related to deeper networks since the vanishing gradient, as the number of layers in the neural network increases, the accuracy gets saturated and starts to degrade after a certain point. These residual blocks behave as shortcut connections that perform identity mapping [27]. Two residual neural network architectures from this family were tested: ResNet50 and ResNet15V2.

EfficientNet is a convolutional network architecture that uses a new scaling approach that uniformly scales all depth / width / dimensions using a composite coefficient [28]. The EfficientNetV2L and EfficientNetB7 architectures were used in the tests.

The VGG network family is mainly characterised by its simplicity, which uses 3x3 convolutional layers stacked on top of each other [29]. From this family, VGG19 was the pre-trained neural network evaluated.

DenseNet201 was also compared with the remaining architectures. DenseNet uses dense interlayer connections via Dense Blocks. Each layer receives extra inputs from all previous layers and passes its own features to all following layers.

InceptionResNetV2 is a convolutional neural architecture based on the Inception family of architectures which incorporates residual connections [30]. And finally, the Xception is also inspired by Inception architectures, but instead of using full convolutions, it replaces the standard Inception modules with depth wise separable convolutions [31].

In summary, eight neural networks were tested and compared:

1. Resnet50
2. EfficientNetV2L
3. EfficientNetB7
4. VGG19
5. DenseNet201
6. Xception
7. ResNet15V2
8. InceptionResNetV2

These network architectures were tested because they are all available in Keras applications and because of their good performance on general image classification problems. They usually have a strong capability of generalisation for images and problems outside the ImageNet dataset [32].

Additionally, multiple experiments were performed for each tested model architecture to test different parametrisation for the last dense layers placed in the network (in a transfer learning context).

3.2.2 Illegal Graffiti Detector

The goal of this model is to correctly identify the coordinates where illegal graffiti is located on a figure. As input, the model receives a picture, and as output, it returns the coordinates of the bounding boxes identified as graffiti locations.

Similarly to the classifier, different architectures were tested. These three architectures were tested as they were the supported architectures for the python library used, the *detecto*[3], this way, its implementation and evaluation were simpler and more straightforward.

The architectures tested were:

1. *Fasterrcnn resnet50 fpn*
2. *Fasterrcnn mobilenet v3 large fpn*
3. *Fasterrcnn mobilenet v3 large 320 fpn*

All tested architectures correspond to faster R-CNN architectures, short for region-based convolutional neural networks. Fast R-CNN tries to overcome some issues in R-CNN, one of them being, as the name suggests, its speed. As shown in Figure 6, these architectures are composed of two networks: a Region Proposal Network (RPN) and a fast R-CNN detector.

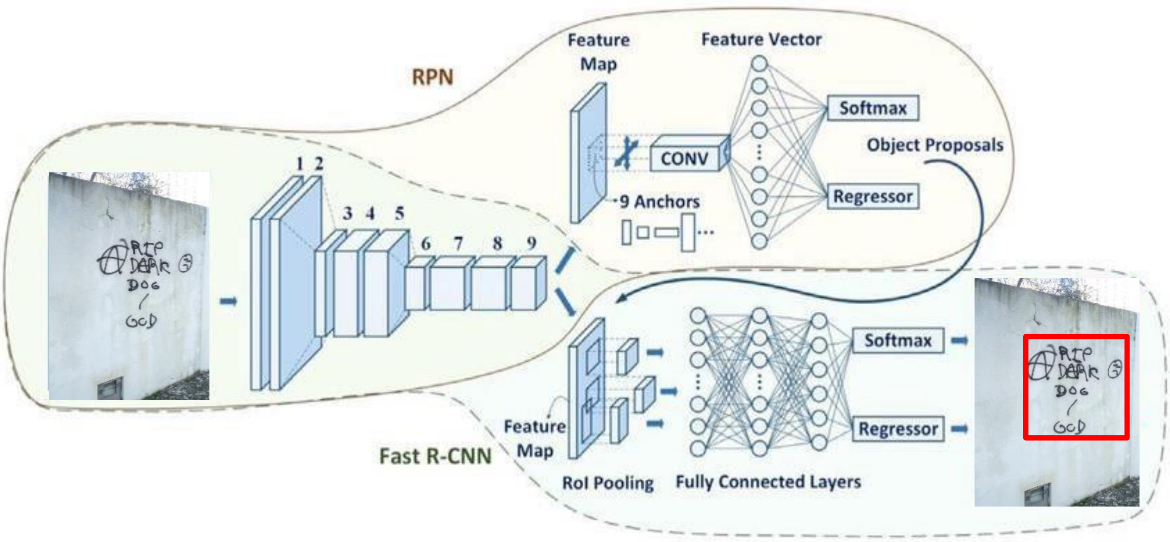

Figure 6. The architecture of Faster R-CNN (image adapted from [33])

[3] https://detecto.readthedocs.io/en/latest/index.html

As Ren describes in [33], a region proposal network "is a fully convolutional neural network that simultaneously predicts object boundaries and objectivity scores at each position". In other words, the purpose of the RPN is the generation of region proposals with various scales and aspect ratios that will be passed to the Fast R-CNN to guide it into where to look for the detection in the image.

Then, the Fast R-CNN detection network will implement object detection using the proposed regions. The output of the RPN, the feature map, is fed to a ROI Pooling layer that uses the max pooling operation on the RoI (Regions of Interest) to extract a fixed-length feature vector from each region proposal. This vector is then passed through Fully connected (FC) layers, and its output is split in two branches: 1) Softmax layer - to predict the class scores, 2) FC layer - to predict the bounding boxes and detected objects.

All the tested architectures use Feature Pyramid Networks (FPN), which, in short, is a feature extractor that generates multiple layers of feature maps with better quality information instead of just one [34]. The most important feature of this type of architecture is that, at each level of an image pyramid, it produces a multi-scale feature representation (as illustrated in Figure 7) which introduces more robustness to scale differences in the objects to be located. This feature improves accuracy and speed in most cases.

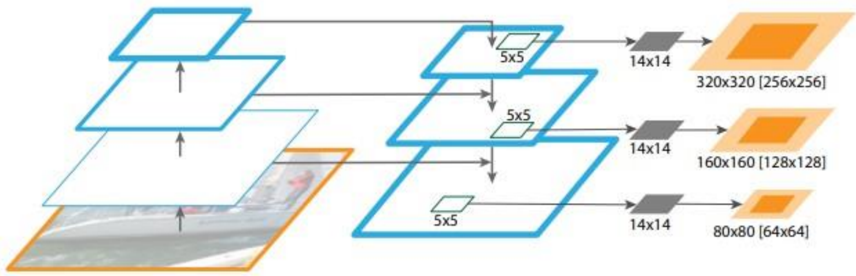

Figure 7. Illustration of the multi-layers feature maps [34]

Two different model architectures were tested: Resnet and MobileNet. Unlike ResNet, MobileNet are neural networks with a smaller size, lower latency, and lower power, hence they are considered suitable for mobile devices [35].

The weights used in either architecture come from the use of the COCO dataset[4]. COCO stands for 'Common Objects in Context' and is mainly used for object detection and segmentation due to its largescale labelled dataset.

[4] https://cocodataset.org

3.3. Data Description

This section describes the three sets of images used for training, evaluating, and testing the models: Images with illegal graffiti (mainly tags), pictures with street art graffiti, and images without graffiti. The selection of images in each class was established based on the images provided by the members of the Lisbon camera.

The first set of images (examples in Figure 8), with illegal graffiti, were collected by various means and sources, such as members of the City Council's urban hygiene and inspection teams or images submitted by Lisbon residents through the "Na minha rua" application[5]. This set of images was used for both proposes: illegal graffiti location and classification of graffiti into illegal or street art.

Figure 8. Examples of images with illegal graffiti

The second set of examples, depicted in Figure 9, contains street art graffiti images extracted from the urban art gallery website[6] from the Lisbon City Council. This set was used for the classification model.

[5] https://naminharualx.cm-lisboa.pt/
[6] http://gau.cm-lisboa.pt/galeria.html

Figure 9. Examples of images with street art graffitis

Lastly, the set of images without graffiti (examples in Figure 10) were obtained from two sources: some of them were downloaded from the internet, and others were provided by the Lisbon City Council and correspond to images captured after the removal of some graffiti from walls and streets in Lisbon.

Figure 10. Examples of images without graffiti

3.3.1. Data Preparation

Initially, due to the great diversity of types and formats of images shared by the Lisbon City Council, a great deal of work was required in filtering and processing the images to obtain a set of images suitable for the training of the initial graffiti detection model. From the initial set, several images were removed for the object detection model image set because they seemed to cause confusion and bias to the model, such as images where the graffiti delimitation was almost impossible, images with low quality, or with minimal graffiti hardly visible such as the examples shown in Figure 11.

Figure 11. Examples of images removed from the set of images used for the graffiti detection model

To standardise the types of images obtained, all pictures of the set were converted to .png before the labelling process.

Subsequently, the images containing illegal graffiti were labelled using the LabelImg tool[7]. LabelImg is a graphical image annotation tool that allows the definition of the bounding boxes referring to the graffiti and saves the annotations as XML files.

Figure 12 represents a labelled image after using the labelling tool.

Figure 12. Example of a labelled image

[7] https://github.com/heartexlabs/labelImg

The initial data set was split into three sets: 70% images for the training set, 15% for validation, and 15% for the test set.

Table 3 represents the number of images used to train, evaluate, and test each model:

Table 3. Number of images used for each model

Model type	Type of image	Number of images
Classification Model	Illegal graffiti	898
	Street art	898
	Without graffiti	341
	Total	**2137**
Detection Model	Illegal graffiti	639
	Total	**639**

3.3.2. Data Augmentation for Classification

In an effort to increase the accuracy of the graffiti classification task, data augmentation techniques were used. These techniques allow the creation of new images based on existing ones and thus increase the size of the data set and its diversity, this way, decreasing the chances of overfitting [36]. Two different types of data augmentation were used:

- Random Flip that flips the images horizontally.
- Random Rotation that rotates the image at 20 degrees.

Since graffiti can have various shapes and orientations, using new images from their rotations will increase diversity and generalise the problem regarding the position of the graffiti.

Figure 13 represents four outputs of the same image when the data augmentation techniques are used repeatedly.

Figure 13 . Example of the output of the data augmentation layers when running them repeatedly to the same image.

CHAPTER 4

Experimental Setup and Results

As mentioned earlier, after the image pre-processing and selection process, several experiments were conducted. Different models were trained to reach two reliable and practical models. One is to classify the image based on the type of graffiti present, and the other is to detect the coordinates of the graffiti in the image.

Since we have a pre-trained model, it isn't necessary to train the entire model. Only the final layers are trained with the images in question so that the model understands the specifics of the problem at hand.

4.1. Classification between Street Art and Illegal Graffiti

To identify the type of graffiti more quickly and therefore define the following steps, an image classification algorithm was developed to classify an image into the following categories: illegal graffiti, street art graffiti or no graffiti.

First, different pre-trained models available in Keras Applications were tested with similar conditions and parameters to understand which model better suited the problem. For all these first tests, only four new training layers were added to the end of the model in question (explained in session 3.2.1).

The models were trained for 50 epochs, with the Adam optimiser having a learning rate of 0.01. The *ReduceLROnPlateau* technique was used to reduce the learning rate when the validation loss stopped improving, i.e., when it reached a plateau.

Table 4 presents the classification metrics obtained for each tested model.

Pre-trained model	Accuracy	Precision (weighted)	Recall (weighted)	F1-score Street-art	F1-score Graffiti illegal	F1-score Without graffiti
Resnet50	0.715	0.763	0.758	0.833	0.750	0.581
EfficientNetV2L	0.680	0.728	0.697	0.769	0.715	0.513
EfficientNetB7	0.606	0.632	0.603	0.667	0.605	0.478
VGG19	0.686	0.713	0.700	0.790	0.678	0.551
DenseNet201	0.503	0.581	0.617	0.715	0.617	0.112
Xception	0.473	0.542	0,572	0.661	0.578	0.142
ResNet15V2	0.406	0.496	0.506	0.505	0.578	0.051
InceptionResNetV2	0.390	0.467	0.481	0.628	0.312	0.063

Table 4. Metrics from different classification models

According to Table 4, we can see that the Resnet50 architecture was the one that presented the best results overall and that InceptionResNetV2 presented the worst performance with only 39% accuracy.

It is also worth noting that, for any of the architectures, the results for the F1-score metric always presented worse results on the "without graffiti" class, most likely due to the dataset balance regarding the number of images used to train this class compared to the others, or due to the vast diversity of images that can be labelled as "without graffiti" (images of houses, buildings, roads, benches, walls with posters, streets, etc.). However, the “street art” class showed the best F1-score metric values in all architectures tested except for Resnet15V2.

Once the initial best architecture model was found, the Resnet50, a fine-tuning was performed to find the best parameters and optimisers for the classification. The fine-tuning was done using a python library called "Keras Tuner", which allows the definition of the hyperparameters and their values to be tested. These hyperparameter combinations are used for training and verifying which combination provides the best metrics.

The following Table 5 presents the different parameters tested.

Table 5. Hyperparameters tested in Keras tuner

Layer	Hyperparameter	Values Tested
Dense	Units	[8, 55, 150, 300, 500]
Dense	Activation Function	['linear', 'relu', 'tanh']
Dropout	Rate	[0.0, 0.15]
Optimiser	Learning Rate	[0.01, 0.001]

Among the combinations tested in the random search of the Keras tuner, the parametrisation that achieved the best results is the one presented in Table 6:

Table 6. Best Hyperparameters found

Layer	Hyperparameter	Best hyperparameters found
Dense	Units	55
Dense	Activation Function	‘tanh’
Dropout	Rate	0.15
Optimiser	Learning Rate	0.001

4.1.1 Results from the Classification Model

Figure 14 presents some cases correctly classified using the trained model.

True: street-art | Predicted: street-art

True: illegal graffiti | Predicted: illegal graffiti

True: without graffiti | Predicted: without graffiti

True: street-art | Predicted: street-art

True: illegal graffiti | Predicted: illegal graffiti

True: without graffiti | Predicted: without graffiti

Figure 14. Examples of correctly classified images

Two metrics were measured and tracked to evaluate the results: accuracy and categorical crossentropy. The accuracy calculates how often predictions are equal to the labels. The categorical crossentropy is one of the most commonly used functions for deep learning multi-class classification problems [37] because it computes the cross-entropy loss between the labels and predictions, i.e. it measures the difference between the two probability distributions (predicted and actual).

Figure 15 shows the accuracy and categorical cross entropy for the training and validation sets over the fifty epochs used for training.

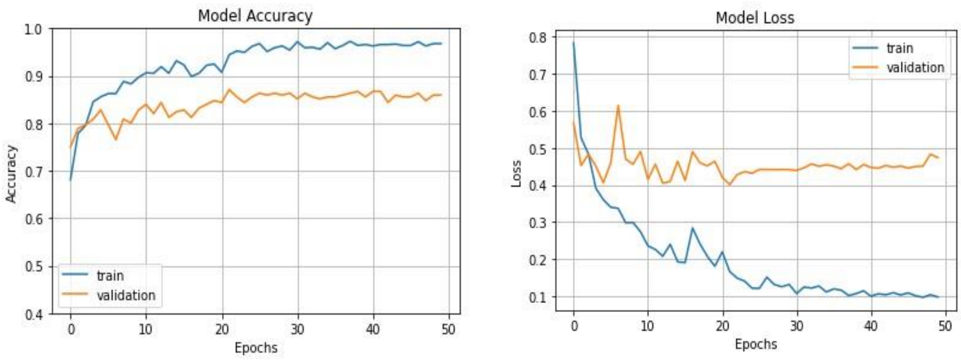

Figure 15. Accuracy and categorical cross entropy over the epochs

As shown in Figure 14, the metrics stabilize, and the best values for the validation set (lowest loss and highest accuracy) are observed at epoch 21. The weights corresponding to this epoch were saved and used for testing the model and calculating the following metrics.

After training the model, the model was tested with new images (the test set), and the following confusion matrix (Table 7) was obtained.

Table 7. Confusion matrix obtained with images from the test set

		Predicted Class		
		Illegal Graffiti	Street Art	Sem Graffiti
Actual Class	Illegal Graffiti	183	37	4
	Street Art	13	202	9
	Sem Graffiti	4	33	48

This matrix allows a deeper analysis of the type of errors made by the classifier and the number of incorrect images misclassified in each class. Table 8 shows the classification metrics obtained for the trained model.

Table 8. Classification metrics for the test set

Metric		**Value**
Accuracy		0.81
Balanced Accuracy		0.76
Precision	Street-art	0.92
	Illegal graffiti	0.74
	Without graffiti	0.79
Recall	Street-art	0.82
	Illegal graffiti	0.90
	Without graffiti	0.56
F1-score	Street-art	0.86
	Illegal graffiti	0.81
	Without graffiti	0.66

Overall, the model achieved an accuracy of 81.4%. However, since the data set is imbalanced due to the much lower number of images without graffiti, the balanced accuracy, which corresponds to the average recall obtained in each class, has slightly decreased to 77.2%.

The precision explains how many images predicted with a positive class were correctly classified. This metric indicates that the class 'illegal graffiti' is the one that is more times mispredicted, and the class with the lowest false positive rate is 'street art'.

The recall metric corresponds to the ratio between correct predicted positive observations by the total number of observations. This metric indicates that from all the images without graffiti, only 56% were correctly classified. The pictures with illegal graffiti have the highest proportion of true positives.

However, the metric that considers false positives and false negatives is the F1-score, the harmonic mean of precision and recall. We can see that the class with the best predictions overall is 'street art', and the worst are the images without graffiti, most likely due to the smaller number of pictures used without graffiti in the model training.

4.1.2 Image Classification Limitations

Due to the subjectivity concerning the differences between illegal graffiti and street art, this distinction sometimes becomes a bit blurred and contradictory. Sometimes the distinction between these two groups of graffiti is already difficult to classify by a human. Usually, it depends on the person and their idea of art, which can vary considerably.

Some images belonging to the set of street art (for example, the ones in Figure 16), taken from the website of the urban art gallery, may raise some doubt due to their resemblance to some images defined as illegal graffiti, as in the examples displayed in Figure 17.

Figure 16. Images labelled as street art

Figure 17. Images labeled as illegal graffiti

Since the model is trained based on the classes of these images, it is expected that there will be some misclassifications in some cases, as in Figure 18.

True: Street art | Predicted: Graffiti illegal

Figure 18. Misclassified graffiti due to similarities between the two classes

Then there are other images, such as those in Figure 19, containing street art elements and illegal graffiti. However, since they were images taken from the Lisbon urban art gallery, they were classified as street art.

Figure 19. Images classified as street art that also contain illegal graffiti

This leads to some images being incorrectly classified (or not, because they also contain illegal graffiti), as in Figure 20.

True: Street art | Predicted: Graffiti illegal

True: Street art | Predicted: Graffiti illegal

Figure 20. Images incorrectly classified as containing more than one type of graffiti

There are also other cases where the image has a significant amplitude, and the graffiti is at a considerable distance, and therefore may go unnoticed, as the pictures in Figure 21, that result in errors such as the one in Figure 22.

Figure 21. Examples of pictures where the graffiti is at a large distance

True: Street art |
Predicted: Without graffiti

Figure 22. Example of an incorrect classification

Furthermore, the relevant difference between the number of images containing graffiti (street art and illegal) and without graffiti is quite significant, as displayed in Table 3, which may justify the differences between the performances obtained for these classes.

4.2. Illegal Graffiti Detector

In cases where the classifier's output is 'illegal graffiti', the image goes through a new model, but in this case, to detect the coordinates of the graffiti in the picture. This model gives an idea of the feasibility of a graffiti detector through videos taken around the city.

Thus, the goal is to obtain a model that can identify, accurately as possible, the coordinates of the graffiti present in an image. For this purpose, a python library called 'detecto'[8], created on top of PyTorch, was used to build a graffiti detection model.

Three different Faster R-CNN model architectures (*faster cnn resnet50 fpn*, *faster cnn mobilenet v3 fpn*, *faster cnn mobilenet v3 large 320 fpn*) were tested for different hyperparameters.

Table 9 presents the parameters that retrieved the minimum validation loss.

Table 9. Values for the hyperparameters tried

Hyperparameter	Value
Model architecture	*faster cnn resnet50 fpn*
Learning Rate	0.005
Momentum	0.7
Weight Decay	0.001
Gamma	0.3
Learning Rate Step Size	3

[8] https://detecto.readthedocs.io/en/latest/

4.2.1. Bounding Box Predictions Post-processing

After the identification of the graffiti coordinates in the image, in the cases where more than one bounding box is identified in the same image, it is necessary to check if the bounding boxes need to be grouped.

It was found that the model often separates graffiti in more than one bounding box, thus decreasing the intersection area with the coordinates of the actual bounding boxes, even if the prediction is correct. To tackle this, the bounding boxes are grouped when the intersection area between two bounding boxes is greater than or equal to 0.35 times the area of one of the two bounding boxes in question.

For example, Figure 23 exemplifies a case where it was necessary to combine two bounding boxes because the model detected two different bounding boxes for the same graffiti. As these overlapped, they were merged.

Figure 23. An example of an image where the joining of bounding boxes was necessary

4.2.2. Results from the Graffiti Detection Model

Figure 24 displays some successful graffiti detections through the model.

Figure 2 4. Examples of graffiti detections

The Intersection over Union (IoU) was used to evaluate the model's performance. This metric is used to measure the accuracy of an object detector because it is calculated by dividing the overlap area by the union area between the predicted bounding box and the ground-truth bounding box.

The closer this metric is to 1, the greater the overlap between the prediction and the actual coordinates of the object. Therefore, the better the prediction and hence the model are, as exemplified in Figure 25.

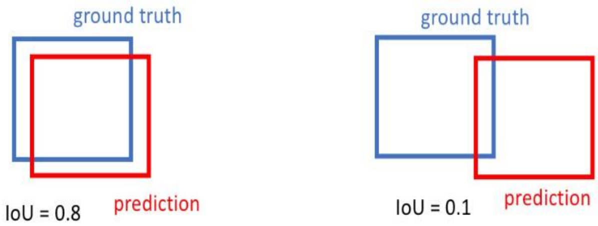

Figure 25. Ilustractive figure to explain IoU metric

As shown in Table 10, the IoU reported a value greater than 70% for all image sets.

Table 10. IoU for each set of images

Image set	Intersection over Union
Train set	0.897
Validation set	0.721
Test set	0.703

4.2.3. Graffiti Detection Limitations

In Figure 26, it is possible to observe some images with low IoU scores. Sometimes, it is due to the presence of words or posters in the picture that are easily confused with possible graffiti. Others correspond to the false detection of objects (usually when they present a more irregular shape) or images with low quality or with more distant graffiti.

Figure 26. Graffiti detections with low IoU

CHAPTER 5

Conclusions

This dissertation proposed a machine learning-based graffiti identifier on images. It has been developed to support the process, currently done manually by members of the Lisbon City Council, of identifying areas that need to be painted due to the presence of illegal graffiti. Furthermore, it also identifies graffiti that may be considered as street art and, consequently, a potential cultural asset to the city.

The diagram depicted in Figure 27 synthesises the workflow followed in this Dissertation. It reached the most suitable models for the problem previously described within those tested with the help of transfer learning and data augmentation.

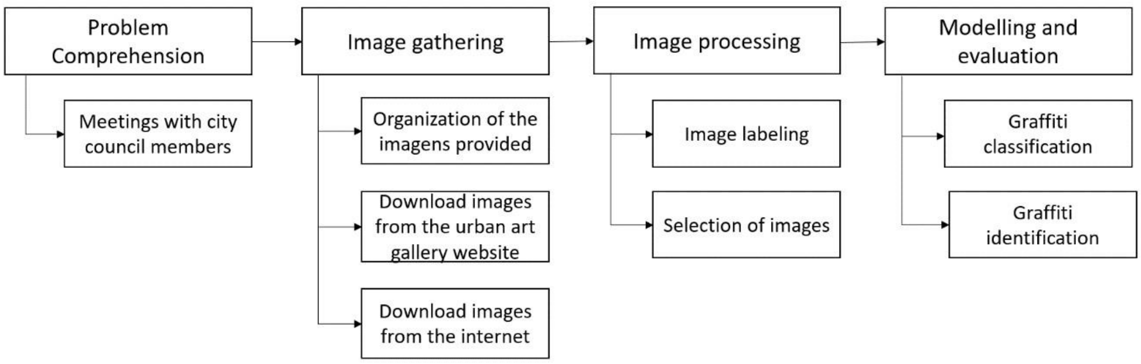

Figure 27. Pipeline implemented for reaching the most suitable models

The workflow allowed the development and training of two models: 1) an image classification model that classifies an image according to three classes: street art, illegal graffiti, and no graffiti classes; and 2) a graffiti identification model that provides the location of illegal graffiti identified by the classification model. For the image classification problem, Resnet50 showed the best results, presenting an overall accuracy of 81% and an F1-score of 86%, 81%, and 66% for the street art, illegal graffiti, and no graffiti classes, respectively. While for the illegal graffiti identification model, the best performance was obtained using a Faster R-CNN architecture resnet50fnn, getting 0.703 of the IoU in the test set.

Answering the first research question elaborated at the beginning of the dissertation, *Can a deep learning model successfully discriminate the differences between street art and illegal?,* the results showed that it was possible to implement a deep learning architecture good enough to discriminate the differences between street art and illegal graffiti. Among the 435 images labelled as street art or illegal graffiti, only 11% were misclassified within the opposite graffiti class.

However, this leads us to the first concerning point of this thesis. In the images taken from the Lisbon Urban Art Gallery website, there, are images with street art designation very similar to some images provided by the Lisbon City Council, defined as illegal graffiti. In other words, the image database used for training and testing the models contains images that a human himself may have difficulties defining as street art or illegal graffiti. Furthermore, it can depend a lot on the individual's likes and personality, and for this reason, there can be different opinions from different people. Since this is sometimes a complex problem in some cases for humans, it is natural and predictable that it will also be challenging for a model to distinguish between the two classes in some images

Regarding the second research question, *How accurate is the automatic identification and location of illegal graffiti on images acquired in the streets of Lisbon under loose controlled conditions?* the IoU metric, which measures the degree of overlap between the ground-truth bounding boxes (defined in the labelling process) and the predicted ones, achieved a value of about 70%, using randomly chosen images not used in the training process and containing a wide diversity of objects and locations.

As for limitations, the tested and implemented classification models showed poorer results when dealing with images without any graffiti, presenting an F1-score of only 66%. However, this behaviour is most likely due to the small number of images used during the training process and the great variety among them. As for the illegal graffiti detection model, it was observed that sometimes objects with more irregular shapes or images with phrases or words were misinterpreted as graffiti. Also, lowerquality images were prone to flaws in the identification of graffiti.

Nevertheless, based on the results and metrics calculated and mentioned above, this Dissertation demonstrates that it is safe and feasible to invest in a system capable of automatically detecting, in real-time, the places that need painting for illegal graffiti removal and to identify new works of street art through the use of cameras placed in cars navigating the city. For this, it would be necessary to adapt the system to use videos instead of images.

As for future work, a different approach could be applied, using a detection model to detect the existence of any graffiti on an image. If a graffiti is detected, the image could be cropped based on the coordinates identified as graffiti and only then go through the classification algorithm to distinguish between street art and illegal graffiti. This way, more than one type of graffiti could be detected in the same image, which is currently a limitation since an image can only be classified with a single class. Furthermore, if the workflow remains the same as the one described in this dissertation, it would be important to include additional images for the training of the models, mainly images without any graffiti, to try to improve the metrics obtained regarding the classification of these images. Another change that would improve the results obtained with the graffiti detector is to use a labelling tool that allows the generation of irregular bounding boxes instead of rectangular ones since graffiti can have very different and unregular shapes.

It is also important to develop an application that allows the City Council employees to use the models in an easier and more user-friendly way, preferably with the possibility of running several images simultaneously.

Bibliographic References

[1] "When does graffiti become heritage | Audio Trails." http://audiotrails.co.uk/whendoes-graffiti-become-heritage/ (accessed Oct. 24, 2021).

[2] S. Bacharach, "Street Art and Consent," *British Journal of Aesthetics*, vol. 55, pp. 481–495, 2015, doi: 10.1093/aesthj/ayv030.

[3] J. R. Alzate, M. S. Tabares, and P. Vallejo, "Graffiti and government in smart cities: A deep learning approach applied to Medellín city, Colombia," *ACM International Conference Proceeding Series*, pp. 160–165, 2021, doi: 10.1145/3460620.3460749.

[4] J. I. Ross and B. S. Wright, "'I've Got Better Things to Worry About': Police Perceptions of Graffiti and Street Art in a Large Mid-Atlantic City," *Police Q*, vol. 17, no. 2, pp. 176–200, 2014, doi: 10.1177/1098611114531491.

[5] R. Campos, "Urban art in Lisbon: opportunities, tensions and paradoxes," *Cultural Trends*, vol. 30, no. 2, pp. 139–155, 2021, doi: 10.1080/09548963.2021.1897779.

[6] "GAU: Onde pintar." http://gau.cm-lisboa.pt/onde-pintar.html (accessed Oct. 03, 2022).

[7] T. Novack, L. Vorbeck, H. Lorei, and A. Zipf, "Towards detecting building facades with graffiti artwork based on street view images," *ISPRS Int J Geoinf*, vol. 9, no. 2, 2020, doi: 10.3390/ijgi9020098.

[8] D. Angiati, G. Gera, S. Piva, and C. S. Regazzoni, "A novel method for graffiti detection using change detection algorithm," *IEEE International Conference on Advanced Video and Signal Based Surveillance - Proceedings of AVSS 2005*, vol. 2005, pp. 242–246, 2005, doi: 10.1109/AVSS.2005.1577274.

[9] S. Wang, J. Z. Gao, W. Li, Y. Li, K. Wang, and S. Lu, "Building smart city drone for graffiti detection and clean-up," *Proceedings - 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People and Smart City Innovation, SmartWorld/UIC/ATC/SCALCOM/IOP/SCI 2019*, pp. 1922–1928, 2019, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00337.

[10] F. Tombari, L. di Stefano, S. Mattoccia, and A. Zanetti, "Graffiti detection using a timeof-flight camera," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5259 LNCS, pp. 645– 654, 2008, doi: 10.1007/978-3-540-88458-3_58.

[11] L. di Stefano *et al.*, "Graffiti Detection Using Two Views To cite this version : HAL Id : inria-00325654 Graf fi ti Detection Using Two Views," 2008.

[12] J. Capucho, “Lisboa gasta um milhão por ano para limpar graffiti,” *Diário de Notícias*, 2015. https://www.dn.pt/sociedade/lisboa-gasta-um-milhao-por-ano-para-limpargraffiti-4922699.html (accessed Oct. 21, 2021).

[13] “Gastos com graffitis numa década davam para CP comprar um comboio novo.” https://www.dinheirovivo.pt/empresas/gastos-com-graffitis-numa-decada-davampara-cp-comprar-um-comboio-novo-12695179.html (accessed Oct. 21, 2021).

[14] R. Wirth and J. Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining”.

[15] P. Nahar *et al.*, “Autonomous UAV forced graffiti detection and removal system based on machine learning,” *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017 -* , pp. 1–8, 2018, doi: 10.1109/UIC-ATC.2017.8397582.

[16] E. K. Tokuda, R. M. Cesar, and C. T. Silva, “Quantifying the Presence of Graffiti in Urban Environments,” *2019 IEEE International Conference on Big Data and Smart Computing, BigComp 2019 - Proceedings*, pp. 1–4, 2019, doi: 10.1109/BIGCOMP.2019.8679113.

[17] G. R. Munsberg, P. Ballester, M. F. Birck, U. B. Correa, V. O. Andersson, and R. M. Araujo, “Towards Graffiti Classification in Weakly Labeled Images Using Convolutional Neural, Networks,” *Communications in Computer and Information Science*, vol. 720, pp. 39–48, 2017, doi: 10.1007/978-3-319-71011-2_4.

[18] H. Li, J. Kim, and E. J. Delp, “Deep gang graffiti component analysis,” *IS and T International Symposium on Electronic Imaging Science and Technology*, pp. 1–4, 2018, doi: 10.2352/ISSN.2470-1173.2018.15.COIMG-201.

[19] W. Tong, J. E. Lee, R. Jin, and A. K. Jain, “Gang and moniker identification by graffiti matching,” *MM’11 - Proceedings of the 2011 ACM Multimedia Conference and CoLocated Workshops - Multimedia in Forensics and Intelligence Workshop, MiFor’11*, pp. 1–6, 2011, doi: 10.1145/2072521.2072523.

[20] A. Parra, M. Boutin, and E. J. Delp, “Location-aware gang graffiti acquisition and browsing on a mobile device,” *Multimedia on Mobile Devices 2012; and Multimedia Content Access: Algorithms and Systems VI*, vol. 8304, no. Cil, p. 830402, 2012, doi: 10.1117/12.906469.

[21] C. Yang, P. C. Wong, W. Ribarsky, and J. Fan, “Efficient graffiti image retrieval,” *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, ICMR 2012*, 2012, doi: 10.1145/2324796.2324840.

[22] A. Parra, B. Zhao, J. Kim, and E. J. Delp, “Recognition, segmentation and retrieval of gang graffiti images on a mobile device,” *2013 IEEE International Conference on Technologies for Homeland Security, HST 2013*, pp. 178–183, 2013, doi: 10.1109/THS.2013.6698996.

[23] A. K. Jain, J. E. Lee, and R. Jin, “Graffiti-ID: Matching and retrieval of graffiti images,” *1st ACM Workshop on Multimedia in Forensics - MiFor’09, Co-located with the 2009 ACM International Conference on Multimedia, MM’09*, pp. 1–6, 2009, doi: 10.1145/1631081.1631083.

[24] R. Patel and S. Patel, “A comprehensive study of applying convolutional neural network for computer vision,” *International Journal of Advanced Science and Technology*, vol. 29, no. 6 Special Issue, pp. 2161–2174, 2020.

[25] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” *Comput Intell Neurosci*, vol. 2018, 2018, doi: 10.1155/2018/7068349.

[26] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, *A survey of transfer learning*, vol. 3, no. 1. Springer International Publishing, 2016. doi: 10.1186/s40537-016-0043-6.

[27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, Accessed: Oct. 23, 2022. [Online]. Available: http://imagenet.org/challenges/LSVRC/2015/

[28] M. Tan and Q. v Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”.

[29] K. Simonyan and A. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGESCALE IMAGE RECOGNITION,” 2015, Accessed: Oct. 23, 2022. [Online]. Available: http://www.robots.ox.ac.uk/

[30] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.” [Online]. Available: www.aaai.org

[31] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions.”

[32] “ImageNet: VGGNet, ResNet, Inception, and Xception with Keras - PyImageSearch.” https://pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xceptionkeras/ (accessed Oct. 23, 2022).

[33] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Accessed: Nov. 01, 2022. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/results

[34] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection”.

[35] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”.

[36] J. Wang and L. Perez, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”.

[37] E. Gordon-Rodriguez, G. Loaiza-Ganem, G. Pleiss, and J. P. Cunningham, “Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning”, Accessed: Oct. 11, 2022. [Online]. Available: https://github.com/cunningham-lab/cb_and_cc.