



Article

Deep Learning-Based Graffiti Detection: A Study Using Images from the Streets of Lisbon

Joana Fogaça ¹, Tomás Brandão ¹  and João C. Ferreira ^{1,2,*} ¹ Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal² Inov Inesc Inovação Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal* Correspondence: jcafa@iscte-iul.pt

Abstract: This research work comes from a real problem from Lisbon City Council that was 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, with a georeferenced position. We contribute also a city graffiti database to share among the scientific community. 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. 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 of 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.

Keywords: graffiti; street art; classification; detection; computer vision



Citation: Fogaça, J.; Brandão, T.; Ferreira, J.C. Deep Learning-Based Graffiti Detection: A Study Using Images from the Streets of Lisbon. *Appl. Sci.* **2023**, *13*, 2249. <https://doi.org/10.3390/app13042249>

Academic Editor: Corrado Santoro

Received: 10 January 2023

Revised: 3 February 2023

Accepted: 7 February 2023

Published: 9 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 where the artists can apply to create street art, encouraging the creation of more of these works by making them legal and publishing them on 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], have already begun to focus on identifying and analyzing 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].

To try to control and minimize 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 research work 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.

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 of 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 research work developed a contribution/proof of concept to understand the viability of the identification and classification of graffiti and, in 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 existing graffiti in the city of Lisbon.

Our major contributions are (1) a deep learning model that successfully discriminates the differences between street art and illegal; and (2) determining how accurate the automatic identification and location of illegal graffiti on images acquired in the streets of Lisbon under loose controlled conditions.

The system development methodology was based on the Cross Industry Standard Process 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 [12].

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 consist of images, which require a different kind of processing and collection. These adjustments can be observed in Figure 3.

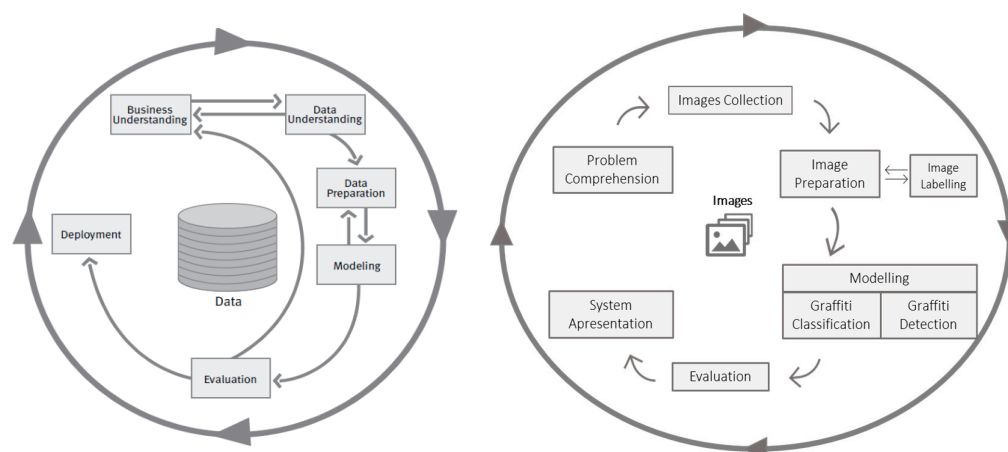


Figure 3. CRISP-DM (left) and CRISP-DM proposed adjustments (right).

2. Literature Review

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/analyzing 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 analyzing 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, 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 analyzing visible changes due to variations of the appearance, also analyzing 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 64×64 pixels for both intensity and depth images, which did not allow a distance of more than 1.5/2 m between the camera and the wall under test. However, the experimental results of [11] allowed us 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. [13] 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 MobileNetV2 transfer learning model pre-trained on the COCO dataset 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 recognizes 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 [14], 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 ResNet101 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 optimization 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. [15] 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 was considered art or not, such as the approach applied to Medellín City [3] that used the PyTorch library to implement an R-CNN, the ResNet50 classifier, already pre-trained on the ImageNet dataset. This research allowed the construction of a visualization tool through heat maps that, besides helping define measures to improve sectorial policies, also allow 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 [16–19].

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 [20] was a research project conducted at Michigan State University that aimed 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 was based on calculating the similarities via Euclidean distances between the critical points of the two images. Local geometric constraints were added to try to reduce false associations.

A CNN model was adopted by He Li et al. [16] 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. [21] proposed a model-based system that, by analyzing graffiti images, can present relevant information about the gangs associated with the graffiti. This system is composed of three methods: color recognition (taking advantage of the capabilities of the modern mobile device's touch screen and the user's help to trace the path along

the color region), segmentation of the color 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 quantization, thus making the comparison dramatically less expensive than previous methods in the literature” [21], 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	Percentage 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’ behaviors, locations, roots, and ambitions, since the number of gang-related crimes has increased in the US [21]. None of the selected articles were written by Portuguese authors.

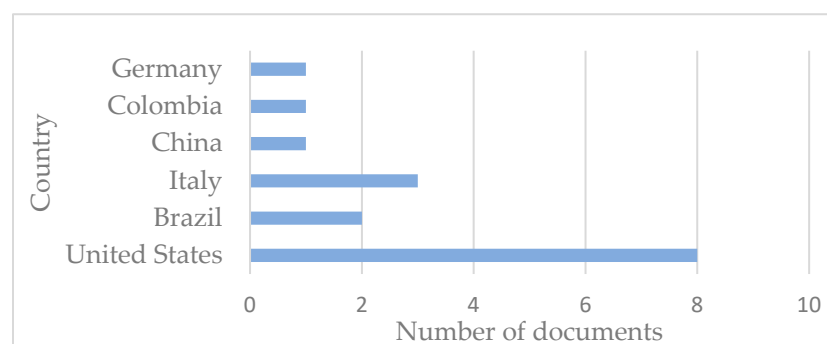


Figure 4. Number of documents by country.

Although the detection and classification of graffiti is a topic that can help not only to minimize the damage related to illegal graffiti but also to spread the benefits of urban art, the literature review allowed us to realize 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 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 minimize the costs related to the issue has become more and more certain.

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

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 [18,21].

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.

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 [22,23].

A total of 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 [24].

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, publicizing 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 us 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 and 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.

Considering this literature review, we raise two research questions: RQ1—“can a deep learning model successfully discriminate the differences between street art and illegal graffiti?” and RQ2—“how accurate is the automatic identification and location of illegal graffiti on images acquired in the streets of Lisbon under loose controlled conditions?” In addition, we label a graffiti dataset, which we have provided for the scientific community on the request of a graffiti image dataset.

3. Graffiti Identifier

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

Initially, the images go through an image classification model that tries to identify the type of graffiti in 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 in 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.

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. 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. Figure 5 shows a diagram of the developed system.

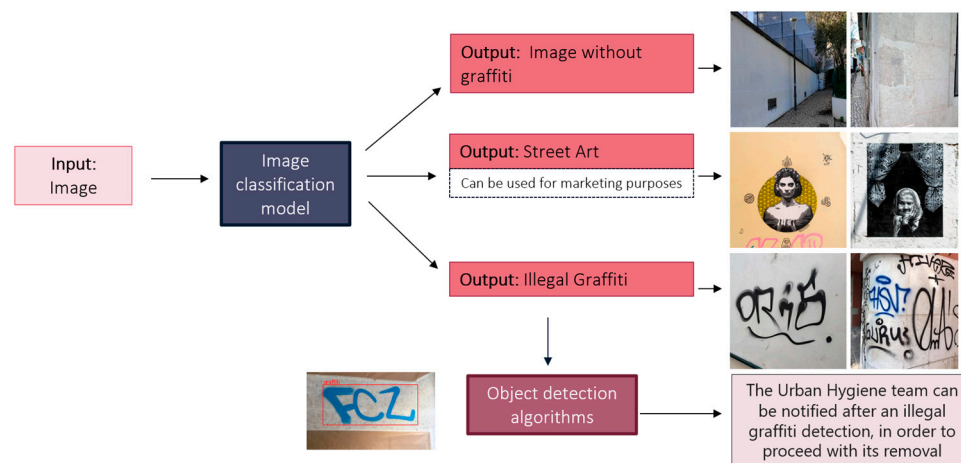


Figure 5. Summary of the implemented system.

3.1. 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 it 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 the output to be converted into a probability distribution over the predicted output classes: illegal, street art, or without graffiti.

The weights used in the tested architectures were obtained using the ImageNet dataset, a large dataset organized according to the WordNet hierarchy, comprising over 14 million images categorized into about 22 thousand different object categories. Although the images included here exhibit considerable differences relative to the images used in this research work (regarding graffiti), pre-trained networks with weights optimized for this large dataset can be useful as feature extractors. For 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 research work 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, as a result of the vanishing gradient, as the number of layers in the neural network increases, the accuracy becomes saturated and starts to degrade after a certain point. These residual blocks behave as shortcut connections that perform identity mapping [25]. 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 [26]. The EfficientNetV2L and EfficientNetB7 architectures were used in the tests.

The VGG network family is mainly characterized by its simplicity, which uses 3×3 convolutional layers stacked on top of each other [27]. 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 [28]. Finally, Xception is also inspired by Inception architectures, but, instead of using full convolutions, it replaces the standard Inception modules with depth-wise separable convolutions [29].

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 generalization for images and problems outside the ImageNet dataset [30].

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

3.2. Illegal Graffiti Detector

The goal of this model is to correctly identify the coordinates where illegal graffiti is located in 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*. This way, its implementation and evaluation were simpler and more straightforward.

The architectures tested were:

1. Faster R-CNN ResNet50 FPN
2. Faster R-CNN MobileNetV3 large FPN
3. Faster R-CNN MobileNetV3 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. These architectures are composed of two networks: a Region Proposal Network (RPN) and a Fast R-CNN detector.

As Ren describes in [31], 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 Faster R-CNN detection network will implement object detection using the proposed regions. The output of the RPN, the feature map, is fed to an 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 into two branches: (1) the SoftMax layer—to predict the class scores and (2) the FC layer—to predict the bounding boxes and detected objects.

All the tested architectures use Feature Pyramid Networks (FPN), which, in short, are feature extractors that generate multiple layers of feature maps with better quality information instead of just one [32]. 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, which introduces more robustness to scale differences in the objects to be located. This feature improves accuracy and speed in most cases.

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 [33].

The weights used in either architecture come from the use of the COCO dataset. COCO stands for ‘Common Objects in Context’ and is mainly used for object detection and segmentation due to its large-scale labeled dataset.

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 6), 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. This set of images was used for both purposes: illegal graffiti location and classification of graffiti into illegal or street art.

The second set of examples, depicted in Figure 7, contains street art graffiti images extracted from the urban art gallery website from the Lisbon City Council. This set was used for the classification model.



Figure 6. Dataset samples of images with illegal graffiti.

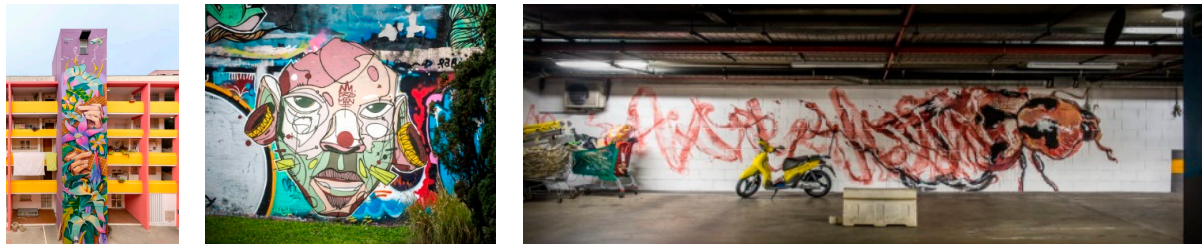


Figure 7. Dataset samples of images with street art graffiti.

Lastly, the set of images without graffiti (examples in Figure 8) 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 8. Dataset samples of images without graffiti.

3.4. 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 images with minimal graffiti hardly visible, such as the examples shown in Figure 9.



Figure 9. Examples of images removed from the dataset used for the graffiti detection model.

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

Subsequently, the images containing illegal graffiti were labeled using the LabelImg tool. 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 10 represents a labeled image after using the labeling tool. The initial dataset was split into three sets: 70% images for the training set, 15% for validation, and 15% for the test set.



Figure 10. Example of a labeled image.

3.5. 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 dataset and its diversity, this way, decreasing the chances of overfitting [34].

Two different types of data augmentation were used: horizontal flips and random rotations of 20 degrees. Since graffiti can have various shapes and orientations, using new images from their rotations will increase diversity and generalize the problem regarding the position of the graffiti. Figure 11 represents four outputs of the same image when the data augmentation techniques are used repeatedly.



Figure 11. Example of the output of the data augmentation layers when running them repeatedly to the same image. The first image is the original one.

Table 3 represents the number of images used to train, evaluate, and test each model after increasing the number of images in the dataset with the augmented versions.

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

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 is not 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 Section 3.2).

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

Table 4. Performance metrics achieved with different classification models.

Pre-Trained Model	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score		
				Street Art	Illegal Graffiti	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

It is also worth noting that, for any of the architectures, the results for the F1-score metric always presented worse results in 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 labeled 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. Table 4 presents the classification metrics obtained for each tested model.

Once the initial best architecture model was found, the Resnet50, a fine-tuning was performed to find the best parameters and optimizers 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.

Table 5 presents the different parameters tested. Among the combinations tested in the random search of the Keras tuner, the parametrization that achieved the best results is the one presented in Table 6.

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]
Optimizer	Learning Rate	[0.01, 0.001]

Table 6. Best hyperparameters found.

Layer	Hyperparameter	Best Hyperparameters Found
Dense	Units	55
Dense	Activation Function	'tanh'
Dropout	Rate	0.15
Optimizer	Learning Rate	0.001

4.1.1. Results from the Classification Model

Figure 12 presents some cases correctly classified using the trained model. Two metrics were measured and tracked to evaluate the results: accuracy and categorical cross-entropy. The accuracy calculates how often predictions are equal to the labels. The categorical cross-entropy is one of the most commonly used functions for deep learning multi-class classification problems [35] 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 12. Examples of correctly classified images.

Figure 13 shows the accuracy and categorical cross entropy for the training and validation sets over the fifty epochs used for training. As shown in the figure, 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.

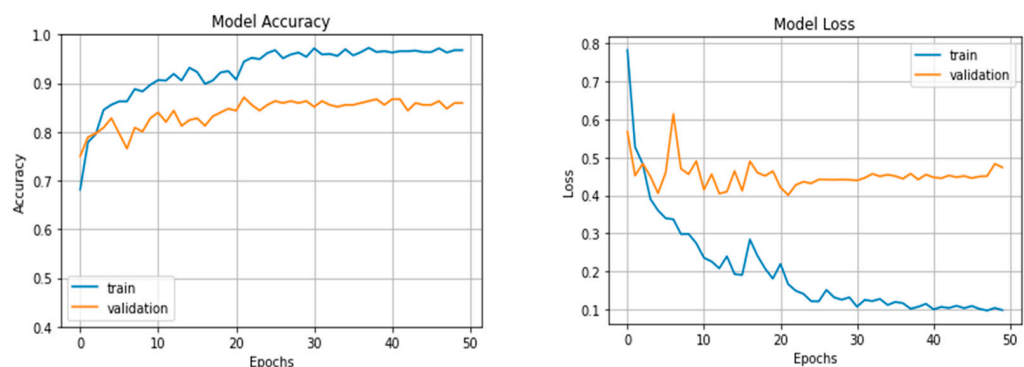


Figure 13. Accuracy and categorical cross-entropy over the epochs.

After training the model, the model was tested with new images (the test set), and the confusion matrix in Table 7 was obtained. This matrix allows a deeper analysis of the types of errors made by the classifier and the number of incorrect images misclassified in each class.

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

		Predicted Class		
		Illegal Graffiti	Street Art	W/o Graffiti
Actual Class	Illegal Graffiti	183	37	4
	Street Art	13	202	9
	W/o Graffiti	4	33	48

Table 8 shows the classification metrics obtained for the trained model. Overall, the model achieved an accuracy of 81.4%. However, since the dataset 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%.

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

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 and 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 14), 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 15.

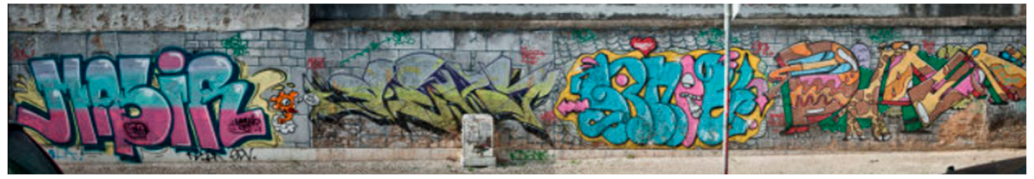


Figure 14. Images labeled as street art.



Figure 15. 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 16 (left image). Other images, such as those in Figure 16 (central and right image), contain street art elements and illegal graffiti. However, since they were images taken from the Lisbon urban art gallery, they were classified as street art.

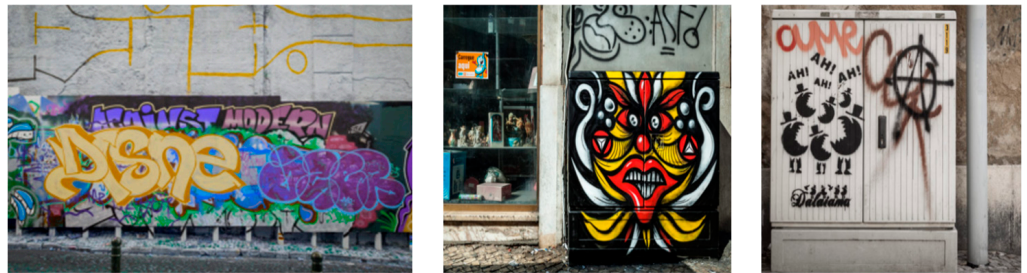


Figure 16. Misclassified graffiti due to similarities between the two classes (left) and due to images also containing illegal graffiti ((central) and (right) images).

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 17.



Figure 17. Examples of pictures where the graffiti is at a large distance.

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, as accurately as possible, the coordinates of the graffiti present in an image. For this purpose, a python library called "detecto", created on top of PyTorch, was used to build a graffiti detection model.

Three different Faster R-CNN model architectures (Faster R-CNN ResNet50 FPN, Faster R-CNN MobileNetV3 FPN, and Faster R-CNN MobileNetV3 large 320 FPN) were tested for different hyperparameters. Table 9 presents the parameters that retrieved the minimum validation loss.

Table 9. Values for the optimized hyperparameters.

Hyperparameter	Value
Model Architecture	Faster R-CNN ResNet50 FPN
Learning Rate	0.005
Momentum	0.7
Weight Decay	0.001
Gamma	0.3
Learning Rate Step Size	3

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 into 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 18 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 18. An example of an image where the joining of bounding boxes was necessary.

4.2.2. Results from the Graffiti Detection Model

Figure 19 displays some successful graffiti detections through the model.

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.



Figure 19. Examples of graffiti detections.

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

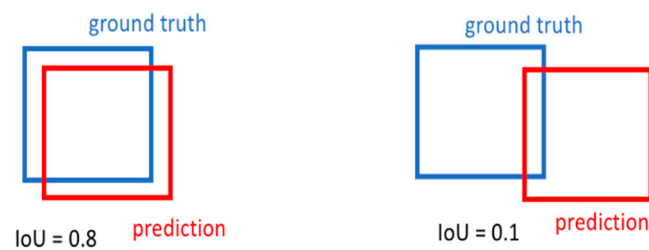


Figure 20. Illustrative figure to explain the IoU metric.

Table 10. IoU for each set of images.

Image Set	Average IoU
Train set	0.897
Validation set	0.721
Test set	0.703

4.2.3. Graffiti Detection Limitations

In Figure 21, 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 21. Graffiti detections with low IoU.

The diagram depicted in Figure 22 synthesizes the workflow followed in this research work. It reached the most suitable models for the problem previously described within those tested with the help of transfer learning and data augmentation.

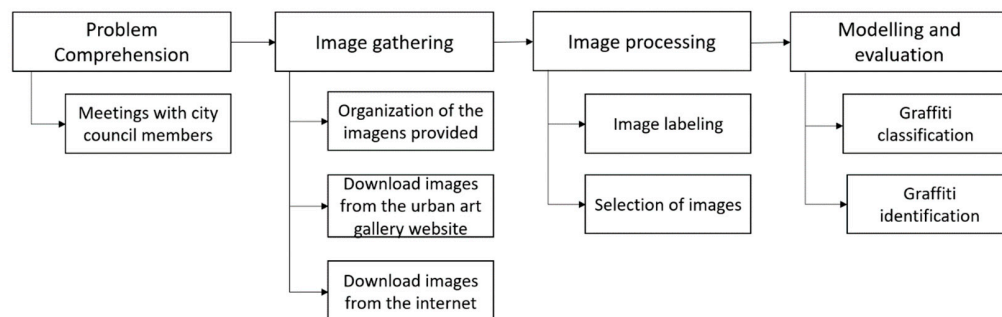


Figure 22. Pipeline implemented for reaching the most suitable models.

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. Meanwhile, for the illegal graffiti identification model, the best performance was obtained using a Faster R-CNN architecture ResNet50 FPN, having 0.703 of the IoU in the test set.

5. Conclusions

This research work proposed a machine learning-based prototype to automatically detect, classify, and annotate graffiti from the input images. It can support the process, currently done manually by the Lisbon City Council's staff, of identifying walls that require painted for removing illegal graffiti. Furthermore, it also identifies graffiti that may be considered as street art and, consequently, a potential cultural asset to the city.

Answering the first research question elaborated, “can a deep learning model successfully discriminate the differences between street art and illegal graffiti?”, the results showed that the proposed deep learning-based architecture successfully discriminates the differences between street art and illegal graffiti. Among the 435 images in the test set, an accuracy of 89% has been achieved. Furthermore, we observed that most of the misclassified images were illegal graffiti whose characteristics were quite similar to other street art images registered on the Lisbon Urban Art Gallery website. In other words, the image databases used for model training and testing contain images that a human himself may have difficulties in distinguishing between the street art and the illegal graffiti cases. 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 of the illegal graffiti object detector achieved an average value of about 70%, using test images containing a wide diversity of graffiti samples and locations.

As for the limitations of our study, all tested image classification models showed poor performance when dealing with images without the presence of graffiti. The best configuration achieved an F1-score of 66%. This behavior is most likely due to the relatively small number of images without the presence of graffiti used during the training process and the great variety among them. For the illegal graffiti detection model, it was verified that its performance was negatively affected by spurious detections due to objects whose patterns somewhat resemble illegal graffiti tags, such as words on billboards

and posters. In addition, lower-quality images on the dataset were prone to flaws in the detection of graffiti.

As for further research, a different approach could be followed by using an integrated solution for a street art and illegal graffiti object detection model (instead of using a separate image classification model for this task). Regardless of the approach followed, it would also be important to increase the number of images in the dataset, especially of images without the presence of graffiti. A larger image dataset is expected to improve the training of the models and, subsequently, to achieve better results. Considering the achieved results, it is worth to invest in a more complete system that will use video cameras mounted in the City Council vehicles already navigating in the city, as well as georeferencing data, to produce automatic annotations for registering and locating the walls subject to graffiti removal and to identify potential street art graffiti.

Author Contributions: J.F. performed all technical work related to their master thesis, T.B. and J.C.F. were supervisors and helped in the writing process. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, under the Information Sciences, Technologies, and Architecture Research Center (ISTAR) projects UIDB/04466/2020 and UIDP/04466/2020. Also part of support from Fish2Fork EEAGrants PT-Innovation 0069.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Images dataset belongs to Lisbon Municipality. Can be shared based on a formal request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. When does graffiti become heritage | Audio Trails. Available online: <http://audiotrails.co.uk/when-does-graffiti-become-heritage/> (accessed on 24 October 2021).
2. Bacharach, S. Street Art and Consent. *Br. J. Aesthet.* **2015**, *55*, 481–495. [CrossRef]
3. Alzate, J.R.; Tabares, M.S.; Vallejo, P. Graffiti and government in smart cities: A deep learning approach applied to Medellín city, Colombia. In Proceedings of the DATA 21: International Conference on Data Science, E-learning and Information Systems 2021, Petra, Jordan, 5–7 April 2021; pp. 160–165. [CrossRef]
4. Ross, J.I.; Wright, B.S. 'I've Got Better Things to Worry About': Police Perceptions of Graffiti and Street Art in a Large Mid-Atlantic City. *Police Q.* **2014**, *17*, 176–200. [CrossRef]
5. Campos, R. Urban art in Lisbon: Opportunities, tensions and paradoxes. *Cult. Trends* **2021**, *30*, 139–155. [CrossRef]
6. GAU: Onde Pintar. Available online: <http://gau.cm-lisboa.pt/onde-pintar.html> (accessed on 3 October 2022).
7. Novack, T.; Vorbeck, L.; Lorei, H.; Zipf, A. Towards detecting building facades with graffiti artwork based on street view images. *ISPRS Int. J. Geoinf.* **2020**, *9*, 98. [CrossRef]
8. Angiati, D.; Gera, G.; Piva, S.; Regazzoni, C.S. A novel method for graffiti detection using change detection algorithm. In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance—Proceedings of AVSS 2005, Como, Italy, 15–16 September 2005; Volume 2005, pp. 242–246. [CrossRef]
9. Wang, S.; Gao, J.Z.; Li, W.; Li, Y.; Wang, K.; Lu, S. Building smart city drone for graffiti detection and clean-up. In Proceedings of the 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, Leicester, UK, 19–23 August 2019; pp. 1922–1928. [CrossRef]
10. Tombari, F.; di Stefano, L.; Mattoccia, S.; Zanetti, A. Graffiti detection using a time-of-flight camera. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5259, pp. 645–654. [CrossRef]
11. Di Stefano, L.; Tombari, F.; Lanza, A.; Mattoccia, S.; Monti, S. Graffiti Detection Using Two Views to Cite This Version: HAL Id: Inria-00325654 Graffiti Detection Using Two Views. In Proceedings of the Eighth International Workshop on Visual Surveillance-VS2008, Marseille, France, 17 October 2008.
12. Wirth, R.; Hipp, J. CRISP-DM: Towards a Standard Process Model for Data Mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 11–13 April 2000.

13. Nahar, P.; Wu, K.H.; Mei, S.; Ghoghari, H.; Srinivasan, P.; Lee, Y.L.; Gao, J.; Guan, X. Autonomous UAV forced graffiti detection and removal system based on machine learning. In Proceedings of the 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, San Francisco, CA, USA, 4–8 August 2017; pp. 1–8. [\[CrossRef\]](#)
14. Tokuda, E.K.; Cesar, R.M.; Silva, C.T. Quantifying the Presence of Graffiti in Urban Environments. In Proceedings of the 2019 IEEE International Conference on Big Data and Smart Computing, BigComp 2019, Kyoto, Japan, 27 February–2 March 2019; pp. 1–4. [\[CrossRef\]](#)
15. Munsberg, G.R.; Ballester, P.; Birck, M.F.; Correa, U.B.; Andersson, V.O.; Araujo, R.M. Towards Graffiti Classification in Weakly Labeled Images Using Convolutional Neural Networks. *Commun. Comput. Inf. Sci.* **2017**, *720*, 39–48. [\[CrossRef\]](#)
16. Li, H.; Kim, J.; Delp, E.J. Deep gang graffiti component analysis. In *IS and T International Symposium on Electronic Imaging Science and Technology*; Society for Imaging Science and Technology: Springfield, VA, USA, 2018; pp. 1–4. [\[CrossRef\]](#)
17. Tong, W.; Lee, J.E.; Jin, R.; Jain, A.K. Gang and moniker identification by graffiti matching. In Proceedings of the MM'11—2011 ACM Multimedia Conference and Co-located Workshops—Multimedia in Forensics and Intelligence Workshop, MiFor'11, Scottsdale, AZ, USA, 28 November–1 December 2011; pp. 1–6. [\[CrossRef\]](#)
18. Parra, A.; Boutin, M.; Delp, E.J. Location-aware gang graffiti acquisition and browsing on a mobile device. In Proceedings of the Multimedia on Mobile Devices 2012, and Multimedia Content Access: Algorithms and Systems VI, Burlingame, CA, USA, 15 February 2012; Volume 8304, p. 830402. [\[CrossRef\]](#)
19. Yang, C.; Wong, P.C.; Ribarsky, W.; Fan, J. Efficient graffiti image retrieval. In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, ICMR 2012, Hong Kong, China, 5–8 June 2012. [\[CrossRef\]](#)
20. Jain, A.K.; Lee, J.E.; Jin, R. Graffiti-ID: Matching and retrieval of graffiti images. In Proceedings of the 1st ACM Workshop on Multimedia in Forensics—MiFor'09, Co-located with the 2009 ACM International Conference on Multimedia, MM'09, Beijing, China, 19–24 October 2009; pp. 1–6. [\[CrossRef\]](#)
21. Parra, A.; Zhao, B.; Kim, J.; Delp, E.J. Recognition, segmentation and retrieval of gang graffiti images on a mobile device. In Proceedings of the 2013 IEEE International Conference on Technologies for Homeland Security, HST 2013, Waltham, MA, USA, 12–14 November 2013; pp. 178–183. [\[CrossRef\]](#)
22. Patel, R.; Patel, S. A comprehensive study of applying convolutional neural network for computer vision. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 2161–2174.
23. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. *A Survey of Transfer Learning*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; Volume 3. [\[CrossRef\]](#)
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. Available online: <http://image-net.org/challenges/LSVRC/2015/> (accessed on 23 October 2022).
26. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946.
27. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015. Available online: <http://www.robots.ox.ac.uk/> (accessed on 23 October 2022).
28. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv* **2016**, arXiv:1602.07261. [\[CrossRef\]](#)
29. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv* **2017**, arXiv:1610.02357.
30. ImageNet: VGGNet, ResNet, Inception, and Xception with Keras—PyImageSearch. Available online: <https://pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/> (accessed on 23 October 2022).
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Available online: <http://image-net.org/challenges/LSVRC/2015/results> (accessed on 1 November 2022).
32. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. *arXiv* **2017**, arXiv:1612.03144.
33. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
34. Wang, J.; Perez, L. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv* **2017**, arXiv:1712.04621.
35. Gordon-Rodriguez, E.; Loaiza-Ganem, G.; Pleiss, G.; Cunningham, J.P. Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning. Available online: https://github.com/cunningham-lab/cb_and_cc (accessed on 11 October 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.