Neural Networks project

“Analysis of pedestrian activity before and during COVID-19 lockdown”

Lorenzo Nicoletti – 1797464

Contents

Abstract

This project is an implementation of the paper "Analysis of pedestrian activity before and during COVID-19 lockdown, using webcam time-lapse from Cracow and machine learning" that focuses on the understanding the effects that the COVID-19 outbreak has introduced in the daily-life and in everyday activities. The provided work is not only a re-implementation of the original document but also a consistent extension: starting from the previously collected data from webcam time-lapses in several locations in Cracow, this project firstly reconstructs the results explained in the paper and, secondly, it tries to expand them in order to cover other cities all over Europe and to provide interesting analogies (or differences) between the initial achievements and the additional conclusions derived from this further analysis. The study is performed by looking at the detections of one of the major state-of-the-art Neural Networks, YOLOv3, and, therefore, the overall work consists in an object detection task combined with a transfer learning, in particular fine-tuning, application.

Introduction

This period of pandemic has introduced the inhabitants of the cities all over the world to the words "social distancing" to indicate the required separation between people in a public area that governments were forced to legislate in order to try to mitigate the spread of the virus and to secure the inter-personal contacts with potentially infected individuals. These social orders have affected not only the economy in all its contests, but also the everyday activities of the population. This work aims to provide a precise analysis of this affection in Cracow by using the meticulous and huge set of data collected by the webcams located in four strategic locations of the city and then expand the reasoning to describe and include qualitatively the implicit connection between COVID-19 outspread/required movement control measures and the decreasing in the activity of humans in some of the most populated cities in the European continent.

The analysis is accomplished with the usage of YOLOv3, a deep and wide network mainly designed to perform detection of objects belonging to a total of eighty classes of different categories. This project is based on the recognition of pedestrians and some other groups of objects related to the human activity, like cars, motorcycles, bikes, boats and so on.

The main goal of this work is to show consistently how much the increasing restrictions on movements also affect the pedestrian activity and to analyze the way this proportion has been propagated to other countries in Europe by observing the similarities in this trend. At the end, one expected result can be the one that suggests a relevant reduction of individuals presence in public locations starting from some days after the beginning of the stay-at-home orders dictated by the governments and, moreover, a similar diminution in other cities of Europe, where the diffusion of the virus has been almost uniform in this area of the world.

The following report will try to analyze in a formal way the theory and the reasoning which the deep learning section of the project is based on. Then, it will discuss the performed experiments, highlighting the results achieved by parsing the data and the detections collected by the model, verifying, even in a qualitatively way, the consistency of the correctness of the final outcomes of the project.

Related work

Object Detection with YOLOv3: 'You Only Look Once' (YOLO) is a state-of-the-art, real-time object detection system. It is recognized as one of the fastest and most accurate Neural Networks and a strong and robust algorithm to perform object detection tasks. One among the exciting innovations that YOLO has introduced is the following: while prior detection systems repurpose classifiers or localizers to perform detection and then apply the model to an image at multiple locations and scales, this algorithm applies a single neural network to the full image to divide it into regions and predicts bounding boxes and probabilities for each region. This approach brings several advantages like the incredibly superior velocity and the usage of a single network instead of an ensemble of even thousands of networks, like the models before the appearance of YOLO did.

The third version of YOLO, YOLOv3, that is used in this project is the newest edition, characterized by few tricks to improve training and increase performance. The first difference with its predecessor, YOLOv2, is the increasing complexity of the underlying architecture called Darknet, structured in fifty-three layers. To perform detection, fifty-three more layers are stacked onto it; therefore, the entire convolutional underlying architecture of YOLOv3 is composed by a total of one hundred and six layers, that are summarized in the figure 1 below.

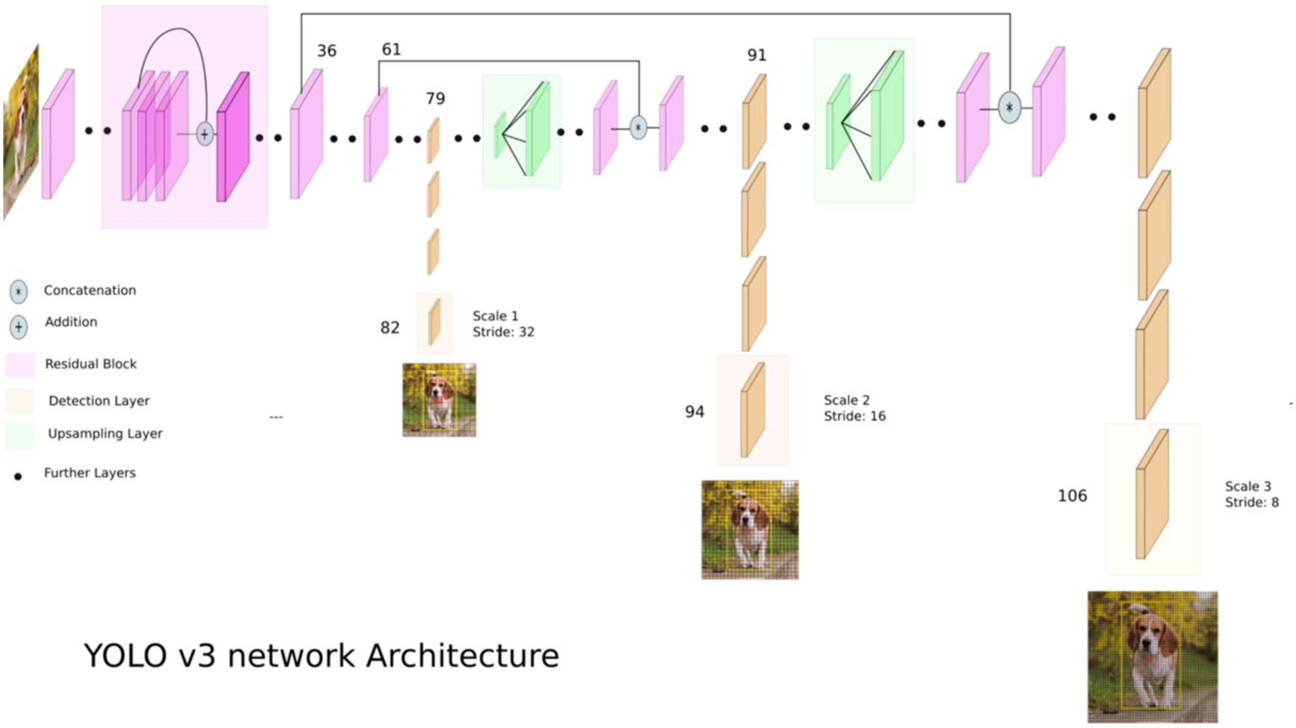

Figure 1) YOLOv3 architecture.

As we can see, the detection is effectuated at three different stages, this aspect is maybe the most interesting revolution the network has introduced.

The detection is done by applying $1 \times 1 \times 255$ detection kernels on feature maps of three different sizes at three different places in the network. YOLOv3 makes prediction at three scales, which are precisely given by downsampling the dimensions of the input image by 32, 16 and 8 respectively. The first detection is made by the 82nd layer. The input image of size $w_{in} \times h_{in}$ is processed by the 81st layer with stride 32 such that its output is of dimension $(w_{in} \div 32) \times (h_{in} \div 32)$. The application of the detection returns consequently a feature map of size $(w_{in} \div 32) \times (h_{in} \div 32) \times 255$. As shown in figure 1, the output of layer 79 is subjected to convolutional operations before it is up sampled by 2x to dimensions of $(w_{in} \div 16) \times (h_{in} \div 16)$ and concatenated with the output of layer 61, previously computed. It is noticeable how the same reasoning is recurrent in the second and in the third detections executed respectively by layers 94 and finally 106. In particular, the second generates a $(w_{in} \div 16) \times (h_{in} \div 16) \times 255$ feature map and the third a $(w_{in} \div 8) \times (h_{in} \div 8) \times 255$ feature map. As before, the convolutional output of layer 91 is first up sampled by 2x and then concatenated with the output of layer 36.

The three-stages detection and the upsampling operation are particularly useful to detect small objects, a frequent criticism in YOLOv2 now corrected in v3. In particular, the $(w_{in} \div 32) \times (h_{in} \div 32)$ layer is responsible for large object detections, the $(w_{in} \div 16) \times (h_{in} \div 16)$ layer for medium object detections and finally the $(w_{in} \div 8) \times (h_{in} \div 8)$ layer for small object detections.

The high complexity of the overall network makes YOLOv3 slower than its predecessor YOLOv2, but, on the other hand, it is capable to predict a number of bounding boxes that overcomes the number in YOLOv2 by a factor of about 10x, reason that guarantees the incredibly elevated performances in terms of accuracy of the network.

To conclude the analysis of the adopted Neural Network in the project, a little mention to the loss function of YOLOv3 is required. Generally, this function in the YOLO networks is articulated and complicated and it is reported in the figure 2 below.

$$\begin{aligned}
&\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\
&+ \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \\
&+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2 \\
&+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \qquad (3)
\end{aligned}$$

Figure 2) YOLO Loss Function.

In YOLOv3, it is almost the same except for the last three terms that are replaced by cross-entropy error terms. These three addends dictate respectively the objectness score prediction penalization for bounding boxes responsible for predicting objects, for bounding boxes predicting no objects and the class prediction penalization for the bounding boxes predicting objects. At the end, object confidence and class predictions are estimated with logistic regression.

Transfer Learning and Fine-Tuning: As mentioned before, a consistent part of this work is a transfer learning application. This type of deep learning is based on the assumption that a model pre-trained on a large and general enough dataset will qualify it as a generic model of visual world, re-adopting its weights without the need of training a large model on an even larger dataset. Usually, transfer learning is divided in two categories:

- Feature Extraction: use the representations learned by a previous network to extract meaningful features from new samples. Then a new classifier will be trained from scratch, on top of the pretrained model so that it is possible to repurpose the feature maps learned previously for the dataset.
- Fine-Tuning: use the same network architecture with pre-trained weights that are simply 'copied' in the network parameters in order to transfer the knowledge already acquired by the pre-trained model into the new one.

This project is an example of fine-tuning: it loads the weights of Darknet and put them in the YOLOv3 structure, so at the end the model does not need to be trained because it uses the past experience of the original network. In particular, Darknet, that composes the first fifty-three layers of the full network, has been trained on ImageNet, a dataset with more than fourteen millions of images belonging to more than twenty thousands of classes. These parameters are frozen and, as there will be explained in the next section of this report, they will produce consistent results even without a second of training.

Model performance evaluation: YOLOv3 accuracy is evaluated after its predictions on images. The detections are all registered in a data file and then compared with the ground truth annotations to verify if the output results of the network are reliable to execute the successive parsing of the data. The results are evaluated by computing the following two error metrics:

- Mean Absolute Error (MAE),

$$MAE = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i| \qquad (1)$$

- Root Mean Squared Error (RMSE),

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2} \qquad (2)$$

where n is the number of observations, y_i the actual number of pedestrians (namely, the ground truth value) and $\hat{y}_i$ the predicted number of pedestrians detected by the network.

It is important to notice that these error metrics may reach their maximum values, and consequently the highest errors rate of the network, with pictures representing crowds or mass of individuals very close to each other. Crowd counting, in fact, is a difficult problem to solve; in this project, the issue is quite mitigated by the strength of YOLOv3 and the results are not so affected. However, in some highly populated images, the difficulty of the network in the identification of every pedestrian is objective, even if the final conclusions are still valid and acceptable.

Experiments and Results

As mentioned in the introductory part of this report, the experiments have focused, firstly, on a re-implementation from scratch of the original work the paper is based, and, secondly, on an extension to include other cities in Europe and to analyze its analogies and differences in the results. The paper is centered on the observations of four locations in Cracow, that are:

1. All Saints' Square, touristic mainly.
2. Grodzka, touristic.
3. Podgorze Market Square, residential.
4. Wawel Castle, touristic/residential.

The work related to the original paper consisted in the analysis of a huge amount of data organized in detections of hourly human activities from 9 June 2016 to 19 April 2020 and annotated in CSV files that include the detections of the network effectuated on time-lapse images derived from the webcams placed in the locations of Cracow listed before.

To test the correctness of predictions of YOLOv3, the paper provided a little set of ten images. As a practical example of detections, some of them are reported in the figures below.

Figure 3) Example of YOLOv3 output images. Each prediction is composed by a bounding box, a label with the predicted class and a confidence value to indicate the probability of that class.

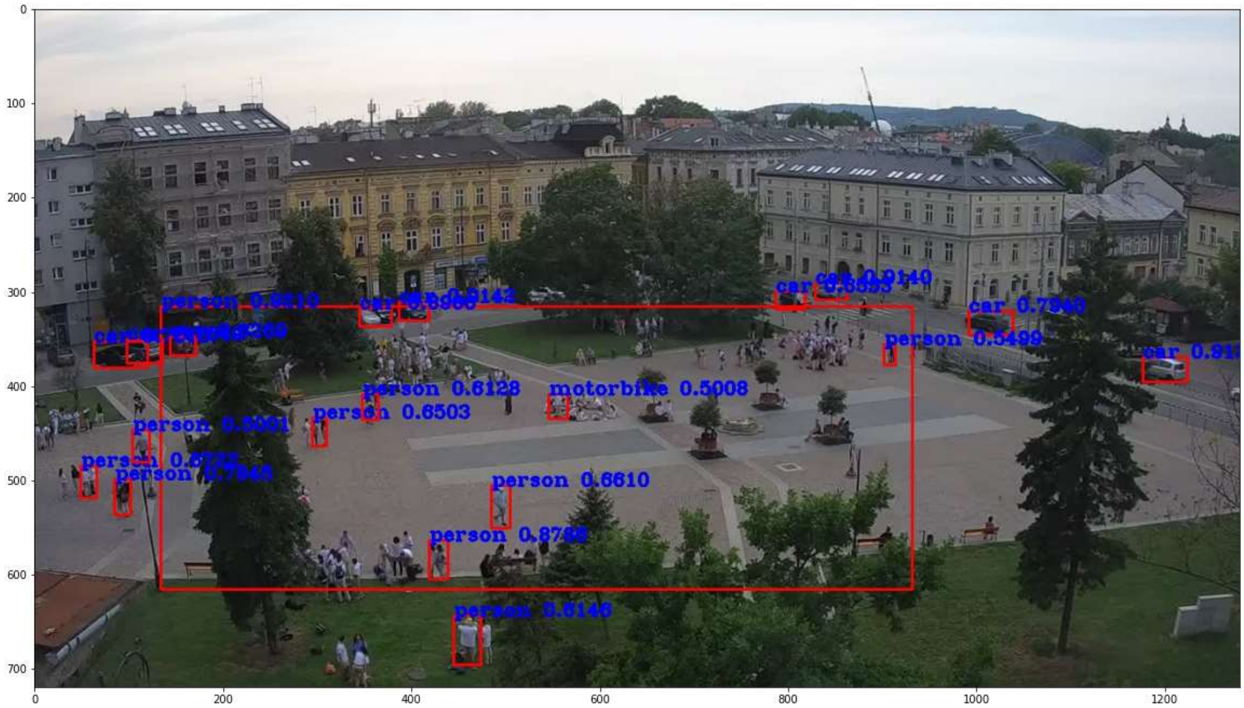

Figure 4) Example of YOLOv3 output images. Each prediction is composed by a bounding box, a label with the predicted class and a confidence value to indicate the probability of that class.

The first image detections are very accurate, almost every pedestrian is detected with a reasonably high confidence; the second image instead can practically show some of the main difficulties for the model: the crowd counting problem makes the network unable to identify the groups of individuals distributed in the picture, moreover, the distance between the camera and the objects aggravates again the prediction capability of YOLOv3 (like the misclassified person in the middle detected as a motorbike). However, these errors can be overcome by a careful choice of the testing images in the successive step of the experiments stage.

The parsing of previously collected data has requested the usage on the DataFrame class provided by the Pandas library and, after a simple manipulation of their content, has generated the result shown in figure 5, corresponding to the same conclusions achieved by the paper.

The Polish government measures to face the COVID-19 outbreak in March have been covered the second half of the month with the following social distancing safety measures:

- March 13, the Minister of Health has announced the state of epidemic threat with the introduction of the restrictions concerned the cancellation of mass events and the closure of cultural institutions such as orchestras, operas, theatres, museums and cinemas.
- March 24, further limitations on the obligation of online classes for schools or others on people leaving their homes, on public gatherings by default to a maximum of two people and on prohibited non-essential travels. Everyday activities with no contact with others and the participation of at most two people.

- March 31, minors were prohibited from leaving their homes unaccompanied by a legal guardian. Parks, boulevards and beaches were closed. Individuals walking in public were obliged to be separated by at least two meters.

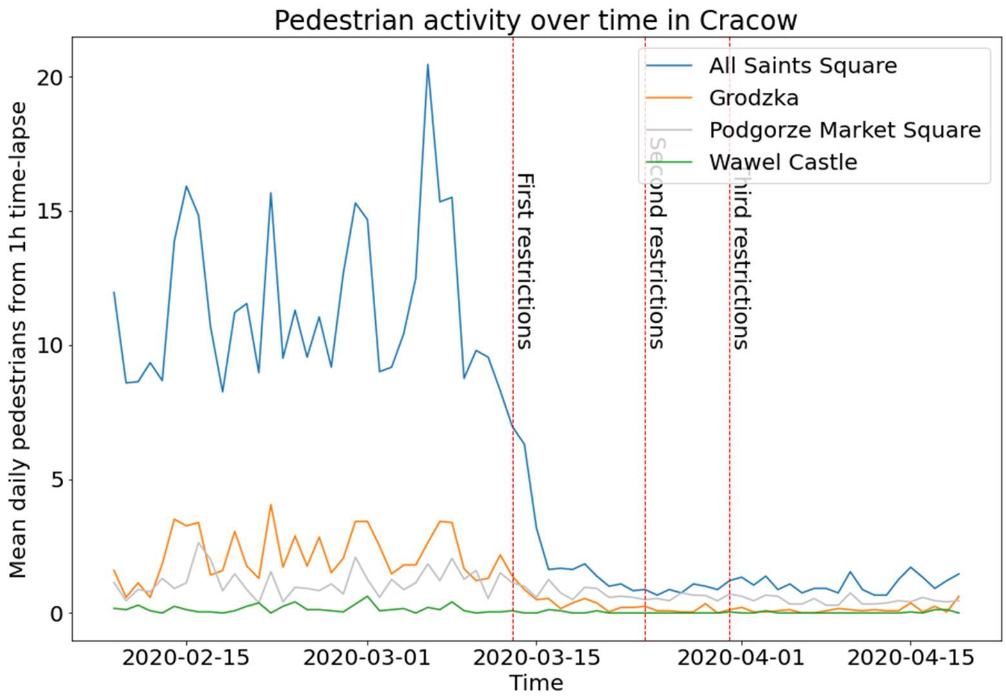

Figure 5) Analysis of pedestrian activity over time in Cracow divided in the four locations of study and highlighting the restriction dates.

As expected, the tightening of the required orders dictated by the authorities have heavily affected human activities in all the four places in the city, especially in the touristic ones, after the first restriction. The second and the third ones have produced a constant low trend in a form similar to a 'plateau'. The result plotted in the figure above shows how invasive the disturbs, that the spread of the virus has introduced in the daily lives of the inhabitants, are, independently on their locations.

The error metrics to evaluate the model performances have registered the following values:

- $MAE = 7.95$.
- $RMSE = 13.57$.

These quantities are reasonably tolerable also considering the huge amount of data stored for the task and the previously described difficulties that the network may have encountered during its predictions on high-populated or object-distanced images.

After the reproduction of this initial outcome, the project has extended its results in order to include other states in Europe. The diffusion of COVID-19, in fact, has been almost uniformly distributed all over the continent and has forced the authorities to introduce other restrictions according to the gravity of the dispersion of the disease. In the month of March, every European country faced this enormous crisis by adopting several and always stricter stay-at-home orders during the weeks. The continuation of the work of this project is to show qualitatively, due to the lack of a defined dataset, the comparison between the decreasing pedestrian activity trend in Cracow with the one of the most populated cities located in several states of Europe.

The proposed extension has required a preliminary effort to search some images representing the city situation before and during the COVID-19 epidemic. The selected images of 'COVID-19 condition' mainly came from newspapers and articles in order to associate them with a date, while the pictures of the period before the virus have been carefully chosen to avoid manifestations or mass public events in order to represent the suggested location in its everyday standard routine. In particular, the locations that have been selected for this part of the experimental stage are (in alphabetical order):

- Amsterdam, Dam Square.
- Barcelona, La Rambla.
- Berlin, Brandenburg Gate.
- London, Trafalgar Square.
- Paris, Louvre Museum.
- Rome, Colosseum.

Since all these areas are touristic, the comparison has been performed by considering only the plot corresponding to All Saints' Square in Cracow of figure 5. Moreover, the detections corresponding to the pre COVID-19 phase has been averaged in order to give a stronger and robust initialization and the resulting values have been assigned to the fictitious date of 15 February 2020, before the launch of social restrictions. Figures 6-11 show the result corresponding to each of the six chosen locations.

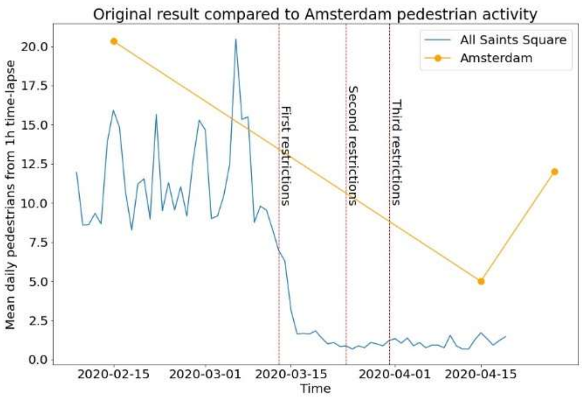

Figure 6) Comparison of result with Amsterdam.

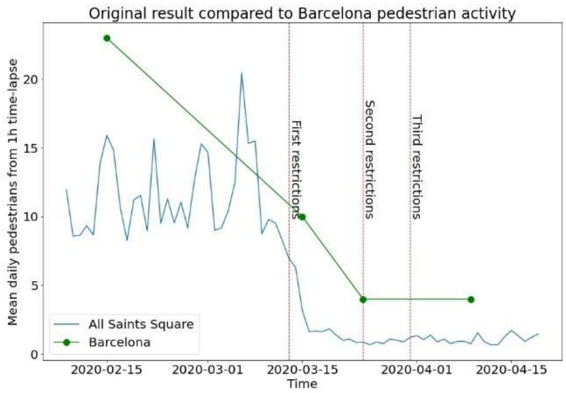

Figure 7) Comparison of result with Barcelona.

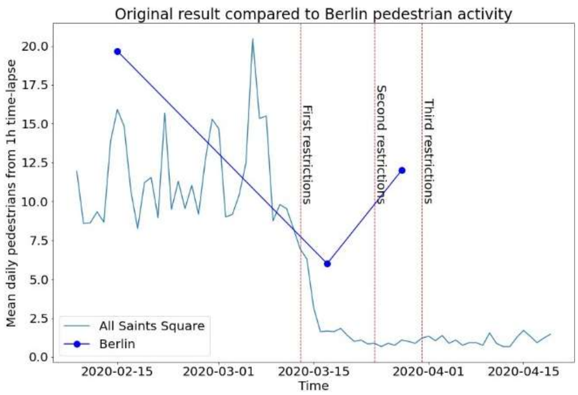

Figure 8) Comparison of result with Berlin.

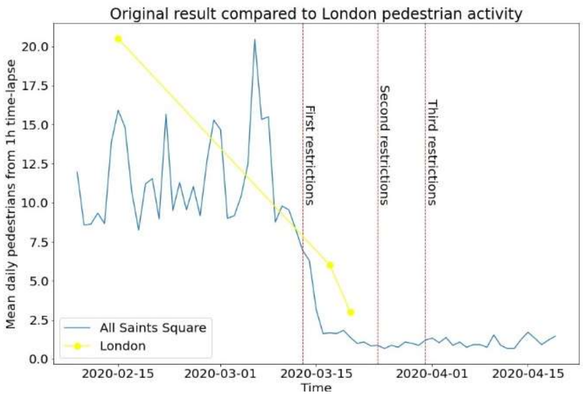

Figure 9) Comparison of result with London.

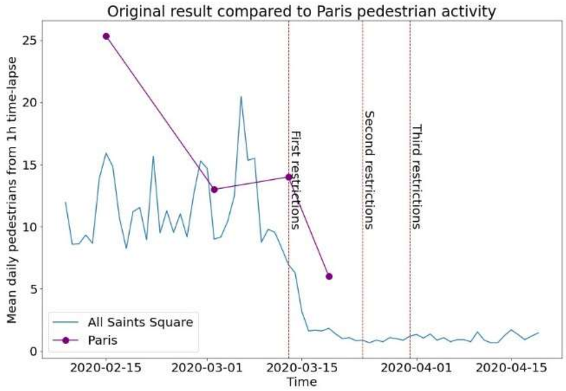

Figure 10) Comparison of result with Paris.

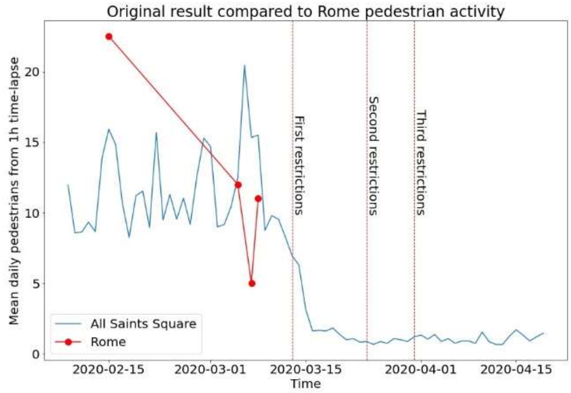

Figure 11) Comparison of result with Rome.

It is incredibly interesting observing how the descending movement is frequent in almost every tested city: some localities like Barcelona, London or Paris follows extremely faithfully the original trend traced by All Saints' Square. The remaining cities also show similarities at the occurrence of the first restrictions and all of them are characterized by the decreasing behavior expected with the diffusion of the virus and the introduction of the control limitations of the movements. A required further observation is needed to highlight that all these plots are qualitative: the results have considered only a little set of images (for each city, two or three images for each period) and, therefore, the similarities are evident but not so precise. However, a reasonable supposition can be assumed to overcome this lack of testing data: the increase of the number of location images is rationally believed to bring an improvement of the outcomes, since it is directly proportional to the accuracy of the data to be parsed.

Finally, to summarize the whole extended work, figure 12 groups all the previously analyzed results together to show even better the analogies in the behavior of pedestrians among the cities included in this study.

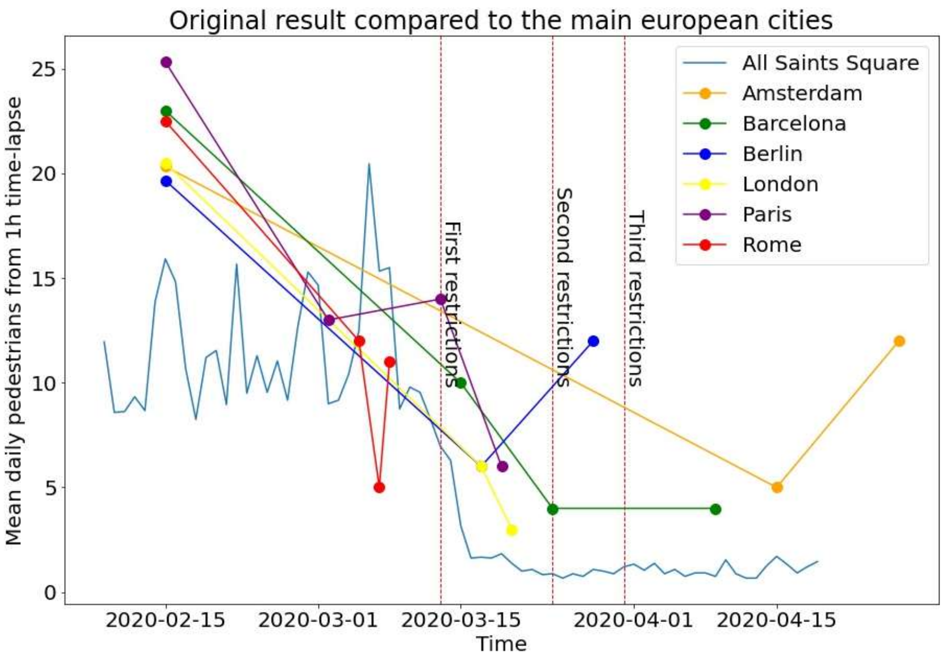

Figure 12) Comparison between all the cities.

At the end, the thesis sustained by this project has been perfectly verified. The COVID-19 outbreak and, consequently, the introduction of social restrictions to mitigate its spread that governments all over the Europe have been forced to apply has affected the everyday activity of pedestrians and the relation with their daily routine. A first study has been performed by analyzing some strategic locations in Cracow and reaching specific results, then it has been extended to verity their integrity for cities all over Europe, highlighting the interesting similarities resulting from this comparison and the further improvements applicable to get even more detailed conclusions.

Conclusions

In conclusion, after having described the situation of emergency due to the diffusion of the virus and the required social restrictions imposed by this crisis to explain the context that has bring this work to be carried on, this report has focused on a detailed explanation of the deep learning resources to perform all the needed operations. A declaration of YOLOv3 has been sustained by the theory behind the design of the network, the model has then been initialized with pre-trained weights according to the fine-tuning principle and the corresponding object detection task has produced the required data to be parsed. The analysis, conducted following the paper but also re-adopting some crucial steps with a personal preparation on the topic, has generated the same results of the original work. Then, the project has extended them to expand the same reasoning to other cities in Europe by using previously collected images, carefully selected to be in line with the related work. The overall achieved outcomes have reported very interesting conclusions about the similarities between countries and locations. At the end, the preliminary thesis on the almost uniform reduction of pedestrian activity in the continent has been confirmed with this further analysis, analysis that has demonstrated the relation between control movement measured and everyday human activity, independently from the location of study.

This report terminates with a final observation of the author: during the COVID-19 pandemic, not only in Cracow but all over Europe, pedestrian activity objectively decreased in an incredibly fast way, confirming the starting thesis of the project, but also highlighting the civic and social responsibility shown by the citizens in almost the whole continent that have quickly react to the restrictions, understanding the emergency situation and revealing a sense of accountability and community altruism.

References

- Szczepanek, R. Faculty of Environmental and Power Engineering, Cracow University of Technology, Cracow, Poland. (2020) Analysis of pedestrian activity before and during COVID-19 lockdown, using webcam time-lapse from Cracow and machine learning. Available at: Analysis of pedestrian activity before and during COVID-19 lockdown, using webcam time-lapse from Cracow and machine learning (peerj.com) (Last Accessed: 7 February 2021).

- Redmon, J. & Farhadi, A. (2018) YOLOv3: An Incremental Improvement. Available at: https://arxiv.org/pdf/1804.02767.pdf (Last Accessed: 6 February 2021).

- Kathuria, A. Towards Data Science Website. (2018) What's new in YOLO v3? Available at: https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b (Last Accessed: 6 February 2021).

- Uncini A. Neural Networks Course Material. (2020) Convolutional Neural Networks. Available at: index of (uncini.com) (Last Accessed: 5 February 2021).

- Uncini A. Neural Networks Course Material. (2020) CNN lessons. Available at: https://youtu.be/4mZ0a9XbldA and https://youtu.be/rNHWq3dwhE0 (Last Accessed: 5 February 2021).

- Tensorflow Core. (2021) Transfer learning and fine-tuning. Available at: https://www.tensorflow.org/tutorials/images/transfer_learning (Last Accessed: 7 February 2021).