Collective Dynamics

Video Analytics for Understanding Pedestrian Mobility Patterns in Public Spaces: The Case of Milan

Lorenzo Lorgna[1,2] · Giulia Ceccarelli[1] · Andrea Gorrini[1] · Michele Ciavotta[2]

[1] Fondazione Transform Transport ETS, Milan, Italy,
E-mail: a.gorrini@transformtransport.org
[2] Department of Computer Science, University of Milan-Bicocca, Milan, Italy,

Received: 3 August 2023 / Last revision received: 12 April 2024 / Accepted: 7 Mai 2024
DOI: 10.17815/CD.2024.172

Abstract The main objective of this research was to characterize public spaces through a mobility study on pedestrian patterns analyzed by means of video analytics (i.e., *object detection, crowd counting, pedestrian tracking*), for the case study of Piazza Duomo (Milan, Italy). The analysis focused on defining different pedestrian profiles through observable behavioural parameters (e.g., density conditions, speeds, trajectories, etc.). The results of the research could support the definition of an evidence-based approach for regeneration projects of urban public spaces.

Keywords Pedestrians · public spaces · video analytics · computer vision

1 Introduction

Field data represents a key starting point to support the study of pedestrian behavior in urban settings. As mentioned in [1, 2], the systematic observation of pedestrian dynamics is a very informative method to record what people do and how they behave in a particular space. In order to get such kind of records researchers can utilize manual or automated techniques for people counting and tracking [3], or exploit sensor records such as GPS [4] and Wi-Fi [5, 6]. In particular, computer vision techniques allow the automated analysis on crowd counting and density estimation, crowd motion detection, crowd tracking, and crowd behavior understanding [7–9].

In this framework, the current study aimed to transform raw observation data from a webcam installed in a large crowded public space (Piazza Duomo, Milan, Italy) into valuable urban planning insights. Object detection and object tracking models were employed to identify pedestrians and track their movements within the square [10–16]. The research explored the advantages of these techniques in urban planning, using specific metrics to analyze individual distributions and trajectories. Additionally, the study attempted to identify different categories of pedestrians (e.g., commuters, tourists, singles, groups).

2 Methodology

The study is composed of five methodological steps: (*i*) Model Selection and Data Set Collection; (*ii*) Fine-tuning and Evaluation; (*iii*) Detection and Tracking; (*iv*) Geo-referencing; and (*v*) Urban Analytics.

2.1 Model Selection and Data Set Collection

For the object detection and object tracking steps, YOLOv7 [17] and SORT [18] algorithms were selected, respectively. The dataset used for the fine-tuning phase of the object detection models consists of videos captured by 500 Pan-Tilt-Zoom traffic cameras installed in the city of Montreal[1] (CGMU dataset).

2.2 Fine-tuning and Evaluation

Fine-tuning was conducted for the YOLO v7 object detection model as follows. First, pre-trained weights of the Microsoft Common Objects in Context (MS-COCO) were sourced. Then, a transfer-learning approach was employed to fine-tune the output layers only, using the CGMU dataset. A performance evaluation was carried out on a test set. This evaluation involved the use of metrics such as Precision, Recall, and mean Average Precision (mAP). Precision quantifies the accuracy of predictions, while recall measures how many predictions match the ground truth. Last, mAP combines these metrics, summarizing the precision-recall curve to represent the average of all precisions.

2.3 Detection and Tracking

Once the training step of the object detection models was completed, the SORT algorithm was selected for tracking pedestrians. The result of the object recognition and tracking process is the creation of a text file with the following information: video frame, object ID, object class (i.e., *pedestrian*), pixel X, and pixel Y. The Piazza Duomo dataset consists of several videos recorded by a stationary camera on Thursday, July 15, 2021, in different 30-minute time slots, five different moments of the day: 8:00 - 8:30, 11:00 - 11:30, 12:45 - 13:15, 15:00 - 15:30, and 18:00 - 18:30. Videos have 1920×1080 pixels resolution and a frame rate equal to 15.01 FPS (see Fig. 1).

[1] https://donnees.montreal.ca/dataset/images-annotees-cameras-circulation

Figure 1 Piazza Duomo dataset - camera view

2.4 Geo-referencing

The geo-referencing process estimated the geographical coordinates of pedestrians in the square to eliminate perspective distortion of the images by locating pedestrians in the target area. The methodology relied on the Thin Plate Spline algorithm as a transformation technique, with Ground Control Points selected to anchor points in the perspective image.

2.5 Urban Analytics

To conduct urban analysis, QGIS software and Python scripts were utilized. The geo-referenced data was analyzed in two distinct ways. Firstly, point pattern analysis was employed to characterize public spaces, while trajectory data mining was utilized to uncover various pedestrian utilization profiles.

3 Results

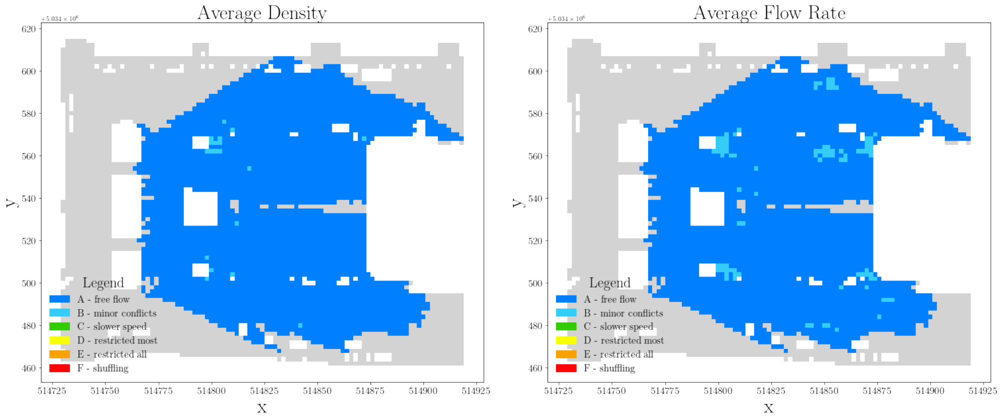
3.1 Model Evaluation

The quality of training results for YOLOv7 [17] were evaluated through metrics as Precision, Recall and mean Average Precision (mAP), calculated on a test set extracted from the CGMU dataset. These were equivalent to 0.860, 0.799 and 0.874 respectively.

3.2 Point Pattern Analysis

The following metrics were calculated to understand and analyze the distribution of geo-referenced pedestrian detections for each time slot: *Density* (ped/sqm), and *Flow Rate* (ped/min/m). Density and Flow Rate are sourced from the Walkway Level of Service Criteria (LOS) [19], these are qualitative measures describing the impact of contextual situations of crowd density on pedestrian circulation dynamics.

The results are presented in Tab. 1. Both Density and Flow Rate show lower values in the early hours of the day and highest values in the time slot 18:00-18:30. Results (see Fig. 2) showed that the use of the square increases over the hours, with the highest values in the afternoon, especially nearby the subway accesses in the square. In general, most



(a) (b)

Figure 2 Density (ped/sqm), and flow rate (ped/min/m)
Image axes correspond to the projected coordinate reference system WGS 84/UTM zone 32N (EPSG:32632).

Table 1 Results of density, and flow rate.

Time Slot	**Density** [ped/m^2]	**Flow Rate** [$ped/min/m$]
08:00 - 08:30	0.007 ± 0.008	0.948 ± 0.861
11:00 - 11:30	0.023 ± 0.021	3.058 ± 2.252
12:45 - 13:15	0.029 ± 0.025	3.653 ± 2.472
15:00 - 15:30	0.026 ± 0.022	3.669 ± 2.531
18:00 - 18:30	0.027 ± 0.025	3.971 ± 2.754
Total	0.021 ± 0.017	2.949 ± 1.973

of the square is characterized by LOS A indicating “free flow”, considering the average scenario of the five time slots. Some areas of the square are nevertheless characterized by LOS B representing the presence of irregular flow under low- to medium density conditions.

3.3 Trajectory Data Mining

Trajectories analysis included two filtering operations. The first consisted of considering only those trajectories lasting more than 20 seconds. The second filtering step consisted of discarding those trajectories containing points with speeds above 2 m/s, which were considered outliers. At the end of the data cleaning operations, 4,326 trajectories were obtained in total (see Fig. 3). For each time slot, average values of time distance, duration, speed, and direction, are presented in Tab. 2. The first time slot, 08:00-08:30, is characterized by the lowest number of trajectories but at the same time by the highest speed value. It can be assumed that there are mostly commuters at this time of day. Over the hours, the number of identified trajectories increases and the average speed decreases. This indicates greater pedestrian congestion in the square. In the last time slot,

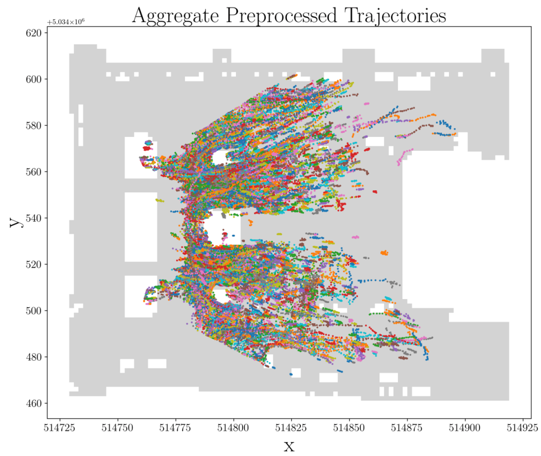


Figure 3 Overall processed trajectories
Image axes correspond to the projected coordinate reference system WGS 84/UTM zone 32N (EPSG:32632).

Table 2 Trajectories summary statistics

Time Slot	Count	Avg. Distance $[m]$	Avg. Duration $[s]$	Avg. Speed $[m/s]$	Avg. Direction $[°]$
08:00 - 08:30	255	18.087 ± 10.472	31.751 ± 15.226	0.610 ± 0.377	120.422 ± 57.002
11:00 - 11:30	922	13.236 ± 8.332	32.150 ± 18.740	0.436 ± 0.274	113.616 ± 46.916
12:45 - 13:15	1,274	13.589 ± 8.259	31.147 ± 14.354	0.450 ± 0.271	116.79 ± 47.385
15:00 - 15:30	970	13.653 ± 8.852	30.046 ± 13.367	0.468 ± 0.296	121.909 ± 53.729
18:00 - 18:30	905	12.523 ± 8.525	31.235 ± 16.279	0.418 ± 0.281	116.899 ± 47.671

18:00-18:30, the lowest values for speed and distance traveled are reported.

3.4 Commuters vs Tourists

Considering the pre-processed trajectories, clustering was performed to identify commuters and tourists. According to [20], it was assumed that the trajectories of commuters are characterized by higher linearity and higher speed, unlike those of tourists which are characterized by greater uncertainty and lower speed. Results in Tab. 3 showed that the tourists' cluster (69% of the observed pedestrians) is characterized by more irregular trajectories and a lower speed, which leads to covering shorter distances in more time. A series of two-tailed t-tests analyses confirmed that the average values of distance, speed, and direction are significantly different for the two typologies of pedestrians, except for the 11:00 and 18:00 time slots.

3.5 Single vs Groups

Further analysis was devoted to investigate the presence of groups of two or three members, namely pedestrians whose trajectories have close start/ intermediate/ end points (no more than 1.5 m) [20]. Results (see Tab. 4) showed that single pedestrians cross the square

Table 3 Clustering results

Time Slot	**Avg. Distance** [m]		**Avg. Duration** [s]	
	Commuters	*Tourists*	*Commuters*	*Tourists*
08:00 - 08:30	26.956 ± 7.269	10.802 ± 6.209	28.012 ± 9.149	34.822 ± 18.274
12:45 - 13:15	22.902 ± 7.210	9.778 ± 4.984	28.895 ± 10.607	32.069 ± 15.541
15:00 - 15:30	24.488 ± 7.990	9.727 ± 5.068	27.430 ± 8.129	30.994 ± 14.705
Time Slot	**Avg. Speed** [m/s]		**Avg. Direction** [°]	
	Commuters	*Tourists*	*Commuters*	*Tourists*
08:00 - 08:30	0.960 ± 0.240	0.322 ± 0.166	156.568 ± 61.191	90.731 ± 29.962
12:45 - 13:15	0.792 ± 0.204	0.310 ± 0.138	155.572 ± 59.831	100.910 ± 28.875
15:00 - 15:30	0.872 ± 0.203	0.322 ± 0.155	177.725 ± 61.486	101.684 ± 32.058

Table 4 Groups detection results

Time Slot	**Avg. Distance** [m]		**Avg. Duration** [s]	
	Single pedestrians	*Groups*	*Single pedestrians*	*Groups*
11:00 - 11:30	14.060 ± 8.664	10.123 ± 6.006	30.829 ± 15.044	37.142 ± 28.192
12:45 - 13:15	14.229 ± 8.408	10.928 ± 7.022	30.034 ± 13.468	35.777 ± 16.822
15:00 - 15:30	14.344 ± 9.137	10.208 ± 6.234	29.559 ± 13.348	32.479 ± 13.237
18:00 - 18:30	12.981 ± 8.698	10.343 ± 7.287	30.589 ± 14.630	34.316 ± 22.345
Time Slot	**Avg. Speed** [m/s]		**Avg. Direction** [°]	
	Single pedestrians	*Groups*	*Single pedestrians*	*Groups*
11:00 - 11:30	0.471 ± 0.285	0.301 ± 0.171	116.754 ± 49.238	101.763 ± 34.457
12:45 - 13:15	0.483 ± 0.276	0.310 ± 0.195	120.643 ± 48.918	100.746 ± 36.307
15:00 - 15:30	0.496 ± 0.301	0.329 ± 0.226	125.572 ± 55.086	103.640 ± 41.973
18:00 - 18:30	0.440 ± 0.290	0.309 ± 0.203	119.304 ± 49.783	105.444 ± 33.791

at a higher speed than groups (17% of the observed pedestrians). A series of two-tailed t-tests analyses confirmed that the average values of the considered metrics (distance, duration, speed, direction) are significantly different for single pedestrians and groups considering the five time intervals, except for the duration value in the 08:00 time slot.

4 Conclusions and Future Work

The aim of the research was to use innovative computer vision techniques to detect and track pedestrians in a large public space and extract meaningful mobility patterns. Urban planning metrics were applied to characterize the urban space and identify different categories of pedestrians, including the presence of commuters, tourists, single pedestrians and groups. Results showed the high potential of computer vision for urban mobility analysis. However, the algorithms used for detection and tracking have limitations in crowded and small pedestrian scenarios. More annotated datasets reflecting complex situations are needed for improved performance. This research could be relevant for developing an evidence-based approach for regeneration projects of urban public spaces [7].

Acknowledgements This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement The analyzed data were treated according to the GDPR-General Data Protection Regulation (EU, 2016/679).

Author Contributions Lorenzo Lorgna: Software, Writing – Original draft preparation, Visualization, Investigation / Giulia Ceccarelli: Conceptualization, Software, Visualization, Investigation/ Andrea Gorrini: Conceptualization, Investigation / Michele Ciavotta: Conceptualization.

References

[1] Zhang, F., Miranda, A., Duarte, F., Vale, L., Hack, G., Liu, Y., Batty, M., Ratti, C.: Urban visual intelligence: Studying cities with ai and street-level imagery (2023). doi:10.48550/arXiv.2301.00580

[2] Hou, J., Chenid, L., Zhang, E., Jia, H., Longid, Y.: Quantifying the usage of small public spaces using deep convolutional neural network. PLoS ONE **15** (2020). doi:10.1371/journal.pone.0239390

[3] Messa, F., Ceccarelli, G., Gorrini, A., Presicce, D., Choubassi, Choubassi: Deep learning video analytics to assess vga measures and proxemic behaviour in public spaces. In: 13th International Space Syntax Symposium (13SSS), 22-24 June 2022, Bergen (Norway), pp. 1–22 (2022)

[4] Lin, M., Hsu, W.J.: Mining gps data for mobility patterns: A survey. Pervasive and mobile computing **12**, 1–16 (2014). doi:10.1016/j.pmcj.2013.06.005

[5] Gorrini, A., Messa, F., Ceccarelli, G., Choubassi, R.: Covid-19 pandemic and activity patterns in milan. wi-fi sensors and location-based data. TeMA-Journal of Land Use, Mobility and Environment **14**(2), 211–226 (2021). doi:10.6093/1970-9870/7886

[6] Farrokhtala, A., Chen, Y., Hu, T., Ye, S.: Toward understanding hidden patterns in human mobility using wi-fi. In: 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), pp. 1–4 (2018). doi:10.1109/CCECE.2018.8447727

[7] Ceccarelli, G., Messa, F., Gorrini, A., Presicce, D., Choubassi, R.: Deep learning video analytics for the assessment of street experiments: The case of bologna. Journal of Urban Mobility (2023). Submitted

[8] Niu, T., Qing, L., Han, L., Long, Y., Hou, J., Li, L., Tang, W., Teng, Q.: Small public space vitality analysis and evaluation based on human trajectory modeling using video data. Building and Environment **225**, 109563 (2022). doi:10.1016/j.buildenv.2022.109563

[9] Tripathi, G., Singh, K., Vishwakarma, D.K.: Convolutional neural networks for crowd behaviour analysis: a survey. The Visual Computer **35**, 753–776 (2019). doi:10.1007/s00371-018-1499-5

[10] Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J.: Object detection in 20 years: A survey. Proceedings of the IEEE **111**(3), 257–276 (2023). doi:10.1109/JPROC.2023.3238524

[11] Aziz, L., Haji Salam, M.S.B., Sheikh, U.U., Ayub, S.: Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. IEEE Access **8**, 170461–170495 (2020). doi:10.1109/ACCESS.2020.3021508

[12] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M.: Deep learning for generic object detection: A survey. International journal of computer vision **128**, 261–318 (2020). doi:10.48550/arXiv.1809.02165

[13] Ciaparrone, G., Sánchez, F.L., Tabik, S., Troiano, L., Tagliaferri, R., Herrera, F.: Deep learning in video multi-object tracking: A survey. Neurocomputing **381**, 61–88 (2020). doi:10.1016/j.neucom.2019.11.023

[14] Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X.: Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems **PP**, 1–21 (2019). doi:10.1109/TNNLS.2018.2876865

[15] Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R.: A survey of deep learning-based object detection. IEEE access **7**, 128837–128868 (2019). doi:10.1109/ACCESS.2019.2939201

[16] Tomè, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., Tubaro, S.: Deep convolutional neural networks for pedestrian detection. Signal processing: image communication **47**, 482–489 (2016). doi:10.1016/j.image.2016.05.007

[17] Jiang, K., Xie, T., Yan, R., Wen, X., Li, D., Jiang, H., Jiang, N., Feng, L., Duan, X., Wang, J.: An attention mechanism-improved yolov7 object detection algorithm for hemp duck count estimation. Agriculture **12**(10), 1659 (2022). doi:10.3390/agriculture12101659

[18] Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B.: Simple online and realtime tracking. In: 2016 IEEE international conference on image processing (ICIP), pp. 3464–3468. IEEE (2016). doi:10.1109/ICIP.2016.7533003

[19] Fruin, J.: Pedestrian Planning and Design. Metropolitan Association of Urban Designers and Environmental Planners (1971). URL https://books.google.it/books?id=AydSAAAAMAAJ

[20] Gorrini, A., Vizzari, G., Bandini, S.: Age and group-driven pedestrian behaviour: from observations to simulations. Collective Dynamics **1**, 1–16 (2016). doi:10.17815/CD.2016.3