USE OF OBJECT RECOGNITION AI IN COMMUNITY AND HERITAGE MAPPING FOR THE DRAFTING OF SUSTAINABLE DEVELOPMENT STRATEGIES SUITABLE FOR INDIVIDUAL COMMUNITIES, WITH CASE STUDIES IN CHINA, ALBANIA AND ITALY

FRANCESCO GRUGNI[1], MARCO VOLTOLINA[2] and TIZIANO CATTANEO[3]
[1,3]Università degli Studi di Pavia, [2] Politecnico di Milano.
[1]grugnifrancesco@gmail.com, 0000-0003-3198-9950
[2]marcovoltolina@outlook.it, 0000-0002-4786-7475
[3]tiziano.cattaneo@unipv.it, 0000-0003-1604-6831

Abstract. In order to plan effective strategies for the sustainable development of individual communities, as prescribed by the United Nations' Sustainable Development Goal 11, it is necessary for designers and policy makers to gain a deep awareness of the bond that connects people to their territory. AI-driven technologies, and specifically Object Recognition algorithms, are powerful tools that can be used to this end, as they make it possible to analyse huge amounts of pictures shared on social media by residents and visitors of a specific area. A model of the emotional, subjective point of view of the members of the community is thus generated, giving new insights that can support traditional techniques such as surveys and interviews. For the purposes of this research, three case studies have been considered: the neighbourhood around Siping Road in Shanghai, China; the village of Moscopole in southeastern Albania; the rural area of Oltrepò Pavese in northern Italy. The results demonstrate that a conscious use of AI-driven technologies does not necessarily imply homogenisation and flattening of individual differences: on the contrary, in all three cases diversities tend to emerge, making it possible to recognise and enhance the individuality of each community and the genius loci of each place.

Keywords. Sustainable Communities; Artificial Intelligence; Object Recognition; Social Media; SDG 11.

1. Introduction

According to the United Nations, "sustainable cities and communities" are a crucial goal to achieve before the year 2030, as prescribed by Sustainable Development Goal 11 (SDG 11). Specifically, some of the targets outlined in this goal are: to "enhance inclusive and sustainable urbanisation and capacity for participatory, integrated and sustainable human settlement planning and management", to "strengthen efforts to protect and safeguard the world's cultural and natural heritage", and to "support positive economic, social and environmental links between urban, peri-urban and rural

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 1, 717-726.

areas by strengthening national and regional development planning" (United Nations General Assembly, 2015, pp. 21-22).

In order to achieve such targets, designers and policy makers need operative tools to analyse the necessities of local communities and the unique relationship that they entertain with the place they inhabit. This research argues that AI-driven technologies, and specifically Object Recognition algorithms, can be effectively used for this purpose.

In recent years, researchers have applied this kind of algorithms to the analysis of cities and neighbourhoods, focusing on parameters related to inequality, safety and liveliness (see for example Salesses et al., 2013, and De Nadai et al., 2016). In these studies, the image dataset was composed of geo-tagged pictures taken from Google Street View or collected manually by the authors. A study by MIT researchers, however, points out that this choice presents some limitations, as "a lot of content related to city perceptions, such as mountains and crowded indoor scenes are missing", and proposes instead to use images taken from photo-sharing platforms such as Flickr and Panoramio (Zhou et al., 2014, p. 522). Through this different strategy, the researchers try to define the *identity* of cities, and analyse millions of geo-tagged images according to seven pre-determined attributes: green space, water coverage, transportation, architecture, vertical building, athletic activity, and social activity (pp. 524-525).

Rather than the *identity* of a city, the present research aims at defining what Kevin Lynch called the *image* of the city (Lynch, 1960), that is how the community lives and perceives the place. The purpose here is to understand the relationship that a community entertains with a place through the lens of the emotional, subjective point of view of its members. In order to do so, this study proposes a new approach, based on:

- A multi-scale strategy through which it is possible to consider a specific neighbourhood, a city or even an entire region, based on the local context, the features of the community and the purpose of the analysis;
- A dataset composed of geo-tagged images taken from a popular social network such as Instagram, as it reflects everyday life and subjective perceptions more accurately than Google Street View or photo-sharing platforms;
- Flexibility and absence of pre-determined attributes, as the purpose in this case is not to compare different places according to fixed parameters, but to allow their specificity and uniqueness to emerge;
- Versatility, as the approach can be successfully applied to places in different parts of the world, with different cultures, demographics and economies (this is demonstrated by the diversity of the proposed case studies).

2. Method

The method followed for this research is composed of five main steps. The first one consists in defining the object of the analysis. As mentioned above, this study follows a multi-scale strategy, so it is possible to select, for example, a specific neighbourhood,

a village, a city or even a larger territory. Another possibility is to work on two different scales at the same time, as experimented in the third case study. What is important in this phase is to define an area that is meaningfully associated with one or more communities of people.

The second step is to define the dataset. All the pictures of this research have been taken from Instagram, as it is a popular image-based social media. The images have been collected through a python script, based on their geo-tag and hashtags. The number of pictures collected for each case study depends on the local context and the scale of the considered area: for the first two case studies (a neighbourhood and a village), 500 images have been selected, while for the third one (a larger-scale territory) the number increased to 1,200. The date range is the same for all the case studies, as in this research have only been used pictures shared on Instagram during the year 2021.

The third step is to apply to the dataset an Object Recognition algorithm that includes the following tools: the SIFT (Scale-Invariant Feature Transform) system and the Lowe method (Lowe, 1999), the R-CNN method (Girshick et al., 2014), the Spatial Pyramid Planning strategy (He et al., 2014), and the SSD model (Liu et al., 2016). For this study, the algorithm has been pre-trained with millions of images and has learnt to recognise around 10,000 different objects ("picturable nouns"). It has also been programmed to select, for each picture, only the five most prominent objects, so as to avoid giving irrelevant or redundant information as output. As the research progressed, the algorithm has been trained to recognise new objects that were peculiar to the analysed areas, thus enabling it to better highlight the uniqueness of each place. This process gives as a result a graph that shows the frequency with which different objects appear in the image dataset (the graphs in this paper include only a selection of objects, but the algorithm actually recognised many more).

For the fourth step, it is necessary to select a limited number of significant objects: in this research, the focus has been put on five of the objects that appeared more frequently in each case study. Then, the positions of all the pictures in which these significant objects appear are located on a map. This is done through the geo-tags and metadata of the pictures or, when the geo-tag is generic, through a software that defines the position of the image relative to a set of known points (e.g. famous buildings, monuments, infrastructures).

Finally, the results are discussed in order to understand how the output data (graphs and maps) can be used as a basis for drafting strategies for the sustainable development of the local communities.

2.1. LIMITATIONS

The effectiveness of strategies based on AI-driven technologies always depends on the quality of input data. As the method proposed in this study takes its dataset from social media, it is important to be aware that these platforms are used by certain demographic groups more than others.

In fact, a research carried out in 2021 by Auxier and Anderson shows that most Instagram users in the United States of America are young, have studied in college and live in urban areas, while categories such as elderly people and residents in rural areas are much less present on the platform.

For this reason, it is important to stress that the method proposed in this research is not intended to be used alone, but rather as part of a wider strategy that also involves other tools and techniques. In fact, it is only by using an adequate array of methods that it is possible to carry out an analysis that reflects the feelings and perceptions of the whole community.

3. The Neighbourhood of Siping Road, Shanghai, China

The first case study that will be discussed is the analysis of the neighbourhood around Siping Road, in the Yangpu district of Shanghai, China. The area is part of one of the most populous cities in the world, with more than 24 million inhabitants. Specifically, the segment of Siping Road on which the analysis has focused is the one close to Tongji University's main campus. A dataset of 500 pictures has been collected from Instagram pictures with geo-tags located in the area and hashtags related to the neighbourhood.

3.1. OUTPUT

The output of the Object Recognition algorithm (fig. 1) shows that many of the elements that appear in the pictures are connected to the presence of the university: "graduation" (42 appearances), "student" (32), "graduation photo" (30), "ceremony" (12), and "lecture" (10). The other themes that emerge from the analysis are related to social activities ("indoor", 36 appearances; "food", 27; "get-together", 15; "night", 12), natural elements ("tree", 33), urban landscape ("architecture", 24; "building", 12) and art ("art installation", 10).

Then, the distribution of five elements has been located on a map (fig. 2): "food", "student", "tree", "graduation" (also including "graduation photo"), and "indoor". The result is a map that shows the areas usually frequented by students, and the places that they consider important enough to have their graduation photos taken there, but also the location of popular bars and restaurants, of the most frequented public buildings (e.g. the university library), and the spots where trees are found.

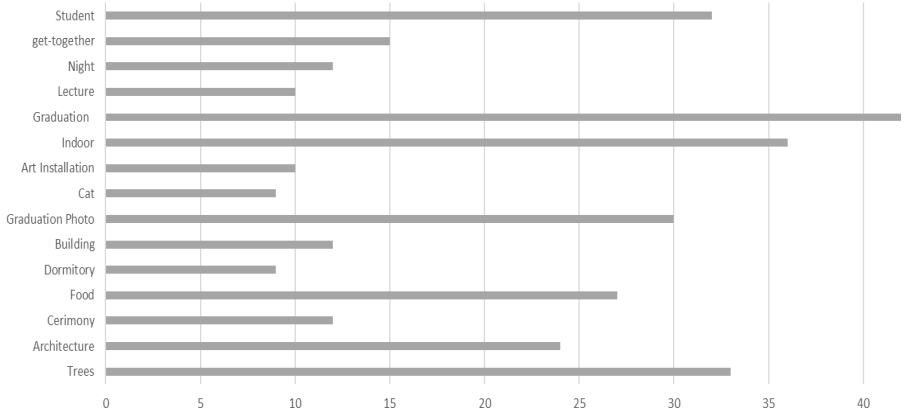

Figure 1. Output of the Object Recognition algorithm, Siping Road.

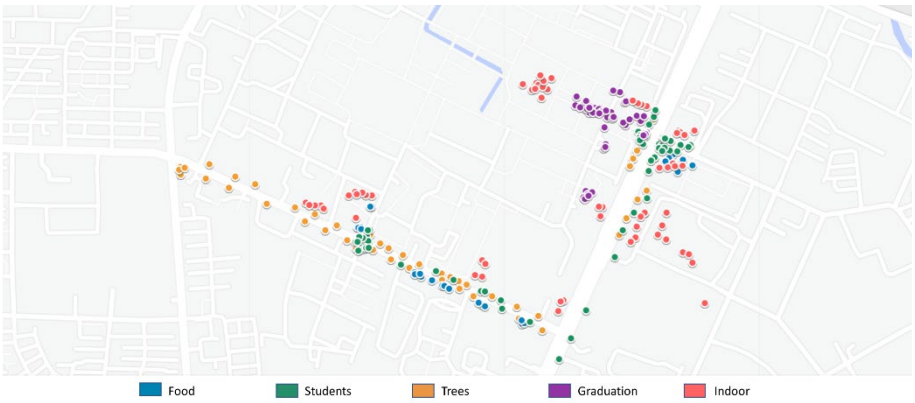

Figure 2. Distribution of five frequent elements in the neighbourhood of Siping Road.

3.2. CONSIDERATIONS

The experiment shows the potential of the Siping Road neighbourhood, especially from the point of view of its student community. Its output describes what kind of social activities are practised in the area and where exactly they are practised. Based on the results, it is possible to imagine strategies for establishing beneficial relationships between students, residents and business owners, and for improving the public spaces of the neighbourhood: on one hand, by enhancing the most frequented areas, adapting them to the activities that they already host; and on the other hand by regenerating less frequented areas, reproducing in them the traits of more successful parts of the same neighbourhood.

4. The Village of Moscopole, Albania

The second case study focuses on Moscopole, a village of little more than 1,000 inhabitants in Korçë County, in southeastern Albania. Despite its small population, however, Moscopole is slowly gaining popularity as a tourist destination, thanks to its mountain landscape and its cultural heritage. The dataset is once again made of 500 Instagram pictures with geo-tags located in the village and its immediate surroundings, and hashtags related to the area.

4.1. OUTPUT

In this case, in the output data (fig. 3) it is possible to clearly recognise two main themes: on one hand nature and landscape, and on the other architecture and history. The former includes “outdoor” (43 appearances), “nature” (37), “forest” (20), “hills” (18), “mountain” (14), “snow” (14), “treehouse” (12), “wild” (11), and “quad” (11). The latter, instead, is related to “mediaeval” (47), “church” (24), “city” (13), and “house” (13).

These two themes also define the map (fig. 4) that shows the position of the following five elements: “forest”, “nature”, “outdoor”, “mediaeval” and “church”. The

map clearly suggests what are the most frequented naturalistic paths and cultural heritage in Moscopole. It also demonstrates that people in the village tend to spend time and perform activities outdoors.

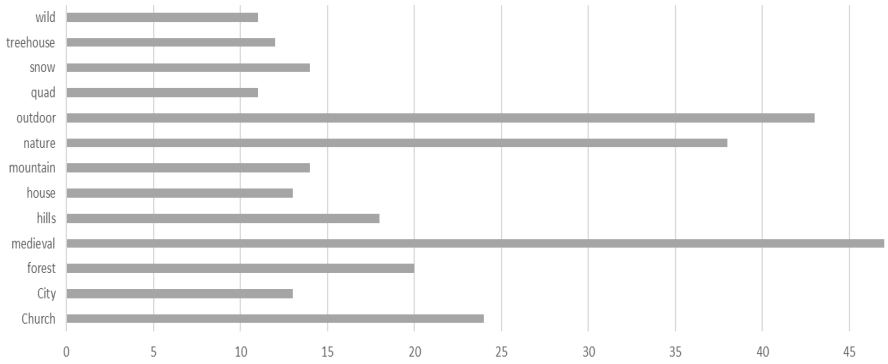

Figure 3. Output of the Object Recognition algorithm, Moscopole.

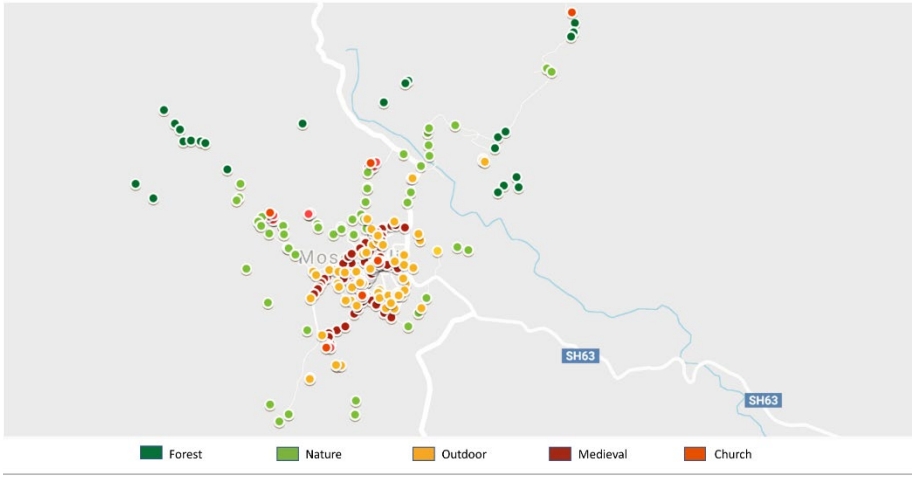

Figure 4. Distribution of five frequent elements in the village of Moscopole.

4.2. CONSIDERATIONS

In this case, the experiment provides a good basis for implementing strategies for sustainable tourism. On one hand, it highlights the potential for naturalistic excursions and sports, also showing existing trails that could be enhanced and promoted. On the other hand, it suggests that Moscopole could also attract tourists interested in its history and architecture, and this could be achieved, for example, by renovating outdoor public spaces, enhancing tourist facilities and preserving its cultural heritage, especially its ancient churches. As these two aspects are interrelated, improvements in any of them would end up being beneficial for the whole tourism sector of Moscopole.

5. The Territory of Oltrepò Pavese, Italy

Finally, the last case study takes into consideration a large rural territory, made up of four municipalities (Corvino San Quirico, Mornico Losana, Oliva Gessi and Torricella Verzate) in an area known as Oltrepò Pavese, in the province of Pavia, in northern Italy. The area is known for its wines and castles, and it is home to around 2,600 people. In this case, the analysis has been carried out at two different scales: that of the whole territory and that of each specific municipality. For this reason, the dataset is larger, as it includes 300 Instagram pictures for each municipality, for a total of 1,200 images, always based on geo-tags and hashtags related to the area.

5.1. OUTPUT

Compared with the previous case studies, the output data for the Oltrepò territory present a greater variety (fig. 5). In fact, the list of objects that appear in the pictures include elements related to many different semantic fields: nature ("landscape", 118 appearances; "flowers", 11, "apiary", 11), wine ("wine", 68; "Defilippi winery", 53; "Monsupello winery", 33; "Cavallini winery", 11; "Tenuta di Oliva winery", 11), historical heritage ("castle", 61; "church", 23), sports ("sport", 42; bike, 27; "swimming pool", 22; "walk", 12; "motocross", 10), weddings ("bride and groom", 28; "wedding", 27), social activities ("hamburger", 13; "show", 12).

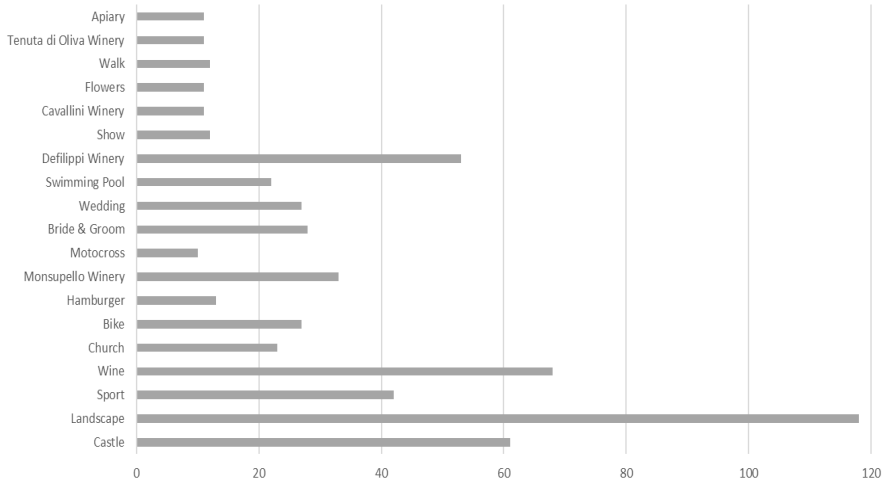

Figure 5. Output of the Object Recognition algorithm, Oltrepò Pavese.

The map with the positions of the most frequent elements (fig. 6) clearly shows how some of them are distributed quite homogeneously on the whole area ("landscape" and "sport"), while other ones are found only in some places ("wine", that also includes all the different wineries), and still others are specific to certain locations ("castle" and "church").

For this reason, in this case it is useful to zoom in and consider the territory at a different scale (fig. 7). In other words, after a first general analysis that highlights the

main vocations of the area, a second step is needed to better identify the unique features of each location. By analysing the data of each of the four municipalities, what emerges is that Mornico Losana is a coveted destination for weddings, but it also hosts a historical castle and a popular swimming pool; Corvino San Quirico and Oliva Gessi are frequented mostly because of their landscape and wine, but also for their apiaries; and Torricella Verzate has a popular winery and a historical church, but also good restaurants and places to practice motocross.

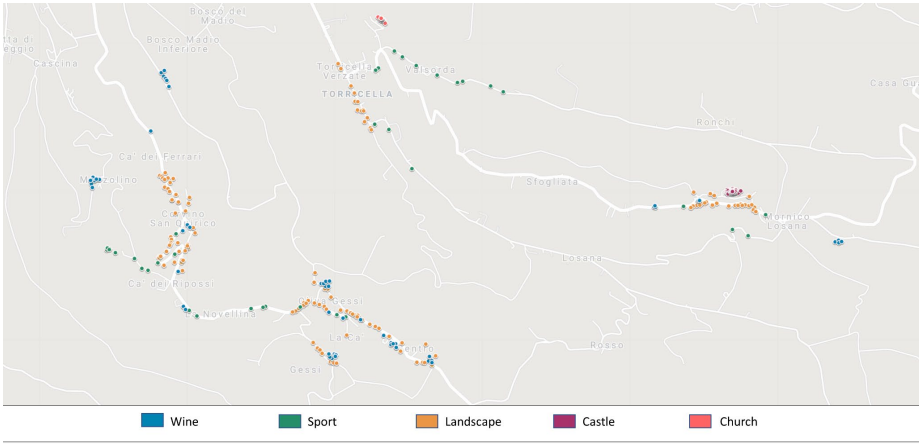

Figure 6. Distribution of five frequent elements in the territory of Oltrepò Pavese.

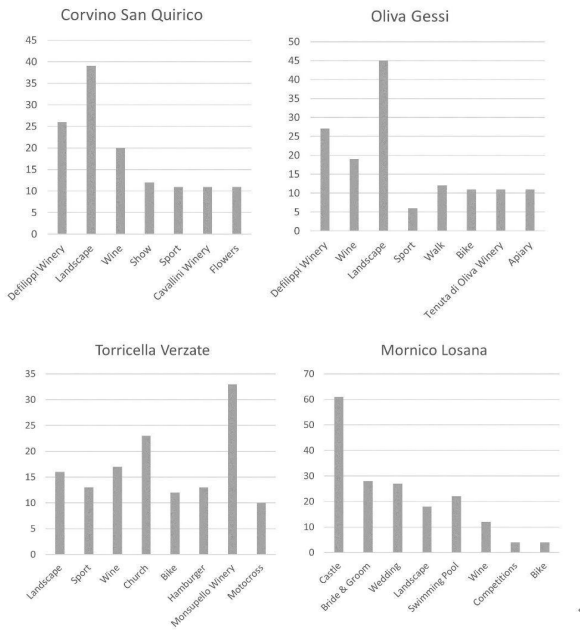

Figure 7. Output of the Object Recognition algorithm: Torricella Verzate, Mornico Losana, Corvino San Quirico, and Oliva Gessi.

5.2. CONSIDERATIONS

As the analysis for this case study has been carried out at two different scales, it is possible to outline two sets of strategies. On one hand, the experiment suggests general strategies for the whole territory: enhancing viticulture, food and wine tourism and naturalistic trails. On the other hand, it highlights the need for specific strategies to implement in selected locations: preserving and promoting the cultural heritage (the castle of Mornico Losana and the church of Torricella Verzate), encouraging certain sports (swimming in Mornico Losana and motocross in Torricella Verzate) and enhancing beekeeping in Oliva Gessi.

6. Conclusions

This research proposes a method for understanding the subjective relationship between a community and its territory. Three different case studies have been analysed, demonstrating how Object Recognition algorithms can be used to recognise diversity, enhancing the individuality of each community and the genius loci of each place (fig. 8).

Despite its limitations, the proposed method can offer a good understanding of the potential of an individual community and the opportunities offered by its territory. It can thus contribute to creating a basis for drafting effective strategies for the sustainable development of the community, in harmony with its feelings and sensitivity: for example, it suggests to enhance the public spaces for the students of Siping Road, to develop sustainable tourism in Moscopole, and to promote viticulture, apiculture, sustainable tourism and the preservation of natural and cultural heritage in Oltrepò Pavese.

This study demonstrates the great potential of Object Recognition algorithms and how they can be used for the purpose of achieving the UN's SDG 11. In particular, they can play a key role in improving indicator 11.a.1, giving local administrators the tools to draft "development plans that (a) respond to population dynamics; (b) ensure balanced territorial development" (United Nations, n.d.). Furthermore, they can indirectly contribute to enhancing the participation of citizens to the planning process (indicator 11.3.2) and to improving the efficiency of public funding for the preservation of natural and cultural heritage (indicator 11.4.1).

Further research will probably expand even more this potential: some possibilities include combining data from different sources (other social media, but also review platforms such as TripAdvisor), combining different datasets (not only images, but also text or music), integrating algorithms for sentiment analysis, or working at greater scales (metropolitan areas, provinces or regions).

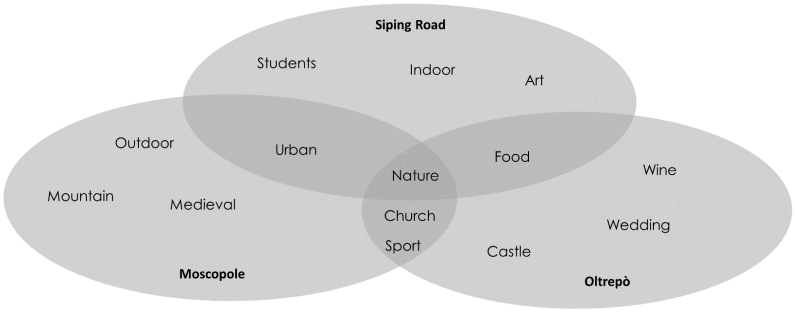

Figure 8. The use of Object Recognition algorithms recognises diversity: the graph shows the main semantic fields associated with each case study.

References

Auxier, B., & Anderson, M. (2021, April 7). *Social Media Use in 2021*. Pew Research Center. Retrieved November 17, 2021, from https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

De Nadai, M., Dragicevic, S., Hidalgo, C.A., Vieriu, R. L., Naik, N., Sebe, N., Zen, G., Caraviello, M., & Lepri, B. (2016, October). Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life. In *MM '16: Proceedings of the 24th ACM International Conference on Multimedia* (pp. 1127-1135). https://doi.org/10.1145/2964284.2964312.

Girshick, R.B., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). Retrieved February 4, 2022, from https://arxiv.org/pdf/1311.2524.pdf.

He, K., Zhang, X., Ren, S., & Sun, J. (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In Computer Vision – ECCV 2014. *Lecture Notes in Computer Science*, vol. 8691. Springer, Cham. https://doi.org/10.1007/978-3-319-10578-9_23

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. (2016). SSD: Single Shot Multi-Box Detector. In Computer Vision – ECCV 2014. *Lecture Notes in Computer Science*, vol. 9905. Springer, Cham. https://doi.org/ 10.1007/978-3-319-46448-0_2

Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.1999.790410.

Lynch, K. (1960). *The Image of the City*. MIT Press.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013, July 24). The Collaborative Image of the City: Mapping the Inequality of Urban Perception. *PLOS ONE*, 10(3), https://doi.org/10.1371/journal.pone.0119352.

United Nations (n.d.). *Goal 11*. United Nations Department of Social and Economic Affairs. Retrieved Februrary 4, 2022, from https://sdgs.un.org/goals/goal11.

United Nations General Assembly (2015, September 25). *Transforming our world: the 2030 Agenda for Sustainable Development. Resolution 70/1*, UNGAOR, 70th Sess, UN Doc A/RES/70/1. Retrieved from undocs.org/A/RES/70/1.

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014). Recognizing City Identity via Attribute Analysis of Geo-tagged Images. In *European Conference of Computer Vision 2014* (pp. 519-534). https://doi.org/10.1007/978-3-319-10578-9_34.