



ICT4SS - A.Y. 2020/2021, Politecnico di Torino - Interdisciplinary Project

Automatic Urban functions identification via social analysis

Baldoni Vincenzo - s277836

Boffa Matteo - s275328

Fabiani Federico - s269124

Lanza Chiara - s277750

Ravera Arianna - s277752

May 2021

Contents

Introduction	3
1 Background	4
2 Methodology	6
2.1 Data Collection	6
2.2 Clustering	8
2.3 Image Processing	9
3 Analysis	11
3.1 Global analysis	12
3.2 Central cluster analysis	13
3.3 Peak analysis	15
3.4 Yearly comparison: 2019 vs 2017	16
3.5 Synthetic Indexes	18
3.5.1 Attraction capacity indexes	18
3.5.2 Cadastral value	21
4 Conclusion	22
Conclusion	22
Appendix	27

Abstract

We live in a connected world: more and more people share various moments of their daily life on social networks, mostly by posting photos, videos and descriptions. Consequently, every day billions of geo-referenced data are generated by user activities shaping what are generally known as *digital footprints* and providing inspiring insights about human activities and behaviors [1].

This research aims at studying the geographical contour of Turin in order to provide the administration a tool for capturing the users' perception about the city. Through the analysis of the geo-tagged contents published on Flickr and Instagram, a pipeline comprehending clustering and classification techniques will provide a powerful link between the technology latent data and the municipalities' need for feedback and guidance to past and new investments.

The results show that both global studies about the city and specific investigations on particular zones and periods are feasible and reliable on respect to both our knowledge of the city and real indexes such as the cadastral value.

Introduction

While Hollands [2] in 2008 proposed a *preliminary critical polemic against some of the more rhetorical aspects of smart cities*, which aimed at identifying the novelties brought by this cutting-edge phenomenon and proposing new suggestions to enhance their progressive and inclusive qualities, Kitchin [3] wisely identified their two founding aspects: firstly, its real-time capacity of monitoring, managing and regulating the city flow processes through the pervasive and ubiquitous computing and collecting; secondly, its being a conjunction between ICT and the human educational and economical development and governance. This trend has tremendously accelerated in the past years with the availability of more and more data: *at some point the Internet of people gave away to the Internet of Things*, already observed by Townsend [4] in 2013 commenting the *torrent of readings* generated by the new technologies.

Social networks are greatly involved to aforementioned stream of information (see Appel et al. [5]); during the last year, in the dramatic context of the Coronavirus pandemic, their contribution, also due to an higher technology awareness, has further augmented. Instagram for example has registered an astonishing growth of 4 million of new users and a 10 per cent increase of the daily average time spent on the platform during 2020 according to We are Social [6]. Italy is indeed characterized by an higher number of 'connected' people than the world average, with 49,48 million of people (82% of the population) connected to Internet and 35 million users (57%) active on social media; in the top 5 ranking of the monthly most visited social channels from a browser in Italy we find in order Facebook, Youtube, Instagram, Twitter and Whatsapp (see SIM [7]).

Unsurprisingly, projects based on the exploitation of ICT technologies and Big Data to create smarter societies are therefore widespread all over the world: London [8], Amsterdam [9] and Singapore [10] are only some of the many examples. Among them, a remarkable one is that of Turin [11], which has been said to be one of the few smart-city cases in Italy. Its *Torino City Lab* (TCL) focuses on themes like *Autonomous Vehicles*; *Drones*; *Electrical monitoring systems*; *Internet of Things* (IoT); *Artificial Intelligence* (AI) [12].

In compliance with this tendency and with the European regulation 2016/679 about

user profiling rules [13], our analysis aims at exploiting the so called users *social footprints* in the area surrounding Turin municipality to create a tool for the administration and investors. The main purpose is providing an instrument to quantify the people perceptions on the various areas of the city, leading them to adopt investment strategies to redevelop areas in those fields that show an increase in social interest. Specifically, collecting data from Instagram and Flickr and feeding them to some clustering and classification processes, we will show interesting applications in the fields of event-detection, period to period comparison and area-discovering. One of the key points of the proposed study is its being completely unsupervised; consequently the results it brings do not derive in any way from previous knowledge of the territory. This guarantees spatial and temporal elasticity to the analysis, which can thus be replicated in any area or city under study, even if completely unknown to the researchers. On the other hand, the knowledge of the chosen city (Turin) by all contributors is a surplus that helps in the performances evaluation of the tool.

The remainder of this paper is organized as follows. On **Section 1** we first discuss some previous works based on the socials analysis. **Section 2** on the other hand gives an outlook on the proposed pipeline, which is further analysed on the followings **2.1**, **2.2** and **2.3**. **Section 3** collects the main results of our work, and eventually **Section 4** ends the paper with some final thought and suggestions for future improvements.

1 Background

A number of researches have investigated the use of geo-referenced social media contents. Already in 2008, Girardin et al. [14] were including Flickr in the list of the so called *explicit footprints*, namely active or volunteered locational information provided by users (see Offenhuber and Ratti [15]), stressing their importance for tasks as investigating the visitors mobility in diverse locations (Girardin et al. [14], Girardin et al. [16]) or measuring the impact of an event (Girardin et al. [17]). Those applications still capture the attention of researchers: Kdr and Gede [18] with a Flickr analysis and Mukhina et al. [19] with Instagram attempted to visualize the most visited places in Budapest and St.Petersburg with important distinctions of locals and tourists profiles, in order to provide the administration a report of the areas of potential interest but still hidden from the international tours; Brandt et al. [20] shifted the analysis to textual contents coming from Twitter to reinforce what they defined the “smart urban tourism” analysing the tourists spatial patterns through a kernel density estimation and a latent Dirichlet allocation; Kdr [21] considered spatial patterns using Flickr data, but enlarged the analysis considering three cities with similar tourist markets (Vienna, Prague and Budapest) to underline similarities and differences.

With the tremendous development of technology-supported networks and popularization of the mobile devices, more and more opportunities were eventually given to researchers of any fields to access to the massive quantities of information produced by and about people, things, and their interactions (see Boyd and Crawford [22]). Thanks to this availability, performing sharper analysis also becomes an option: Deng et al. [23] for example, studied the destination image (DI) perceptions comparing the pictures taken in Shanghai by Western and Eastern groups exploiting their profile information; Donaire et al. [24] were able to identify “groups” of photographers through a cluster analysis and

capturing common stylistic features on samples taken at the Boí Valley in the Pyrenees; Florio et al. [25] and Chiesa [26] succeeded on providing tools for quantifying the visual impact of renewable sources, efficiently preserving the landmarks and point of interest with “smarter” limitations for the stakeholders.

Interestingly, the aforementioned data explosion, together with the latest breakthroughs in terms of training algorithm, also represents raw inputs for techniques of data mining using artificial neural networks; these have been successfully applied to various fields of knowledge such as medicine, environmental studies, information science, and computer graphics (see Gron [27]). Some of these researches indeed combined Machine Learning (ML) and Deep Learning (DL) with geo-referenced content from the socials: for instance ElQadi et al. [28], which worked with Flickr geo-tagged pictures and an image classification algorithm provided by the Microsoft Azure services, examined whether those contents could bring any complementary information to a land-cover analysis of African territories provided by satellites; Samany [29], which exploited the work of 120 “wayfinders” both in terms of collecting the data with a Telegram bot and to validate the results, created a tool for automatically detecting landmarks analysing a picture; van Weerdenburg et al. [30], with a very interesting study on “soft data” (namely perceived and experienced qualities of places), on the other hand analysed Twitter tweets with three Natural Language Processing (NLP) models to extract the potential for urban leisure, tourism research and related city policies and planning in the smart context of the city of Zwolle.

The pipeline (illustration at **Figure 1**) we are presenting in this paper shares with most of the cited researches the goal of creating a data-driven analysis tool: particularly, we use the geo-tagged contents from both Instagram and Flickr to provide the administration and the stakeholders a power instrument for an unsupervised capturing of the user perspective about the “zones” of Turin. Kim et al. [31] also proceeded thought a three phases pipeline which consisted on data-collection, data-agglomeration and image-processing using a two years dataset of the city of Seul. Following a similar approach, we want to point out the following considerations:

- While they used Flickr as a unique source of data, in accordance with Salas-Olmedo et al. [32], which in their analysis of Madrid stressed out the importance of having different data sources using themselves three of them (Panoramio, a photo-sharing services, Foursquare, a social for consumption analysis and Twitter), we enlarged the image-classification task with Instagram. We think that this can bring additional value guaranteeing a more comprehensive research, also considering the different typologies of user characterizing the socials (as Christant [33]).
- Our algorithm makes no use of prior information about the city; each cluster is highly period-dependant and might change shape, size or even stop existing according to the time frame we are analysing. This provides a very dynamic and elastic tool, which only needs the raw data to automatically tuning on them: no hyper-parameters are required for its functioning, meaning that the analysis, as we will later show, can easily move within different years, filter data including or excluding observations (i.e. a weekend-only analysis is possible) or zooming in and out (i.e. deeper focus on the city center) according to the required analysis.
- Thanks to the aforementioned elasticity, we are able to handle multiple considerations (i.e. peaks-detection in correspondence of events; neighborhood comparison

among years; statistics per cluster) which could result in a more complete instruments to guide future urbanistic investments and return feedback on already completed ones.

2 Methodology

The algorithm we propose is structured on a four-stages pipeline which is represented on **Figure 1**; each of them will be independently analysed on the following sections. The examples we will propose come from the Turin's municipalities, but as previously mentioned the pipeline is generic and also suitable for different contexts.



Fig. 1: Proposed pipeline

2.1 Data Collection

The data collection represents the core of the digital footprint mining proposed in this discussion; collecting exhaustively descriptive and representative data here is indeed the key for the following clustering and classification steps. Being a differential analysis among different years one of our goal, we referred to a period which spans from 2017 to 2019. We decided not to include the 2020 in the analysis since, during the Covid-19 time span, lockdowns and limitations in movements have potentially defined another way of living cities and spaces. This is why, in order to avoid the risk to overlap a peculiar situation, the last year was not considered in the current analysis, and might be part of a further comparative studies.

Concerning the source of the pictures, we both considered Flickr and Instagram. The former is a website widely used by professional photographers and well appreciated by developers for its handily API, which allows queries with filters per years and per location and its richness of metadata comprehending the GPS coordinates. On our case, we download contents of the aforementioned years within a radius of 32 km from the $lat = 45.116177$ and $lon = 7.742615$. The latter is one of the most popular social networks in Italy [7] and it is exploited to get a more complete overview of the users' perceptions.

Unfortunately, there are some crucial drawbacks in the Instagram developer API: it only provides very few queries per hour; most of the meta data - including geo reference - are not open, meaning that differently from Flicker, it is not possible to obtain a continuous representation over the territory (**Figure 2**); it is possible only to do searches for *location ids* but there are no complete lists for a given area; the API simply rolls a desired profile's page and the research always starts from the most recent pictures.

With such restrictions, we follow the algorithm presented in **Algorithm 1**: this consists on two loops, one collecting the location ids from pictures of meaningful Instagram pages of Turin and the other downloading their information. The parameter N_{full} and N_{pic} were set respectively to 10.000 and 1.000. Clearly, it is crucial to feed the first loop with an heterogeneous set of pages, meaning that it should include different types of assets

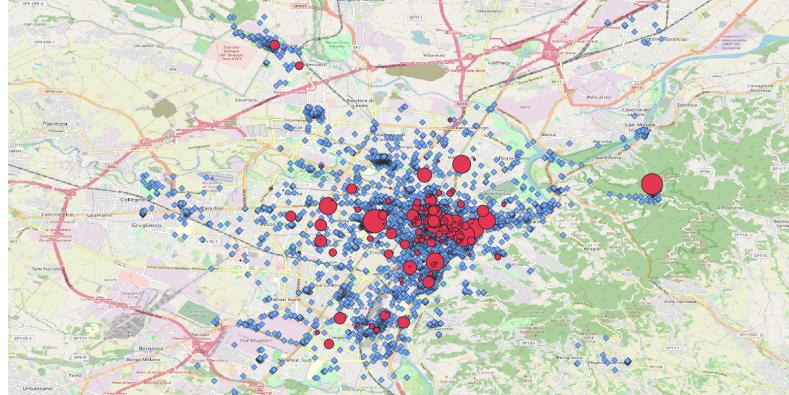


Fig. 2: The red round points represent the Instagram location ids; the size of each point represents the number of pictures for that location. The blue points on the other hand belong to the Flickr dataset

like shops activities, cultural-heritage resources and entertainment centers. This is indeed a delicate task, which risk to bias the analysis favouring only certain aspects if the set is not exhaustively describing.

Algorithm 1: Instagram dowloader

```

Choose a set of famous Turin Instagram pages ( $> N_{foll}$  followers);
while Still pages to analyse do
    while Analysed pictures for page  $< N_{pic}$  do
        location id = photo[location id] ;
        if New Location then
            | Save it
        else
            | Discad information
        end
    end
end
while Location to analyse do
    while Still photo  $\in [2017, 2019]$  do
        Download picture;
        Save metadata;
    end
end

```

When downloaded, all the data related to each photo are included in a document and uploaded into a MongoDB database. MongoDB is a NoSQL cross-platform document-oriented database program, indicated to deal with, as in our case, even quite large collections. Images are instead saved into a shared Dropbox folder on the Cloud and are linked to their relative metadata by an 'Unique Identifier'.

The results of the image mining are represented in terms of pictures' upload date on the respective social network in **Figure 3a** and **Table 3b**. While in overall both distributions

follow an increasing trend, the first thing which comes to attention is the spiky trend of the Flickr collection compared to the more stable behaviour of the Instagram one. As we will show on Section 3 however, those picks often correspond to interesting events organized in the city. This also reflects the different dataset users: while many professional photographers adopt Flickr as a way to publish their works, Instagram pictures are more exemplary for every-day occasions. The latter also suffered from the absence of API made for analytic purpose; this seriously impedes the research of past content, and therefore, our collection results are biased toward more recent pictures.

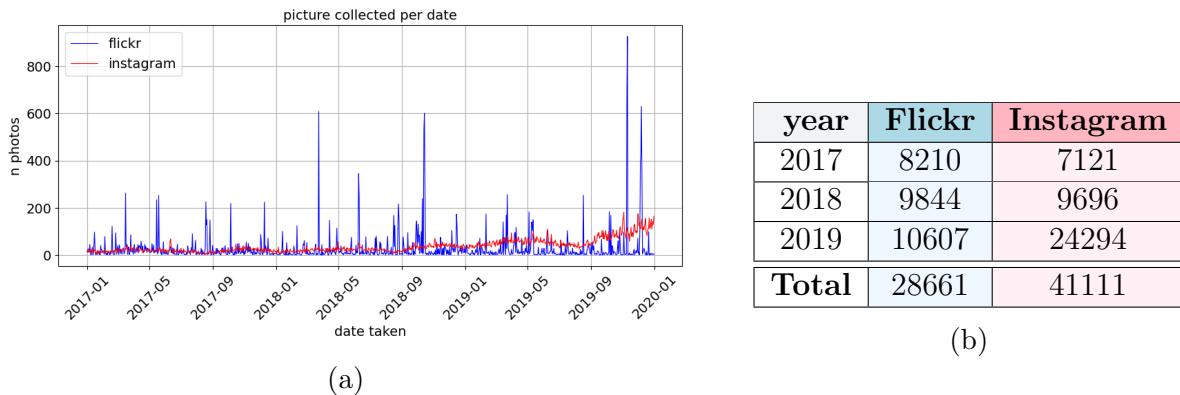


Fig. 3: In 3a the distribution of the two collection among the years; in 3b the total number of pictures downloaded, aggregated by year.

2.2 Clustering

The second step of the pipeline described in **Section 2** regards the detection of the city zones for the period under analysis.

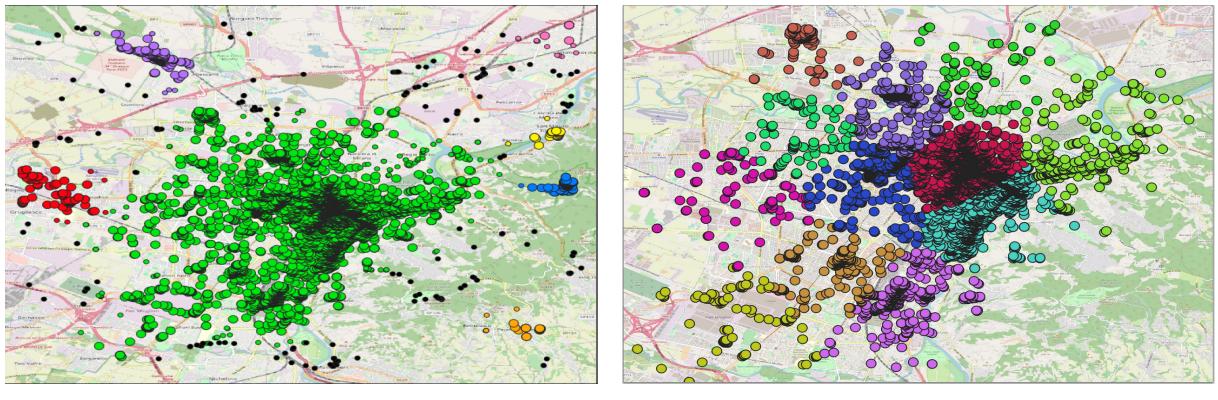
First of all, we have to clearly specify what we means with “zone”. This is not a district nor a census zone; both of them are indeed static elements, which does not change unless manually modified and do not represent an increased or decreased interest on an area. Our “zone” is a dynamic element, only determined by the density of downloaded points in the map (**Figure 4b** gives a visual representation). The main idea and basis for the aforementioned spatial and temporal elasticity is that, under the condition of having a representative collection, any changes in terms of shape, dimension or existence of each zone can be justified by a different perception of the users about that territory.

A practical example might be the construction of the Allianz Stadium (at the time, Juventus Stadium) in Turin [34]: this urbanization work has greatly modified the equilibrium and the importance of that locality, which in the years after has been able to attract new shops and activities. While the geographical boundaries of the district has remained untouched, the zones directly involved on the event has enlarged (new businesses has been created, starting from the supermarket in the nearby or more recently the entire area of *La Continassa* [35]). While official divisions are bounded to the “history of places” and are not able to rapidly handle those changes, the adoption of a data-correlated zoning is more able to adapt and reflect them.

However, as explained in the previous Section, the Instagram discrete distribution of points based on locations (**Figure 2**) prevents us from considering it a uniformly

representative dataset like in the Flickr case. This is the reason why we choose not to use Instagram for the construction of the clusters, which is therefore only resulting from the Flickr points. It is then important to remind that Instagram on the other hand will be precious for the following classification task, where a picture's perception might influence the zonal one depending on location id's position laying in the cluster or not.

The zones detection can be described in two sub-phases: an earliest subdivision of the points mainly aimed at cleaning the data from outliers (namely isolated points which could not be assigned to any zone) and the search of the major clusters, performed through the use of a Density-based spatial clustering (DBSCAN) algorithm (**Figure 4a**). After this first round of clustering, K-Means algorithm is subsequently applied to the largest cluster found by DBSCAN to find more fine-grained clusters (**Figure 4b**). As mentioned in the



(a) *DBSCAN; the major clusters are coloured while the outliers are black. Interestingly, the less numerous cluster corresponds to peripheral sites of interests, like Venaria (violet) and Superga (light blue)*

(b) *Kmeans; it represents a more detailed analysis on the most numerous cluster identified in 4a. Also on such a case, some parallelism with the “real district” can be drawn (i.e. garnet for the center, violet for Lingotto)*

Fig. 4: Results found in the three-year dataset

Background Section, this is a completely unsupervised technique which automatically discovers the best suited hyper-parameters which, on this case, are the radius ϵ and number of neighbours M for the DBSCAN part and the number of cluster K for KMeans. All the technical details of this process are explained in **Section Appendix**. To better understand and visualize the clustering results we use QGIS software; through a *dissolve* function, this tool allows passing from the discrete points distribution to a continuous polygon representation, as shown in (**Figure 5**). In this fashion, the subdivisions of the territory and the searched zones are clearly visible and their evolution and characteristics (number of images, density of images, etc.) are easier to study.

2.3 Image Processing

As outlined in **Figure 1**, the third step of our pipeline consists of the characterization of each cluster through the image classification process. The idea is to exploit one of the existing pre-trained networks on the well known dataset of *ImagNET*. Among them, we preferred *Inception v3* for its computational efficiency as suggested in Szegedy et al. [36]. This is aligned to the choice of Kim et al. [31], which also succeeded at achieving acceptable results without further re-training the model on their specific dataset. For this

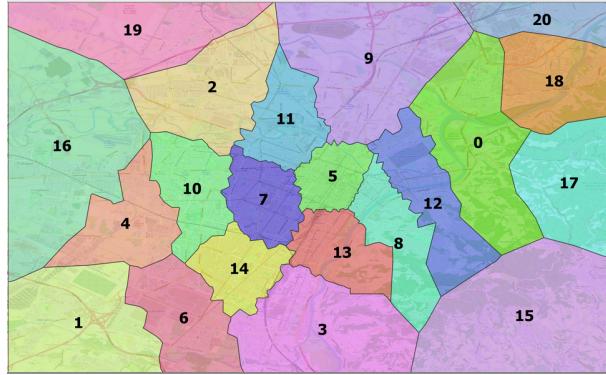


Fig. 5: *Clustered zones on the three years dataset; notice once more that, since no relationship exists with the “real nomenclatures” (except the ones that can be found retrospectively), we maintained numerical identifier to distinguish them*

reason and taking into account we are only interested on a general categorization (i.e. we want our model to label the *Mole Antoneliana* as *Palace/Monument*, not to distinguish it from the *Tour Eiffel*), we also do not refine the model’s weights with a further training.

It is important to notice that these Neural Networks are usually meant to output a single label as a result of the image processing; however, knowing that in the ImagNET contests they are able to achieve up to 0.937 Top-5 accuracy (see Ker [37]), we interpret the probability of each class as a level of confidence of the network of recognizing those classes into the picture. This is similar to what applied in the study of Ribeiro et al. [38], which used a local interpretable model-agnostic explanations (LIME) on the top 3 predicted classes of Google’s Inception, showing that their so called “superpixels” perfectly explain the parts that compose the images. **Figure 6** from the aforementioned paper gives a visual explanation.

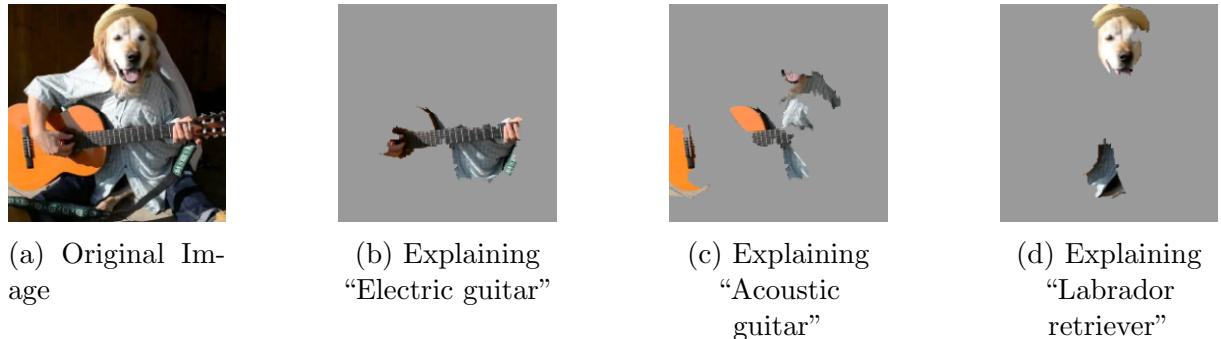


Fig. 6: *Explaining an image classification prediction made by Google’s Inception network. The top 3 classes predicted are “Electric Guitar” ($p = 0.32$), “Acoustic guitar” ($p = 0.24$) and “Labrador Retriever” ($p = 0.21$) [38]*

Doing so, we chose to handle differently the top 3 classes; each of them is added to the total score of that class on the cluster under study weighted for the confidence of the model of having recognized it on the picture. This way of interpreting the outputs of the algorithm seems to be coherent considering that the pre-training phase has been performed on a dataset where there was only an object per picture; on a real world example, this is rarely the case and this policy seems to fit a context where an object in

the foreground receives more attention but also the rest is considered proportionally to its importance in the image. A practical example is made in **Figure 7**.



(a) Photo from the dataset

Micro-classes	Macro-classes	Percentages
Fountain	Urbanscapes	0.888
Palace	Palaces	0.030
Fireboat	Transportation	0.006

(b)

Fig. 7: In this photo the first label identified is Fountain, the second one is Palace (Palazzo Madama is in the background) and then a Fireboat is identified with a very low probability. Notice that the fountain in foreground receives the highest attention, while the palace in the background is noticed but recognized as something of secondary importance. Interestingly, the model confuses the statue with a real person; however, the confidence score is very low, and this recognition will only contribute for 0.006 to the Transportation class, not affecting the quality of the work.

The last step has then to deal with the fact that Inception v3 is trained to output a thousand different labels, which are way too many for our scope. We therefore manually group the resulting sub-classes emitted by the model into macro ones, which are: *Natural views/Flora, Animals, Shopping/shops/clothing/merchandising, Transportation, Urbanscape, Exhibits/sculptures/museums, Religion, Residence, Entertainment, People*.

It is finally important to underline that while the clustering results are dynamic and period-dependant, the classification ones are attributes of each picture and do not change. Therefore, the labels are saved in the database together with their confidence levels.

3 Analysis

This Section aims at showing the possible applications and benefits of the described tool. The first example exploits the results of the image classification process to capture the users perceptions on each of the zones identified in the clustering phase; it both includes a three-years analysis and a focus on the city-center to demonstrate its spatial adaptability.

The second one on the other hand focus on a peak-detection task. Not every zone “is lived” uniformly over a year; a particularly relevant aspect, for example the entertainment coming from a concert, is peaked on a particular day and it might be interesting to also take this into account.

Thirdly, we present a 2017 vs 2019 comparison based on the clustering results only to prove its temporal elasticity. As previously mentioned, an analysis on the differences

of shapes and dimensions of each zone might also bring some information on the users presence and activity. A severe reduction, like in the case we will present, could also shade some lights to the administration to investigate why this emptying process occurred.

Finally, we have created synthetic indexes which can perform inter-zonal analysis, measuring for example the degree of cultural interest of a zone on respect to another or drawing a parallelism with the real cadastral value.

3.1 Global analysis

The first proposed analysis regards the autonomously learnt perception of the territory.

Figure 8 shows the subdivision of the territory into the various clusters and each of them is colored with the color of its main class. Firstly, the number of collected pictures seems coherent with the attractiveness of the area - intuitively related to the closeness to the center and to the presence of monuments or landmarks. It is then visible a predominance of the classes of *Entertainment*, *Palaces* and *Transportation*, which are certainly representative of a large city like Turin.

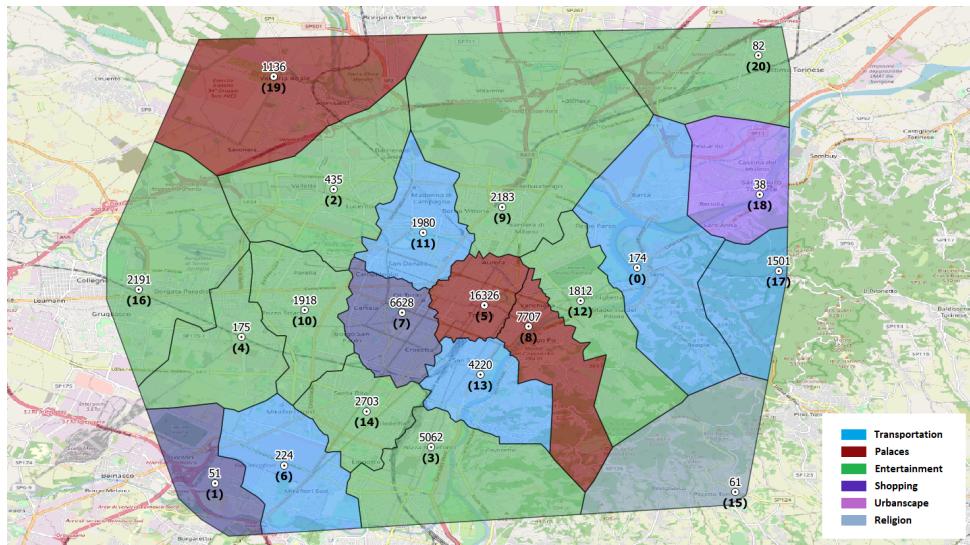


Fig. 8: Results of the clustering and classification applied on the whole period under study (2017-19). Zones are obtained as minor clusters of DBSCAN joined to results of KMEANS on the main one (each one numbered by the number below the point). Each zone is correlated by extra information such as: number of pictures collected there (number on top of point), baricenter of the pictures (white point), and the most represented category (color of the zone).

In general the zones perceptions are very faithful to our knowledge of the city. An example is Cluster (14) shown in **Figure 9a**; here the points are more dense in the district corresponding to the stadium (1097 pictures over the total 2703). Furthermore, according to **Figure 9a**, *Entertainment* is the macro-class with the highest score; considering the building's use for concerts and events (**10a** this seems reliable).

On the contrary, we have noticed that in some other cases a bias exists due to the large number of images mostly derived from events in those areas in the years analysed. As evidence of this, **Figure 9b** shows that in Cluster (13) the presence of the event called

“Salone dell’Auto di Torino” held at the *Parco del Valentino* in 2019, leads to have a small deviation from the expected results. Being this park one of the most important of the city and famous for its historical buildings, fountains and statues, a strong presence of the classes *Urbanscape*, *Palaces* and *Natural Views* was predictable. This is the case - the class *Natural Views* has indeed obtained the highest score among the central clusters - but, due to the high number of pictures taken in correspondence to the event, the main class is *Transportation*. The algorithm therefore not only correctly captures the general perception, but is also able to handle temporary deviation. This could lead to a double potential use of discovering the city and its main assets and/or deepening the knowledge about it with its hidden aspects.

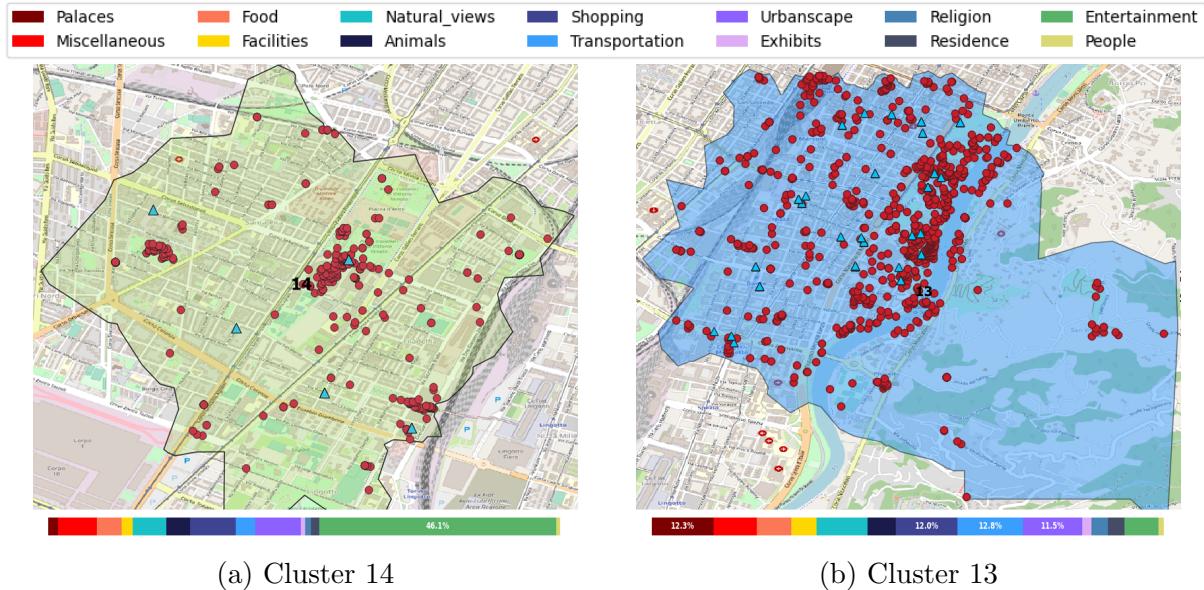


Fig. 9: Zoom of areas 13 and 14 of **Figure 8**; dots represent the location where Flickr pictures were collected while triangles are the Instagram landmarks. The background color reflects the most popular class but a distribution of the classification of all the pictures is attached

From **Figure 8** emerges the great numerosity of photos present in Cluster (5), the so called *main cluster*; for this reason we have decided to analyze it in the detail separately and to apply an ulterior clustering inner to this zone.

3.2 Central cluster analysis

As already observed for **Figure 8**, also **Figure 11** has been subdivided and then colored with respect to the most important class of each cluster. Besides, a picture from that class for the most relevant areas is attached to the map. Here we notice how the distribution of the main zones is focused on the classes of *Food*, *Palaces* and *Shopping*; while the presence of historical building is therefore confirmed from the global study, here food and shopping take over as auxiliary assets provided to visitors.

More precisely, **Figure 12** shows the exact percentages and quantities of classes for each clusters with the respective numerosity. With the assumption that a higher numerosity corresponds to a more confident analysis since able to capture a more complete



(a) Cluster 14, Entertainment



(b) Cluster 13, Transportation

Fig. 10: Collection of some pictures belonging to clusters 14 and 13 matching their most represented category

perception of the area, we will analyze some of the clusters with the largest number of photos.

The central red zone, corresponding to the Palaces class, is composed by some of the heavier clusters like Cluster (11), (13) and (10). In these areas there is a large presence of the most important landmarks of the city; the main streets are brimful of architectural references and magnificent hundred-year-old constructions, especially in the most important squares such as Piazza San Carlo and Piazza Castello and with buildings like Palazzo Madama and Palazzo Carignano.

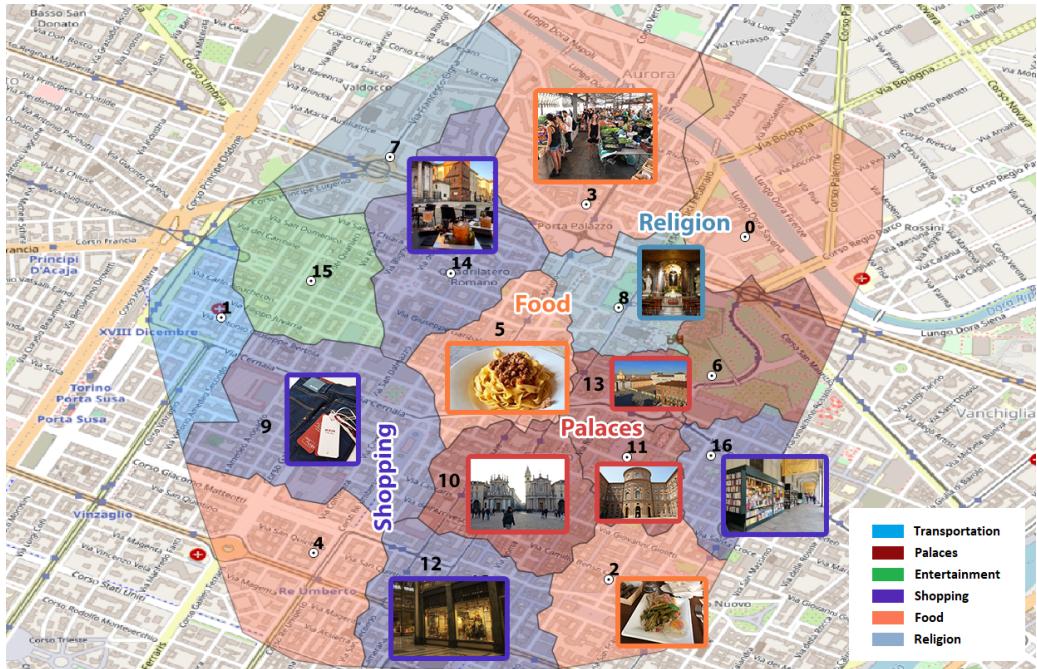


Fig. 11: Clusters of the central zone with the color of their most representative category and example photos for each.

Another large cluster is the (8), in which the *Religion* class is the master, this is easily

associated with the fact that in this zone contains the main churches of Turin, like the dome or the *Chiesa di San Lorenzo*, which definitely are important points for tourists.

In the middle of the map there is the Cluster (5) dominated by the *Food* class. This area includes the *Quadrilatero* district, famous for its bars and restaurant, and also the *Porta Palazzo* zone with its market and all the streets located around *Via Garibaldi* where it is customary to go for lunch or for aperitif. Then, in Cluster (12) there is an important shopping prevalence in front of *Porta Nuova*, which is reliable since it contains the shopping-street of *Via Roma*. Cluster (16) is both characterized by the high number of shops of *Via Po* and the many peculiar stalls selling books. Cluster (15) differs from the others being the only zone with a strong prevalence of the *Entertainment* class; this is highly influenced by a *TedX talk* (see [39]) and by the central library of the city. Finally, Cluster (1) coherently presents a prevalence of the *Transportation* class that corresponds to the presence of the *18 Dicembre* square, stop of buses, trams and metro. Both of these last two clusters however have a very low number of photos; their results might therefore not be so reliable.

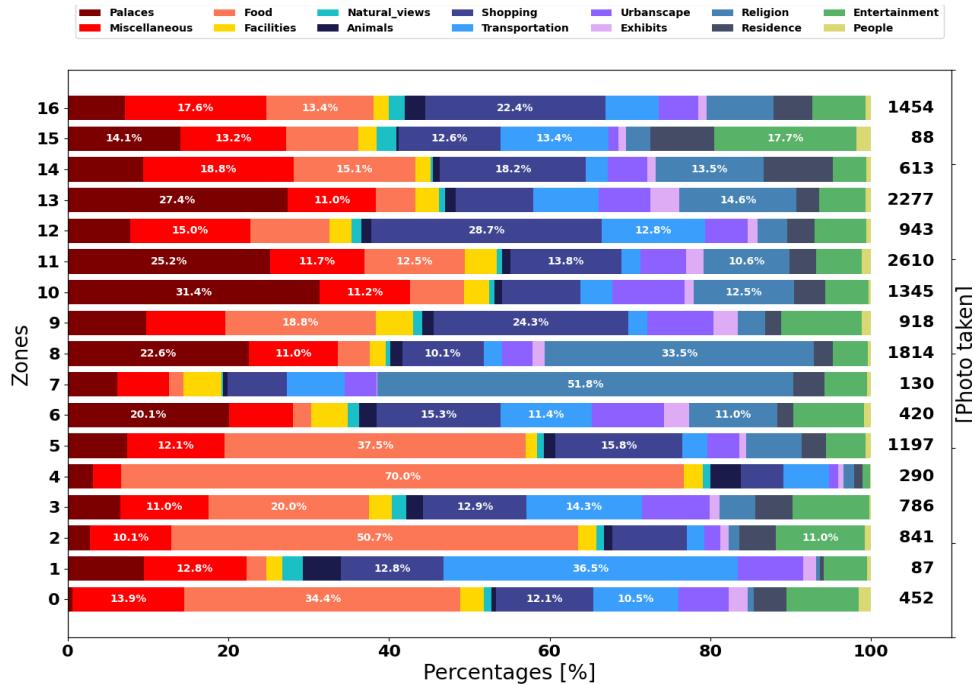
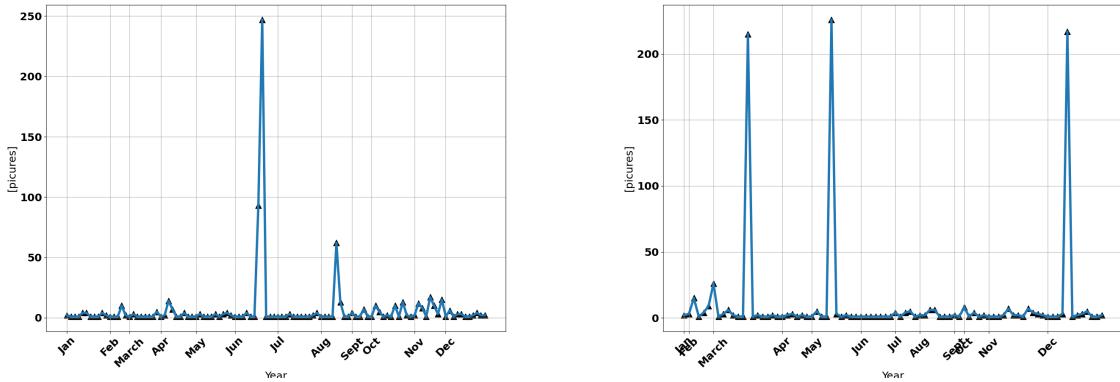


Fig. 12: Weighted results of the picture classification for each cluster of the central zone

3.3 Peak analysis

While the analysis we presented in the previous section successfully captures the general users perception, it fails at identifying which part of it is due to temporary events (like the *Salone dell'Auto* example of cluster 13 on **Figure 10b**) and which on the other hand represent its backbone. However, a good hint might come from the peaky trend already observed on **Figure 3**, especially for what regards the Flickr dataset. An important occasion indeed tends to behave like an outlier on respect to the generic tendency over the year: it is the case of the already mentioned cluster 13 (**Figure 13a**), but also cluster

14 of the *Stadio Grande Torino* and *Pala Alpitur*, often used for sport events and concerts (**Figure 13b**).



(a) *Cluster 13 trend; the peaks corresponding to the even are clearly identifiable*

(b) *Cluster 14 trend; also on this case three concerts (Ed Sheeran, Kiss and Depeche Mode) clearly stands out*

Fig. 13: *Peak-detection task*

The effectiveness of the method on the task of event-capturing is nevertheless very dependent on the completeness of the dataset; for example, while one would expect to find peaks corresponding to sport events on the Allianz Stadium area (cluster 2 of **Figure 8**), that is not the case since only one peak is actually present (Juventus-Young Boys, which curiously is not even a *big match*). In general, this is the main issue of referring mainly to the Flickr dataset for this purpose; Instagram pictures are indeed mostly referred to a every-day behaviour and hardly present peaks. This is why a possible improvement for the work could be using *Instagram stories* instead of pictures, which to our knowledge could better fit the scope.

3.4 Yearly comparison: 2019 vs 2017

The third analysis we present mostly refers to the second step of the pipeline, thereby the clustering phase. As mentioned on **Section 2.2**, different datasets of the same city might result on different zones depending on the period under study. An increased number of pictures might indicate that an area has been “lived” more in terms of events or activities; less feedback on the other hand could be an alert signal for the administration.

The example we propose regards a differential analysis between 2017 and 2019.

Already in the DBSCAN phase, as **Figure 14a** and **Figure 14b** testify, the algorithm picks very different choices: while 2017 is characterized by a best ϵ value of 1.3 km, this is only 0.5 km in 2019. According to the **Appendix 4**, the choice of a smaller radius can be justified by a more dense distribution of points; on numerical terms, while on average a point in 2017 has its closest neighbor 700 m far, in 2019 this distance is reduced to 300 m. Intuitively, since the distances among points are smaller, the model tends to choose a smaller contour for the cluster as well. This eventually results also on identifying more noise, namely isolated points from which nothing else is within the chosen range (black dots on **Figure 14**).

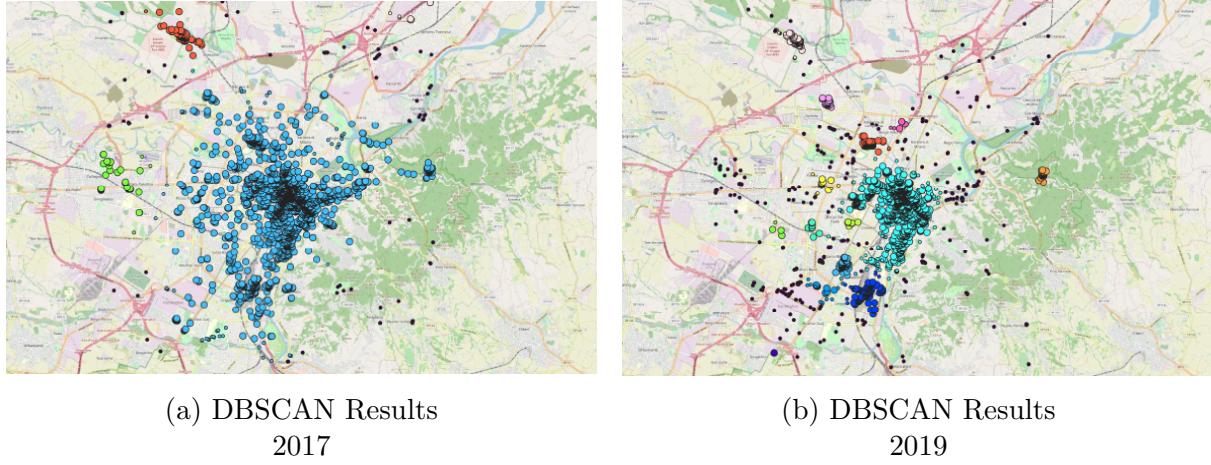


Fig. 14: *Comparison between years*

Figure 15 then, is an heat-map dividing Turin in squares of about 500 m; a square can have red tones if 2019 have more pictures or blue ones on the other case.

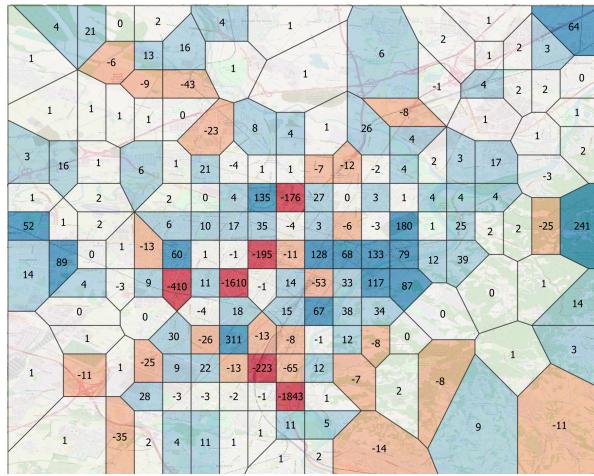


Fig. 15: *Visual representation of the different number of pictures collected on 500m-diameter areas of the city between years 2017 and 2019 ($|2017| - |2019|$)*

Interestingly, 2019 has squares far more represented than 2017; this is mainly due to a couple of sports event widely pictured, namely the *Gran Prix Finals of Ice Skating* and the *Torino Climbing Marathon*. On the other hand, in 2017 an entire area seems to be more represented: this covers the coast of river Po just a few chilometers from the center, and also comprehend the real zone of *Murazzi*, famous for its bars and clubs, and the cultural site of *Villa della Regina*. Thus, while in 2017 those assets have been visited and pictured (**Figure 16a** and **Figure 16b**, in 2019 their scarcity and sparsity even induce the algorithm to label them as noise).

To conclude, the idea is that the administration could therefore benefit from those pieces of information, both keeping promoting very attractive onetime events like in 2019 and trying to reinforce the everyday backbone like in 2017.

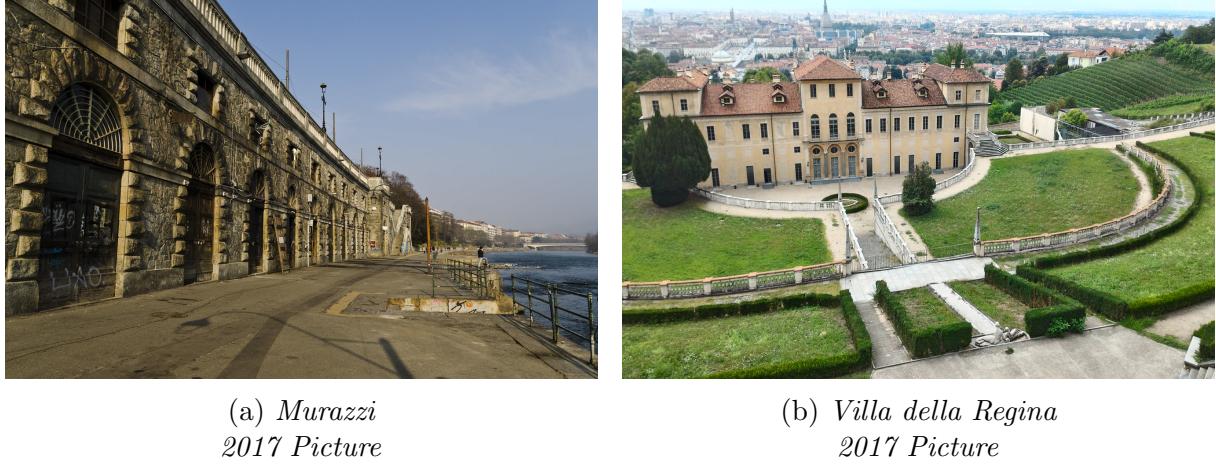


Fig. 16: *2017 vs 2019
locations more represented during 2017*

3.5 Synthetic Indexes

As aforementioned, synthetic indexes can help understanding the attractive strength, in a given context, of a zone in relation to another. Building indices regarding the major attractive forces that the city offers and that the administration may wish to encourage coherently fits the idea of our unsupervised analysis. Among the different possibility, we respectively examine the tourists attraction capacity, the presence of cultural heritage, and finally we propose an estimation of the zonal cadastre value. Each of these features require taking into account different qualities of an area: for instance, while the cultural heritage might be simply described by the presence of historical palaces and museums, the tourist attraction capacity also requires thinking at how much a zone can offer in terms of catering facilities and shopping opportunities. We will demonstrate that not only the model successfully fulfil our expectations based on the knowledge we have of the city, but it also returns reasonable results compared to a real estimator like the cadastre value.

We generally build our indexes as presented on **Alg. 2**; notice that *perception* for each picture refers to the InceptionV3 score described in **Section 2.3**.

The choice of weighting the total score for distances' standard deviation was introduced in order to encourage areas with a good density of points and not focusing on few centers of attraction which, in some zones, can distort the reading of the indices themselves. **Figure 17** gives a visual representation.

3.5.1 Attraction capacity indexes

In order to estimate the attractiveness of an area, we have synthesized a couple of indexes, namely \hat{I}_t for the Tourists attraction capacity and \hat{I}_c for the Cultural attraction capacity. We collected all those macro-classes that intuitively influence the decision of tourists in the choice of their destinations for the first, while all the macro-classes involved in cultural activities for the second.

$$\begin{aligned}\hat{I}_t &= \sum(\text{Palaces}, \text{Food}, \text{Shopping}, \text{Exhibits}) \\ \hat{I}_c &= \sum(\text{Palaces}, \text{Exhibits})\end{aligned}$$

Algorithm 2: Constructing synthetic index

```

 $SumScores = 0;$ 
Collect points of a zone from database;
Sort the points for latitude and longitude ;
while Still points to analyse do
| Pick point  $p$  at position  $i$ ;
| Pick point  $q$  at position  $i+1$ ;
| Evaluate Haversine Distance between  $p$  and  $q$ ;
| Get perception of point  $p$ ;
| if perception  $\in$  macro classes forming the index then
| |  $SumScores += perception$ 
| else
| | Pass
| end
end
Sort distances on ascending order;
 $stdDev$  = standard deviation of distances;
 $FinalScore = \frac{SumScores}{stdDev}$ 

```

As it is possible to notice, \hat{I}_c is incorporated in \hat{I}_t since the presence of cultural heritage is undoubtedly one of the factors of choice that most influence potential tourists.

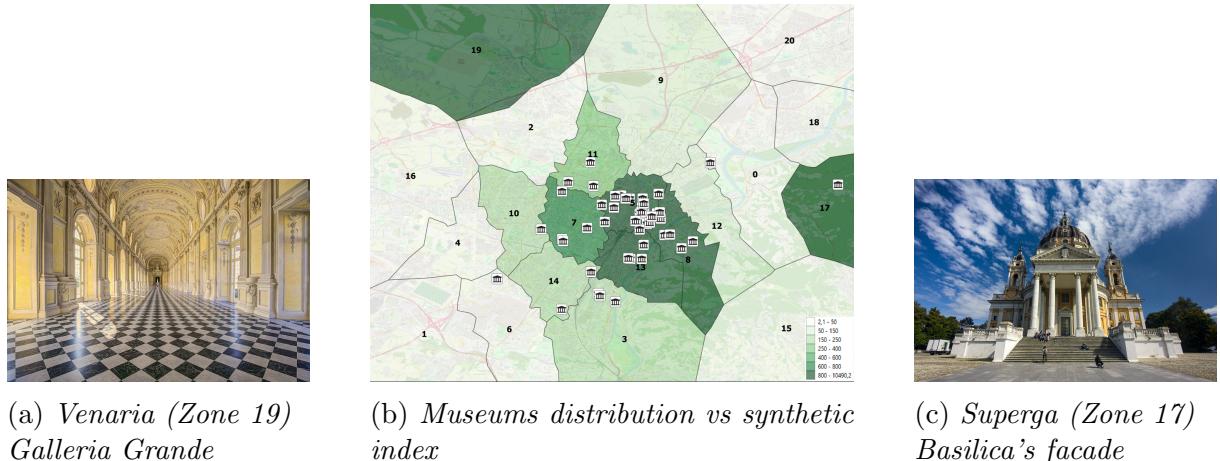


Fig. 18: *Cultural analysis*

At first a simple analysis and comparison on the cultural impact (**Figure 18b**) can be performed examining the distribution of the museums' locations over the studied area. The synthetic index seems to reflect concentration of the museums; furthermore, it also captures important landmarks which are more generally part of the cultural heritage of the city. Despite not containing galleries or having only a limited amount of them, clusters 17 and 19, respectively the area of Venaria and Superga (**Figure 18a** and **Figure 18**), are indeed crucial artistic and historical assets. Both of them are representative of the Savoia family, with the first one being their summer and hunting palace and the second their burial place.

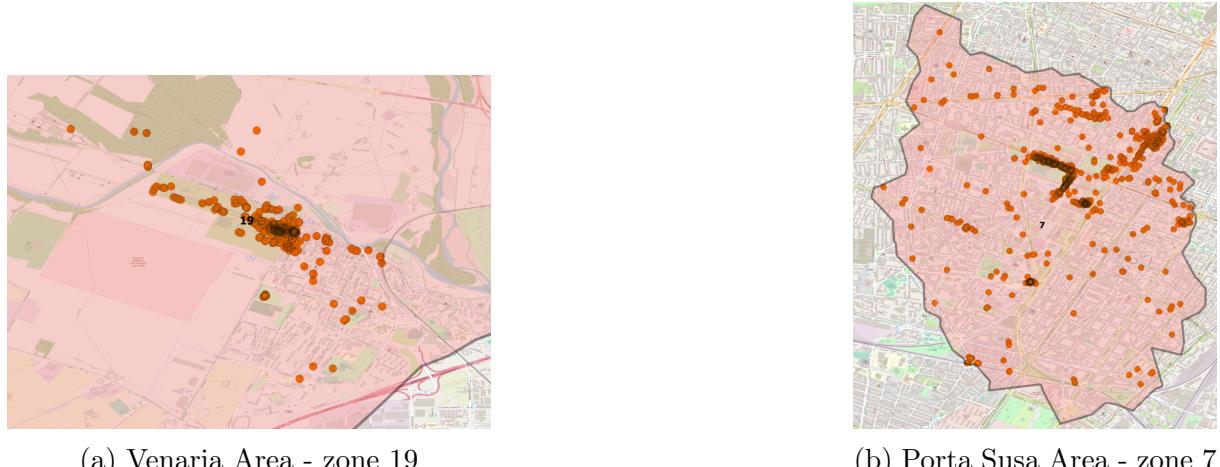


Fig. 17: *Density comparison.* The unweighted scores regarding the cultural heritage were respectively 251.36 and 232.12, with a very tight advantage of the area of Venaria. However, we must consider how well represented is the zone under analysis; on **Figure 17b** the standard deviation of closest points is of 0.39 km, while on **Figure 17a** is only 0.19 km. This means that in the Porta Susa's case points are sparser than the Venaria's counterpart; since few and sparse centers of interest might be giving a deceiving contribution (in the Porta Susa example, the score is obtained mainly thanks to the richness of the building's architecture in that area), we decided to favour more dense representations. On the aforementioned case, this leaded to a score of 647.52 for zone 7 and of 1260.729 for 19, which for our knowledge seems a result far more reliable.

Focusing again on the city center to better analyze the peculiarity of this second index, we remind that this is not only based on *Palaces* and *Exhibits* as the first one, but also includes *Food* and *Shopping* because those are intuitively the complementary facilities to attract tourists.

Also referring to the statistics of **Figure 12**, clusters 8, 10, 11, 13 are the ones with the greatest cultural impact and also present considerable scores in terms of food and shopping; it therefore makes sense that they labeled as the most attractive. On the other hand, clusters 5, 12, 16, despite not being so attractive from the cultural point of view, are active areas in terms of secondary facilities; indeed, zone 12 is the one of *Via Roma*, an hotspot for the shopping in Turin, while zone 16 is the one of *Via Po*, which similarly

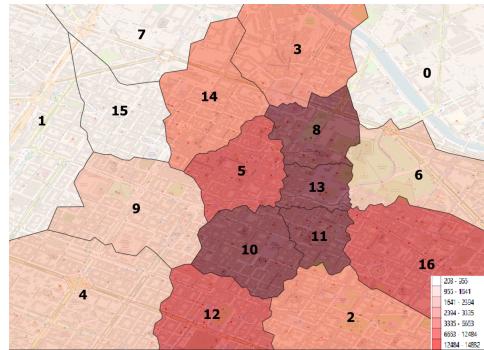


Fig. 19: *Tourists attraction capacity*

contains important shops and catering services.

To conclude, \hat{I}_t and \hat{I}_c are obviously estimates, as there might be other variables involved in measuring those aspects; nevertheless, the results shown seem to be largely in line with our knowledge of the area.

3.5.2 Cadastral value

On the basis of the same approach used before, we create another index, namely I_{cad} , taking into account the macro-classes that possibly most influence the cadastral value in a given zone. The analysis performed by Seth [40], on its section *The Most Important Factors for Real Estate Investing. Property location - Why it's important*, suggests that factors like “green space”, “scenic views”, “closeness to markets, warehouses, transport hubs” might be well representative. This is why we choose to approximate this real cadastral index with the following formulation:

$$\hat{I}_{cad} = \sum(Palaces, Shopping, Transportation, Natural_View)$$

The way we distributed the cadastral values of the Turin’s census areas on our zones simply consists on projecting the census areas’ centroids on the map, spatially joining them with the identified clusters and averaging in case of multiple centroids contribution. The result is shown on **Figure 20a**.

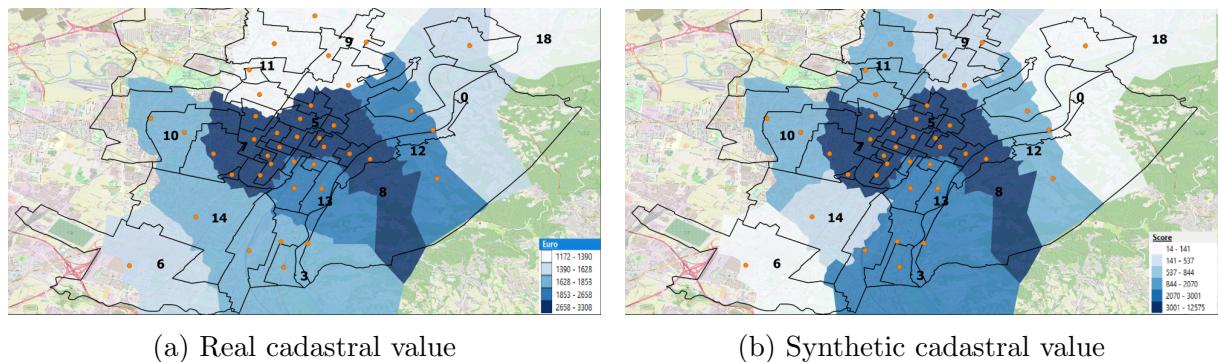


Fig. 20: *Cadastral value estimation; the black contours corresponds to the census zone of Turin ([41]), while the orange points are their centroids.*

The comparison between the real indexes of **Figure 20a** and the synthesized one of **Figure 20b** seems very promising; clusters 5, 7 and 8 are on both cases the most valuable, with the overall price getting lower while moving towards the suburbs. However, there are cases of slight divergences which are worth to be mentioned: for example, the estimated index regarding the Lingotto zone on cluster 3 is biased by the large presence of *Shopping* (due to shopping center it hosts and to its many temporary events) and *Transportation* (it contains a train-station, a metro-terminal and a stop for the Flixbus service); those increase its synthetic value, while the real one is lower because it is actually part of the suburb. Another example is the one of cluster 12; its average real value includes the area of hills of Turin, which historically has always hosted nobles and celebrities residences. This aspect is partially caught by the **Natural_View** macro-class; however, this is clearly not enough to capture such a complex and Turin-related concept, and therefore to some extent our index lacks at estimating its real value.

Finally, it is important saying that a possible improvement might consist on regressing a monetary value from the features regarding the perceptions exploiting machine learning techniques; this could allow having a real-monetary vs estimated monetary comparison, which would indeed result on a more fair analysis.

4 Conclusion

This research focused on providing the administration and the stakeholders a generic pipeline to guide their effort and investments. The tool is completely unsupervised: no prior temporal or spatial information are needed, and it automatically tunes its weights on the provided data. This means that it can elastically analyse periods and territories of different sizes.

The obtained results on the city of Turin are promising both compared to real indicators, like the cadastral value or the museums' map, and to our knowledge of the city. The model can capture the general perception of a zone (i.e. strong presence of *palaces* and touristic facilities for the centre; remarkable of *nature* for the city-park) and its peculiar and more hidden aspects (i.e. big events like the “Salone dell’auto”); furthermore, it ranges among a considerable set of potential studies like *intra* and *inter zonal* comparisons, *peak detection* and *period to period* analysis. These kind of results could have concrete applications: for what concern the city of Turin and the current projects in the moment of the studies, they could be starting points for the regional consultation about the *DSU* in the context of the *PNRR* and the European programming for the 2021-2027 [42]. Most of their key points such as inclusion, connection or sustainability can be analyzed through our tool.

Nevertheless, further efforts might be needed for increasing both the quality and the quantity of collections. Finding solution to perform more exhaustive downloads from Instagram would indeed help having more uniform data over the different years (biased in our case towards the last one), and preventing sparse points, which mostly represent only partial information that often polarizes the perception of an area far from its real one. Moreover, finding a way of adding Instagram “stories” to the posts would guarantee an even more consistent every-day backbone provided by the social. Finally, since more punctual analysis like a weekend vs weekdays one requires a great amount of data for being meaningful, considering additional source of data (Twitter, Tripadvisor, etc.) could improve the overall quality of the results.

References

- [1] Zolzaya Dashdorj, Stanislav Sobolevsky, SangKeun Leea, and Carlo Ratti. Deriving human activity from geo-located data by ontological and statistical reasoning. *Knowledge-Based Systems*, 12 2017. URL <https://www.journals.elsevier.com/knowledge-based-systems>.
- [2] Robert Hollands. Will the real smart city please stand up? *City*, 12:303–320, 12 2008. doi: 10.1080/13604810802479126.
- [3] R. Kitchin. The real-time city? big data and smart urbanism. *Urban Research eJournal*, 2013.
- [4] A. Townsend. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. 2013. ISBN 978-0-393-08287-6.
- [5] Gil Appel, Lauren Grewal, Rhonda Hadi, and Andrew Stephen. The future of social media in marketing. *Journal of the Academy of Marketing Science*, 48, 10 2019. doi: 10.1007/s11747-019-00695-1.
- [6] Hootsuite We are Social. Report digital 2020, 2020.
- [7] Similarweb, 2020. URL <https://www.similarweb.com/>.
- [8] London. Smart london, 2021. URL <https://www.london.gov.uk/what-we-do/business-and-economy/supporting-londons-sectors/smart-london>.
- [9] Amsterdam. Amsterdam smart city. URL <https://amsterdamsmartcity.com/>.
- [10] Singapore. Transforming singapore through technology. URL <https://www.smartnation.gov.sg/>.
- [11] Turin. Welcome to to city lab. URL <https://www.torinocitylab.it/en/>.
- [12] Iot al servizio della smart city: l'esempio di torino, 2020. URL <https://www.lumi4innovation.it/iot-al-servizio-della-smart-city-lesempio-di-torino>.
- [13] Official journal of the european union, 2016. URL <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>.
- [14] Fabien Girardin, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE*, 7:36 – 43, 01 2009. doi: 10.1109/MPRV.2008.71.
- [15] Dietmar Offenhuber and Carlo Ratti. *Decoding the City: Urbanism in the Age of Big Data*. 09 2015. ISBN 978-3038215974.
- [16] Fabien Girardin, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *J. Location Based Services*, 2:41–56, 03 2008. doi: 10.1080/17489720802261138.

- [17] F. Girardin, Andrea Vaccari, A. Gerber, A. Biderman, and C. Ratti. Quantifying urban attractiveness from the distribution and density of digital footprints. *Int. J. Spatial Data Infrastructures Res.*, 4:175–200, 2009.
- [18] Bálint Kádár and Mátyás Gede. Where do tourists go? visualizing and analysing the spatial distribution of geotagged photography. *Cartographica The International Journal for Geographic Information and Geovisualization*, 48:78–88, 06 2013. doi: 10.3138/carto.48.2.1839.
- [19] Ksenia Mukhina, Stepan Rakitin, and Alexander Visheratin. Detection of tourists attraction point using instagram profiles. 06 2017.
- [20] Tobias Brandt, Johannes Bendler, and Dirk Neumann. Social media analytics and value creation in urban smart tourism ecosystems. *Information & Management*, 54(6):703–713, 2017. ISSN 0378-7206. doi: <https://doi.org/10.1016/j.im.2017.01.004>. URL <https://www.sciencedirect.com/science/article/pii/S0378720617300125>. Smart Tourism: Traveler, Business, and Organizational Perspectives.
- [21] Bálint Kádár. Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*, 16(1):88–104, 2014. doi: 10.1080/14616688.2013.868029. URL <https://doi.org/10.1080/14616688.2013.868029>.
- [22] Danah Boyd and Kate Crawford. *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*. Taylor & Francis, information, communication & society edition, May 2012. URL <https://www.microsoft.com/en-us/research/publication/critical-questions-for-big-data-provocations-for-a-cultural/>.
- [23] Ning Deng, Jiayi Liu, Ying Dai, and Hong Li. Different cultures, different photos: A comparison of shanghai’s pictorial destination image between east and west. *Tourism Management Perspectives*, 30:182–192, 2019. ISSN 2211-9736. doi: <https://doi.org/10.1016/j.tmp.2019.02.016>. URL <https://www.sciencedirect.com/science/article/pii/S2211973619300273>.
- [24] José A. Donaire, Raquel Camprubí, and Nuria Galí. Tourist clusters from flickr travel photography. *Tourism Management Perspectives*, 11:26–33, 2014. ISSN 2211-9736. doi: <https://doi.org/10.1016/j.tmp.2014.02.003>. URL <https://www.sciencedirect.com/science/article/pii/S2211973614000087>.
- [25] Pietro Florio, Maria Cristina Munari Probst, Andreas Schüler, and Jean-Louis Scartezzini. Visual prominence vs architectural sensitivity of solar applications in existing urban areas: an experience with web-shared photos. *Energy Procedia*, 122:955–960, 2017. ISSN 1876-6102. doi: <https://doi.org/10.1016/j.egypro.2017.07.437>. URL <https://www.sciencedirect.com/science/article/pii/S1876610217333921>. CISBAT 2017 International ConferenceFuture Buildings & Districts – Energy Efficiency from Nano to Urban Scale.

- [26] Giacomo Chiesa. Social indicators to localize renewable energy sources considering their visual impacts. *Energy Procedia*, 122:529–534, 2017. ISSN 1876-6102. doi: <https://doi.org/10.1016/j.egypro.2017.07.315>. URL <https://www.sciencedirect.com/science/article/pii/S1876610217329132>. CISBAT 2017 International ConferenceFuture Buildings & Districts – Energy Efficiency from Nano to Urban Scale.
- [27] Aurlien Gron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 1st edition, 2017. ISBN 1491962291.
- [28] Moataz Medhat ElQadi, Myroslava Lesiv, Adrian G. Dyer, and Alan Dorin. Computer vision-enhanced selection of geo-tagged photos on social network sites for land cover classification. *Environmental Modelling & Software*, 128:104696, 2020. ISSN 1364-8152. doi: <https://doi.org/10.1016/j.envsoft.2020.104696>. URL <https://www.sciencedirect.com/science/article/pii/S1364815219308059>.
- [29] Najmeh Neysani Samany. Automatic landmark extraction from geo-tagged social media photos using deep neural network. *Cities*, 93:1–12, 2019. ISSN 0264-2751. doi: <https://doi.org/10.1016/j.cities.2019.04.012>. URL <https://www.sciencedirect.com/science/article/pii/S0264275118316470>.
- [30] Demi van Weerdenburg, Simon Scheider, Benjamin Adams, Bas Spierings, and Egbert van der Zee. Where to go and what to do: Extracting leisure activity potentials from web data on urban space. *Computers, Environment and Urban Systems*, 73:143–156, 2019. ISSN 0198-9715. doi: <https://doi.org/10.1016/j.comenvurbssys.2018.09.005>. URL <https://www.sciencedirect.com/science/article/pii/S0198971518302138>.
- [31] Dongeun Kim, Youngok Kang, Yerim Park, Nayeon Kim, and Juyoon Lee. Understanding tourists' urban images with geotagged photos using convolutional neural networks. *Spatial Information Research*, 28, 07 2019. doi: 10.1007/s41324-019-00285-x.
- [32] María Henar Salas-Olmedo, Borja Moya-Gómez, Juan Carlos García-Palomares, and Javier Gutiérrez. Tourists' digital footprint in cities: Comparing big data sources. *Tourism Management*, 66:13–25, 2018. ISSN 0261-5177. doi: <https://doi.org/10.1016/j.tourman.2017.11.001>. URL <https://www.sciencedirect.com/science/article/pii/S0261517717302406>.
- [33] Ferdy Christant. The rise, fall and resurrection of flickr, 2019. URL <https://ferdychristant.com/the-rise-fall-and-resurrection-of-flickr-ca1850410ee1>.
- [34] N. Lekakis. The urban politics of juventus' new football stadium. *Articulo - Journal of Urban Research*, 2018. doi: <https://doi.org/10.4000/articulo.3349>.
- [35] Progetto continassa juventus. URL <https://continassa.jimdofree.com/>.
- [36] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision, 2015.

- [37] Keras applications, 2019. URL <https://keras.io/api/applications/>.
- [38] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [39] Tedx torino salon - game over / learn over, 2019. URL <https://www.tedxtorino.com/evento/game-over/>.
- [40] Shobhit Seth. The most important factors for real estate investing. URL <https://www.investopedia.com/articles/investing/110614/most-important-factors-investing-real-estate.asp>.
- [41] Osservatorio immobiliare città di torino. URL http://www.oict.polito.it/microzone_e_valori.
- [42] Piemonte 2021-2027. URL <https://piemonte2021-2027.eu/>.
- [43] Dbscan - wikipedia. URL <https://en.wikipedia.org/wiki/DBSCAN>.
- [44] Nadia Rahmah and Imas Sitanggang. Determination of optimal epsilon (eps) value on dbscan algorithm to clustering data on peatland hotspots in sumatra. *IOP Conference Series: Earth and Environmental Science*, 31:012012, 01 2016. doi: 10.1088/1755-1315/31/1/012012.
- [45] Kevin Arvai. Knee/elbow point detection, 2020. URL <https://www.kaggle.com/kevinarvai/knee-elbow-point-detection>.
- [46] Tara Mullin. Dbscan parameter estimation using python, 2020. URL <https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd>.
- [47] S. Joel Franklin. Effect of outliers on k-means algorithm using python, 2019. URL <https://medium.com/analytics-vidhya/effect-of-outliers-on-k-means-algorithm-using-python-7ba85821ea23>.
- [48] Imagenet. URL <http://www.image-net.org/index>.
- [49] Nghia Duong-Trung, Luyl-Da Quach, and Chi-Ngon Nguyen. Learning deep transferability for several agricultural classification problems. *International Journal of Advanced Computer Science and Applications*, 10(1), 2019. doi: 10.14569/IJACSA.2019.0100107. URL <http://dx.doi.org/10.14569/IJACSA.2019.0100107>.
- [50] Sik-Ho Tsang. Review: Inception-v3 — 1st runner up (image classification) in ilsvrc 2015. URL <https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77>

Appendix

Appendix A

After the first run, both datasets still required some fixes: for what concerns Flickr (blue line in **Figure 21a**) there was a noticeable hole in the collection rate in 2018. This was found out to be a limitation of the API that, without warnings, caps the maximum number of pictures obtained per query. A finer second run was performed; the light blue line represents the obtained results. The opposite problem was found in the Instagram dataset, where a policy to downscale the number of pictures downloaded per year was introduced once the 2019 related ones were already taken. Therefore a correction randomly discarding some of them was performed to reduce the final peak of the distribution.

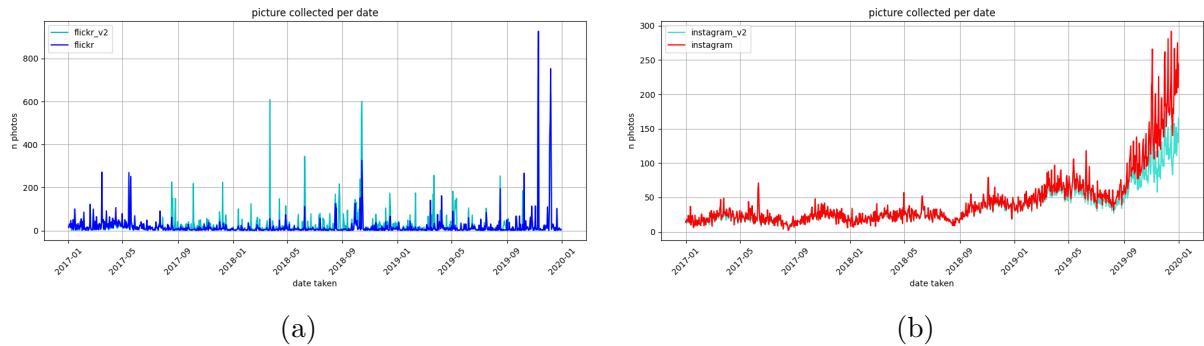


Fig. 21: *Flickr* (a) and *Instagram* (b) download rates before and after collections. Note that the change in download policy provided Flickr dataset with more data in the central periods, while for Instagram the effect was that of reducing the peak of recent months.

Appendix B

As presented on (ref. **Section 2.2**), the clustering algorithm we propose consists on a first DBSCAN run followed by a KMeans one.

Particularly, it might be of help remembering that DBSCAN stands for *Density-Based Spatial Clustering of Applications with Noise*. Its fundamental idea can be summarized as following [43]:

- A point p is a *core point* if at least $\text{min}Pts$ are within distance ϵ from it (p included)
- A point q is *directly reachable* from p if point q is within distance ϵ from the core point p . Notice that the concept of *directly reachability* is only applicable from core-points.
- A point q is *reachable* from point p if a path p_1, p_2, \dots, p_N exists, where $p_1 = p$ and $p_N = q$ and p_{i+1} is *directly reachable* from p_i . Note that this implies that all points p_i except $p_N = q$ must be core points.
- All points which are not reachable from any others are said to be *noise points*
- If p is a core point, then it forms a cluster with all points (core points or not) connected with it

Fig 22 allows a visual representation.

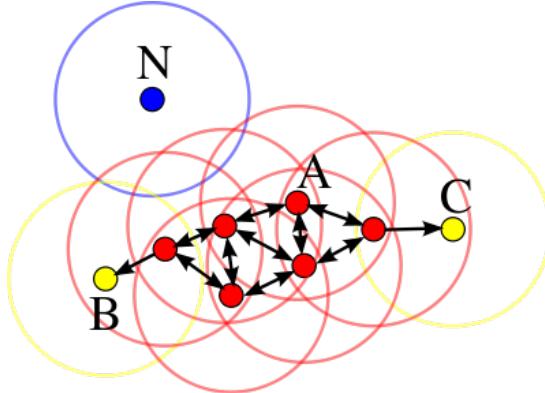


Fig. 22: With $\text{minPts} = 4$, point A and the other red points are core points since they satisfy the requirements of having at least minPoints neighbors distant less than a radius ϵ . Points B and C on the other hand are not core points but still are reachable from the red points; the yellow and red points therefore form a cluster. Point N on the other hand is not reachable from any core point, and is therefore a noise point. By Chire - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=17045963>.

The key hyper-parameters are therefore the radius ϵ , which determines the distance within which two points are considered neighbors, and the minPoints number M , which on the other hand establish how many neighbors must be within ϵ for being a core point.

Intuitively, especially with geo-referenced data, the radius ϵ is the parameter which mostly reflects the physical behaviour of the model. Depending on how points are distributed, large values will allow the model recognizing neighbors where points are actually far (diminishing the total number of clusters possibly), while small ones might result on the opposite effect. An interesting solution was proposed by [44], which exploits the concept of *diminishing returns* (picking a small radius value choosing more clusters both results in a better fit of the model and a higher risk of over-fitting) and the so called *crook of the elbow* method to obtain an approximation of its optimal value. Following their intuition, our algorithm is structured like **Alg. 3**.

Algorithm 3: Finding optimal ϵ for each level of density in the dataset

Result: ϵ radius

Collect points from database;

Sort the points for latitude and longitude values;

while Still points to analyse **do**

| Pick point p at position i ;

| Pick point q at position $i+1$;

| Evaluate Haversine Distance between p and q;

end

Sort distances on ascending order;

Find critical change in the curve;

A python-class proposed by [45] was eventually used to obtain the *elbow value*, as **Fig 23a** shows.

For the other hyper-parameter, according to [46] the *MinPts* value should be set using domain knowledge and familiarity with the data set; however, this involves some prior knowledge on the problem, limiting the tool's flexibility. Instead, we propose a fine-tuning based on the *silhouette coefficient*. This index, which can be expressed as $\frac{b(i)-a(i)}{\max(a(i), b(i))}$, represents how similar an object is to its own cluster (cohesion, expressed by the $a(i)$ term) compared to other clusters (separation, expressed by the $b(i)$ one); the more it tends to one, the more an object is well matched with its cluster. The idea is that, once the optimal radius is found, the DBSCAN procedure is run with different values of *minPts* and eventually the one resulting in the best silhouette is kept.

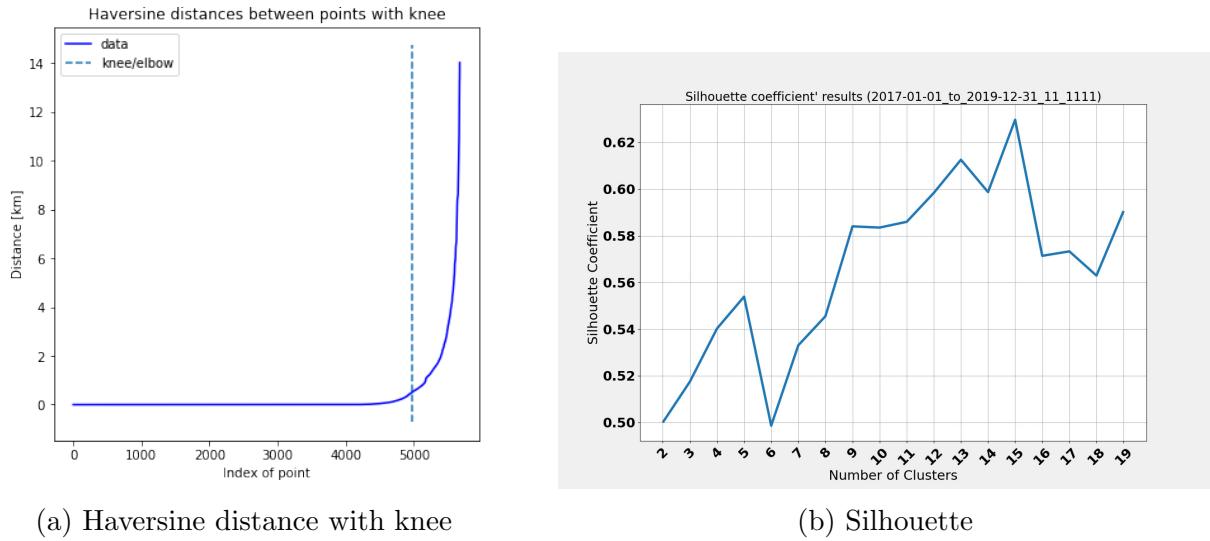


Fig. 23: Clustering tuning on the three years dataset

Through this procedure we are able to unsupervisedly define outliers (black dots), core points (colored and larger dots) and border ones (colored and smaller dots), as shown in **Fig 4a**.

Once identified the primary cluster as the most numerous, we perform a deeper analysis on it applying a further K-Means algorithm. This is another clustering technique which, differently from DBSCAN, starts from a random initialization of *centroid points* (namely points which represent the barycenter of a cluster) and then performs multiple iterations in which points are assigned to those centroids according to some metric (i.e. minimum distance) and the centroids position is re-evaluated according to an interpolation of the assigned points (i.e. the mean of their positions). Its main advantage is its being faster than the DBSCAN counterpart; one of its main weaknesses is its being non-robust to outliers [47], which anyway we filtered with the first DBSCAN run. With K-Means the only hyper-parameter is the number of clusters k representing the initial number of random centroids: a fine-tuning similar to the one for *minPts* is therefore adopted and once more the parameter corresponding to the highest silhouette score is picked. **Fig 23b** gives a visual representation.

Figure 24 eventually shows how the results of the two applied algorithms (**Figure 4**) are combined and projected into a QGIS representation.

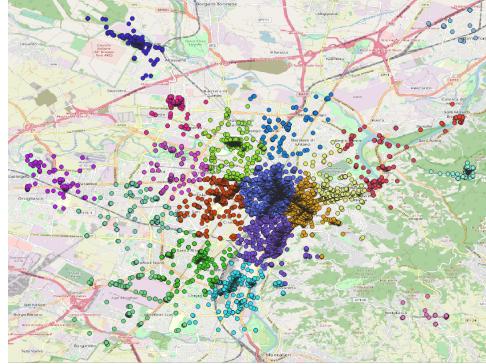


Fig. 24: DBSCAN and KMeans combined results

Appendix C

ImagNet is an image dataset organized according to the WordNet hierarchy, composed of 1331167 images which are split into training and evaluation datasets containing 1281167 and 50000 images respectively, labelled in 1000 different classes [48].

Among all existing pre-trained networks over *ImagNet* the choice of InceptionV3 is due to its accuracy on the image-classification task; its more functional use of parameters on respect than other models (see **Tab. 25**); its spread in the literature in the field.

Model	Top-1 Accuracy	Top-5 Accuracy	Parameters
Xception	0.790	0.945	22,910,480
VGG16	0.713	0.901	138,357,544
VGG19	0.713	0.900	143,667,240
ResNet50	0.749	0.921	25,636,712
ResNet101	0.764	0.928	44,707,176
ResNet152	0.766	0.931	60,419,944
ResNet50V2	0.760	0.930	25,613,800
ResNet101V2	0.772	0.938	44,675,560
ResNet152V2	0.780	0.942	60,380,648
InceptionV3	0.779	0.937	23,851,784
InceptionResNetV2	0.803	0.953	55,873,736
MobileNet	0.704	0.895	4,253,864
MobileNetV2	0.713	0.901	3,538,984
DenseNet121	0.750	0.923	8,062,504
DenseNet169	0.762	0.932	14,307,880
DenseNet201	0.773	0.936	20,242,984
NASNetMobile	0.744	0.919	5,326,716
NASNetLarge	0.825	0.960	88,949,818

Fig. 25: Tabular data taken from

Thus, the network is composed by *Factorized Convolutions* layers, which, maintaining the same computational efficiency as the previous versions, both reduce the total number of connections/parameters and the possibility to overfit; *Smaller Convolutions* layers,

which replace bigger ones in order to have faster training; *Asymmetric Convolutions* and *Grid size reduction*.

Its 42 layers deep lowers the error rate and makes it the 1st Runner Up for image classification in ILSVRC 2015 (Challenges over [48]) (Szegedy et al. [36]).

More specifically (illustration of **Fig 26**):

- **Inception Module A:** Two 3×3 convolutions replaces one 5×5 convolution: 1 layer of 5×5 filter requires $5 \times 5 = 25$ parameters, while 2 layers of 3×3 filters only $3 \times 3 + 3 \times 3 = 18$ parameters. With a reduction of the 28%, they are 2.78 times less computationally expensive (Duong-Trung et al. [49])
- **Inception Module B:** One 3×1 convolution followed by one 1×3 convolution replaces one 3×3 : 3×3 filter requires $3 \times 3 = 9$ parameters while 3×1 and 1×3 filters only $3 \times 1 + 1 \times 3 = 6$ parameters, with an overall reduction of the 33%
- **Inception Module C:** Modules which expand the filter bank output. Used only on the coarsest (8×8) grids to promote high dimensional representation, since, according to (Szegedy et al. [36]), it is the place where obtaining that high dimensional sparse representation is mostly complex.
- **Auxiliary Classifier:** Only one auxiliary classifier with batch normalization is used on the top of the last 17×17 layer instead of the two of the past versions
- **Grid Size Reduction:** In order to achieve a less expensive but still efficient network, 320 feature maps are done by conv with stride 2 and 320 feature maps are obtained by max pooling. These 2 sets of feature maps are concatenated as 640 feature maps and go to the next level of Inception module.

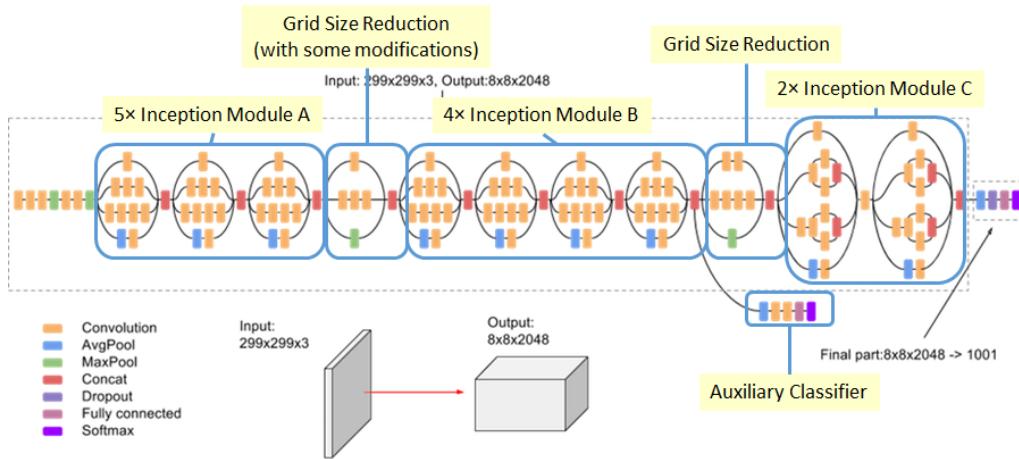


Fig. 26: InceptionV3 Architecture (Batch Norm and ReLU are used after Conv) [50]