

# Venice through the Lens of Instagram

## A Visual Narrative of Tourism in Venice

Luca Rossi

Aston University

Birmingham, United Kingdom

[l.rossi@aston.ac.uk](mailto:l.rossi@aston.ac.uk)

Eric Boscaro

Università Ca' Foscari

Venice, Italy

[eboscaro90@gmail.com](mailto:eboscaro90@gmail.com)

Andrea Torsello

Università Ca' Foscari

Venice, Italy

[torsello@unive.it](mailto:torsello@unive.it)

### ABSTRACT

The last decade has seen a huge expansion in the use of social media to extract data about human behaviour. While metadata and textual information have taken the lion's share as data sources for social media analysis, geotagged image-based platforms represent an unprecedented and as yet almost untapped source of data to analyse human behaviour and characterise the physical space we live in. In this paper we investigate the use of Instagram photos to analyse tourism consumption. We take the city of Venice (Italy) as a case study and we collect a dataset of about 90k photos taken between January 2014 and December 2015. Using computer vision techniques, we build a supervised classifier which assigns each photo to one of six different categories. We then observe how the frequency and spatial distribution of these categories varies with time. This in turn allows us to confirm the existence of a number of touristic hotspots associated with different events, such as Venice Carnival and Biennale. Our analysis also uncovers the existence of touristic flows associated with these events, such as the *Folklore Line* that marks the path of tourists from "Santa Lucia" railway station to "San Marco" square during the Carnival period. Overall, our findings confirm the effectiveness of the proposed framework to investigate tourism consumption using Instagram data.

### KEYWORDS

Social media analysis; Urban informatics; Instagram

#### ACM Reference Format:

Luca Rossi, Eric Boscaro, and Andrea Torsello. 2018. Venice through the Lens of Instagram: A Visual Narrative of Tourism in Venice. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3184558.3191557>

### 1 INTRODUCTION

Tourism is one of the fastest growing industries in many countries around the world, with more than 1 billion international tourist arrivals in 2015 according to the United Nations World Tourism Organization<sup>1</sup>. Eager to share their experiences with others, tourists

<sup>1</sup><http://www.e-unwto.org/doi/pdf/10.18111/9789284418145>

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

*WWW '18 Companion, April 23–27, 2018, Lyon, France*

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191557>

make an increasing use of photo-sharing social media like Instagram<sup>2</sup> and Flickr<sup>3</sup> [4, 8]. Therefore, it is not surprising that these platforms are attracting an increasing interest from researchers that aim to study and understand tourism [7–9].

In this paper, we propose to use images collected from Instagram to investigate tourism consumption in the city of Venice (Italy). Italy was the 5th most popular destination worldwide in 2015<sup>4</sup>, and Venice is consistently listed among the most visited cities in Italy<sup>4</sup>. We collect 90k photos taken between January 2014 and December 2015 and we design a simple-yet-effective supervised classification framework to automatically assign them to 6 different categories. We study the temporal and spatial distribution of these categories and we find that:

- The number of pictures taken increases over the observed period, with local peaks around major events like Carnival, the Venice Film Festival, and the Biennale exhibitions;
- Approximately 25% of the photos are of architectural elements such as bridges, churches, squares, highlighting the rich architectural heritage of Venice. However, the fraction of pictures in each categories varies with time, showing a clear dependency on events like art exhibitions and folklore festivals. For example, we observe a marked increase in the frequency of pictures of subjects related to Venetian folklore during the Carnival period (17%, as opposed to 14% in the period after the Carnival).
- The spatial distribution of the photos' categories closely reflects the function of the associated venues. For example, we find that most artworks pictures are located in well-known art venues, such as Biennale venues.
- Pictures of Venetian masks and other objects related to Venetian folklore are mainly distributed along the *Folklore Line*, a straight path that connects Venice "Santa Lucia" railway station (one of the main entry points for the city of Venice) to "San Marco" square, passing through "Rialto" bridge (together with "San Marco" square, one of the area with the highest concentration of Venetian masks). This effect is particularly pronounced during the Carnival period.

To the best of our knowledge, this work represents the first effort toward an automated spatio-temporal analysis of tourism using image-based social media. We believe that such a framework could find applications in several areas, from urban planning to marketing and the general tourism industry. In particular, we envision the

<sup>2</sup><https://www.instagram.com>

<sup>3</sup><https://www.flickr.com>

<sup>4</sup><http://blog.euromonitor.com/2016/01/top-100-city-destinations-ranking-2016.html>



**Figure 1: Six sample photos from the categories (Lagoon, Townscape, Art, Folklore, Food, and Other) selected for this study. We collect a total of 90k photos taken between January 2014 and December 2015.**

possibility of using a similar analysis to understand tourism consumption over a certain geographical area and to support marketing campaigns as well as general tourism investments.

The remainder of this paper is organised as follows: §2 reviews the related work, and §3 describes the dataset and its collection. §4 introduces and evaluates the image classification framework used in this paper, while §5 and §6 present the main findings and discuss the limitations of our study. Finally, §7 concludes the paper.

## 2 RELATED WORK

The increasing availability of large scale datasets generated from users online activity has spawned a flurry of research looking at modelling and understanding human behaviour. In particular, a significant amount of research has focused on the analysis of geotagged data, as this allows to draw a powerful parallel between the virtual and physical worlds. Colombo et al. [1] looked at Foursquare activity to analyse human mobility patterns. The authors found the visits to both home and workplace are highly regular and thus easily predictable. Song et al. [19] explored the boundaries of predictability of human mobility, and De Domenico et al.[2] showed how integrating information on the movements of a user’s friend can make it easier to predict the movement of the user him/herself. De Montjoye et al. [3] and Rossi et al. [15–17], on the other hand, investigated the extent to which these datasets of geotagged activity can be used to compromise users privacy.

Image-based social media like Instagram and Flickr allow users to geotag and share photos with friends. In contrast to purely location-based social networks, the focus is not on the location but rather on the image being shared. This in turn provides a remarkably rich source of information to understand human behaviour in connection with the world around us. Hochman and Schwartz [5] collect Instagram photos from New York City and Tokyo and study the differences in hue and brightness between the images collected over the two cities. They identify recurring *visual rhythms* that provide insights into different cultural practices on a local and global scale. Hu et al. [6] provide one of the first quantitative analysis of the Instagram social network. Similarly to our work, they use Scale-Invariant Feature Transform (SIFT) [10] features and Bag-of-Words (BOW) [18] to describe the images, and they use  $k$ -means to cluster them into 8 categories. With these categories to hand, they identify different type of Instagram users and they study the relation between the number of a user’s followers and the photos he/she shares on Instagram. Souza et al. [20] gather a dataset of Instagram

selfies, i.e., digital self-portraits, and study the characteristics of this dataset as well as the demographics of the users who post them. They confirm the increasing popularity of selfies and show that there exists a relationship between the number of selfies taken in a country and the country’s culture of individuation and connectedness. Photos from Instagram has also been used to understand food habits around the world [11] as well as to automatically capture the smells and colours that characterise urban environments [12, 14].

In this work we propose to use Instagram photos to characterise tourism consumption. Latorre-Martinez et al. [8] and Iniguez et al. [7] also use photos collected from both Flickr and Instagram to study tourism consumption (in Zaragoza, Spain). However their analysis only looks at geolocation indicators and hashtags, and it does not make use of the visual information provided by the images themselves. We take a complementary approach and decide to look at the content of the images, which in turn allows us to tap a much richer source of information than the hashtags alone. Our work also bears some similarities with that of Hu et al. [6], who also employ computer vision techniques to automatically group images into categories. In contrast to them, we take a supervised approach and we classify the images using a more sophisticated framework. This in turn allows us to define more meaningful classes as well as to measure the quality of the resulting categorisation.

## 3 DESCRIPTION OF THE DATASET

We retrieved a set of approximately 90k geotagged photos taken in Venice between January 2014 and December 2015 using the Instagram API<sup>5</sup>. We made use of the *Media Search* endpoint, which allowed us to search a circular area on a certain date by specifying a pair of latitude/longitude coordinates, a radius, and a date. More specifically, we called this method on a grid of coordinates to maximise the coverage of the rectangular area corresponding to the geographical zone of Venice while minimising the overlap. An alternative solution would have been to select a single circular area with a radius large enough to contain all the city of Venice, however this solution was rejected because of the API limitation on the maximum number of results per request. Each query to the *Media Search* endpoint returns a JSON file containing, among other attributes: 1) the univocal ID of the image, 2) the url of the image,

<sup>5</sup><https://www.instagram.com/developer/>

	Lagoon	Townscape	Art	Folklore	Food	Varies	
Lagoon	73	11	2	6	0	8	Lagoon
Townscape	11	72	2	8	0	7	Townscape
Art	1	12	71	4	1	11	Art
Folklore	0	10	11	64	7	8	Folklore
Food	0	2	5	2	86	5	Food
Varies	3	15	12	16	12	42	Varies
(a) $M = 1$							

	Lagoon	Townscape	Art	Folklore	Food	Varies	
Lagoon	68	5	1	4	2	20	Lagoon
Townscape	9	68	1	7	0	15	Townscape
Art	0	7	65	0	0	28	Art
Folklore	0	8	8	55	5	24	Folklore
Food	0	0	4	0	77	19	Food
Varies	0	9	4	2	6	79	Varies
(b) $M = 2$							

**Figure 2: Confusion matrix for different values of the  $M$  parameter which controls the probability of assigning an image to the *Other* category.**

which we use to retrieve the actual image, 3) latitude and longitude of picture, and 4) tags, comments and other information. Note that the Instagram API has changed since the 1st of June 2016<sup>6</sup>. Among other things, the update introduced a number of important restrictions and limitation to the *Media Search* endpoint.

The photos retrieved through the Instagram API can be very different from each other, with variations in the subject, scale, orientation, and luminance. For the purpose of this study, we decide to group the images into 6 categories: 1) *Lagoon*, 2) *Townscape*, 3) *Art*, 4) *Folklore*, 5) *Food*, and 6) *Other*. Fig. 1 shows a sample of images from each of these categories. The *Lagoon* category is composed of photos taken near or inside the lagoon, which are characterised by the presence of water, typically in the bottom part of the image, as well as other objects, such as boats. The *Townscape* category includes photos taken in the city of Venice, and it is characterised by the presence of architectural elements such as bridges, churches, squares. The *Art* category is composed of photos taken inside museums and art galleries, while photos belonging to the *Folklore* category are related to Carnival events or Venetian folklore elements, e.g., Venetian masks. The *Food* category includes photos of dishes and drinks. Finally, the *Other* category includes all those images that do not fit in any other previous categories. The manual annotation of 90k images would be a very time consuming task. Therefore, in this study we propose to use a simple-yet-effective framework to automatically annotate our dataset given a training set of manually annotated images.

It should be noted that these 6 categories were chosen as they capture well known landmarks of Venice (e.g., *Lagoon*). Similarly, the *Folklore* category was chosen as we expected this to become particularly relevant during the Carnival period. To conduct a similar study on a different city, potentially different categories tailored to the geographical area and period being studied would have to be chosen first. A possible way to select these categories could be to use tags or other textual descriptors associated to the images. However, the methodology underpinning this study would remain essentially the same. Finally, an alternative unsupervised approach could be to extract a number of categories by clustering the images,

<sup>6</sup><http://developers.instagram.com/post/133424514006/instagram-platform-update>



**Figure 3: Two examples of misclassified photos. (a) is a *Folklore* picture of gondolas that has been incorrectly classified as a *Townscape* picture (0.37 probability vs 0.33 for *Folklore*), while (b) is a *Townscape* picture that has been incorrectly classified as a *Other* picture (0.46 probability vs 0.35 for *Townscape*).**

however the resulting categories would still need to be validated to select the most relevant ones.

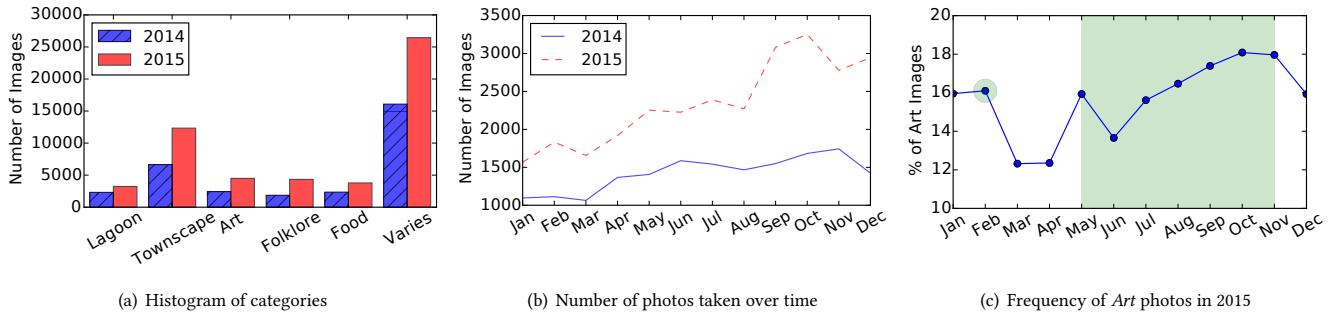
#### 4 IMAGE CLASSIFICATION FRAMEWORK

The image classification framework is as follows. We use a version of the dense SIFT algorithm [10] implemented in the VLFeat library [22] to extract a set of keypoints for each image in the dataset. A SIFT keypoint is a circular region of the image with an associated feature vector which describes the local patch appearance. While being fast to compute, SIFT features are designed to provide invariance to image translation, scaling, and rotation, as well as partial invariance to illumination changes and robustness to local geometric distortions.

With these feature vectors to hand, we use  $k$ -means to cluster them and extract a set of codewords that form the dictionary of our BOW representation [18]. Given an image with an associated set of SIFT features and a dictionary of codewords, we assign each feature to the closest codeword, allowing us to represent the original image as an histogram over the codewords.

For a set of labelled images, the corresponding histograms are used to train a set of one-vs-one Support Vector Machine (SVM) classifiers using a Radial Basis Function (RBF) kernel [23]. Recall that in the one-vs-one approach we need to train a set of binary classifiers, one for each pair of classes. Each classifier then votes for one of two classes, and the final class label is assigned by (weighted) majority voting. An alternative approach would have been to use a one-vs-all approach, where for each class we define a single classifier, with the samples of that class positive samples and the sample from all the other classes as negative samples. In this paper, we opt for the one-vs-one approach as this allows for more flexibility in the decision boundary between classes.

With this framework to hand, given an unlabelled image, its BOW representation is passed to the SVM classifiers. The output is a probability distribution over the classes [23], and the category with the maximum probability is finally used to label the original image.



**Figure 4: Histograms of the number of photos taken for each category in 2014 and 2015 (a), line plot showing the growth of the total number of photos taken in 2014 and 2015 (b), and line plot showing the fraction of photos in the *Art* category over 2015 (c). In (c), the shaded circular and rectangular areas highlight the period in which Carnival and the Biennale take place, respectively.**

#### 4.1 Image Classifier Evaluation

In order to train the classifier, we manually label 100 images for each of the 6 classes, for a total of 600 images. With this set to hand, the performance of the classifier is evaluated using 5-fold cross validation. Fig. 2(a) shows the resulting confusion matrix. The average classification accuracy is 68% and the f-measure is 68%. However, while for the first 5 classes the number of misclassified images is low, in the case of the *Other* class this number increases sharply, with over 50% of the photos belonging to this class being misclassified as belonging to another category. This is an important issue, as we expect a large number of images in our dataset to belong to this class. In fact, in our dataset *Other* is essentially a placeholder for anything that does not belong to the 5 classes we are interested in.

To address this problem, we decide to multiply the probability of an image being assigned to *Other* by a parameter  $M$ , optimised through 5-fold cross validation on the training set. Fig. 2(b) shows the confusion matrix for the optimal value  $M = 2$ . The corresponding average accuracy is 69% and the f-measure is 70%. Note that the number of false positives for the first 5 classes is drastically reduced, while the number of false positives for *Other* is significantly increased. Since the main classes we are interested in are the first 5 classes, we aim to keep the number of false positives for these classes as low as possible. Therefore, we use the classifier with  $M = 2$  to assign the labels to the rest of the images in the dataset.

Fig. 3 shows two examples of incorrectly classified images. Fig. 3(a) shows an image of gondolas that belongs to the *Folklore* category but has been incorrectly classified as *Townscape*. This is particularly a challenging image, which could be easily classified by a human as belonging to either category, as well as the *Lagoon*. The classifier assigns the highest probability (0.37) to the *Townscape* class, while the second highest probability is assigned to the *Folklore* class (0.33). Fig. 3(b), on the other hand, shows an image that belongs to the *Townscape* category but has been incorrectly classified as *Other*. In this case, the error is due to the  $M$  correction, which artificially increases the probability of assigning the image to the class *Other*.

from 0.23 to 0.46. The second most probable class is indeed *Townscape* (0.35). Note that without this correction the image would have been classified correctly.

### 5 DATA ANALYSIS

In this section we present the result of our experimental analysis. After some preliminary analyses, we study the distribution of the pictures over time and space, and we conclude by focusing on the Carnival period (February 2015).

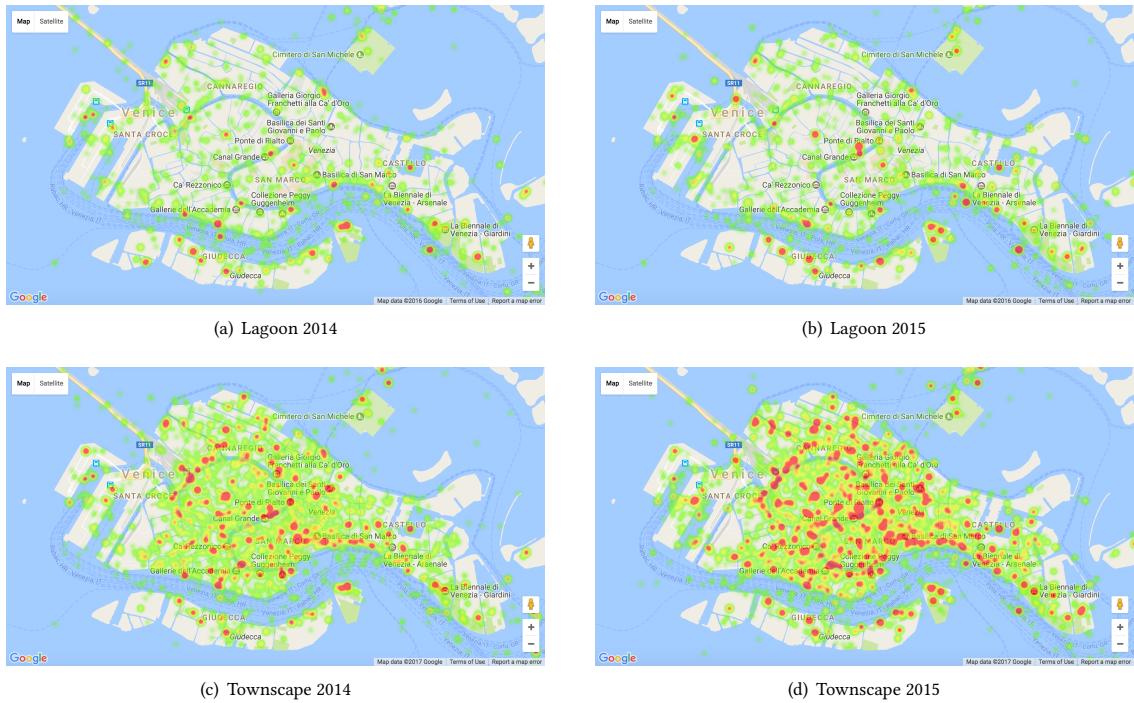
#### 5.1 Temporal Analysis

Using the proposed classifier, we assign a class label to each of the 90k images in our dataset. Fig. 4(a) shows the distribution of categories in 2014 and 2015. Note that in both years about 50% of the images are assigned to *Other*, for a total of about 44k images. Note that in the remainder of the paper we will not consider the images belonging to this category. In general, we observe that the number of images taken during 2015 is significantly higher than in 2014.

Fig. 4(b) also shows a clear rise in Instagram activity over Venice. This is in line with the worldwide growth in Instagram active users, which have more than doubled from the beginning of 2014 to the end of 2015<sup>7</sup>. We also observe the presence of a peak in Instagram activity in the periods of Carnival in 2014 and 2015, reflecting the increased number of tourists during Venice Carnival.

Finally, Fig. 4(c) shows the variation over the year 2015 of the percentage of photos belonging to the *Art* category. The circular shaded area shows the Carnival period, while the rectangular shaded area highlights the months during which the 56th “Art Biennale” exhibitions took place (from the 9th May to the 22nd November 2015). We observe a clear increase in the percentage of images belonging to the *Art* category. Note that this is not influenced by the increase of Instagram users observed during the same period, as here we measured the fraction rather than the absolute number of *Art* images. While the increase of *Art* pictures during the “Art Biennale” does not come as a surprise, we argue that the relative increase of

<sup>7</sup><https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/>



**Figure 5: Geographical distribution of the photos in the *Lagoon* (a,b) and *Townscape* (c,d) categories. The areas of highest density for the *Lagoon* category are located near waterways, while *Townscape* photos are mainly distributed over the central area of Venice.**

the rate of *Art* pictures during Carnival is due to the large number of cultural events that are organized in museums and art galleries during this period.

## 5.2 Geospatial Analysis

We now focus on the spatial distribution of the data. For each of the 5 main categories (we exclude *Other*), we plot the heatmap of the geographical distribution of the data.

**5.2.1 Lagoon and Townscape Heatmaps.** Figs. 5(a) and 5(b) show the distribution of photos in the *Lagoon* categories for 2014 and 2015, respectively. As expected, the majority of pictures belonging to this class is found near waterways, both canals (e.g., “Grand Canal” and “Giudecca Canal”) and the lagoon (e.g., “San Marco” and “Giudecca”). Similarly, we expect photos in the *Townscape* category to be distributed all over the urban area, with the highest density points being located near the most touristic spots, e.g., “Rialto” and “San Marco”. Figs. 5(c) and 5(d), where the highest density of photos is shown to be over the central area of Venice, supports this intuition.

These findings provide a strong qualitative validation of the correctness of the image classifier, which further supports the quantitative evaluation of the previous sections.

**5.2.2 Art, Folklore, and Food Heatmaps.** The Biennale Foundation, based in Venice, is one of the most prominent cultural organizations in the world. It is responsible for organising, among other

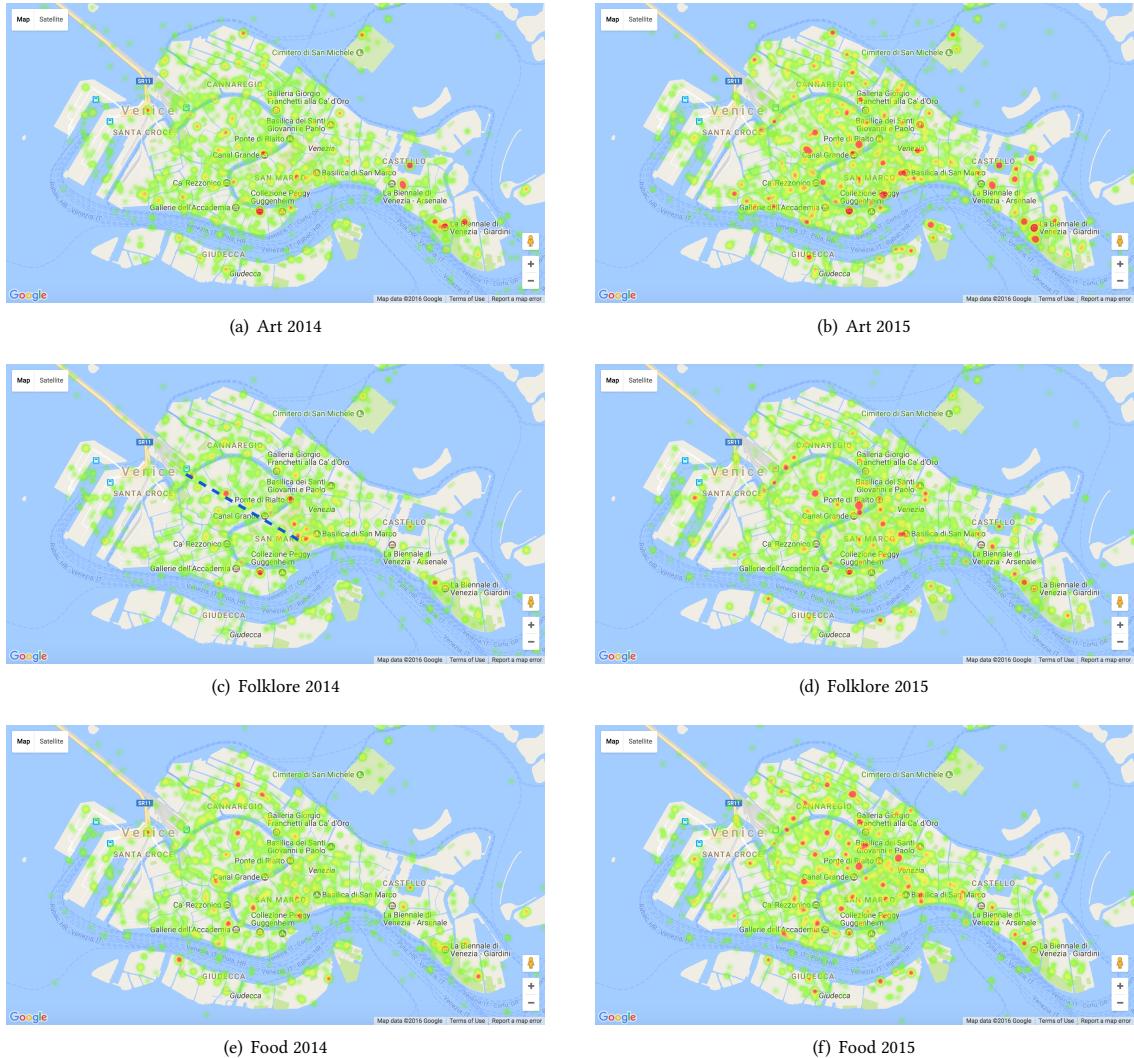
events, the “Art Biennale” (in odd-numbered years), the “Venice Biennale of Architecture” (in even-numbered years), and the “Venice Film Festival” (annually)<sup>8</sup>. It is not surprising, therefore, to find that the buildings where these events take place are among the hotspots of the *Art* heatmap (see Figs. 6(a) and 6(b)). Other notable areas of high density include the “Venetian Arsenal” and the “Guggenheim Museum”. In addition to these locations, the higher volume of data recorded for 2015 allow us to retrieve more areas characterised by a high activity in the *Art* category, including “Palazzo Grassi” (a famous art gallery overlooking the Grand Canal<sup>9</sup>) and “San Giorgio Maggiore” (a 16th century church designed by the Italian architect Andrea Palladio, where several paintings of Jacopo Tintoretto can be found<sup>10</sup>).

For the *Folklore* category, both Fig. 6(c) and Fig. 6(d) show a high concentration of photos taken in the most central and touristic areas of Venice. More interestingly, these hotspots seem to be aligned along the shortest path connecting Venice “Santa Lucia” railway station and “San Marco” square, passing through “Campo San Polo”, “Campo San Silvestro”, and “Rialto”. This suggests the existence of a preferred route for tourists visiting Venice during folklore related events, such as Carnival. We call this the *Folklore Line*. Indeed, during Carnival most masks are found in the “Rialto” and “San Marco” areas, and the shortest path to these places from the bus

<sup>8</sup><http://www.labriennale.org/en/biennale/organization/>

<sup>9</sup>[https://en.wikipedia.org/wiki/Palazzo\\_Grassi](https://en.wikipedia.org/wiki/Palazzo_Grassi)

<sup>10</sup>[https://en.wikipedia.org/wiki/Church\\_of\\_San\\_Giorgio\\_Maggiore](https://en.wikipedia.org/wiki/Church_of_San_Giorgio_Maggiore)



**Figure 6: Geographical distribution of the photos in the *Art* (a,b), *Folklore* (c,d), and *Food* (e,f) categories. Note that the areas of highest density for the *Art* category are located near well-known art venues, and that the majority of the pictures in the *Folklore* category lie along a straight path connecting Venice railway station (the main entry point to the island) and “San Marco” square (one of the areas with the highest density of Venetian masks).**

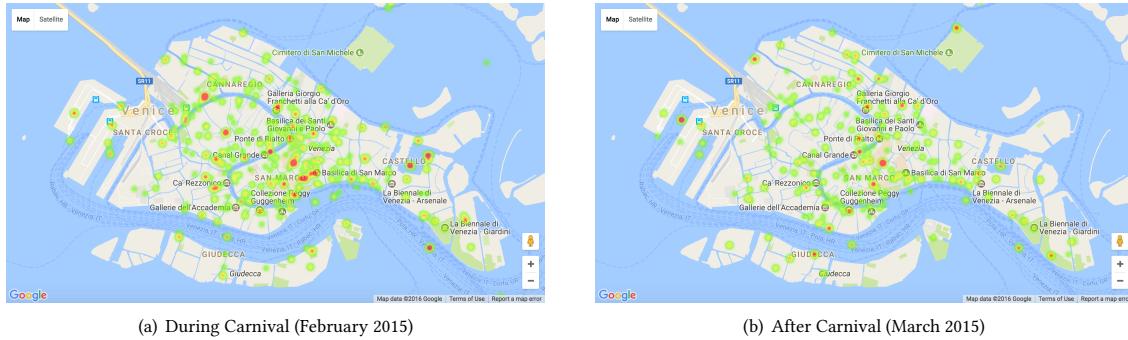
and train stations (the main points of entry to Venice) coincides with the *Folklore Line*. Our analysis also shows that the “San Marco” area has the highest rate of *Folklore* pictures throughout most of the year, with over 19% of the pictures taken in this area during Carnival 2014, the second most popular areas being “San Polo” (which is connected to “San Marco” by the “Rialto” bridge), where approximately 14% of the *Folklore* pictures are taken.

The last category we consider is *Food*. The heatmaps for the 2014 and 2015 data are shown in Figs. 6(e) and 6(f), respectively. While at first the interpretation of these plots may appear less clear, a closer inspection shows that the locations with the highest density of pictures correspond to well-known gastronomical venues. Notable examples include “Al Timon”, “Tonolo”, and “Al Paradiso Perduto”.

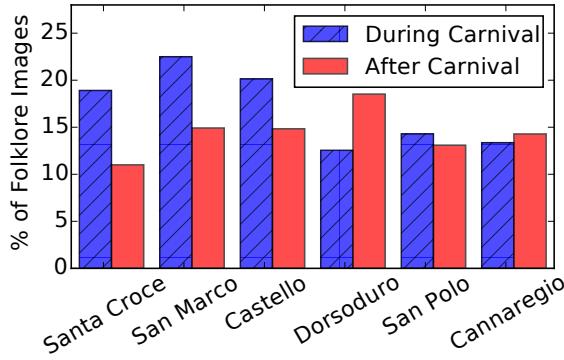
It is also interesting to note that some areas with a remarkable presence of notable restaurants are not discovered through our analysis. We argue that this is due to their position, being outside the main touristic routes.

### 5.3 A Case Study: Venice Carnival 2015

We conclude our study with a more detailed analysis of the two weeks of Carnival 2015, from the 1st of February 2015 to the 14th of February 2015. We compare the data collected during this period with that collected on the same days of the following month, from the 1st of March 2015 to the 14th of March 2015. As expected, during the Carnival period the frequency of *Folklore* pictures (17%)



**Figure 7: Geographical distribution of the photos in the *Folklore* category during the Carnival period (a) and after it (b). Note the emergence of the *Folklore Line* in (a).**



**Figure 8: Percentage of photos in the *Folklore* category per neighbourhood during and after Carnival 2015.**

is significantly higher than during March (14%). Similarly, the *Folklore* category is the second most commonly photographed category during Carnival, but it drops to being the 4th most photographed one after Carnival. Figs. 7(a) and 7(b) show the geographical distribution of *Folklore* pictures. Note the emergence of the *Folklore Line* during the Carnival period, which instead disappears during the two weeks of March we considered.

Finally, Fig. 8 shows a bar chart of the percentage of photos belonging to the *Folklore* category taken in the six main districts of Venice during and after the Carnival period. As in Fig. 7, we observe a drop in the number of *Folklore* pictures after Carnival, as opposed to the Carnival period. The largest drops are observed in "Santa Croce" (from 19% to 11%) and "San Marco" (from 15% to 23%), respectively the start and end districts of the *Folklore Line*.

## 6 DISCUSSION AND LIMITATIONS

Photo-sharing social media provide a rich source of information to investigate human behaviour. The ability to geotag this type of data opens up even more possibilities of analysis. This richness, however, comes at the cost of an increased difficulty in interpreting the content of an image as opposed to a string of text or a list of keywords. Here we have proposed a simple-yet-effective computer

vision pipeline to analyse images shared on Instagram and assign them to a set of predefined categories. While there exist more accurate techniques for classifying images (e.g., [24]), we believe that the proposed framework provides a good trade-off between accuracy and computational complexity. We should also stress, however, that while the experimental results have highlighted the quality of the classifier and the ability of the overall framework to reveal cultural and food hotspots, our study presents a number of important limitations.

For example, we have seen that in some cases the classifier is unable to reach a clear consensus on the class membership of some pictures (see Fig. 3(b)). Our approach, in these cases, is to assign the image to *Other*. While this ensures that an image assigned to another category is unlikely to have been misclassified, this means also that some images are wrongly assigned to the *Other* class, thus reducing the amount of data that we can analyse. One way to overcome this problem may be to integrate hashtags and other textual attributes in the classification process when the class membership is unclear.

In some cases, however, an image may actually belong to more than one category (see Fig. 3(a)). This could be addressed by extending the proposed classifier to handle multi-label categorisation [21]. A relatively straightforward solution would be to use off-the-shelf software libraries like the one used in this paper<sup>11</sup>. However note that this would require relabeling the entire training set assigning multiple labels to each image.

The results of our work could have applications in a number of areas, from urban planning to marketing and advertisement campaigns. One possibility would be to use this framework to guide the development of urban areas and in particular the allocation of investments to boost and support tourism. Neighbourhoods in need of development could be identified by looking at the digital traces left by tourists on platforms like Instagram or Flickr. This would be particularly interesting if augmented with semantic information from tags or other metadata that captures the perceived attractiveness of a place. One can also envision a system that creates a personalised tourist guide by linking the representation of a city given by our framework to the preferences of a user, similarly to what proposed by Quercia et al. [12–14]. The preferences of a user

<sup>11</sup><http://scikit-learn.org/stable/modules/multiclass.html>

could be determined from his/her shared photos, and itineraries over a certain area that match his/her preferences could be suggested by looking at aggregated data collected over that same area from other users.

## 7 CONCLUSION AND FUTURE WORK

In this paper we explored tourism consumption through the lens of social media. We have collected a dataset of 90k Instagram photos taken in the city of Venice between January 2014 and December 2015. We have designed a simple-yet-effective supervised classification framework to automatically assign each image to one of six predefined categories.

Our analysis highlighted the presence of a number of touristic hotspots, associated with popular cultural as well as gastronomical venues. We have observed that this signal is influenced by a number of external events, such as folklore and art festivals. Interestingly, we have found that the spatial distribution of the photos can reveal the presence of preferred touristic routes during some events, as in the case of the *Folklore Line* during the Carnival period.

An obvious extension of this work would be to improve the classifier, for example by replacing SIFT features with those extracted by a Convolutional Neural Network. Particularly interesting would be to consider also the textual information associated with a picture, e.g., the descriptions, comments, and tags. This in turn would allow us to design a more robust classifier. More interestingly, one could use natural language processing techniques to understand the mood associated with a certain picture, adding an additional layer of analysis.

## REFERENCES

- [1] Gualtiero B Colombo, Matthew J Chorley, Matthew J Williams, Stuart M Allen, and Roger M Whitaker. 2012. You are where you eat: Foursquare checkins as indicators of human mobility and behaviour. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*. IEEE, 217–222.
- [2] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. 2013. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing* 9, 6 (2013), 798–807.
- [3] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3 (2013).
- [4] David L Groves and Dallen J Timothy. 2001. Photographic techniques and the measurement of impact and importance attributes on trip design: a case study. *Loisir et Société/Society and Leisure* 24, 1 (2001), 311–317.
- [5] Nadav Hochman and Raz Schwartz. 2012. Visualizing instagram: Tracing cultural visual rhythms. In *Proceedings of the Workshop on Social Media Visualization (SocMedVis) in conjunction with the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM-12)*, 6–9.
- [6] Yuheng Hu, Lydia Manikonda, Subbarao Kambhampati, et al. 2014. What We Instagram: A First Analysis of Instagram Photo Content and User Types.. In *ICWSM*.
- [7] T Íñiguez-Berrozpe, M Plumed-Lasarte, and MP Latorre-Martínez. 2016. Flickr: tool for a market analysis of tourism consumption. *Managing the Environment: Sustainability and Economic Development of Tourism* (2016), 20.
- [8] María Pilar Latorre-Martínez, Tatiana Íñiguez-Berrozpe, and Marta Plumed-Lasarte. 2014. Image-focused social media for a market analysis of tourism consumption. *International Journal of Technology Management* 64, 1 (2014), 17–30.
- [9] Iris Sheungting Lo, Bob McKercher, Ada Lo, Catherine Cheung, and Rob Law. 2011. Tourism and online photography. *Tourism Management* 32, 4 (2011), 725–731.
- [10] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, 2 (2004), 91–110.
- [11] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. 2016. Fetishizing Food in Digital Age:# foodporn Around the World. *arXiv preprint arXiv:1603.00229* (2016).
- [12] Daniele Quercia, Luca Maria Aiello, and Rossano Schifanella. 2016. The Emotional and Chromatic Layers of Urban Smells. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016*. 309–318.
- [13] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 116–125.
- [14] Daniele Quercia, Rossano Schifanella, Luca Maria Aiello, and Kate McLean. 2015. Smelly Maps: The Digital Life of Urban Smellscape. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*. 327–336.
- [15] Luca Rossi and Mirco Musolesi. 2014. It's the way you check-in: identifying users in location-based social networks. In *Proceedings of the second ACM conference on Online social networks*. ACM, 215–226.
- [16] Luca Rossi, James Walker, and Mirco Musolesi. 2015. Spatio-temporal techniques for user identification by means of GPS mobility data. *EPJ Data Science* 4, 1 (2015), 1.
- [17] Luca Rossi, Matthew J Williams, Christoph Stich, and Mirco Musolesi. 2015. Privacy and the city: User identification and location semantics in location-based social networks. *arXiv preprint arXiv:1503.06499* (2015).
- [18] Josef Sivic and Andrew Zisserman. 2009. Efficient visual search of videos cast as text retrieval. *IEEE transactions on pattern analysis and machine intelligence* 31, 4 (2009), 591–606.
- [19] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.
- [20] Flávio Souza, Diego de Las Casas, Vinícius Flores, SunBum Youn, Meeyoung Cha, Daniele Quercia, and Virgilio Almeida. 2015. Dawn of the Selfie Era: The Whos, Wheres, and Hows of Selfies on Instagram. In *Proceedings of the 2015 ACM on Conference on Online Social Networks (COSN '15)*. ACM, 221–231.
- [21] Grigoris Tsoumakas and Ioannis Katakis. 2006. Multi-label classification: An overview. *Dept. of Informatics, Aristotle University of Thessaloniki, Greece* (2006).
- [22] Andrea Vedaldi and Brian Fulkerson. 2010. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 1469–1472.
- [23] Ting-Fan Wu, Chih-Jen Lin, and Ruby C Weng. 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research* 5, Aug (2004), 975–1005.
- [24] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*. 487–495.