

Chapter 5

Towards Improved Air Quality Monitoring Using Publicly Available Sky Images



Eleftherios Spyromitros-Xioufis, Anastasia Moutzidou,
Symeon Papadopoulos, Stefanos Vrochidis, Yiannis Kompatsiaris, Aristeidis
K. Georgoulas, Georgia Alexandri, and Konstantinos Kourtidis

Abstract Air pollution causes nearly half a million premature deaths each year in Europe. Despite air quality directives that demand compliance with air pollution value limits, many urban populations continue being exposed to air pollution levels that exceed by far the guidelines. Unfortunately, official air quality sensors are sparse, limiting the accuracy of the provided air quality information. In this chapter, we explore the possibility of extending the number of air quality measurements that are fed into existing air quality monitoring systems by exploiting techniques that estimate air quality based on sky-depicting images. We first describe a comprehensive data collection mechanism and the results of an empirical study on the availability of sky images in social image sharing platforms and on webcam sites. In addition, we present a methodology for automatically detecting and extracting the sky part of the images leveraging deep learning models for concept detection and localization. Finally, we present an air quality estimation model that operates on statistics computed from the pixel color values of the detected sky regions.

5.1 Introduction

Environmental data are crucial both for human life and the environment. Especially, the environmental conditions related to air quality are strongly related to health issues (e.g. asthma) and to everyday life activities (e.g. walking, cycling). Thus, it is

E. Spyromitros-Xioufis (✉) · A. Moutzidou · S. Papadopoulos · S. Vrochidis · Y. Kompatsiaris
Centre for Research & Technology Hellas – Information Technologies Institute, Thessaloniki,
Greece

e-mail: espyromi@iti.gr; moutzid@iti.gr; papadop@iti.gr; stefanos@iti.gr; ikom@iti.gr

A. K. Georgoulas · G. Alexandri · K. Kourtidis
Democritus University of Thrace, Xanthi, Greece

e-mail: argeor@env.duth.gr; alexang@auth.gr; kourtidi@env.duth.gr

© Springer International Publishing AG, part of Springer Nature 2018
A. Joly et al. (eds.), *Multimedia Tools and Applications for Environmental
& Biodiversity Informatics*, Multimedia Systems and Applications,
https://doi.org/10.1007/978-3-319-76445-0_5

67

necessary to provide citizens with up-to-date notifications regarding environmental conditions. Typically, air quality data are measured by official measurement stations established by environmental organizations and are made available to the public through web sites or web services. However, official stations are few and mainly located in urban areas, thus motivating use of crowdsourcing solutions to improve the geographical coverage and density of air quality measurements. To this end, a number of air quality monitoring initiatives (e.g. luftdaten.info¹) have emerged that promote the establishment of personal environmental stations by citizens, based on low-cost and relatively easy-to-use hardware sensors. At the same time, the increasing popularity of social media has resulted in massive volumes of publicly available, user-generated multimodal content that can often be valuable as a sensor of real-world events [1]. This fact coupled with the rise of citizens' interest in environmental issues and the need for direct access to environmental information everywhere (both urban and rural areas) and without any extra specialized equipment, has triggered the development of applications that make use of social data for collecting environmental information and creating awareness about environmental issues. In this context, this paper presents a framework that involves the collection of publicly available images from social media platforms and public webcams, their processing using image analysis techniques, and the application of a method for mapping image color statistics to an air quality index. The proposed framework is part of a platform developed by the hackAIR project² that gathers and fuses environmental data and specifically particulate matter (PM) measurements from official open sources and from user generated content.

5.2 Related Work

Several initiatives attempt to provide citizens with environment-oriented information collected from different data sources. Examples of such initiatives are: (a) iSCAPE³ that encapsulates the concept of smart cities by promoting the use of low cost sensors and the use of alternative solution processes to environmental problems, (b) the Amsterdam Smart Citizens Lab⁴ that uses smartphones, smart watches, and wristbands, as well as open data and DIY sensors for collecting environmental data, (c) CITI-SENSE⁵, which provides air quality information based on portable

¹<http://luftdaten.info>.

²<http://www.hackair.eu>.

³<http://horizon2020projects.com/sc-climate-action/h2020-making-cities-sustainable>.

⁴<https://waag.org/en/project/amsterdam-smart-citizens-lab>.

⁵<http://www.citi-sense.eu>.

and stable sensors, (d) CAPTOR⁶, which aims at engaging a network of local communities for monitoring tropospheric ozone pollution using low-cost sensors, and (e) AirCasting⁷, which is an open-source platform that consists of wearable sensors that detect changes in your environment and physiology, including a palm-sized air quality monitor, an Android app, and wearable LED accessories.

The aforementioned projects use sensors, open data and smart watches as sources. Another source for estimating air quality that has received recently a lot of attention is photos due to their abundance and the fact that no specialized equipment is required. Initiatives that use photos as source for estimating air quality are: (1) the AirTick⁸ application which estimates air quality in Singapore by analyzing large numbers of photos posted in the area, (2) the Visibility⁹ mobile application that encourages users to upload images of sky to get response regarding visibility which is an indicator of the air pollution of the area and (3) the hackAIR project's air quality platform that combines data from various sources including images posted in social media and retrieved from public webcams.

The AirTick application [28] is a mobile app that can turn any camera enabled mobile device into an air quality sensor. AirTick leverages image analytics and deep learning techniques to produce accurate estimates of air quality following the Pollutant Standards Index (PSI). AirTick first extracts the haziness from a single photo and then converts it into an appropriate PSI value. With haziness extracted from a given image, AirTick passes the haziness information to a Deep Neural network Air quality estimator (DNA) to learn to associate given haziness matrices with PSI values. DNA is designed based on the Boltzmann Machine (BM), which is a neural network of symmetrically coupled stochastic binary nodes. The conducted experiments showed that AirTick achieves, on average, 87% accuracy in day time operation and 75% accuracy in night time operation. Although results are encouraging, a limitation of the AirTick approach is that low light conditions prevent the successful extraction of the haziness component of the images and lead to accuracy deterioration.

Regarding the Visibility application, it is based on the work of [30] that requires users to take pictures of the sky while the sun is shining, which can be compared to established models of sky luminance to estimate visibility. Visibility is directly related to the concentration of harmful "haze aerosols", tiny particles from dust, engine exhaust, mining or other sources in the air. Such aerosols turn the blue of a sunlit clear sky gray. The Visibility app uses the accelerometers and the compass incorporated on smartphones to capture its position in three dimensions while the

⁶<http://captor-project.eu>.

⁷<http://aircasting.org>.

⁸<https://www.youtube.com/watch?v=111abvYgvBY>.

⁹<http://robotics.usc.edu/~mobilesensing/Projects/AirVisibilityMonitoring>.

GPS data and time are used to compute the exact position of the sun. The system has been tested in several locations in the US, including Los Angeles and Phoenix. However, a drawback of the method is that it requires the images to depict only or mostly sky, thus depending a lot on human judgement. Also users are requested to specify explicitly the part of the image that contains sky pixels which adds considerable manual effort.

Apart from the applications mentioned, several studies were carried out regarding the estimation of air quality from images. In [21], the authors utilize six image features together with additional information such as the position of the sun, date, time, geographic information and weather conditions, etc., to estimate the amount of $PM_{2.5}$ (particles with aerodynamic diameter less than 2.5 micrometers) in the air. Experimental results have shown that the image analysis method is able to estimate the $PM_{2.5}$ index accurately. Nevertheless, the method relies on the manual selection and labelling of the regions of interest in order to operate effectively. This step requires the users to precisely label the buildings in the photos they have taken, which incurs significant overhead. Furthermore, the additional information required by the method on top of the photos and labels of buildings may not always be available, especially in outdoor locations without Internet access.

Another work is that of [46] that proposes an effective CNN-based model tailored for air pollution estimation from raw images. Specifically, the proposed model involves the use of a negative log-log ordinal classifier to fit the ordinal output well, and the use of a new activation function for photo air pollution level estimation. The proposed approach was validated with qualitative and quantitative evaluations on a set of images taken in Beijing against several state-of-the-art methods and it was found to incur smaller error in the air quality estimation.

Finally, in [20], the authors propose a system to estimate haze level based on a single photo. The method proposed involves estimating a transmission matrix generated from a haze removal algorithm, and estimates the depth map for all pixels in the photo. A haze level score is computed by combining the transmission matrix and depth map, and can be calibrated to estimate the $PM_{2.5}$ level. The method was evaluated both on synthetic and real photos providing promising results especially in the synthetic database. Regarding the real photos, further research is required in order to make large scale monitoring based on online user photos more reliable.

Saito and Iwabuchi [32] recently introduced a new method for measuring aerosol optical properties from digital twilight photos. Their method allows for the estimation of tropospheric and stratospheric aerosols, being very promising, despite the fact that it focuses on twilight photos only. Zerefos et al. [44] had previously introduced a simpler approach to retrieve aerosol loadings from paintings from the period 1500–1900. It was found that aerosol concentrations increased in the atmosphere following major volcanic eruptions. These eruptions inserted huge amounts of aerosols in the stratosphere which remained there for years leading to more reddish sunsets. Zerefos et al. [45] extended the research from Zerefos et al. [44], covering the period 1500–2000.

A method close to that of Zerefos et al. [44] is followed in this work to estimate the aerosol load in the atmosphere as described in detail in Sect. 5.6. However, the

method is not limited to sunset conditions, is extended to images from users, social media and public webcams and furthermore uses a better representation of the local atmospheric characteristics. The methodology described in this chapter is part of the framework developed within the hackAIR project and constitutes an update of the system presented in [25] that overcomes several of its limitations (e.g. need for more images, better sky localization methods).

5.3 Overall Air Quality Monitoring Framework

Figure 5.1 depicts the proposed framework. The framework produces PM measurement estimations using recent (i.e. within the last 24 h) publicly available images. These images are retrieved from media sharing platforms such as Flickr and public webcams. The use of different sources aims to address the need for measurements that are both large in number and cover a large area. Specifically, images retrieved

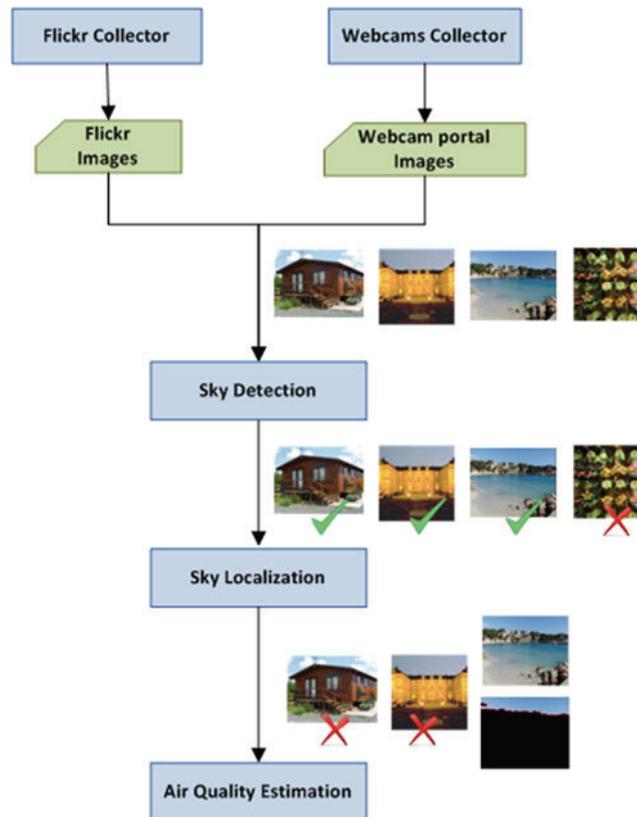


Fig. 5.1 Overview of hackAIR image-based air quality monitoring framework

from media sharing platforms offer the advantage of abundance and high geographic coverage (user generated images are expected to be captured in both rural and urban areas) while images coming from webcams offer the advantage of standard delivery of data on a daily basis, with known and standard quality and with fixed location (webcams are usually installed in urban areas). With regard to social media platforms, we use solely Flickr due to API usage restrictions imposed by other popular image sharing platforms that we considered (see Sect. 5.4.1). The other source of images is public webcams that depict parts of the skyline of an area of interest.

After having collected the images from the aforementioned sources, a series of steps is realized that aim at producing air quality estimations. Initially, a sky concept detector is applied that detects whether sky is depicted in the image by using low-level visual features and a classifier. In the sequel, sky localization detects the sky regions within the image. Two approaches are considered for sky localization, one based on deep learning techniques and the other on heuristic rules provided by air quality estimation experts. The methods are used in a complementary way in order to achieve better results compared to the results produced by either of the two approaches alone. The parts of the images that are identified as sky are used for measuring pixel color statistics, specifically the red to green (R/G) and green to blue (G/B) ratios. The last step involves using these ratios for providing information about the aerosol content of the atmosphere, which can be translated to air quality estimation in the form of air quality index (e.g. low, medium, high).

5.4 Public Image Collection

Social media platforms and webcam sites constitute the sources for collecting regularly updated publicly available images for the proposed framework. These images should be geotagged to be usable from the proposed framework. In this section we present which social media are suitable for collecting images, how we retrieve data from them, as well as the webcams repositories that include webcams dispersed around the world.

5.4.1 Social Media Platforms

Users upload billions of images on a daily basis in social media. However, not all social media are suitable or equally popular for posting images. The KPCB Internet Trends Report 2016¹⁰ provides an overview of the trends related to image

¹⁰<http://www.kpcb.com/blog/2016-internet-trends-report>.

sharing/posting for 2005–2015. Users upload more than three billion images per day in social networks, and the top platforms for photo sharing are Snapchat, Facebook Messenger, Instagram, WhatsApp and Facebook. Unfortunately, a careful examination of these platforms reveals that Snapchat, Facebook Messenger, and WhatsApp do not distribute the user-contributed images through a free API. Instagram, on the other hand, added in June 2016 strict limitations on the apps that could access the data and the number of data they could retrieve, and finally, Facebook allows access only to images from public pages and not from personal user profiles which significantly limits the number of available images.

According to KPCB Internet Report 2014¹¹, Flickr is the next social network in terms of image uploads with more than 3.5 million new images uploaded daily in 2013¹². Flickr provides an open API that enables gathering all public images users share through their profiles. Given the specifications and strict limitations of the other social media platforms as well as the considerable amount of data uploaded to Flickr, we conclude that Flickr is the most appropriate source of publicly available social images.

The Flickr collector periodically calls the Flickr API in order to retrieve the URLs and necessary metadata (e.g. timestamp, geolocation) of images captured within the last 24 h. The collection of geotagged images is conducted by submitting geographical queries to the `flickr.photos.search` API method, using the `woe_id` parameter as input. This parameter allows geographical queries based on WOEID¹³ (Where on Earth Identifier), a 32-bit identifier that uniquely identifies spatial entities and is assigned by Flickr to all geotagged images. Moreover, to retrieve only photos taken within the last 24 h, the `min/max_date_taken` parameters of the `flickr.photos.search` endpoint are used, which operate on the image's Exif metadata. For the geographical area of Europe, the Flickr API returns about 5000 geotagged images per day on average.

5.4.2 Webcam Image Collector

Another source of sky images is public outdoor webcams. Compared to images from social networks, webcams offer the advantage of providing a continuous stream of images from fixed and a priori known locations. As sources of public outdoor webcams, two large-scale repositories are used, AMOS¹⁴ [17] and webcams.travel¹⁵.

¹¹<http://www.kpcb.com/blog/2014-internet-trends>.

¹²<https://en.wikipedia.org/wiki/Flickr>.

¹³<https://en.wikipedia.org/wiki/WOEID>.

¹⁴<http://amos.cse.wustl.edu>.

¹⁵<https://www.webcams.travel>.

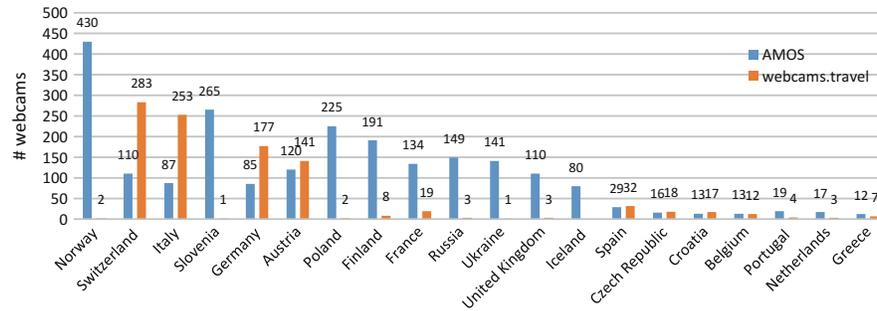


Fig. 5.2 Geographical distribution of webcams from AMOS and webcams.travel

Based on a set of experiments that we conducted, we found that many of the webcams discovered using standard search engines (e.g. Google or Bing) for a specific location are already contained in either AMOS or webcams.travel. Therefore, we believe that these two repositories cover adequately the needs of the proposed framework and there is no need for a specialized webcam discovery framework. Combined, these sources provide data from more than 25,000 webcams in Europe, which is our main area of interest. Figure 5.2 depicts the geographical distribution of webcams stored in the two repositories (top 20 countries are shown).

5.4.2.1 Collecting Images from AMOS Repository

In the case of AMOS, a web data extraction framework was developed that involves downloading and parsing the web page of each webcam and retrieving the images captured within the last 24 h. In order to identify the web page URLs of the webcams located in Europe, we use a search form provided by the AMOS web site that allows performing geographical queries by specifying bounding box coordinates. The number of webcams located in Europe is 4893; however we should note that not all matching webcams are active. The results page is parsed to extract the URLs of the webcams and each page is downloaded and parsed to extract the necessary information. The AMOS image collector is executed four times per day. An analysis of the images collected for a period of 2 months showed that 2246 of the 4893 webcams are active.

5.4.2.2 Collecting Images from webcams.travel Repository

Webcams.travel is a very large outdoor webcams directory that currently contains 64,475 landscape webcams worldwide. Webcams.travel provides access to webcam data through a comprehensive and well-documented free API¹⁶. The provided API

¹⁶<https://developers.webcams.travel>.

is RESTful (i.e. the request format is REST and the responses are formatted in JSON) and is available via Mashape¹⁷. In order to collect data from European webcams, an image collector application is implemented that uses the webcams.travel API. In this type of queries the `/webcams/list/` endpoint is exploited along with the `continent=EU` explicit modifier which narrows down the complete list of webcams to contain only webcams located in Europe. Moreover, two implicit modifiers are used: (a) `orderby` which enforces explicit ordering of the returned webcams, and (b) `limit` which is used for slicing the list of webcams by limit and offset. The use of the `limit` modifier is necessary because the maximum number of results that can be returned with a single query is 50. The last part of the query (`show=webcams:basic,image,location`) is used so that in addition to the basic information for each webcam (`id`, `status`, `title`), the returned webcam objects also contain the URL of the latest image captured from the webcam (and its timestamp) as well as the webcam's exact geographical location. Similarly to the AMOS image collector, the webcams.travel image collector is executed four times per day.

5.4.3 Image Collection Statistics

The three image collectors, i.e. the Flickr collector, the AMOS webcams collector and the webcams.travel collector, have been collecting images since 24/2/2017, 6/3/2017 and 2/5/2017, respectively. During this period and until 15/5/2017 1,019,938 images had been collected in total across the whole Europe from all sources. Figure 5.3 shows the number of images collected daily from each source. A close examination of the graph shows that the number of images collected each day by the two webcam image sources is almost stable since an almost fixed number of webcams are visited a fixed number of times each day. In particular, 2246 webcams from AMOS and 1000 webcams from webcam.travel are visited exactly four times per day and, as a result, about 9000 and 4000 images, respectively, are collected daily from these sources. On the other hand, the number of images collected daily from Flickr exhibits a large variability since it depends on the number of geotagged images (in Europe) that are uploaded daily by Flickr users. As expected, the number of images collected from Flickr increases significantly during Saturday and Sunday, since users tend to capture and upload more images during weekends. On average, about 5500 images are collected daily from Flickr.

¹⁷<https://market.mashape.com/webcams-travel/webcams-travel>.

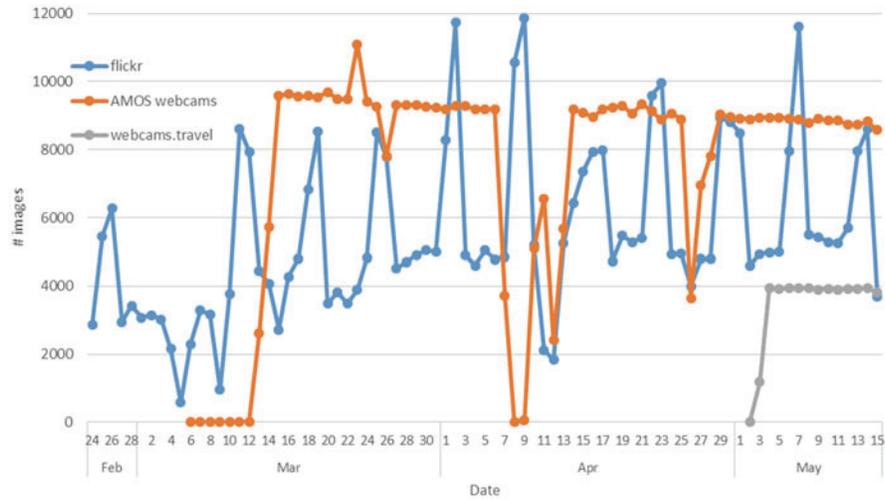


Fig. 5.3 Number of images collected daily from each source

5.5 Image Analysis for Sky Detection and Localization

The next step after image collection is image analysis. This comprises two procedures that are based on sophisticated machine learning and computer vision algorithms; sky detection and sky localization. Given an input image, sky detection is first applied to determine whether sky is depicted in the image, and in case it does, sky localization is applied to determine its exact position (i.e. image pixels). In the sequel, we present an overview of state of the art methods for *concept detection and localization*¹⁸, and then present the proposed framework.

5.5.1 Sky Detection

5.5.1.1 State of the Art

Concept detection in images aims at annotating them with one or more semantic concepts (e.g. sky, trees, road, shadows, etc.) that are chosen from a pre-defined concept list [38]. In general concept detection systems follow a process that first

¹⁸Although in our work we are interested only in the “sky concept”, the discussed methods have been designed to work for a wide range of visual concepts and are therefore widely known as concept detection/localization methods.

performs extraction of visual features, then training of classifiers for each concept using a ground-truth annotated training set, and finally, application of the trained classifiers to the features extracted from the unlabeled images that return a set of confidence scores for the presence of the different concepts.

Feature extraction from images refers to methods that aim at the effective description of the visual content of images. Many descriptors have been introduced for representing various image features and they can be divided in two main groups: hand-crafted and Deep Convolutional Neural Network (DCNN)-based descriptors. It should be noted that DCNN-based features outperform the hand-crafted features in most applications [5].

Hand-crafted features are divided into global and local descriptors. Global descriptors capture global characteristics of the image (e.g. the MPEG-7 [36] descriptor). Instead, local descriptors represent local salient points or regions and the most widely used are the SIFT descriptor [23] and its extensions (e.g. RGB-SIFT [33]), and the SURF descriptor [4] and its variations (e.g. CSURF [39]).

The most recent trend in feature extraction and image representation is learning features directly from the raw image pixels using DCNNs. These consist of many layers of feature extractors and can be used both as standalone classifiers, i.e., unlabeled images are passed through a pre-trained DCNN that performs the final class label prediction directly, or as generators of image features, i.e., the output of a hidden layer of the pre-trained DCNN is used as a global image representation [24, 37]. The latter type of features is referred to as DCNN-based and these features are used in the proposed framework due to their high performance in terms of both accuracy and efficiency.

Classification is the last step of the concept detection process. For learning the associations between the visual features and concept labels, algorithms such as Support Vector Machines (SVM) and Logistic Regression are used [24]. SVMs are trained separately for each concept, on ground-truth annotated corpora, and when a new unlabeled image arrives, the trained concept detectors return confidence scores that show the belief of each detector that the corresponding concept appears in the image.

5.5.1.2 Sky Detection Framework

In the employed framework, we train a 22-layer GoogLeNet [41] network on 5055 concepts, which are a subset of the 12,988 ImageNet concepts. Then, this network is applied on the TRECVID SIN 2013 development dataset and the output of the last fully-connected layer (5055 dimensions) is used as the input space of SVM classifiers trained on the 346 TRECVID SIN concepts. Among these classifiers, we use the one trained on the sky concept.

In order to evaluate the accuracy of the employed sky detection framework, we manually annotated (for the sky concept) 23,000 Instagram images (collected during preliminary past data collection activities) that were captured in the city of Berlin during the time period between 01/01/2016 and 15/04/2016. Sky detection was then applied on each image and the generated confidence scores were recorded in order to facilitate the selection of a decision threshold that provides a good trade-off between precision and recall. Based on this analysis, we opted for a 0.6 threshold (i.e. the sky concept is considered present if the confidence score is ≥ 0.6) which led to 91.2% precision and 80.0% recall.

5.5.2 Sky Localization

Sky localization is an important computer vision problem which refers to the detection of all pixels that depict sky in an image. In this section, we first present the state of the art in sky localization (Sect. 5.5.2.1) and then describe the adopted sky localization approach which consists of the fusion of two diverse approaches, a deep learning-based one (Sect. 5.5.2.2) and one based on a set of heuristic rules (Sect. 5.5.2.3), that were found to work in a complementary manner (Sect. 5.5.2.4).

5.5.2.1 State of the Art

An approach that was proposed by Zhijie et al. [47] suggests measuring the sky border points. The authors propose several modifications of the original sky border position function, namely the determination of multi-border points for detecting complex sky regions in images. In [16], the authors suggest using blue color for localizing and tracking RGB color in different applications of image processing. Specifically, they propose a pixel based solution utilizing the sky color information. The success of deep networks on several domains led to their application in semantic segmentation as well. Specifically, several recent works have applied Convolutional Neural Networks (CNNs) to dense prediction problems, including semantic segmentation such as [8, 26, 29]; boundary prediction for electron microscopy by Cireşan et al. [6] and for natural images by a hybrid convnet/nearest neighbor model by Ganin and Lempitsky [9]. Moreover, Hariharan et al. [13] and Gupta et al. [12] adapt deep CNNs to semantic segmentation, but do so in hybrid detection-segmentation models. These approaches fine-tune a Regional-CNN system [11] by sampling bounding boxes and/or region proposals for detection, semantic segmentation, and instance segmentation. Finally, fully convolutional training is rare, but was used effectively by Tompson et al. [42] to learn an end-to-end part detector and spatial model for pose estimation.

5.5.2.2 FCN for Sky Localization

In the proposed framework, we employ the *fully convolutional network (FCN)* approach [22], which draws on recent successes of deep neural networks for image classification (e.g. [19]) and transfer learning. Transfer learning was first demonstrated on various visual recognition tasks (e.g. [7]), then on detection, and on both instance and semantic segmentation in hybrid proposal classifier models [11–13]. The work in [22] was the first to adapt deep classification architectures for image segmentation by using networks pre-trained for image classification and fine-tuned fully convolutionally on whole image inputs and per pixel ground truth labels. Importantly, it was shown [22] that the FCN approach achieves state-of-the-art segmentation performance in a number of standard benchmarks, including the SIFT Flow dataset where the FCN-16 variant achieved a pixel precision of 94.3% on the set of geometric classes, which include sky.

To measure the performance of the approach specifically on the task of sky localization, we used the SUN Database¹⁹ [43], a comprehensive collection of annotated images covering a large variety of environmental scenes, places and the objects within. More specifically, we used the pre-trained (on the SIFT Flow dataset) FCN-16 model made available²⁰ by Long et al. [22], to predict the sky region of the 2030 SUN images for which the polygons capturing the sky part are provided. We measured a pixel precision of 91.77% and a pixel recall of 94.25%. It should be noted, that we are interested mainly in the precision of the approach given that what is required by the air quality estimation approach presented in Sect. 5.6 is recognizing accurately even a small part of the sky inside the image.

5.5.2.3 Sky Localization Using Heuristic Rules

The second approach for sky detection is based on heuristic rules that aim at recognizing the sky part of the images. The algorithm is based on identifying whether the pixels meet certain criteria involving their color values and the size of color clusters they belong to. The output of the algorithm is a mask containing all pixels that capture the sky. Figure 5.4 presents the pseudocode of the proposed method. It should be noted that the heuristic algorithm is far stricter than the FCN-based since sun and clouds are not considered part of the sky. Similarly to the FCN-based, the heuristic rule-based method was evaluated on the SUN database obtaining a mean precision of 82.45% and a mean recall of 59.22%.

¹⁹<http://groups.csail.mit.edu/vision/SUN>.

²⁰<https://github.com/shelhamer/fcn.berkeleyvision.org/tree/master/siftflow-fcn16s>.

Algorithm: Sky localization using heuristic rule-based approach

Input: Image (size MxN)

Output: Image Mask

1. Select upper 50% of the image
2. **for** each pixel p **do**
3. **if** $(0.5 \leq \frac{R}{G} \leq 1) \wedge (0.5 \leq \frac{G}{B} \leq 1) \wedge (\frac{B}{R} > 1.25)$ **is false, then**
4. p is not part of the output mask
5. **else**
6. p is candidate as part of the output mask
7. **end if**
8. **end for**
9. **for** each p **do**
10. **if** there is at least one neighboring p' that satisfies condition from step 3, **then**
11. p is candidate as part of the output mask
12. **else**
13. p is not part of the output mask
14. **end if**
15. **end for**
16. Finding Connected Components of remaining pixels
17. **for** each pixel p candidate of the output mask **do**
18. **if** p belongs to connected component with size over $(M \times N) / 400$ **is true, then**
19. p is part of the output mask
20. **else**
21. p is not part of the output mask
22. **end if**
23. **end for**
24. Initialize $Sp = 0$
25. **for** each p belonging to the output mask **do**
26. $Sp = Sp + 1$
27. **end for**
28. **if** $Sp \geq \frac{M \times N}{100}$ **then**
29. Image does not contain significant part of clear sky
30. **exit**
31. **end if**
32. **for** each p belonging to the mask **do**
33. **if** $(\frac{R}{G} > \frac{mean_R}{\bar{c}} + 4 \cdot \frac{std_R}{\bar{c}}) \&\& (\frac{R}{G} < \frac{mean_R}{\bar{c}} - 4 \cdot \frac{std_R}{\bar{c}})$ **is true, then**
34. p is not part of the output mask
35. **goto** step 24
36. **end if** **end for** **Initialize** $i = 0$
37. **for** $i < 20$ **do**
38. **if** G/B for pixels in vertical line i increases monotonically **is false, then**
39. Image does not contain significant part of sky and is discarded
40. **exit**
41. **end if**
42. **end for**
43. Image Mask produced from remaining pixels

Fig. 5.4 Flowchart of the heuristic sky localization algorithm

5.5.2.4 Comparison of Sky Localization Methods

As already mentioned both localization methods were evaluated on the SUN database and the results showed that the *FCN* approach performed significantly better than the *heuristic* approach. However, a visual inspection of the ground truth annotations of the collection's images, revealed that the image region that is annotated as "sky" is not always suitable for air quality (AQ) estimation as in many cases the sky part is not clear (e.g. contains clouds, the sun, small objects). For these reasons, a specialized evaluation of the two sky localization methods that focuses explicitly on their ability to correctly identify sky regions that are suitable for AQ estimation is presented. To this end, out of about one million images that were collected with the Flickr and the webcam image collectors, we filtered out those in which the detection confidence of the sky concept is not very high (<0.8) to ensure that most of the remaining images will depict sky and then we took a random sample of 100 Flickr and 100 webcam images. For each image, sky masks were extracted using both approaches and the following questions were collaboratively answered by the authors:

- Q1-a: Does the image contain a sky region usable for AQ estimation? (Y/N)
- Q1-b: Please shortly describe the reason if you answered No to Q1-a.
- Q2: Is the sky region selected with the *FCN* approach usable for AQ estimation? (Y/N)
- Q3: Is the sky region selected with the *heuristic* approach usable for AQ estimation? (Y/N)

To ease the task, annotators were provided with horizontally aligned composite images where the masks generated by each approach were placed next to the original image (see Fig. 5.5).

The first question (Q1-a) aims at helping us identify images with a sky region usable for AQ estimation, so that we can subsequently evaluate the different sky localization methods only on images with a usable sky region. The responses to Q1-a revealed that both for Flickr and webcams images about 60% of the images contain a sky region that is usable for AQ estimation ("Yes" to Q1-a), while looking at the distribution of responses to Q1-b, we see that in most cases and for both types of images, it is the presence of clouds or cirrus clouds or the fact that the image is captured too early in the morning or too late in the evening that render images unusable for AQ estimation, despite the existence of a sky region.

Having identified images with usable sky regions, we focused on the ability of each sky localization approach to extract these regions. The results are presented in Table 5.1, which shows the percentages of correctly detected image regions using the *FCN* (Q2) and the *heuristic* (Q3) approach for Flickr and webcam images. At a first glance, the performance of the two methods appears much worse than the performance obtained on the SUN database. Note, however, that the evaluation performed here is much stricter as even if a small percentage of the region recognized as sky includes non-sky elements, then the whole region is marked incorrect. We observe that in contrast to the results obtained when the evaluation was performed on the SUN database, the *heuristic* approach performs

Table 5.1 Percentages of correctly/incorrectly detected sky regions using each sky localization approach for Flickr and webcam images

Method	Q2-FCN (Y/N)	Q3- <i>heuristic</i> (Y/N)
Flickr	28.8%/71.2%	45.8%/54.2%
webcams	20.7%/79.3%	50.0%/50.0%

**Fig. 5.5** Comparison of the masks generated by the *FCN* approach (2nd column) with the masks generated by the *heuristic* approach (3rd column) for the images of the 1st column. The 4th column shows the masks generated by the *FCN+heuristic* approach

better than the *FCN* approach as it manages to correctly detect the sky region in 45.76%/50.00% of the Flickr/webcam images versus only 28.81%/20.69% for the *FCN* approach.

To better understand the merits of each approach, we performed a visual comparison of the generated masks (two examples are shown in Fig. 5.5). The comparison reveals that the masks generated by the *heuristic* approach are more-fine grained (e.g. small objects and text overlays that are common in webcam images are successfully filtered out), leading to more cases where all pixels identified as sky are actually sky (“Yes” to Q2) compared to the *FCN* approach (“Yes” to Q3). The *FCN* approach, on the other hand, is much better at avoiding “big” mistakes (e.g. recognizing sea, buildings or windows as sky). Motivated by the complementarity of the two approaches, we decided to develop a sky localization approach that combines them (*FCN+heuristic*). More specifically, we first calculate the sky mask using the *FCN* approach and then apply the *heuristic* algorithm, considering only those pixels that have been recognized as sky by the *FCN* approach. This way, we exploit the effectiveness of the *FCN* approach in roughly recognizing the sky region of the image and then utilize the *heuristic* approach to discard small non-sky elements. As can be seen in the right-most column of Fig. 5.5, *FCN+heuristic* generates much better sky masks than either of the two approaches alone.

Besides this qualitative evaluation, we also performed a quantitative evaluation of *FCN+heuristic*, as we did for the *FCN* and *heuristic* approaches, i.e. we collected responses to the question: “Q4: Is the sky region selected with the *FCN+heuristic* approach usable for AQ estimation? (Yes/No)” for the same set of 100 Flickr and

Table 5.2 Comparison of *FCN*, *heuristic* and *FCN+heuristic* sky localization approaches

	<i>FCN</i>	<i>heuristic</i>	<i>FCN+heuristic</i>
Accuracy	24.8%	47.9%	80.3%
Time/hardware	103 ms/Nvidia GTX1070	125 ms/Intel i7-3770	128 ms

100 webcam images. The results of this evaluation are presented in Table 5.2, which shows the percentages of correctly, when considering all images (Flickr and webcams). As expected, there is a very large improvement as **80.34%** of the sky regions are correctly recognized by *FCN+heuristic*, compared to 47.86% for the *heuristic* approach and 24.79% for the *FCN* approach. Table 5.2 also reports the average (over 200 images) running time of the methods, when images are first downscaled to a maximum size of 250,000 pixels (respecting the aspect ratio). We see that both *FCN* and *heuristic* take slightly more than 100 ms per image on average, while *FCN+heuristic* has a running time that is only slightly higher, due to the fact that *heuristic* has to operate only on the pixels that are recognized as sky by *FCN*.

5.6 Air Quality Estimation Based on Sky Color Statistics

Aerosols are tiny particles suspended in the atmosphere which are emitted by natural as well as human activities (volcanoes, desert dust, forest fires, sea salt biomass burning, combustion of fossil fuel, industrial activities, etc.) [35]. Apart from impairing the quality of the air, they determine the levels of surface solar radiation by scattering and absorbing the light that comes from the sun [14]. Their scattering and absorbing efficiency depends on their macrophysical, microphysical and microchemical properties. So, aerosols, depending on their abundance and type, leave their mark on the radiation that reaches the ground.

A number of passive remote sensing instruments (e.g., sunphotometers, spectrophotometers) are capable of retrieving aerosol optical properties such as Aerosol Optical Depth (AOD) by measuring the radiation that reaches the ground at specific wavelengths. As the instruments originally measure light intensities in order to assign the measured light intensities to a specific AOD usually a Look-Up-Table (LUT) approach is followed. LUTs are produced with the use of a radiative transfer model (RTM). RTMs calculate the intensity of the light transferred within the atmosphere under different user-input scenarios that include information about the position of sun (solar zenith angle) relative to Earth and various atmospheric parameters (e.g., clouds, aerosols, water vapor, ozone, surface albedo, etc.). This way, one knows what the expected light intensity for specific atmospheric conditions is. By comparing these measured light intensities with those from a LUT, an estimate of the AOD can be retrieved.

According to the discussion above, the color (RGB) of the sky is expected to be determined partly by the amount and type of aerosols in the atmosphere. To date

there have been some scientific efforts around the world to retrieve atmospheric aerosol properties from images taken from different types of digital cameras (e.g., [15, 27, 32]) and from paintings (e.g., [44, 45]). These efforts have returned promising results so far and further improvement is ongoing. The method followed in this work is based on the use of the ratio of the red and green band of the light (R/G) and the ratio of green and blue band of the light (G/B) from images. The main idea is that R/G and G/B depend on the amount and type of aerosols in the atmosphere [32, 44].

We decided to follow this method for a number of reasons. First of all, as discussed above, the method has already been validated in previous studies. It is based on the physics of light propagation through the atmospheric medium, contrary to approaches based on statistics or machine learning. This allows for a better understanding of the atmospheric processes that lead to high and low ratios in the images and makes it easier to understand the uncertainties and limitations of the method and proceed to corrections. In addition, the use of ratios instead of single-band RGB values compensates for biases emerging from factors such as the camera type, exposure time, sky viewing angle, etc. The LUT approach (see below) constitutes the basis of aerosol retrievals in atmospheric remote sensing, from ground-based instruments to satellite sensors. The same LUT could be used for retrieving the same quantities with images from passive remote sensing instruments in the future, allowing for a more direct validation of the method. Finally, the method is also fast, allowing its use on an operational basis.

The procedure that was followed for the production of the LUT is similar to the one described in [44] but more detailed as it takes into account the special characteristics of each region on a monthly basis, namely the optical properties of the aerosols such as the single scattering albedo and the asymmetry parameter, the ozone total column, the water vapor and the surface albedo. First, a LUT with the R/G and G/B was produced in order to assign R/G and G/B values to various aerosol loads. We use the aerosol optical depth at 550 nm (AOD_{550}) as a measure of the aerosol load in the atmosphere.

To produce the LUT we implemented radiative transfer simulations using the SBDART (Santa Barbara DISORT Atmospheric Radiative Transfer) radiative transfer model [31]. The radiative transfer equation is solved using the DISORT (Discrete Ordinate Radiative Transfer) method [40]. Sixteen streams were used. An IDL (Interactive Data Language) code that “feeds” SBDART with the necessary input data and executes the radiative transfer model for clear sky conditions was developed [2, 3]. The diffuse radiance (radiant flux received by a surface per unit solid angle per unit projected area) for the visible wavelength range (400–700 nm) was calculated. The diffuse radiance values at 700 nm (Red) were divided by the diffuse radiance values at 550 nm (Green) to get the R/G values and the radiance values at 550 nm (Green) were divided by the radiance values at 450 nm (Blue) to get the G/B values. Our tests showed that for specific sky viewing angles and azimuth angles (direction relative to the sun) in summer one should use G/B instead of R/G ratios as it is difficult to distinguish medium from high aerosol conditions with R/G ratios.

The globe was divided into 2592 grid cells with a resolution of $5^\circ \times 5^\circ$ and a sub-LUT was created for each cell in order to take into account the special characteristics of each region (optical properties of the aerosols, the ozone total column, the water vapor column and the surface albedo) (Fig. 5.6). The radiative transfer model was executed for clear-sky conditions for the central coordinates of each grid cell. This was done for various days within a year, times within a day, sky viewing angles, azimuth angles and for various AOD₅₅₀ bins, taking into account the special characteristics of each grid cell (input data). All these parameters are crucial for the radiative transfer calculations and taking into account their spatial and temporal variability increases the accuracy of the results. The core input data come from global climatologies and reanalysis projects. The aerosol optical properties (single scattering albedo and asymmetry parameter) come from the MACv1²¹ (Max-Planck-Institute Aerosol Climatology version 1) climatology [18], the total ozone column, the water vapor column and surface albedo come from the ECMWF's ERA-interim reanalysis dataset²² and the elevation data used in the calculations come from the U.S. Geological Survey GTOPO30 product.²³

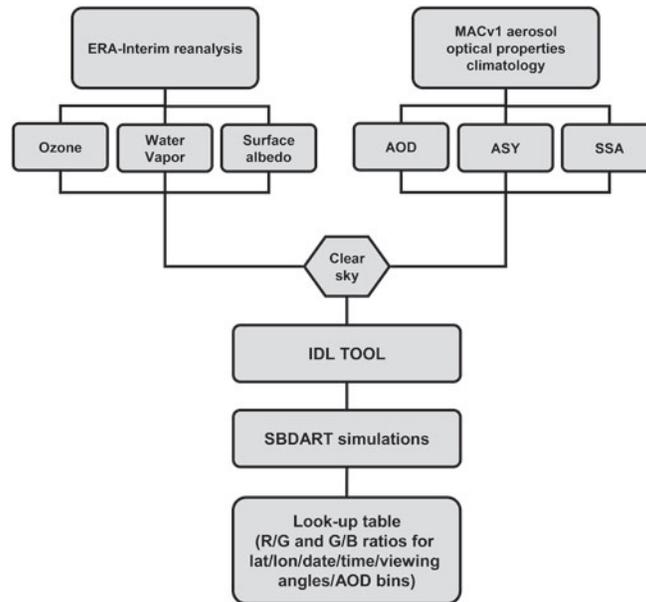


Fig. 5.6 Flowchart of the method followed for the production of the LUT

²¹<ftp://ftp-projects.zmaw.de>.

²²<http://apps.ecmwf.int>.

²³<http://earthexplorer.usgs.gov/>.

The result of the radiative transfer calculations is a LUT consisting of 2592 ASCII columnar files (sub-LUTs), one for each $5^\circ \times 5^\circ$ grid cell. The sub-LUTs include the R/G ratio (where Red: 700 nm and Green: 550 nm), and the G/B ratio (where Green: 550 nm and Blue: 450 nm) for different days within a year, hours within a day and AOD₅₅₀ bins. After going through a number of tests for specific spots it was decided that sky viewing angles of 45° and azimuth angles of 90° should be used as the majority of user generated photos are close to this scenario.

The final step of the retrieval procedure includes the calculation of the AOD₅₅₀ that corresponds to the photo R/G ratio. This is done by calculating the difference of the LUT R/G ratio values that appear in the sub-LUT that corresponds to the geographical coordinates of the photo with the photo R/G values and selecting the AOD₅₅₀ value from the sub-LUT that minimizes this difference. As discussed above, only for summer and for specific sky viewing and azimuth angles G/B ratios are used instead of R/G ratios. Similarly to [45], the errors in AOD₅₅₀ should be less than 0.05 for values around 0.1 and can be up to 0.18 for AOD₅₅₀ values greater than 0.5. To avoid the uncertainties inserted in cases of large solar zenith angles the method is not applied to images taken close to the sunrise or sunset.

So far, the method has been tested for various places in Greece and in Europe. Results from three tests implemented for the city of Thessaloniki, Greece (an aerosol hot spot for the region of Eastern Mediterranean: [10]) and Europe as a whole are presented here, showing that the use of R/G (G/B) ratios is capable of revealing urban as well as regional particle pollution features.

5.6.1 Test 1

On 10/6/2016 from 18:10 to 18:55 (local timezone) we crossed Thessaloniki, Greece (1.5 million inhabitants) using the bus from one side of the city to the other following the coastline. A photo was taken each time the bus stopped in front of a bus stop (see Fig. 5.7 for the position of the 39 bus stops). The 39 photos were taken at a viewing angle of $\sim 45^\circ$ and an azimuth angle of $\sim 30^\circ$ relative to the sun. The photos were processed in order to calculate the R/G ratio. The results show that the R/G ratio increases gradually as one gets into the city centre. The R/G ratio decreases for an extended area covered with green and trees in the centre of the city, then increases again and finally decreases gradually as the bus leaves the city centre. As the distance covered by the bus is nearly 16 km and the R/G levels have a reasonable variability taking into account the expected emissions in the city (busy streets, parks, etc.), the method seems to be adequate to characterize the aerosol variability within an urban centre. According to these results the method is expected to have a spatial representativeness of 1–2 km.

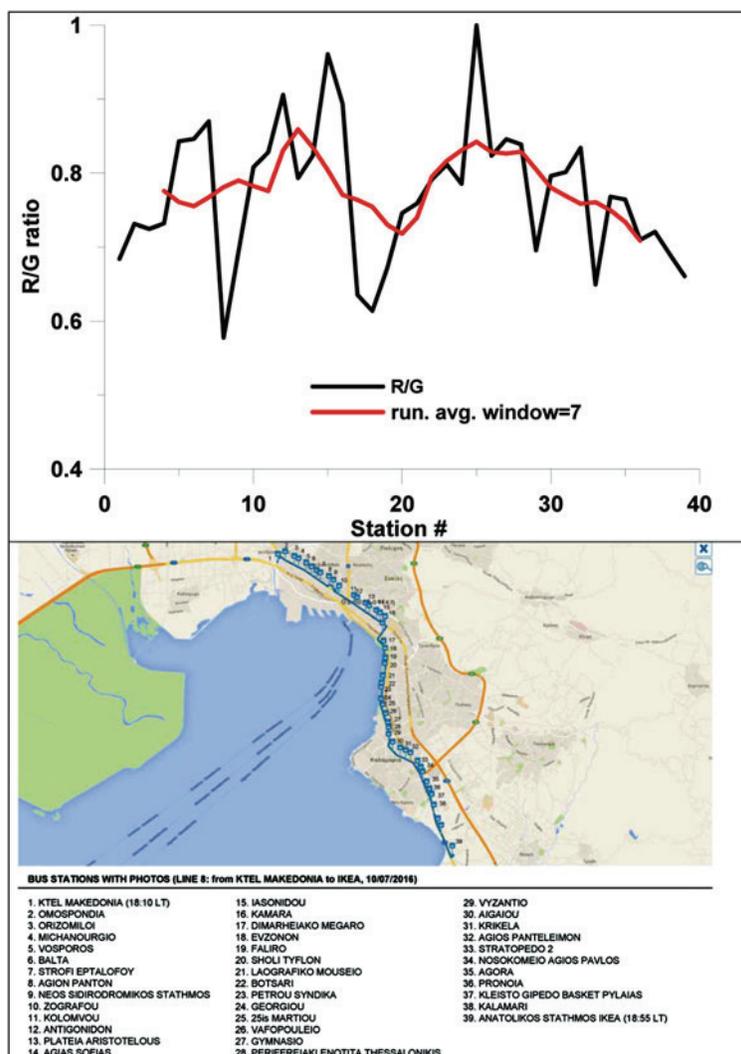


Fig. 5.7 Bus stops where photos were taken during the Thessaloniki transect experiment (lower panel) and the corresponding R/G values for each one of the 39 stations (upper panel)

5.6.2 Test 2

Annual AOD₅₅₀ maps for Thessaloniki were produced using 435 Flickr images for the year 2012 (Fig. 5.8). Figure 5.8 was created using ordinary kriging for interpolation. The results were compared against PM_{2.5} maps from [34] for Thessaloniki. In [34], the authors used a data assimilation algorithm coupling dispersion modeling and ground station data. The resulting PM_{2.5} map of the metropolitan area

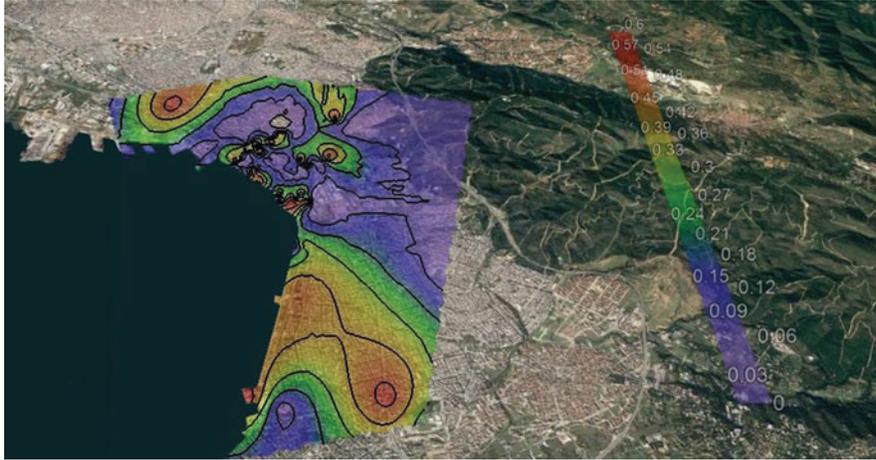


Fig. 5.8 AOD levels over the city of Thessaloniki, Greece as retrieved from Flickr images for the year 2012

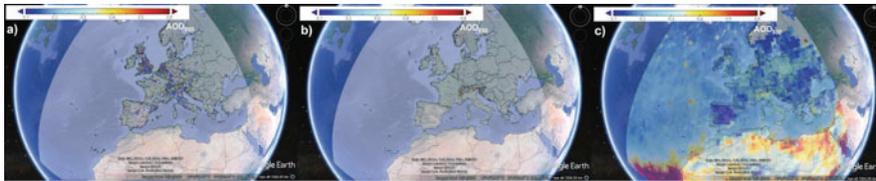


Fig. 5.9 Comparison of results from Flickr (a) and webcam (b) images with MODIS/Terra satellite retrievals (c) from Europe. The data cover the period 24/2/2017 to 13/4/2017

of Thessaloniki reveals similar features with Fig. 5.7 (high pollution to the left of the port and pollution hot spots at the same locations) which adds further credibility to our results.

5.6.3 Test 3

We computed annual AOD₅₅₀ maps for Europe using ~31,000 Flickr and ~25,000 webcam images for the period March–April 2017 (Fig. 5.9a, b). Our results are compared against AOD₅₅₀ maps (Fig. 5.9c) with data from the MODIS/Terra satellite sensor (Col. 6, L3 data) which were acquired from NASA's Giovanni web database.²⁴ These maps show consistently high and low values over specific regions. All the maps share the same features (e.g., high values in N. Italy, Pays Bas, etc.). However, the Flickr images show better details than MODIS or webcam images. Hence, in the Flickr map several major cities are also seen.

²⁴<https://giovanni.gsfc.nasa.gov>.

5.7 Conclusions and Future Work

The proposed framework comprises all the steps required for estimating air quality from publicly available images. The sources used for image retrieval are social media platforms and webcams. As far as the social media platforms that could be used, a study was realized that covered all the popular platforms that are used for image sharing. The results of the study revealed that Flickr is the most appropriate candidate due to the strict limitations on data usage imposed by the other social media platforms as well as the considerable amount of data uploaded to Flickr. A set of experiments regarding the images returned by Flickr covering Europe shows that the average number of geotagged images collected daily are approximately 5000. As far as webcams are concerned, two very large repositories of webcam images were analyzed, AMOS and webcams.travel. Both repositories were found to contain a significant number of webcams and, at the same time, offer a relatively simple way of retrieving images and other required information (location and time) from them. Consequently, two specialized collectors were implemented, facilitating the collection of images from approximately 3500 different European locations at regular time intervals.

All the collected images are processed using a three-step procedure. The first step involves sky detection, the second sky localization, and the third air quality estimation. Sky localization involves detecting the sky part of the image and two methods were studied. One based on Fully Convolutional Networks and one based on heuristic rules proposed by air quality experts. An evaluation of the two techniques was realized, showing that the two methods achieve better results when applied in a complementary way. Eventually, for the sky part of the images the R/G and G/B ratios are calculated and air quality estimation is realized. A number of atmospheric aerosol measurements using personal photos, images from Flickr and from webcams for the city of Thessaloniki, Greece and Europe was produced to study the ability of the method to reveal local and regional pollution features. The first comparisons with results from previous studies and with satellite observations highlight the potential of the method.

The evaluation of the proposed framework showed that results are promising. However, there is still room for improvement with respect to the accuracy of the sky detection and localization methods and the spatial and temporal resolution of the LUTs. It has been shown that the presence of cirrus clouds is in many cases the reason why an image is considered unsuitable for air quality estimation. Even though in many cases it is difficult to decide whether an image is unsuitable for air quality estimation due to the presence of cirrus clouds even with a naked eye, a possible direction for future work would be the development of a specialized concept detector that would automatically recognize and filter sky-depicting images where sky is covered by this type of clouds or the use of haze as proposed in other works for estimating air quality.

As a final remark, we would like to point out that the very promising results of the proposed framework as well as results of a number of other recent works

on image-based air quality estimation, on one hand highlight the potential of using images as cheap air quality sensors but on the other hand highlight the importance of evaluating all these approaches under a common evaluation framework in order to draw more reliable conclusions with respect to their relative merits. The development of such a benchmark is a promising direction for future work.

Acknowledgements This work is partially funded by the European Commission under the contract number H2020-688363 hackAIR.

References

1. Aiello, L.M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., Jaimes, A.: Sensing trending topics in twitter. *IEEE Transactions on Multimedia* **15**(6), 1268–1282 (2013)
2. Alexandri, G., Georgoulas, A., Meleti, C., Balis, D., Kourtidis, K., Sanchez-Lorenzo, A., Trentmann, J., Zanis, P.: A high resolution satellite view of surface solar radiation over the climatically sensitive region of eastern mediterranean. *Atmospheric Research* **188**, 107–121 (2017)
3. Alexandri, G., Georgoulas, A., Zanis, P., Katragkou, E., Tsikerdekis, A., Kourtidis, K., Meleti, C.: On the ability of regcm4 regional climate model to simulate surface solar radiation patterns over europe: an assessment using satellite-based observations. *Atmospheric Chemistry and Physics* **15**(22), 13,195–13,216 (2015)
4. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). *Computer Vision and Image Understanding* **110**(3), 346–359 (2008). Similarity Matching in Computer Vision and Multimedia
5. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531* (2014)
6. Cireşan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: *Advances in neural information processing systems*, pp. 2843–2851 (2012)
7. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: *International conference on machine learning*, pp. 647–655 (2014)
8. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence* **35**(8), 1915–1929 (2013)
9. Ganin, Y., Lempitsky, V.: N^c 4-fields: Neural network nearest neighbor fields for image transforms. In: *Asian Conference on Computer Vision*, pp. 536–551. Springer (2014)
10. Georgoulas, A.K., Alexandri, G., Kourtidis, K.A., Lelieveld, J., Zanis, P., Pöschl, U., Levy, R., Amiridis, V., Marinou, E., Tsikerdekis, A.: Spatiotemporal variability and contribution of different aerosol types to the aerosol optical depth over the eastern mediterranean. *Atmospheric Chemistry and Physics* **16**(21), 13,853 (2016)
11. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587 (2014)
12. Gupta, S., Girshick, R., Arbeláez, P., Malik, J.: Learning rich features from rgb-d images for object detection and segmentation. In: *European Conference on Computer Vision*, pp. 345–360. Springer (2014)

13. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Simultaneous detection and segmentation. In: *European Conference on Computer Vision*, pp. 297–312. Springer (2014)
14. Haywood, J., Boucher, O.: Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. *Reviews of geophysics* **38**(4), 513–543 (2000)
15. Igoe, D., Parisi, A., Carter, B.: Characterization of a smartphone camera’s response to ultraviolet a radiation. *Photochemistry and photobiology* **89**(1), 215–218 (2013)
16. Irfanullah, K.H., Sattar, Q., Sadaqat-ur Rehman, A.A.: An efficient approach for sky detection. *IJCSI International Journal of Computer Science Issues* **10** (2013)
17. Jacobs, N., Roman, N., Pless, R.: Consistent temporal variations in many outdoor scenes. In: *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pp. 1–6. IEEE (2007)
18. Kinne, S., O’Donnell, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T.F., Stevens, B.: Mac-v1: A new global aerosol climatology for climate studies. *Journal of Advances in Modeling Earth Systems* **5**(4), 704–740 (2013)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
20. Li, Y., Huang, J., Luo, J.: Using user generated online photos to estimate and monitor air pollution in major cities. In: *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, p. 79. ACM (2015)
21. Liu, C., Tsow, F., Zou, Y., Tao, N.: Particle pollution estimation based on image analysis. *PLoS one* **11**(2), e0145955 (2016)
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
23. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**(2), 91–110 (2004)
24. Markatopoulou, F., Mezaris, V., Patras, I.: Cascade of classifiers based on binary, non-binary and deep convolutional network descriptors for video concept detection. In: *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 1786–1790. IEEE (2015)
25. Moutzidou, A., Papadopoulos, S., Vrochidis, S., Kompatsiaris, I., Kourtidis, K., Hloupis, G., Stavrakas, I., Papachristopoulou, K., Keratidis, C.: Towards air quality estimation using collected multimodal environmental data. In: *International Workshop on the Internet for Financial Collective Awareness and Intelligence*, pp. 147–156. Springer (2016)
26. Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., Barbano, P.E.: Toward automatic phenotyping of developing embryos from videos. *IEEE Transactions on Image Processing* **14**(9), 1360–1371 (2005)
27. Olmo, F.J., Cazorla, A., Alados-Arboledas, L., López-Álvarez, M.A., Hernández-Andrés, J., Romero, J.: Retrieval of the optical depth using an all-sky ccd camera. *Applied optics* **47**(34), H182–H189 (2008)
28. Pan, Z., Yu, H., Miao, C., Leung, C.: Crowdsensing air quality with camera-enabled mobile devices. In: *AAAI*, pp. 4728–4733 (2017)
29. Pinheiro, P., Collobert, R.: Recurrent convolutional neural networks for scene labeling. In: *International Conference on Machine Learning*, pp. 82–90 (2014)
30. Poduri, S., Nimkar, A., Sukhatme, G.S.: Visibility monitoring using mobile phones. *Annual Report: Center for Embedded Networked Sensing* pp. 125–127 (2010)
31. Ricchiazzi, P., Yang, S., Gautier, C., Soble, D.: Sbdart: A research and teaching software tool for plane-parallel radiative transfer in the earth’s atmosphere. *Bulletin of the American Meteorological Society* **79**(10), 2101–2114 (1998)
32. Saito, M., Iwabuchi, H.: A new method of measuring aerosol optical properties from digital twilight photographs. *Atmospheric Measurement Techniques* **8**(10), 4295–4311 (2015)

33. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9), 1582–1596 (2010)
34. Sarigiannis, D.A., Karakitsios, S.P., Kermenidou, M.V.: Health impact and monetary cost of exposure to particulate matter emitted from biomass burning in large cities. *Science of The Total Environment* **524**, 319–330 (2015)
35. Seinfeld, J.H., Pandis, S.N.: *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons (2016)
36. Sikora, T.: The mpeg-7 visual standard for content description-an overview. *IEEE Transactions on Circuits and Systems for Video Technology* **11**(6), 696–702 (2001). 10.1109/76.927422
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
38. Snoek, C., Cappallo, S., Fontijne, D., Julian, D., Koelma, D.C., Mettes, P., van de Sande, K., Sarah, A., Stokman, H., Towal, R., et al.: Qualcomm research and university of amsterdam at trecvid 2015: Recognizing concepts, objects, and events in video. In: *NIST TRECVID Workshop* (2015)
39. Spyromitros-Xioufis, E., Papadopoulos, S., Kompatsiaris, I., Tsoumakas, G., Vlahavas, I.: A comprehensive study over vlad and product quantization in large-scale image retrieval. *IEEE Transactions on Multimedia* (2014)
40. Stamnes, K., Tsay, S.C., Wiscombe, W., Jayaweera, K.: Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Applied optics* **27**(12), 2502–2509 (1988)
41. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9 (2015)
42. Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: *Advances in neural information processing systems*, pp. 1799–1807 (2014)
43. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo. In: *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pp. 3485–3492. IEEE (2010)
44. Zerefos, C., Gerogiannis, V., Balis, D., Zerefos, S., Kazantzidis, A.: Atmospheric effects of volcanic eruptions as seen by famous artists and depicted in their paintings. *Atmospheric Chemistry and Physics* **7**(15), 4027–4042 (2007)
45. Zerefos, C., Tetsis, P., Kazantzidis, A., Amiridis, V., Zerefos, S., Luterbacher, J., Eleftheratos, K., Gerasopoulos, E., Kazadzis, S., Papayannis, A.: Further evidence of important environmental information content in red-to-green ratios as depicted in paintings by great masters. *Atmospheric Chemistry and Physics* **14**(6), 2987–3015 (2014)
46. Zhang, C., Yan, J., Li, C., Rui, X., Liu, L., Bie, R.: On estimating air pollution from photos using convolutional neural network. In: *Proceedings of the 2016 ACM on Multimedia Conference*, pp. 297–301. ACM (2016)
47. Zhijie, Z., Qian, W., Huadong, S., Xuesong, J., Qin, T., Xiaoying, S.: A novel sky region detection algorithm based on border points. *International Journal of Signal Processing, Image Processing and Pattern Recognition* **8**(3), 281–290 (2015)