

UPDATING A ROAD NETWORK DATASET EXPLOITING THE RESULTS OF SEMANTIC SEGMENTATION TECHNIQUES APPLIED TO STREET-LEVEL IMAGERY

A. Ajmar^{1,*}, E. Arco², P. Boccardo², F. Giulio Tonolo³, Janine Yoong⁴

¹ ITHACA, Via Pier Carlo Boggio 61, 10138 Torino, Italy - andrea.ajmar@ithaca.polito.it

² Politecnico di Torino – DIST, Torino, Italy - (emere.arco, piero.boccardo)@polito.it

³ Politecnico di Torino, Department of Architecture and Design, Italy - fabio.giuliotonolo@polito.it

⁴ Mapillary Inc., 134 North 4th Street, Brooklyn NY 11249-3296, USA - janine@mapillary.com

Commission IV, WG IV/4

KEY WORDS: Road network, Traffic sign, Semantic segmentation, Street-level imagery, Data fusion, Topology

ABSTRACT:

Traffic (or road) signs are an important component for applications in the mobility domain. When integrated with a road network, traffic signs, e.g. speed limits, restricted access, breakthrough prohibition signs, provide information that can be exploited in determining impedances, travel times and routing options. Additionally, the availability of a traffic sign geospatial dataset is considered crucial, especially when installation and maintenance tasks are strictly managed by road concessionaires. Unfortunately, public and private road concessionaires not always have this kind of dataset, and its generation could be highly demanding in terms of both human and time resources.

The focus of this paper is on exploring the fit for purpose of semantic segmentation techniques to feed and update existing road network datasets and traffic sign censuses, exploiting free and open mapping initiative like Mapillary (possibly including commercial derivative products) and OpenStreetMap (OSM). More specifically, the authors are analysing the best approaches for integrating the results of map features extraction into road network data with the objective to develop a semi-automated procedure for supporting this task and made it feasible over large urban areas.

1. INTRODUCTION

Traffic (or road) signs, intended as signs erected at the side of or above roads to give instructions or provide information to road users, are an important component for applications in the mobility domain. Availability of traffic sign geospatial dataset is considered crucial, especially when installation and maintenance tasks are strictly managed by road concessionaires. Additionally, if this dataset is part of a road network, it provides information that can be exploited in determining impedances, and to estimate travel times and routing options. Additionally, points of interest (PoI) signs (e.g. directions to public offices, schools, hospitals, parking lots, etc.) can enable infomobility services tailored for the citizens and/or the concerned users. Unfortunately, public and private road concessionaires not always have this kind of dataset and its generation could be highly demanding in terms of both human and time resources. The focus of this paper is therefore on exploring the fit-for-purpose of semantic segmentation techniques to feed and update existing road network datasets, exploiting free and open mapping initiative like Mapillary¹ (possibly including commercial derivative products) and OpenStreetMap (OSM). OSM is a map of the world, created by volunteers and free to be used under an open license². It is an initiative founded in 2004 in response to the general difficulty in accessing geographic dataset, by filling the gap in terms mainly of licensing conditions, data accessibility and inhomogeneous data structures. As of 31/01/2019, OSM database contains more than 7 billion of uploaded GPS points, more than 5 billion of nodes

and more than 5.6 billion of ways (ordered list of nodes)³. Number of registered users, nodes and ways are constantly growing. As for many others map features⁴, traffic signs are managed in OSM as tags associated to a specific way. Restriction traffic signs, in particular, are well encoded, as they are mandatory for implementing navigation rules⁵. Parking spaces are also well codified, including capacity differentiated per category, e.g. including spaces reserved to disabled people. Open Transport Map (OTM⁶) is a road network suitable for routing derived from OSM, available over the whole Europe and in a schema compatible to INSPIRE.

Semantic segmentation models are a technology for automatically assign to each pixel (of almost any sort of imagery) a categorical tag. This technique is becoming disrupting considering the huge potential, in terms of data volumes and technical characteristics, of mobile devices equipped with a camera and a GPS receiver in imagery acquisition. The identification of road network elements from geotagged images using semantic segmentation techniques represents a very effective solution for quick field surveys of road networks and subsequent update of relevant datasets, particularly in urban areas where a comprehensive census of traffic signs is most often not available. Same techniques represent a fundamental enabler for self-driving cars. Mapillary has a well-recognized capacity in supporting effective imagery acquisition and mapping update. They provide a plugin⁷ for one

¹ <https://www.mapillary.com/>

² <https://www.openstreetmap.org>

* Corresponding author

³ https://www.openstreetmap.org/stats/data_stats.html

⁴ https://wiki.openstreetmap.org/wiki/Map_Features

⁵ https://wiki.openstreetmap.org/wiki/Map_Features#Restrictions

⁶ <http://opentransportmap.info/>

⁷ <https://wiki.openstreetmap.org/wiki/JOSM/Plugins/Mapillary>

of the main OSM editor (JOSM) that allows the users to work with street-level imagery hosted on their server, e.g. to retrieve on the ground imagery of a specific spot from which deriving information not visible from satellite imagery, the other main source of content creation in OSM. Mapillary is also in the front line in developing semantic segmentation techniques, with high efficiency (Rota Bulò et al., 2018), performances⁸ and accuracy⁹ (Mancini et al., 2018; Cariucci et al., 2017). Mapillary technology recognizes objects in street-level imagery and places them on the map, supporting over 40 object classes and more than 1,500 different traffic signs¹⁰.

In the framework of a project of National Interest funded by the Italian Ministry of Education¹¹, the authors are analysing the best approaches for integrating the results of map features extraction into road network data. In particular, the authors focus on the main issues related to updating OSM with information that can be derived from knowing the position of different traffic signs, with the objective to develop a semi-automated procedure for supporting this task and made it feasible over large urban areas.

2. MAPILLARY TRAFFIC SIGNS DETECTION ALGORITHM

Mapillary adopt the well-established photogrammetric algorithm Structure from Motion (SfM) to create and reconstruct surfaces in 3D (Pumarola et al., 2018). By automatically matching points between different images, SfM is able to estimate their three-dimensional coordinates therefore determine their location on a map. The SfM method solves the camera position and image orientation simultaneously and automatically, using a highly redundant bundle adjustment based on matching features in multiple overlapping images (Westoby et al., 2012). Mapillary library also integrates external sensor (e.g. GPS, accelerometer) measurements to estimate the initial camera positions and attitude and to increase the robustness of the solution. The more images there are available for a specific point, the more accurately its position can be estimated. In Figure 1, original image locations (red dots) are swinging from one side of the road to the other, occasionally crossing buildings. Exploiting overlapping areas in photo sequences, positions can be re-estimated, eventually returning corrected positions (green dots, aligned with roads).

Mapillary also runs semantic segmentation, based on computer vision and deep neural networks, on the images. The network tries to understand what is in the image and assigns a category tag to each pixel, enabling the detection of different objects (such as buildings, pedestrians, cars etc.). Semantic segmentation together with 3D reconstruction enables to extract 3D coordinates (in a known reference system) of the detected objects such as traffic signs, and display them on the map.

Traffic signs recognition is a priority for Mapillary, since they are one of the key map features for navigation (Figure 2). Nevertheless, traffic sign recognition on a global level is challenging even for the most advanced computer vision technology, mainly due to the following issues:

- traffic signs that are in the same class and therefore have the same meaning, such as stop signs, don't all look the same. This becomes even more problematic when considering that there are shifts in what traffic signs look like depending on the weather or possible occlusions by other objects
- traffic signs can be visually similar to other objects in the scene, e.g. tail lights on vehicles or even random patterns formed by tree branches
- a traffic sign taxonomy is needed to organize the different traffic signs into semantic classes over different countries. Current Mapillary traffic sign recognition algorithm supports 1,500 traffic sign classes across over 100 countries, and has a detection rate of 98%, with false positives on vehicles, buildings and vegetation properly skipped thanks to the availability of a large and diverse (e.g. different lighting conditions) annotated street-level imagery dataset (Neuhold et al., 2017).



Figure 1. Original photo locations (in red) and corrected locations (in green) estimated through a photogrammetric approach

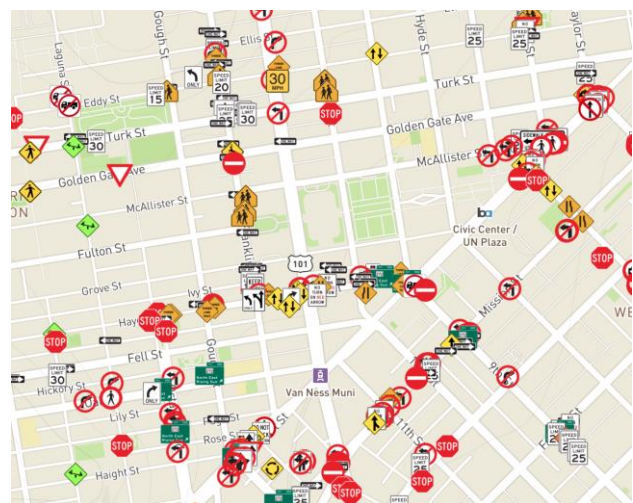


Figure 2. Example of traffic signs dataset automatically recognised by Mapillary

Traffic sign recognition system consists of two basic components: detection and classification. Detection is the step of finding where traffic signs are in images. Classification is the step of classifying each detected traffic sign into the corresponding traffic sign class.

To tackle the problem of traffic sign detection under varying capture conditions, Mapillary has integrated the semantic segmentation as part of the detection step at global level,

⁸ <https://blog.mapillary.com/update/2017/12/22/boosting-semantic-segmentation-performance-with-nvidia-and-amazon.html>

⁹ <https://blog.mapillary.com/update/2018/01/11/new-benchmarks-for-semantic-segmentation-models.html>

¹⁰ <https://www.mapillary.com/developer/api-documentation/#data-types>

¹¹ URBAN-GEO BIG DATA (URBAN GEOmatics for Bulk Information Generation, Data Assessment and Technology Awareness)

completed with a refinement process on the traffic sign segmentation to identify individual traffic sign instances as image patches.

The classification module is a light-weighted neural network trained for classification of image patches. Its key role is to tell the difference between different sign classes and assign a traffic sign class to each image patch. The training of this classification network benefits immensely from the data verified by the Mapillary community members who have contributed 1,000,000 verified traffic signs to be used as ground truth. The verified data from the community identifies hard cases of false detections, confusing classification, and occluded signs, which are essential perspectives for the neural network to learn from human-verified data. Figure 3 highlights the impact of manually verified data on the Recall and Precision accuracy metrics.

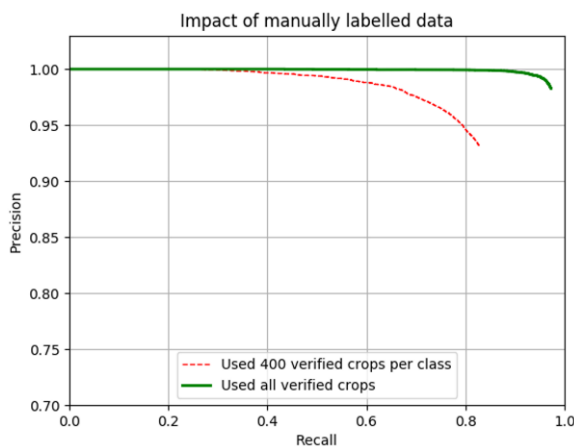


Figure 3. Significant improvements in recognition accuracy when using more verified signs¹²

3. ROAD NETWORK UPDATING: OPERATIONAL ISSUES

As detailed in the introduction section, the main goal of this research is to update/integrate the existing OSM road network dataset exploiting the Mapillary traffic signs dataset (see Chapter 2 for details). The following sections are aimed at describing the most common operational issues that have been encountered, that are mainly related to three categories: i) duplication of traffic sign features, ii) positional inaccuracies and iii) national and local traffic signs.

3.1 Duplication of traffic sign features

One of the main issues limiting the operational exploitation of automatically extracted traffic signs is the generation of multiple point features placed in different location whilst actually related to the same traffic sign.

In the example shown in Figure 4 the same “Keep right” traffic sign (bottom left) was automatically identified in different frames (from several tracks) and extracted as five point features in different locations (white boxes). A simple proximity approach can’t be adopted to merge the aforementioned features in a single traffic sign, since additional point features can be actually related to a different “Keep right” traffic sign (e.g. yellow boxes in Figure 1) in the opposite road lane.

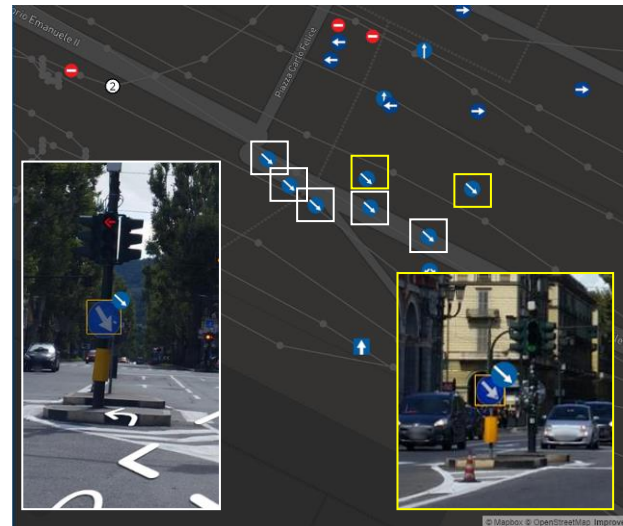


Figure 4. Multiple point features extracted by different frames/tracks related to the same “Keep right” traffic sign (snapshots from the Mapillary desktop app)

Ancillary information as the track direction of the frames used to extract the traffic sign features could be exploited to develop algorithms aimed at minimizing the traffic signs duplication issue. Nevertheless, a time consuming manual editing is currently required to disambiguate the correct location and aggregate the duplicated point features, especially considering that the distance among the different locations can be of several meters.

3.2 Positional accuracy

Positional inaccuracies of traffic signs features are a bottleneck in relating the traffic sign information to the relevant arc or node of the road network dataset. The following sections provide examples of possible geoprocessing workflows aimed at fixing positional inaccuracies in pre-defined scenarios.

Sign positioning supported by one-way restriction limitation available information. In this example, the procedure for correctly locating a “go straight or turn left” traffic sign, exploiting road network information, is described and discussed (Figure 5).

From the initial position, as generated by the Mapillary semantic segmentation tool, the traffic sign position is snapped to the nearest road network edge. The nearest crossing to the snapped point is then identified and road network traffic limitations at the level of the crossing are analysed. In the example of Figure 3 it becomes evident that a go straight or turn left sign cannot be placed on network element having ID 4291323, as it would be inconsistent with the one-way direction of network element having ID 31381685. In this specific context, the only possible assignment of the traffic sign is to the network element having ID 31381685. The traffic sign position is therefore moved on this network element, at a pre-defined distance from the crossing.

The correctness of the new positioning is visually confirmed by a street-level frame (Figure 6).

¹² <https://blog.mapillary.com/update/2018/06/15/global-sign.html>

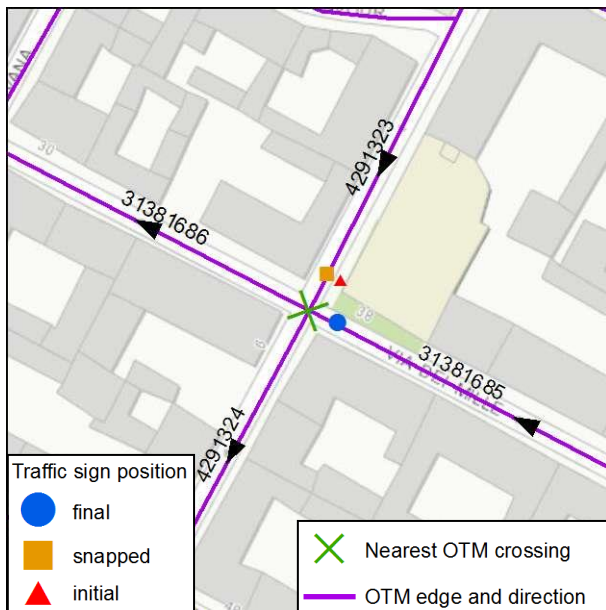


Figure 5. "Go straight or turn left" traffic sign positioning exploiting road network properties



Figure 6. Screenshot from the Mapillary web application where, both the traffic sign captured in the frame and its representation (generated by the Mapillary semantic segmentation tool) are displayed in the red rectangle.

Sign positioning not supported by one-way restriction limitation information. In the following example the possibility of enhancing traffic sign positioning and updating road network restrictions is discussed. Figure 7 shows the position of a "Turn right" and a "No entry" signs as generated by the Mapillary semantic segmentation tool. The nearest road edge to those signs can be therefore identified, together with the nearest road crossing. The situation is then analysed, based on the crossing geometry, to check for all possible assignment of traffic signs to other road edges. The "Turn right" traffic sign cannot be assigned to the network element having ID 4222330 as, approaching to the specific crossing, no road network elements are present on the right side: therefore this edge is excluded. For both the remaining 2 network elements, a compatible positioning option is identified and, consequently, two options of enhanced sign positioning are generated, together with associated network restrictions:

- option 1: the "Turn right" sign is attributed to road network element with ID 4302692, the "No entry" sign to ID 4222329: this last road element is also updated in terms of

traffic restrictions, by adding a one-way restriction (Figure 8, green outline);

- option 2: the "Turn right" sign is attributed to road network element with ID 4222329, the "No entry" sign to ID 4222330: this last road element is also updated in terms of traffic restrictions, by adding a one-way restriction (Figure 8, blue outline).

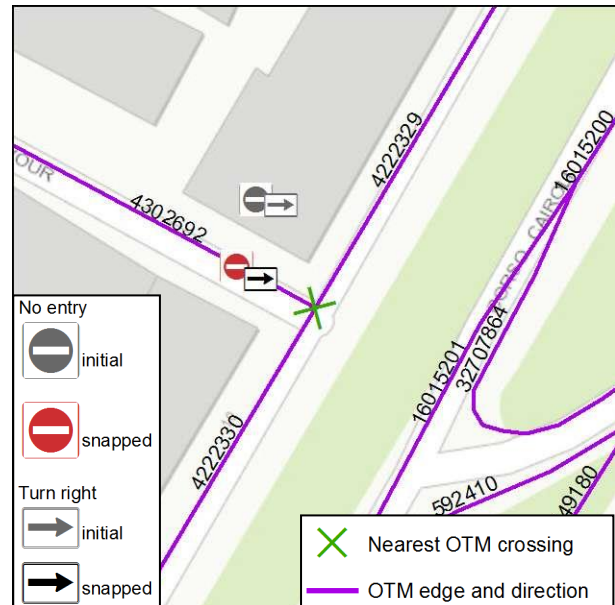


Figure 7. "No entry" and "Turn right" traffic signs in the initial and snapped positions

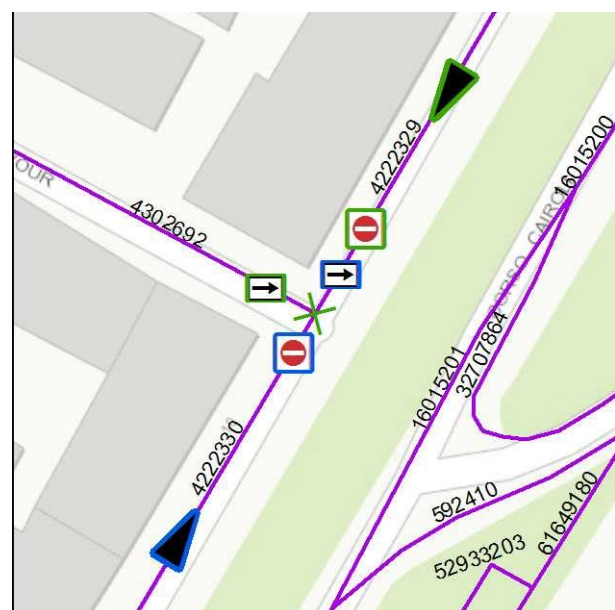


Figure 8. Enhanced traffic sign positioning and network restriction update based on option 1 (green outline) or option 2 (blue outline)

As in this case there is no possibility to unambiguously identify the correct solution, the procedure will then require the intervention of an operator to solve the situation. The relevant Mapillary frame is identified and displayed (Figure 9), allowing the operator to confirm that option 1 (Figure 8, green outline) represents the correct situation.



Figure 9. Mapillary frame captured on road element having ID 4302692

3.3 National and local traffic signs

Traffic signs can vary significantly in different countries: to setup a system with a global coverage it is important to take these differences into account. Mapillary is progressively updating its algorithm to cover different country-specific traffic sign sets, to unify traffic sign similarities and variations globally (Figure 10).



Figure 10. Country specific “No parking” signs currently managed by the Mapillary algorithm

Multiple traffic sign in the same panel. Multiple traffic sign in the same panel are quite common in Italy, especially in urban areas. Examples include the standard combination of no-parking, parking and disabled person icons (Figure 11, left) to identify a parking lot reserved to disabled people. This type of traffic sign is issued by the municipal authority on request by a disabled person living or working nearby. Currently this traffic sign is handled by Mapillary as separate point features, even if very close one to the other (Figure 11, right). This is not a major issue if this dataset is used for car navigation and parking management. But if exploited for creating or updating an inventory of traffic signs, this may cause issues due the wrong number of signs (e.g. if exploited to plan maintenance or to manage disabled people parking authorization).



Figure 11. A standard traffic sign identifying a parking lot reserved to a disabled person (left) and how it is recognised by Mapillary algorithm (right)

Another country specific panel with multiple icons, that change the meaning of the panel, is the one depicted in Figure 12: the parking sign is coupled with motorcycle, motorbike and bike icons, identifying a parking area reserved to these type of vehicles. In this case the current Mapillary algorithm flag it as a standard parking sign, generating a misleading information. Adding these multiple sign panels in the country specific training dataset would solve the aforementioned issues.



Figure 12. Traffic sign indicating a parking lot reserved to motorcycle, motorbike and bike and the related Mapillary detection.

No parking sign on private vehicle entrances. In Italy, authorization can be asked to the local authority for placing a no parking sign in front of private driveways. The specific traffic sign is composed by the specific sign reporting the name of the municipality (with or without the municipality coat of arms), the no-parking sign, the text “PASSO CARRABILE” (i.e. “Car Access”) and the municipality authorization number (Figure 13a). This sign is normally mounted next or on the vehicle entrance gate. The public authorities ask for issue costs and for an annual compensation for the occupation of public land. It is therefore necessary for a municipality to have a unique and easily accessible registry of all authorizations. This information is in most cases only stored in authorization archives, without direct geographical content.

Mapillary traffic sign recognition algorithm handles this specific traffic sign as a standard no-parking panel, not allowing to perform a distinction that would be necessary for the public authorities. Furthermore, the position of this sign, in most cases oriented at 90° with respect to the driving direction, makes its recognition critical.

Despite the Mapillary algorithm is able to detect these signs even with unfavourable geometry acquisition settings (Figure 13b and c), a limited test on 2 roads (for a total of 3 blocks each), suggests that only a portion (6 out of a total of 39, respectively green and red dots in Figure 14) of the “PASSO CARRABILE” signs are correctly identified.



Figure 13. No parking sign in front of private driveways (a) and successful traffic sign recognition by the Mapillary algorithm with unfavourable geometry acquisition settings (b) and (c)

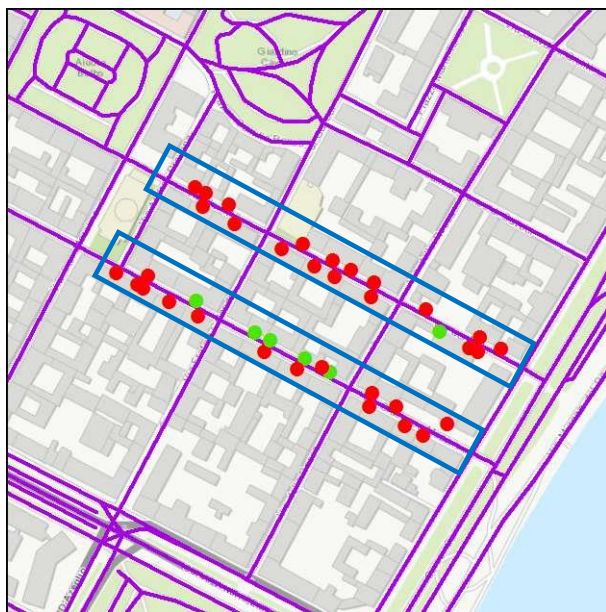


Figure 14. “PASSO CARRABILE” traffic signs in the test area (blue rectangle): actual positions (red dots) vs Mapillary detections (green dots)

4. CONCLUSIONS

The traffic sign dataset currently generated by the Mapillary recognition process is an interesting and unique dataset. There are already documented experiences in exploiting such dataset for comprehensive traffic signs inventories devoted to signs management¹³. Nevertheless, recognition and restitution can be further improved, especially for streamlining the ingestion and integration with other dataset, i.e. road networks.

Improvement may include:

- underlying algorithms (photogrammetric processing, semantic segmentation, deep neural network classification),

e.g. taking track directions of the processed frames into consideration should minimise traffic signs duplication;

- benchmarking studies to characterize the error margins of the output data using a variety of devices compared against data with known accuracy standards (e.g. ground control points or ground truth, high-resolution satellite imagery and LIDAR data);
- the quality of the input data (in terms of imagery and positional accuracy) through:
 - market penetration of new devices, e.g. mobile phones with dual frequency GPS¹⁴, better image quality in sub-optimal conditions (e.g. night acquisitions);
 - Mapillary contributors training/education on the best set-up/operational procedures, e.g. driving more slowly to maximise frames per second for denser reconstruction of the scene.

ACKNOWLEDGEMENTS

This research has been performed in the framework of a Project of National Interest (PRIN) named URBAN-GEO BIG DATA (URBAN GEOMatics for Bulk Information Generation, Data Assessment and Technology Awareness) , funded by the Italian Ministry of Education, University and Research (MIUR) – id. 20159CNLW8.

REFERENCES

- Cariucci, F. M., Porzi, L., Caputo, B., Ricci, E., Rota Bulò, S., 2017. AutoDIAL: Automatic Domain Alignment Layers. IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 5077-5085. doi: 10.1109/ICCV.2017.542
- Mancini, M., Porzi, L., Rota Bulò, S., Caputo, B., Ricci, E., 2018. Boosting Domain Adaptation by Discovering Latent Domains. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3771-3780
- Neuhold, G., Ollmann, T., Rota Bulò, S., Kotschieder, P., 2017. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 5000-5009. doi: 10.1109/ICCV.2017.534
- Pumarola, A., Agudo, A., Porzi, L., Sanfeliu, A., Lepetit, V., Moreno-Noguer, F., 2018. Geometry-Aware Network for Non-Rigid Shape Prediction From a Single View. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4681-4690
- Rota Bulò, S., Porzi, L., Kotschieder, P., 2018. In-Place Activated BatchNorm for Memory-Optimized Training of DNNs. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5639-5647
- Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., Reynolds, J. M., 2012. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*. 179: 300–314. doi:10.1016/j.geomorph.2012.08.021

Revised March 2019

¹³ <https://blog.mapillary.com/update/2018/11/14/streamlining-traffic-sign-inventory-in-clovis.html>

¹⁴ <https://www.gsa.europa.eu/newsroom/news/world-s-first-dual-frequency-gnss-smartphone-hits-market>