Multi-Feature Sample Database for Enhancing Deep Learning Tasks in Operational Humanitarian Applications

GI_Forum 2021, Issue 1
Page: 209 - 219
Research Paper
Corresponding Author:
stefan.lang@sbg.ac.at
DOI: 10.1553/giscience2021_01_s209

Stefan Lang[1], Lorenz Wendt[1], Dirk Tiede[1], Yunya Gao[1], Vanessa Streifender[1], Hira Zafar[1], Adebowale Adebayo[1], Gina Schwendemann[2] and Peter Jeremias[1]

[1]University of Salzburg, Salzburg, Austria
[2]Spatial Services GmbH, Salzburg, Austria

Abstract

Amongst the many benefits of remote sensing techniques in disaster- or conflict-related applications, timeliness and objectivity may be the most critical assets. Recently, increasing sensor quality and data availability have shifted the attention more towards the information extraction process itself. With promising results obtained by deep learning (DL), the notion arises that DL is not agnostic to input errors or biases introduced, in particular in sample-scarce situations. The present work seeks to understand the influence of different sample quality aspects propagating through network layers in automated image analysis. In this paper, we broadly discuss the conceptualisation of such a sample database in an early stage of realisation: (1) inherited properties (quality parameters of the underlying image such as cloud cover, seasonality, etc.); (2) individual (i.e., per-sample) properties, including *a.* lineage and provenance, *b.* geometric properties (size, orientation, shape), *c.* spectral features (standardized colour code); (3) context-related properties (arrangement Several hundred samples collected from different camp settings were hand-selected and annotated with computed features in an initial stage. The supervised annotation routine is automated so that thousands of existing samples can be labelled with this extended feature set. This should better condition the subsequent DL tasks in a hybrid AI approach.

Keywords: humanitarian action, earth observation, deep learning, data assimilation, hybrid AI, sample quality, automation

1 Shifting demands in operational humanitarian EO

1.1 Time criticality vs reliability

Remote sensing and Earth observation (EO) derived products play a growing role in providing relevant and up-to-date information for humanitarian operations (Lang et al., 2019). Amongst the many benefits of remote sensing techniques in disaster- and conflict-related applications, timeliness and objectivity may be regarded as the most critical assets (Denis et al., 2016; Voigt et al., 2016). This applies, for example, to refugee camp mapping or dwelling extraction routines in deprived urban areas for population estimation, where otherwise such figures are missing or largely outdated (Quinn et al., 2018). About a decade ago, when the humanitarian community started to adopt EO technologies in operations, both aspects of actuality (i.e., up-to-date and trustworthy information) were mainly referred to image (source) quality. Reluctance with respect to image provision, image manipulation, limited spatial resolution, cloud contamination, etc., was the main concern with respect to operational use. Recently, increasing sensor quality, data fusion techniques, and data availability have shifted the attention more towards the information extraction process itself (Lang, Füreder, et al., 2019). The increasing acceptance has experienced a synchronous shift in attention of the larger EO community from data management to data exploitation (Giuliani et al., 2017; Sudmanns et al., 2019; Voigt et al., 2016).

In highly demanding operational settings, timeliness and reliability may be considered mutually exclusive, if not contradictory, even. In demanding tasks, innovation in the automation process is limited, making the information extraction process 'stagnant' and dominated by and manual delineation. Recently, the community saw many approaches labelled as "semi-automated", attempting to best implement computer vision with (GE-)OBIA techniques (Lang, Hay, Baraldi, Tiede, & Blaschke, 2019) and to overcome the tedious delineation process of small features which occurs in large frequencies and diversities (Füreder et al., 2015). In particular, in well-structured camp arrangements with distinct structures, the performance of region-based segmentation routines are satisfying and – once the delineation of dwellings has been achieved – object features such as size, colour, shape, etc., can be used to categorise them. The process is challenging when a clear distinction of individual dwellings is hampered by the complexity of the arrangement, and even visual inspection reaches its limits, and inter-subject objectivity is no longer guaranteed among experts.

With promising results obtained by deep learning (DL) in various applications (Ghorbanzadeh, Tiede, Wendt, Sudmanns, & Lang, 2020; Quinn, et al., 2018; Tiede, Wendt, Schwendemann, Alobaidi, & Lang, 2021) humanitarian community adopts to data science techniques as well. This also applies to computer vision, which gradually evolves from static rule-based strategies to a more dynamic, self-adaptable machine learning-based approach. The limitation for the latter, however, is the existence and quality of samples. Despite the inherent improvement capabilities of machine learning, DL is not agnostic to input errors (Ghorbanzadeh, Tiede, Dabiri, Sudmanns, & Lang, 2018) or biases introduced, in particular in sample-scarce situations. Sample scarcity in humanitarian applications may be attributed to the complexity and required level of detail (e.g., complex urban settings or organically grown refugee camps), for which samples on a generalised level do not exist in sufficient number or quality. Even though tents and other dwelling types can be generalised and described according to standard building codes, the confusion with other and similar features, mixed in and intermingled, is high and the appearance on EO imagery greatly depends on seasonal conditions (dry vs. humid periods, overgrown by vegetation, etc.). The detection and correct interpretation of different dwelling types, in a degree relevant to humanitarian organisations, is a dedicated expert task. While support is increasingly available through community-mapping approaches such as Humanitarian Open Street Map (HOT OSM or Missing Maps), utilising crowd-sourced information needs, therefore, to be curated and the existing dwelling delineations need to be evaluated and characterised (Albuquerque, Herfort, & Eckle, 2016; Elia, Balbo, & Boccardo, 2018).

The challenge remains: compromising reliable results for the sake of increased automation is a tricky decision in operational humanitarian settings, where actions and decisions may have severe implications for human lives and wellbeing. Taking the 'best of two worlds', we try to apply hybrid approaches, which are aware of the physical properties of the target dwellings, which rests upon the experience of operational mapping task of the last ten years. Based on this legacy, an annotated dwelling sample database is foreseen, which documents sample provenance and characteristics in a way that observations and dwelling models are well attuned in a hybrid AI and data assimilation approach.

1.2 Hybrid AI and data assimilation

Artificial intelligence (AI) simulates processes characteristic to human intelligence and thereby mimics human actions. Among the cognitive relevant AI tasks includes knowledge representation, automated reasoning, machine learning (ML) and teaching. Types of AI are distinguished by adaptability, performance, proficiency as compared to the human brain ranging from narrow AI and general AI to Super AI. Physics-aware or hybrid AI is a strategy to better condition ML/DL tasks by employing physical models, principles, or even laws. These principles into consideration using general conditions and constraints utilising machine teaching as an enabler (Lang, Hay, et al., 2019). One strategy is data assimilation.

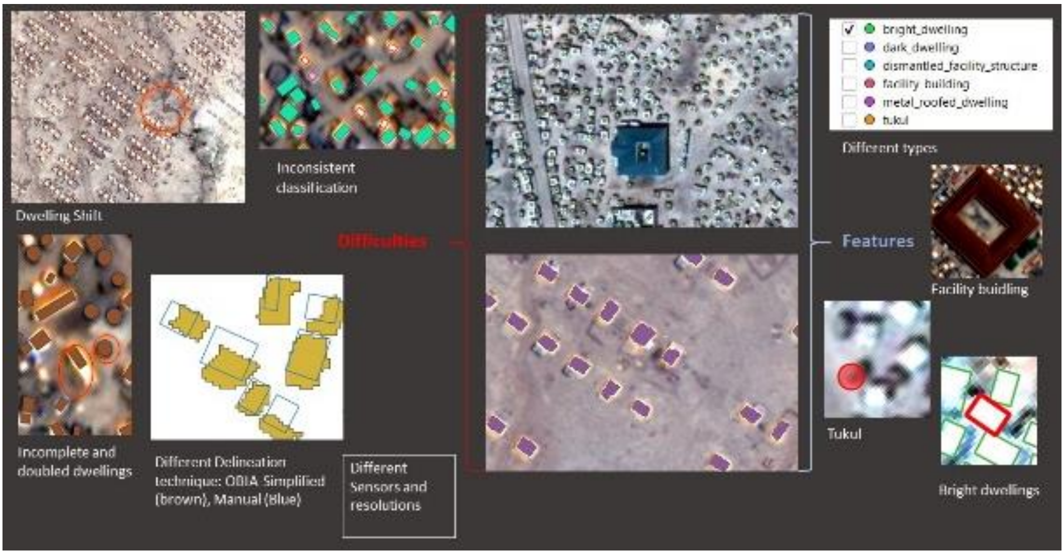
Data assimilation aims to foster data integration and data harmonisation in a bi-directional way between the measured and the modelled reality (Lahoz and Schneider, 2014). In Earth observation, data assimilation compensates for the fact that a specific site may be observed in a variety of measurements by satellites with different sensor types, at different dates, different angular geometries and viewing directions, illumination conditions (solar time), observation frequencies, etc. (Verhoef and Bach, 2003). In particular, for monitoring purposes, measurements over time need to assure to actually measure the status of the system or object and not the divergence in observation. For vegetation and crop type monitoring, radiative transfer modelling (RTF) is being used as an example (Graf, Papp, & Lang, 2020; Verhoef and Bach, 2003). In general, when interpreting images and overcoming the semantic gap, rigorous classifiers based on solid spectral models, acting across sensors, are available. Semantic enrichment of satellite data (Augustin, Sudmanns, Tiede, Lang, & Baraldi, 2019). For satellite image time-series (SITS), the seasonal dynamics and the variability the appearance of the target classes are relevant. Data assimilation can also bridge non-availabilities of EO data and other observations to provide estimates or prediction for geographical variables. A related aspect is data imputation, i.e. filling gaps in observations, e.g. by other, complementary data sets (e.g. Radar imagery in the absence of VHR data under cloudy weather conditions).

2 Quality-controlled samples

2.1 Rationale in the context of dwelling extraction

A better understanding of sample quality is a critical requirement to improve automated DL tasks in image analysis. Our aim is to investigate systematically how various imperfections in the delineation and provision of samples affect the result of machine learning. We, therefore, in a first step, defined and tested a set of quality indicators, computed and recorded in a database next to each sample's label (dwelling category). These indicators comprise: (1) inherited properties (quality parameters of the underlying image such as cloud cover, seasonality, etc.); (2) individual (i.e., per-sample) properties, including a. lineage and provenance, b. geometric properties (size, orientation, shape), c. spectral features (standardized colour code); (3) context-related properties (spatial arrangement Currently, the approach is 'static' and does not consider the temporal dynamics of dwelling evolution, meaning we record quality indicators per image timeslot (epoch). Several hundred samples collected from different camp settings were selected and annotated using the expert-based selection of quality indicators in an initial stage. It is foreseen that thousands of existing samples and future delineations are labelled automatically with this set of quality-relevant features.

The following figures illustrate the challenges encountered in documenting the quality indicators of the samples using mixed methods for dwelling delineation in an operational production environment. We deal with different sensors and image resolutions, limitations due to cloud cover and atmospheric conditions, problems of geometric correction (shifts), incomplete interpretation or extraction (false positives and negatives), different delineation techniques (segmentation-based vs. manual delineation), and inconsistencies in the classification and labelling (see figure 1). Some of these aspects influence the quality of the samples globally, i.e. per image. Atmospheric conditions and cloud contamination or any other aspects of image correction introduce a global bias to the extraction process. While difficult to estimate, this bias is an important aspect of data provenance in the process of turning primary (continuous image) data into secondary (discrete object) data.



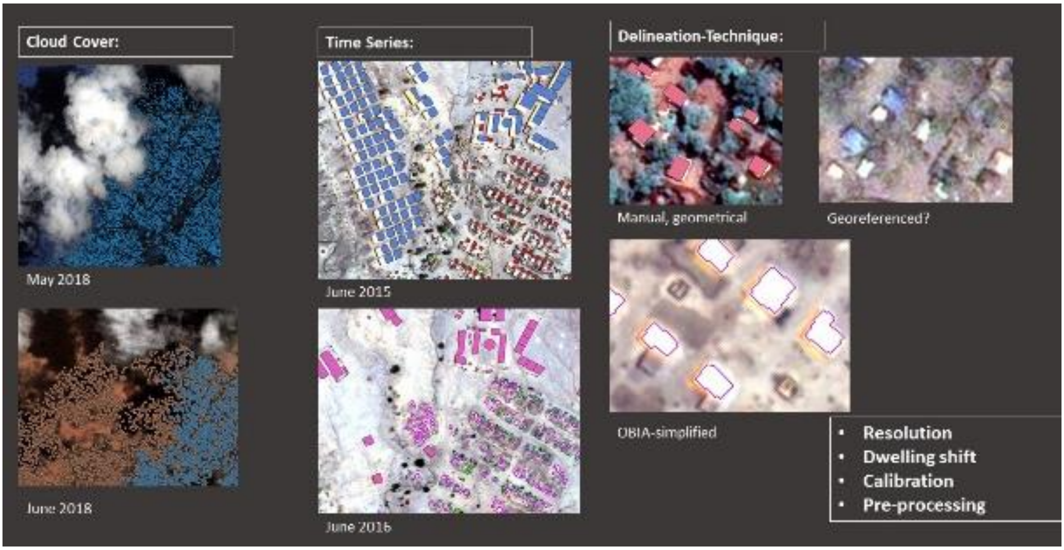


Figure 1: Various aspects to consider in estimating quality parameters of extracted dwellings (see text for further explanation).

Global (per-image) quality indicators. We aim to use quality-proven polygon data as input for training tasks of ML algorithms rather than labelled imagery. We, therefore, have to ensure that the extracted dwelling polygons have a unique image source assigned. What sounds trivial is sometimes impaired by the complexity of operational tasks, using multiple input data (e.g. VHR satellite images, drones), or observing dwellings over a certain longer time period with updated image data (monitoring). Once a unique match of source and dwelling object is assured, the produced dwelling data will inherit the global image quality indicators, like cloud cover, incidence angle, atmospheric conditions, geometric shift, etc. This is another crucial aspect of data provenance and reproducibility because only polygon data with a unique source should be considered a matching pair to be used as training samples.

In the absence of an alternative option, we are currently experimenting with a global quality score for judging the data provenance as a combination of image resolution, cloud cover and feature delineation (including shifts). Table 1 shows a draft version of such a grading scheme, which would attest all samples taken from one input image at a certain epoch a global quality score. Those with a quality score 1 could be used as testing samples to start the model training without any bias. Further, samples with a lesser quality can be used for training to increase the robustness of the model. Samples of quality scores 3 or 4 might suffer inconsistencies in spatial registration or delineation type but may still be used for sample augmentation purposes. Quality score 6 would indicate a status of non-correspondence between image and extracted dwellings.

Table 1: Grading scheme for assessing the overall quality of data provenance (draft)

Delineation	Image resolution	Cloud cover	Quality score
Nearly perfect delineation, minor differences in delineation	High	none	1
Minor differences in delineation	high-medium	none	2
Larger difference in delineation; repairable dwelling shift	High-medium	partly	3
Dwellings partly missing; repairable dwelling shift	High-medium	partly	4
Many dwellings missing; shift not repairable	Medium-Low	large	5
no correspondence in delineation	-	-	6

2.2 Quality features per dwelling

The present work is a precursor to a larger investigation that aims at documenting systematically how different aspects of quality of samples propagate through artificial neural network layers, thereby judging how this reflects in the result of the DL task. An extensive number of annotated samples, semi-automatically derived and manually revised, are collected, representing features suitable for enumerating and estimating the actual local population; they are stored and made accessible in a dedicated sample database. The samples consist of vector representations (polygons) of dwellings of different types like tents, huts, tukuls, facility buildings, etc.; hence have a different characteristic to be taken into account for the structure of the database. Next to the labels, the dwelling samples are characterised by a set of quality indicators assessing their spectral and spatial properties (table 2).

Table 2: Dwelling sample quality indicators

	Dwelling delineation process
Dwelling shift	Polygons do not match the source image and show offsets
Inconsistency in classification	Different labels depending on the source of the footprints and image
Incomplete and double count of dwellings	Dwellings double-counted or do not totally cover the apparent dwelling on the source image
Delineation strategy	Manual, semi-automated (OBIA), etc.
	Image characteristics (inherited by dwellings)
Cloud cover	Not all dwelling in the imaged scene captured
SITS	Evolution of a camp and seasonal effects
Delineation strategy	(see above)
	Dwelling individual properties
Representation	As polygon, as point (centroid)
Spectral properties	Colour categories (SIAM-based)
Neighbourhood, context	Embeddedness in dwellings of the same type
Geometrical attributes	Size, position (centroid)
Shape	Compactness, regular fit, orientation, etc.
SITS	Dwelling dynamics (emergence, disappearance, etc.)

Spectral features. Spectral characteristics are recorded based on standardised colour categories using a knowledge-based feature space partitioning system called SIAM® The idea is to generate standardised categories (semi-concepts) fully automated from VHR imagery (figure 2). This requires calibration (as far as possible), even in operational, demanding application contexts. The advantage is to have stable categories rather than subjective colour impressions ("*light-blue tents*", "*brownish tukul*", "*bright dwellings*", etc.). This helps enrich the sample database because we have a (certain level of) semantic understanding of the global image content (e.g. dominance of dwelling type X) and a per-object uniform colour label to support the classification. The standardised colour categories are based on a fully automated pre-classification of the multi-spectral properties from VHR images; this process involves radiometric calibration of the images into top-of-atmosphere (TOA) reflectance followed by a knowledge-based feature classifier. The image calibration process includes the absolute radiometric calibration factors provided by the VHR image vendors; this ensures the baseline for multiple sensors data integration and fusion in every operational and demanding application contexts. The predefined colour codes consist of a discrete and fixed number of cross-sensor spectral categories (e.g. 33 or 61) whose degree of semantic information – while lower than common land cover classes – is superior to non-semantic image data. This provides a stable and uniform representation of object primitives labelled to support recognition and classification by the model.

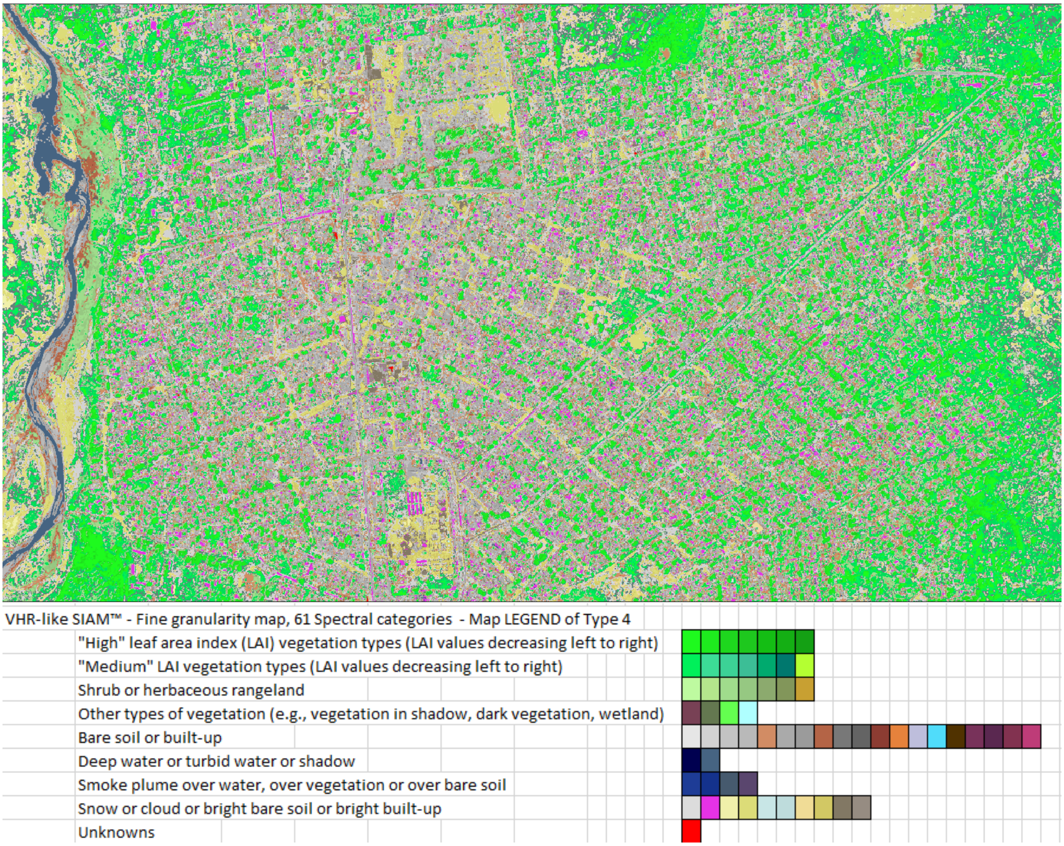


Figure 2: Fully automatic pre-classification of calibrated VHR Pléiades imagery into a fixed set of 61 spectral categories using SIAM®. Different vegetation, water, built-up, and other low-level semantic classes are discerned in a standardised and transferable manner,

Spatial features. Spatial properties comprise geometric properties and spatial arrangement. The azimuth angle is the angle between two points in the Cartesian plane; it is calculated between the centroid of the dwelling itself and the neighbouring dwelling. The orientation angle was calculated to show the orientation of individual dwellings against geographic North. The shape index measures the deviation of a given polygon from the circularity of a perfect circle of the same size. For any set of geometric forms of a given area, the circle has the shortest perimeter in relation to the area; thus its compactness is highest. Any other form exhibiting the same area shows less compactness and a higher shape index. The proximity index is well suited for indicating the embeddedness of an extracted dwelling in its surrounding, i.e. to which degree a dwelling differentiate from neighbours in terms of size and distance. It was calculated for each dwelling by identifying the dwellings that were within the buffer distance of the indexed dwelling and then calculating the size to distance ratio for each of the n dwellings identified within the buffer.

2.3 Annotated sample database – a prototypical implementation

A spatial database in PostgreSQL stores the quality indicators of the image source and the respective dwellings (vector data). This database aims to store spatial and spectral characteristics of the dwellings analysed in the area of investigation. The simplified database schema (figure 3) consists of two main tables: the Image table and the Dwelling table connected via image ID as primary and foreign key. The dwellings delineated from each image were originally stored in separate individual tables per country and now collated in one single schema. The Image table serves as the main table containing information about individual dwellings in all images in one place. The Image table contains information about image characteristics of the image (radiometric, etc.) properties. The Dwelling table contains information about spatial characteristics of individual dwellings, as described above.

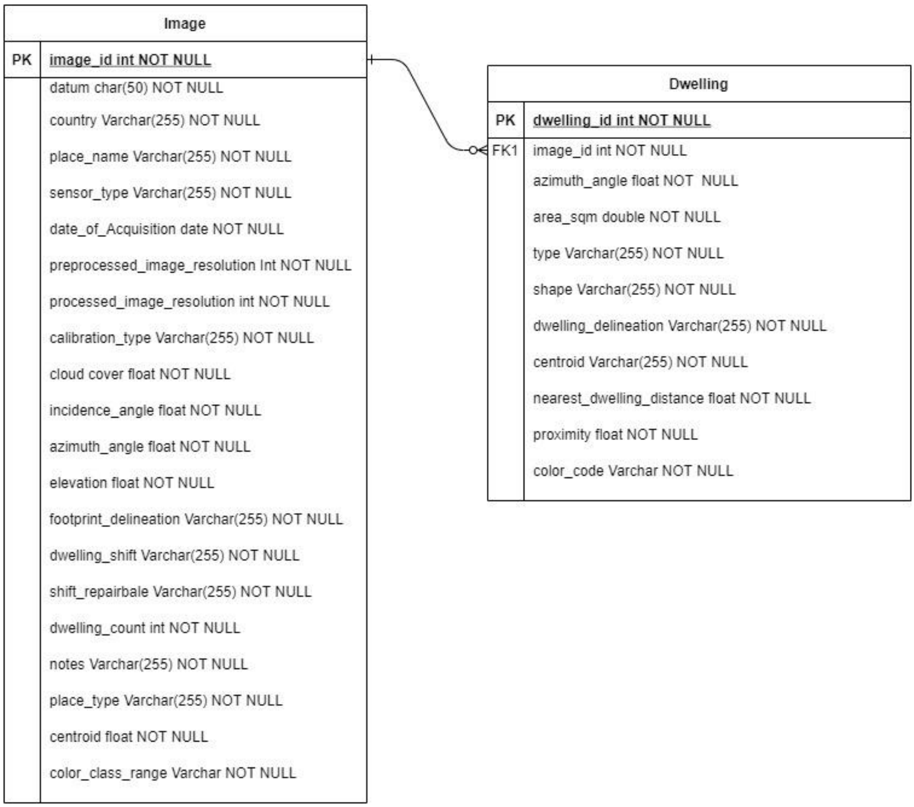


Figure 3: Simplified ER diagram of sample DB

3 Outlook

The described conceptual framework for quality indicators of dwelling extraction is currently investigated and expert-evaluated in terms of impact on the performance of various DL tasks. Being in an early stage of development, we plan to consider temporal aspects of dwelling evolution by deriving quality indicators not from single epochs but from multi-temporal (and potentially semantically enriched) data cubes.

Prospectively, hundreds of thousands of existing samples are going to be labelled automatically with this extended set of quality indicators. This should better condition the subsequent mapping tasks using a hybrid AI approach and improve existing operational mapping routines. It may also serve as a stimulating reference dataset for benchmark contests.

References

Albuquerque, J. P., Herfort, B., & Eckle, M. (2016). The tasks of the crowd: a typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. *Remote Sensing, 8*(10), p 859.

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2019). Semantic Earth observation data cubes. *Data, 4*(3), pp. 102-115. doi:10.3390/data4030102

Baraldi, A. (2011). Satellite Image Automatic Mapper™ (SIAM™). A turnkey software button for automatic near-real-time multi-sensor multi-resolution spectral rule-based preliminary classification of spaceborne multi-spectral images. *Recent Patents on Space Technology*(1), pp. 81-106.

Denis, G., de Boissezon, H., Hosford, S., Pasco, X., Montfort, B., & Ranera, F. (2016). The evolution of Earth Observation satellites in Europe and its impact on the performance of emergency response services. *Acta Astronautica, 127*, pp. 619-633.

Elia, A., Balbo, S., & Boccardo, P. (2018). A quality comparison between professional and crowdsourced data in emergency mapping for potential cooperation of the services. *European Journal of Remote Sensing, 51*(1), pp. 572-586.

Füreder, P., Lang, S., Hagenlocher, M., Tiede, D., Wendt, L., & Rogenhofer, E. (2015). Earth observation and GIS to support humanitarian operations in refugee/IDP camps. In Palen, Büscher, Comes & Hughes (Eds.), *Geospatial Data and Geographical Information Science*. Kristiansand.

Ghorbanzadeh, O., Tiede, D., Dabiri, Z., Sudmanns, M., & Lang, S. (2018). Dwelling extraction in refugee camps using CNN - first experiences and lessons learnt. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII*(1), pp. 161-168.

Ghorbanzadeh, O., Tiede, D., Wendt, L., Sudmanns, M., & Lang, S. (2020). Transferable instance segmentation of dwellings in a refugee camp - integrating CNN and OBIA. *European Journal of Remote Sensing*, pp. 1-14. doi:10.1080/22797254.2020.1759456

Giuliani, G., Chatenoux, B., Bono, A. D., Rodila, D., Richard, J.-P., Allenbach, K., . . . Peduzzi, P. (2017). Building an Earth observations data cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data, 1*(1-2), pp. 1-18. doi:doi.org/10.1080/20964471.2017.1398903

Graf, L., Papp, L., & Lang, S. (2020). OBIA4RTM – towards an operational open-source solution for coupling object-based image analysis with radiative transfer modelling. *European Journal of Remote Sensing*, pp. 1-12. doi:10.1080/22797254.2020.1810132

Lahoz, W. A., & Schneider, P. (2014). Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science, 2*, p 16. doi:doi.org/10.3389/fenvs.2014.00016

Lang, S., Füreder, P., Riedler, B., Wendt, L., Braun, A., Tiede, D., . . . Hochschild, V. (2019). Earth observation tools and services to increase the effectiveness of humanitarian assistance. *European Journal of Remote Sensing*, pp. 1-19. doi:10.1080/22797254.2019.1684208

Lang, S., Hay, G., Baraldi, A., Tiede, D., & Blaschke, T. (2019). GEOBIA achievements and spatial opportunities in the era of big Earth observation data. *ISPRS International Journal of Geo-Information, 8*, pp. 474-483. doi:doi.org/10.3390/ijgi8110474

Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A, 376*(2128)

Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., . . . Blaschke, T. (2019). Big Earth data: disruptive changes in Earth observation data management and analysis? *International Journal of Digital Earth*, pp. 1-19. doi:10.1080/17538947.2019.1585976

Tiede, D., Wendt, L., Schwendemann, G., Alobaidi, A., & Lang, S. (2021). Mask R-CNN-based building extraction from VHR satellite data in operational humanitarian action. *Transactions in GIS, (in press)*

Verhoef, W., & Bach, H. (2003). Remote sensing data assimilation using coupled radiative transfer models. *Physics and Chemistry of the Earth, 28*(1-3), pp. 3-13.

Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., . . . Guha-Sapir, D. (2016). Global trends in satellite-based emergency mapping. *Science, 253*(6296), pp. 247-252. doi:doi:10.1126/science.aad8728