

Automated building characterization for seismic risk assessment using street-level imagery and deep learning

Patrick Aravena Pelizari^{a,*}, Christian Geiß^a, Paula Aguirre^{b,d}, Hernán Santa María^{b,c}, Yvonne Merino Peña^{b,c}, Hannes Taubenböck^{a,e}

^a German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Münchner Straße 20, 82234 Weßling-Oberpfaffenhofen, Germany

^b National Research Center for Integrated Natural Disaster Management (CIGIDEN), Vicuña Mackenna 4860, Macul, Santiago, Chile

^c Department of Structural and Geotechnical Engineering, Pontifical Catholic University of Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile

^d Institute for Mathematical and Computational Engineering, School of Engineering, Pontifical Catholic University of Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile

^e University of Würzburg, Institute of Geography and Geology, Chair of Remote Sensing, Oswald-Kuelpe-Weg 86, 97074 Würzburg, Germany



ARTICLE INFO

Keywords:

Building characterization
Street-level imagery
Visual-structural criteria
Deep convolutional neural networks
Image classification
Seismic risk assessment

ABSTRACT

Accurate seismic risk modeling requires knowledge of key structural characteristics of buildings. However, to date, the collection of such data is highly expensive in terms of labor, time and money and thus prohibitive for a spatially continuous large-area monitoring. This study quantitatively evaluates the potential of an automated and thus more efficient collection of vulnerability-related structural building characteristics based on Deep Convolutional Neural Networks (DCNNs) and street-level imagery such as provided by Google Street View. The proposed approach involves a tailored hierarchical categorization workflow to structure the highly heterogeneous street-level imagery in an application-oriented fashion. Thereupon, we use state-of-the-art DCNNs to explore the automated inference of Seismic Building Structural Types. These reflect the main-load bearing structure of a building, and thus its resistance to seismic forces. Additionally, we assess the independent retrieval of two key building structural parameters, i.e., the material of the lateral-load-resisting system and building height to investigate the applicability for a more generic structural characterization of buildings. Experimental results obtained for the earthquake-prone Chilean capital Santiago show accuracies beyond $\kappa = 0.81$ for all addressed classification tasks. This underlines the potential of the proposed methodology for an efficient *in-situ* data collection on large spatial scales with the purpose of risk assessments related to earthquakes, but also other natural hazards (e.g., tsunamis, or floods).

1. Introduction

Many of the world's largest urban agglomerations are located in regions with high seismic vulnerability (Gu et al., 2015). Due to population growth paired with rapid urbanization it is expected that populations in earthquake-prone regions will face greater losses in the future (Bilham, 2009; Tucker, 2013). Besides the seismic hazard itself, earthquake resistance of exposed buildings is a key parameter for determining the level of seismic losses (Wyss and Rosset, 2013). Thus, an up-to-date model of the exposed built environment including its physical seismic vulnerability is a prerequisite for earthquake risk analyses and loss estimations to initiate pre- and post-event emergency management (Geiß and Taubenböck, 2013).

Traditionally, required information is collected via *in-situ* surveys, where structural engineers assess construction features building-by-building while considering a standardized protocol (e.g., FEMA, 2015). These surveys are expensive in terms of labor and time, making them prohibitive for a continuous inventorization on larger scales (e.g., entire cities). Institutional databases such as census or tax assessor data are another source of information (e.g., Sarabandi and Kiremidjian, 2007). Such data, however, are often inappropriate, out-of-date, or not available (Geiß et al., 2016).

1.1. Seismic risk assessment

The risk R_b (i.e., the probability to suffer a certain degree of damage

* Corresponding author.

E-mail address: patrick.aravenapelizari@dlr.de (P. Aravena Pelizari).

within a specific future time period) for a building E_b exposed to a seismic hazard H_i (i.e., an earthquake with a certain probability of occurrence within a certain time period) with a local intensity i can be regarded as a function:

$$R_b = f(H_i, E_b, V_{(E_b, H_i)}), \quad (1)$$

with $V_{(E_b, H_i)}$ representing the physical vulnerability of E_b , i.e., its propensity to suffer damage from H_i due to its constructional design (Coburn and Spence, 2002). In a seismic risk model, $V_{(E_b, H_i)}$ is represented by a vulnerability model that relates H_i to the damage probability of a building (e.g., a fragility curve). This allows assessing the vulnerability of E_b with respect to a specific level of H_i (Calvi et al., 2006).

To estimate the vulnerability distribution of building inventories, individual structures are classified according to their key seismic resistance determining structural properties. This results in groups with similar expected behavior under seismic forces – the Seismic Building Structural Type (SBST; Coburn and Spence, 2002). In this sense, the SBST characterizes the main load-bearing system of a structure, generally considering the material and configuration of the lateral load resisting system (LLRS), along with the building height as the main influencing factors. When it comes to risk assessment, the SBST is deployed to interlink exposed structures E_b with a specific representative vulnerability model $V_{(E_b, H_i)}$.

1.2. Remote sensing (RS) for mapping of building structural vulnerability

Due to its diverse modalities and high spatiotemporal coverage, the use of RS data holds a high potential to address a large-area collection of up-to-date information on exposed buildings (Geiß and Taubenböck, 2013). In particular very high spatial resolution (VHR) RS data allows for the spatially continuous derivation of vulnerability-related structural characteristics at building level. Taubenböck et al. (2009) and Borzi et al. (2011) indirectly infer building vulnerability from VHR optical RS data. First, they reconstruct and characterize the individual buildings in an object-based fashion (Blaschke, 2010). Then, they assign specific fragility functions to distinct manifestations of derived building attributes (e.g., height, extent, or roof type) based on expert knowledge.

Other studies directly deploy the correlations between features encoded in the RS data and building vulnerability within data-driven machine learning frameworks (Aravena Pelizari et al., 2018). In this sense, Sarabandi and Kiremidjian (2007) calculate building attributes from VHR multispectral RS imagery, align this feature set to occupancy type as well as building age from tax assessor data, and predict the SBST using supervised learning techniques. Geiß et al. (2015) step upon this path but in contrast they fully substitute auxiliary geo-data by a comprehensive set of descriptive features derived from complementary multi-sensor RS data. Such supervised learning approaches particularly depend on appropriate and sufficiently available reference information being either gathered within *in-situ* surveys or derived from institutional data (see also, e.g., Panagiota, et al., 2012; Geiß et al., 2014, 2016; Liuzzi et al. 2019). As outlined above, however, such data is expensive, barely available and often spatially scarce or outdated, which severely limits their flexible and large-area application.

1.3. Remote rapid visual screening (RRVS)

To facilitate and speed-up systematic *in-situ* building information gathering, Wieland et al. (2012) deploy a mobile mapping system. An omnidirectional camera mounted on a vehicle allows for an automated compilation of georeferenced imagery on street-level. Image sequences are captured and structural attributes are collected via visual image interpretation by experts. This is referred to as RRVS and enables operators to inspect a large number of structures in short time in a decentralized and location-independent manner (Geiß et al., 2017a). To finally obtain spatially consistent area-wide inventories, collected data

can be fed to extrapolation procedures based on RS data (Wieland et al., 2012; Geiß et al., 2017a; Section 1.2). Thereby, well elaborated routing strategies for a representative yet efficient data capturing are required (Geiß et al., 2018). However, such location-specific data collection campaigns are still costly in terms of time and money.

Nowadays, the large-area availability of systematically collected and regularly updated street-level data due to global initiatives such as Google Street View (GSV; Anguelov et al., 2010) but also Mapillary¹ or Apple's Look Around service², offer new possibilities for the gathering of *in-situ* information. In covered areas, RRVS can now be completely decoupled from physical on-site presence (e.g., Santa María et al., 2017, Pittore et al., 2018). RRVS based on GSV enabled Santa María et al. (2017) to compile detailed exposure models for the residential structures of three Chilean cities. They derive the LLRS material and the number of storeys among others, based on which residential buildings are categorized into 18 SBSTS. The survey involved a visual inspection of ~21 K individual buildings and was performed within thirty working days, by three structural engineering students and a supervisor (Rivera et al. 2017). These figures underline the complexity and the high efforts still associated with RRVS.

1.4. Objective of this study: automatization of building surveys on street-level with Deep Convolutional Neural networks (DCNNs)

Is there any realistic scenario to obtain *in-situ* building information for seismic risk assessment with large-area coverage in an automated low-cost fashion? If it is possible for image interpreters to infer seismic risk-related building attributes from geotagged street-level images, we believe DCNNs could do likewise. Thus, the presented study explores the potential of custom trained DCNNs for the collection of such information.

Recent developments in machine learning and computer vision – in deep learning in particular – enabled great advances in solving perceptual tasks in the image domain (LeCun et al., 2015). Raising efforts were set on the well-thought engineering of modular organized DCNNs (e.g., Simonyan and Zisserman, 2014; He et al., 2016; Szegedy et al., 2016b) successively evoking improved accuracies on popular benchmark datasets. As a follow up step, research has been spent to automate the engineering of DCNN architectures by incorporating Neural Architecture Search (NAS) into model optimization. Taking this path, NAS approaches, e.g., NASNet (Zoph et al. 2018) showed their potential to create model architectures that outperform purely hand-crafted DCNNs.

Several studies have already explored the automated gathering of information on urban environments using geotagged street-level images and DCNNs. Among them, Kang et al., 2018 explore the use of GSV images and DCNN scene classification to assign building instances as delineated by Open Street Map to one of eight use-classes. Srivastava et al. (2019) as well as Hoffmann et al. (2019) predict urban-land-use classes of Open Street Map building objects, but, in addition to GSV, deploy aerial images. They propose multi-stream DCNN models to fuse both modalities enabling significant improvements in classification accuracies on their benchmark data. Branson et al. (2018) deploy GSV panoramas jointly with aerial imagery to inventorize street-trees. Individual trees are located with multimodal object-detection and subsequently classified according to tree species. On top, building-upon the multi-temporal dimension of GSV, street-level tree views are subject to tree change tracking using a siamese DCNN. Similarly, Illic et al. (2019) use GSV data of different time-steps to detect gentrification-like changes in built-environments. Hu et al. (2020) deploy multi-task learning which enables them to simultaneously predict three different target attributes to efficiently characterize street-canyon geometry. Gebru et al. (2017)

¹ <https://www.mapillary.com/>

² <https://maps.apple.com/imagecollection>

apply DCNNs on big street-level data. They classify cars detected within 50 Million GSV images taken across the U.S. into 2657 categories to map demographic attributes.

This study, in contrast, investigates the potential of DCNNs to infer key structural building characteristics for seismic risk assessment from street-level imagery. This is done to enable an efficient as well as flexible large-area *in-situ* data collection. Thereby, information derivation is rendered as a scene classification problem.

Supervised learning, particularly with DCNNs, relies on sufficient numbers of labeled training data to result in well generalizing models. However, consistent ground truth data at building-level with respect to our target variables is scarce and not available to us. Therefore, we compile reference data in an RRVS manner. As indicated in Section 1.3, the assignment of buildings with structural attributes based on their façade-views is not trivial. This is because the discriminative information with regard to structural properties, e.g., the material of the LLRS, in a façade view can be subtle. Thus, specific knowledge about the impact of construction practices on the visual appearance of buildings, i.e., about distinct dependencies between structural properties and visually inferable indicators (hereinafter referred to as *visual-structural criteria*) is required. To facilitate a reliable labeling we structured such information, and elaborated a proper scheme that aims at both: *i*) being representative in terms of occurring construction types and vulnerabilities; *ii*) to account for the needs of training and evaluating DCNNs.

Plenty of taxonomies have been developed to categorize building exposure in a standardized way (Pittore et al., 2018). In this study, we align the addressed typologies with the Global Earthquake Model (GEM) building taxonomy (Brzev et al., 2013; hereinafter referred to as GEM taxonomy).

Thereupon, we create a reference dataset based on which we assess the potential of DCNNs to predict a *risk-oriented* SBST. This allows to interlink buildings with a specific vulnerability model. In addition, we individually estimate the LLRS material and the height of buildings. We do so, to exemplify the automated collection of specific attributes of generic *faceted* data schemes (Pittore et al., 2018), e.g., the GEM or the GED4ALL taxonomy (its adaption for multi-hazard risk analysis; Silva et al., 2018).

Finally, we demonstrate the large-area characterization of buildings on a sample of >200 K façade images equally distributed over our test site, the earthquake-prone metropolis of Santiago de Chile.

The remainder of the paper is organized as follows. Section 2 provides an overview of the study site Santiago de Chile, its tectonic setting and building stock. Furthermore, the data used are described here. The deployed methods and the experimental setup are pointed out in Sections 3 and 4 respectively. Results are presented and discussed in Section 5. Section 6 concludes this paper.

2. Study site and data

2.1. Tectonic setting

This study focuses on the Chilean capital Santiago. Located straightly aligned with the Peru-Chile trench, the interface where the Nazca plate subducts the South American Plate (Fig. 1a), Chile is among the most seismically active countries worldwide. In the past 100 years, 11 great earthquakes with a moment magnitude (M_w) of 8 or above occurred³. With regard to the Santiago metropolitan area, seismic hazard is not only directly originating from the offshore subduction interface, but also from inland faults on the western flank of the Andean Cordillera (Ammirati, et al. 2019). Among them the San Ramon Fault, bordering the city to the east (SRF, Fig. 1b; Armijo et al. 2010).

2.2. Santiago de Chile and its building inventory

Our study site refers to the central 35 communes (*comunas*) of Greater Santiago. It covers an estimated built-up area (i.e., GUF area, Section 2.3.2) of 568 km², and is living environment to approximately 6.2 million people⁴. In particular after 1920, Santiago has developed very dynamically being subject to high rates of growth (Greene and Soler, 2005).

Regular earthquakes, the upcoming of new materials and seismic design as well as continually updated building codes have led to a variety of construction types in Chile, which are all present in Santiago. The building inventory can be differentiated considering the following predominant LLRS construction materials (identification shortcuts as specified in the GEM taxonomy are given in brackets and will be used in the following): *Unreinforced adobe masonry* (MUR + ADO), *unreinforced clay brick masonry* (MUR + CB), *reinforced masonry* (MR), *confined masonry* (MCF), *reinforced concrete* (CR), and *wooden* (W) (Santa María et al., 2017).

MUR + ADO buildings refer to traditional one to two floor residential houses featuring wall LLRSs (Fig. 2a) built with sun-dried clay brick and mud mortar stabilized by straw (Alvarez et al., 2016a). MUR + CB construction started in the second half of the nineteenth century and was built with burnt clay bricks in cement/sand mortar. Lacking a structural reinforcement these buildings exhibit high seismic vulnerabilities and were largely damaged in past earthquakes (Astroza et al., 2012a).

MCF construction started during the 1930s for low-rise single-family housing and was also used later for medium-rise apartment buildings (Astroza et al., 2012b; Santa María et al., 2017). MCF buildings feature a wall LLRS consisting of unreinforced masonry walls at regular intervals reinforced with vertically and horizontally reinforced concrete confining elements (i.e., tie-columns and tie-beams, Fig. 2d, e). Past earthquakes in Chile revealed a very good performance of MCF if constructed properly (Moroni et al., 2004).

MR construction emerged in Chile in the 1970s being used for low-rise single-family housing and apartment buildings. The wall elements are hollow concrete or clay block masonry, reinforced with vertical steel bars and horizontal bars in the bed joint (Fig. 2f). Before 1986 there were no seismic design codes for MR in Chile and its performance was poor (Moroni et al., 2014b).

CR construction in Chile started in the first half of the 20th century with few mainly non-residential buildings (Duarte, 2009), and has been widely used since the 1950s in buildings of up to 30 storeys or even more. The LLRS is mainly characterized by CR walls along the entire height of a building, but also CR moment resisting frames in combination with walls (Fig. 2a - c). CR construction exhibits the lowest seismic vulnerabilities (Moroni et al., 2002a, b; Alvarez et al., 2016d).

W houses are typically one or two storey single family buildings exhibiting wooden wall or frame constructions. If designed according to the Chilean building standards, W structures show low seismic vulnerabilities (Alvarez et al., 2016c).

Besides the mentioned construction designs, there are also non-engineered (self-constructed) buildings, where the LLRS cannot be clearly specified. Such constructions are specified hereinafter as *Unknown* (UNK).

For additional details see Section 3.2, Fig. 5.

2.3. Data

2.3.1. Street-level imagery

We use street-level imagery from GSV (Anguelov et al., 2010; Fig. 1c) for building characterization. The GSV system collects 360° street-level panoramas as an equidistant cylindrical projection of the environment

³ <http://www.csn.uchile.cl>

⁴ <https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda/poblacion-y-vivienda>



Fig. 1. Overview on the study site Santiago de Chile and used data: a) Tectonic setting and seismicity: the points reflect the epicenters and depths of seismic events with a $M_w \geq 3$ recorded by the Chilean National Network³ within the 15 years period 06/2004–05/2019 (i.e., > 25,000 events); topography is reflected by the SRTM digital elevation model; b) Built-up area as given by the GUF and administrative boundaries of comunas; the SRF is also indicated and underlines the direct spatial proximity to the settlement; c) Filtered GSV façade view positions used for SBST classification (triangles pointing in the direction of the camera's center heading) with examples of gathered façade imagery (R = left, L = right hand views) and remaining shooting positions not included in sample or filtered as "other" (Section 3.1).

with an omnidirectional camera. It is equipped with an internal GPS measurement unit and mounted on top of a car. A panorama is taken approximately every 10 m along the driving route and stored together with the camera's position including geographic coordinates and center compass heading in degrees as metadata. Static street-view imagery and metadata can be accessed via Google's corresponding Street View Static API. Thereby, several parameters on the properties of a requested image can be defined: *size* sets the output size in pixels; *heading* the horizontal

viewing direction on the panorama; the horizontal field of view (*fov*), and *pitch* specify the vertical viewing angle relative to the vehicle⁵.

2.3.2. The global urban footprint

To restrict street-level image acquisition and processing to built environments, we use the Global Urban Footprint (GUF; Fig. 1b), a globally available binary information layer which discriminates "built-up" and "non-built-up" areas. The GUF is the outcome of a fully automated image

⁵ <https://developers.google.com/maps/documentation/streetview>

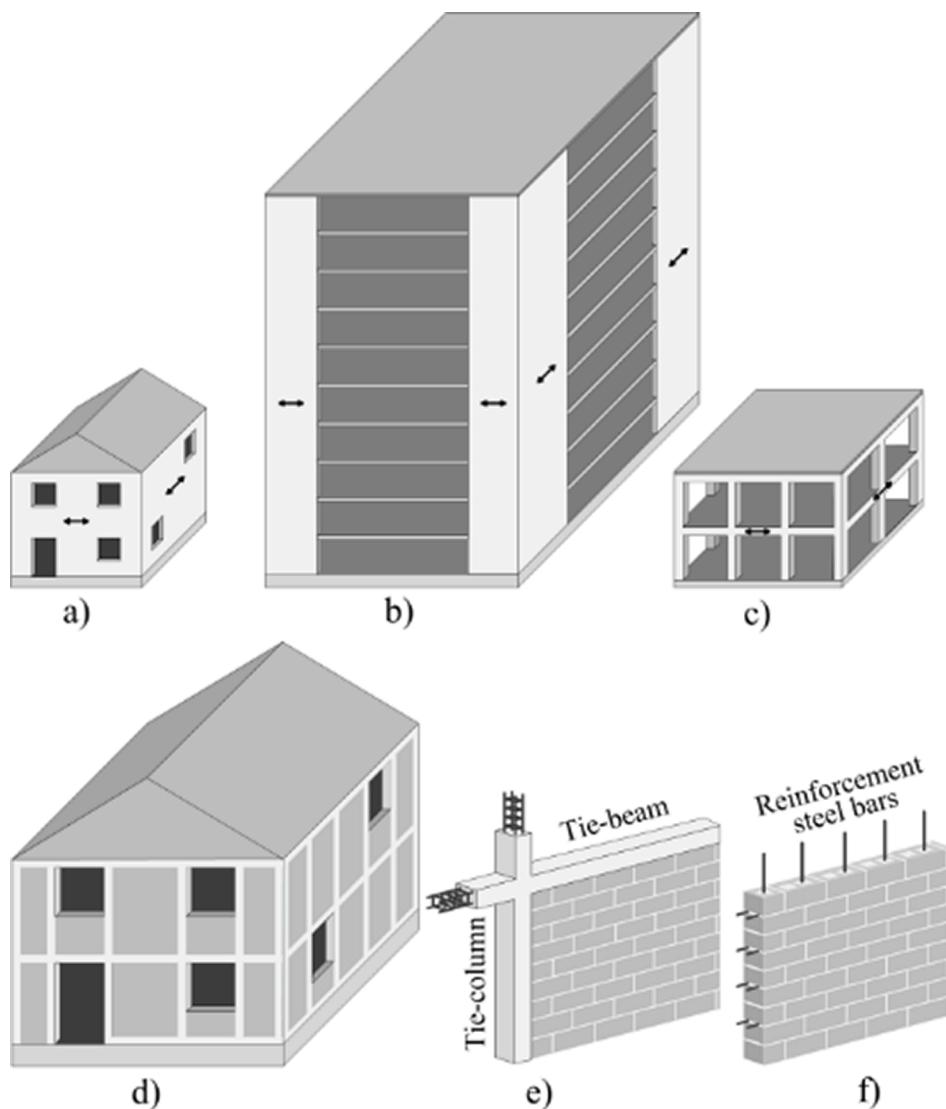


Fig. 2. Schematic illustrations exemplifying LLRSs (a – c; the arrows indicate the main lateral-load resisting structural elements) and masonry reinforcement techniques (d – f): a) single-party wall building, b) multi-party wall building, c) moment-frame building, d) MCF wall building, e) MCF details, f) RM details.

processing chain built upon high resolution synthetic aperture radar (SAR) imagery from the TanDEM-X mission and provides an overall absolute thematic accuracy beyond 85% (Esch et al., 2017; Klotz et al., 2016).

2.3.3. Institutional data

Two spatial polygon layers from Chile's 2017 census data, as provided by the National Statistics Institute of Chile (INE, 2018) are used to implement additional spatial constraints on data sampling. The first are the administrative boundaries of *comunas* (Fig. 1b), which are considered for a spatial stratification within the street-level imagery acquisition workflow (Section 3.1). The second are block level spatial entities which served for the spatial sampling of DCNN training and evaluation data (Section 4.1).

To conduct plausibility checks with regard to our façade image labeling outcomes (Section 3.2), we made use of property cadaster data from the Chilean Internal Revenue Service (Servicio de Impuestos Internos, SII). These include the address, an indication of the LLRS material, and the construction year of the listed taxed building properties (Aguirre et al., 2018). To account for tax exempt public social housing we utilized information provided by the Ministry of Housing and Urbanism (Ministerio de Vivienda y Urbanismo, MINVU); i.e., we

employ *i*) a georeferenced polygon-layer that delineates and locates social housing condominiums and includes corresponding years of building permits⁶; *ii*) a catalog containing photos and construction details on typical social housing buildings in Chile (MINVU, 2010).

3. Methods

In this section, the methodological background of the proposed workflow is presented according to its individual steps (Fig. 3). First, we describe the acquisition of street-level façade imagery. This is followed by detailed insights on the elaboration and implementation of our hierarchical image labeling procedure. We then give a short general introduction to DCNNs, after which we go into the specifics of transfer learning and the employed DCNN architectures.

3.1. Acquisition of street-level façade imagery

In a first step, we query the metadata of image panoramas available

⁶ <http://www.ide.cl/descarga/capas/item/catastro-nacional-de-condominios-sociales-2015.html>

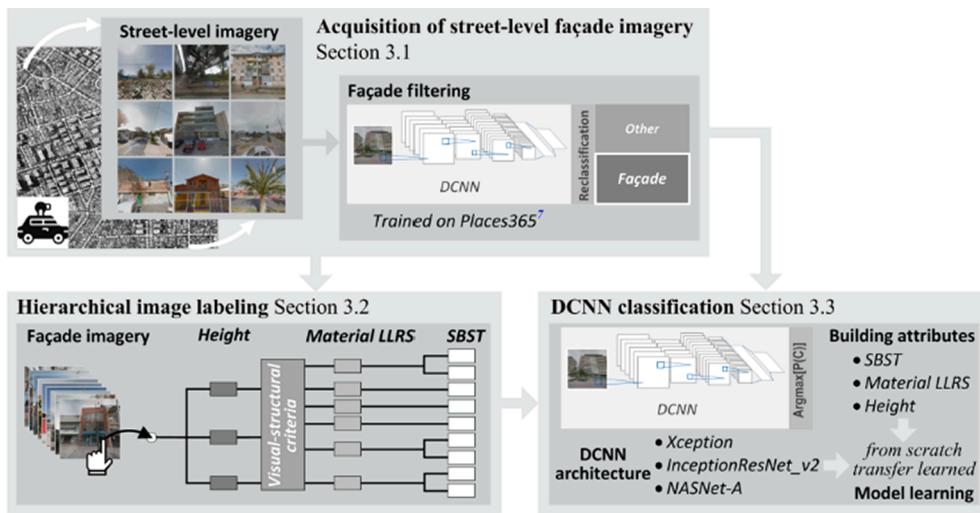


Fig. 3. Overview on the general workflow and the background methodology.

within our study area. These include their unique panorama ID (id_i), the camera's position (lat_i and lon_i), and the camera's center heading (h_c). Next, a representative sample of street-level imagery is drawn. Therefore, we use the GUF and the administrative boundaries of comunas (Fig. 1b) to *i*) constrain image acquisition to built-up areas and *ii*) perform a spatial stratification. Precisely, we use the spatial intersections of the GUF and the comunas as source entities for sampling panorama shooting points randomly for data acquisition. Thereby, each sample is weighted by the share of its source entity in the total GUF area within all considered comunas. Subsequently, street-level imagery are acquired with a size of 640×640 pixels which showed to be a good trade-off between data load and the visibility of structural details in the façade imagery. FOV and pitch are set to 100° and 15° respectively to allow for also capturing larger buildings. To obtain frontal views towards buildings we deploy h_c to define the *heading* and add and subtract 90° to obtain its perpendicular, i.e., we set the heading to $h_c + 90^\circ$ to get right hand side views and to $h_c - 90^\circ$ to get left hand side views.

Despite the constraints on image collection, there still remain non-façade views in the sampled imagery (Fig. 4). Analogous to Kang et al. (2018), we deploy a DCNN pretrained on the places2 data set (Zhou et al., 2018) to automatically filter the data for façade views. Specifically, we use a VGG16 DCNN (Simonyan and Zisserman, 2014) that was trained on the Places365-standard data⁷. The network was trained with more than 1.8 million photos from 356 scene classes. From these we use a set S of 24 outdoor classes that clearly indicate building façades: $S = \{\text{apartment building, beach house, building façade, chalet, church, cottage, courthouse, embassy, fire station, hangar, hospital, hotel, house, hunting lodge, mansion, manufactured home, motel, office building, palace, schoolhouse, shed, skyscraper, synagogue, tower}\}$. We feed the sampled street-view images to the DCNN and apply the subsequent filter rule:

$$L_i = \begin{cases} \text{"façade"} & \text{if } |C_i \cap S| \geq 2 \\ \text{"other"} & \text{otherwise} \end{cases}, \quad (2)$$

with L_i representing the resulting label assigned to an image i and C_i the 4 classes assigned as most probable. Exclusively the street-level imagery labeled as *façade* is kept.

3.2. Hierarchical labeling of reference data

We aim at creating a meaningful and representative reference data set. Therefore, the acquired façade imagery was subject to a properly tailored hierarchical labeling procedure. As a basis, a typologization

scheme (Fig. 5) was elaborated jointly by domain experts; Chilean structural engineers who are familiar with local construction practices and who have already conducted GSV-based digital remote building surveys in Chile (e.g., Santa María et al., 2017), and image analysts. Furthermore, relevant literature on building construction, the damages occurred in past earthquake events, and existing vulnerability models was considered. Requirements for the scheme were the provision of: *i*) a labeling workflow strictly based on visually inferable indicators (*visual-structural criteria*) to objectify the assignment of the required structural information; *ii*) a typology representative for Chile, which is aligned with the GEM taxonomy as well as existing vulnerability models (e.g., Villar-Vega et al., 2017); *iii*) final class frequencies that permit a meaningful learning and evaluation of DCNN models. Thereby, the employed ontology implies the combination of directly visible structural features (e.g., height or wall thickness) as well as indirect construction features (e.g., the architectural style period, indicating a particular construction period and associated building practices). As a complement to the scheme a catalog with example images was elaborated. It provided additional guidance and basis for discussion within the labeling process. Examples thereof are depicted in Fig. 6.

Considering the central building entity depicted on the images (i.e., the completely captured building with the largest area share, or if no building is completely captured, the building with the largest area share), labeling involves three steps. First, images are grouped according to the number of storeys (in the following referred to as *height*) and whether they represent *single*- (i.e. detached, semi-detached, or terraced single-family housing units) or *multi-party* buildings (i.e., apartment buildings and all buildings specifically constructed for non-residential use). This initial, rather rough categorization is carried out relatively quickly and intuitively without much attention to detail. Thereby, this inherently already leads to a useful pre-grouping with regard to the targeted SBSTs, since construction practice relates in particular to the height but also to the use of buildings (e.g., non-residential buildings usually require larger free spaces and have high code requirements to meet). The next step is to assign the material type of the LLRS to the pre-grouped images, according to the *visual-structural criteria*. This results in the *pre-SBSTs*, which represent the occurring combinations of the initial height and LLRS material type. According to the LLRS material types occurring in Santiago (Section 2.2), we differentiate among *MUR + CB*, *MUR + ADO*, *MCF*, *RM*, *W*, *CR*, as well *UNK* structures. Furthermore, we consider two additional categories, *Commercial* and *Industrial* (*COM1_2_IND*) and *Office buildings* (*COM3*). These refer to façade images which cannot be unambiguously assigned to the aforementioned categories. This rarely applies to residential buildings but is the case for non-

⁷ <https://github.com/CSAILVision/places365>

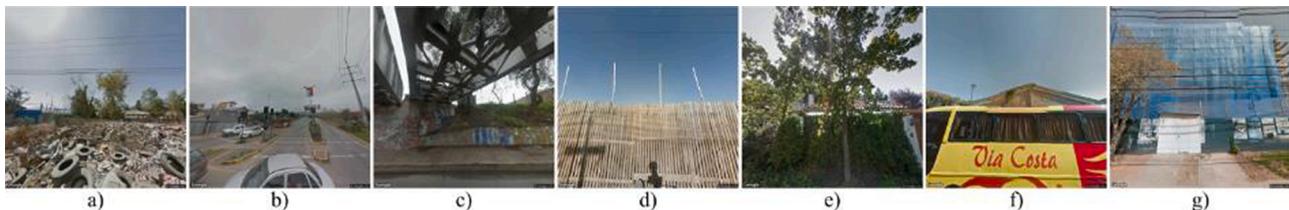


Fig. 4. Examples for excluded non-façade images: a) free spaces; b) crossings and streets; c) non-building structures; d) obscuring fences and walls; e) vegetation; f) vehicles; g) wrapped construction sites.

residential buildings that feature a special structural design adapted to their use or an individual representative design. Hence, these two categories include non-residential buildings, whose vulnerability characteristics are either assumed to be highly individual so that they cannot be described by available generalized vulnerability models, or whose occurrence is too low to be adequately represented in an appropriate additional LLRS material category, considering the learning of DCNN models. Buildings with storeys of different LLRS material types are assigned the LLRS material type of the ground floor. This is typical for two storey single-family buildings, which often have a masonry ground floor (e.g., MCF or MR) and a wooden first floor (e.g., the right of the 04: MR/HBET1,2 examples in Fig. 6). The third and last step consists of a final re-aggregation based on the pre-SBST frequencies to obtain the SBSTS. The objective here is to distinguish SBSTS that are meaningful and consistent with respect to their seismic vulnerabilities, but at the same time do not fall below a certain minimum number of images for the training and evaluation of DCNNs. In this course, MUR + CB and MUR + ADO were merged to MUR as well as W and UNK to W_UNK. Furthermore, the height divisions of the SBSTS were modified.

Following the established scheme, labeling was carried out by undergraduate students guided and counter checked by the expert group. The final aggregation to the 14 SBSTS was performed by an expert. To ensure representativity with respect to the individual target classes and considering occurring intra-class variabilities, labeling was carried out until the least populated class exceeded 800 instances.

To obtain the final reference data, the resulting set of labeled data was subject to additional plausibility checks based on institutional data (Section 2.3.3). The first type of checks was performed using the SII data based on the contained information on construction materials and construction years of taxed buildings; e.g., a MUR building needs to be of material type masonry and have a construction year older than 1940 (see the visual-structural criteria in Fig. 5). However, since we noticed a substantial amount of uncertainties in the attribution as well as the geocoding of the SII data, we did not use the data for a direct re-labeling but rather to select candidates to be revised. In case of ambiguity we retained the manually assigned labels. The second type of plausibility checks was performed on the façade imagery of non-taxed social housing buildings. Here, we used the polygon-layer provided by MINVU. With it, we specifically revised labeled façade images depicting buildings located in a social housing condominium, taking respective years of building permits as well as the construction design of typical social housing buildings in Chile (MINVU, 2010) into account.

3.3. Façade image classification with Deep Convolutional Neural Networks

3.3.1. Deep convolutional neural networks

A DCNN is a trainable architecture (Fig. 7) of multiple hierarchically arranged feature extraction stages each consisting at least of three consecutive types of alternating layers: i) convolutional layers, ii) a non-linearity (or activation) layer and iii) a pooling layer, which are followed by a classification stage containing at least one fully connected layer (LeCun et al., 2015). Within the training stage, this configuration enables the DCNN to learn two dimensional (D) non-linear representations

of the input (Bengio et al., 2013) referred to as low- and high-level features and a classifier based upon them.

The input data of a convolutional layer is a 3D array $X \in \mathbb{R}^{H \times W \times C}$, the first two dimensions H (height) and W (width) reflecting its spatial dimensions and the third dimension C the number of 2D ($H \times W$) feature channels. Each convolutional layer is composed of k learnable kernels which convolve its input to k output feature maps. Specifically, let x_i be the i^{th} feature map of an input X to a convolutional layer, the j^{th} of its k kernels can be defined by its weight w_j and bias b_j and its output y_j denoted as follows:

$$y_j = \sum_i^C f(x_i * w_j + b_j), j = 1, 2, \dots, k \quad (3)$$

where $*$ is a two-dimensional discrete convolution operator and $f(\cdot)$ refers to a point wise nonlinearity function, with the rectified linear unit (ReLU): $f(x) = \max(0, x)$ recently being most common (Li et al., 2019). A feature extraction stage is completed by a spatial pooling layer, which aims to diminish redundant information and to create invariance (LeCun et al., 2015). After the feature extraction stages, in the classification stage, the last multidimensional output is transferred into a one-dimensional feature vector (*flattening*) and fed to the fully connected layer(s) where each output dimension depends on all the input dimensions. The output of the (last) fully connected layer of the DCNN is transferred into probability distributions over the target class labels by the *softmax* function. The parameters of a DCNN, i.e., the weights in the convolutional and fully connected layers are trained based on error backpropagation deploying a gradient decent optimization algorithm that minimizes an objective function (e.g., cross entropy loss; Goodfellow et al., 2016).

3.3.2. Transfer learning

For establishing our DCNNs by training from *scratch*, we have a limited amount of reference data (Section 5.1). Therefore, *transfer learning* (TL) is applied to improve classification performance. Specifically, we cast our classification problem as an instance of *inductive* TL (Pan and Yang, 2010), where knowledge on a source domain D_S and a learning task T_S with plenty of available labeled data is used to improve the learning of the predictive objective function $f_T(\cdot)$ for our target learning task T_T ($T_S \neq T_T$) in our target domain D_T .

Considering, $A^i(\cdot)$ as one of the deployed DCNN architectures (Section 3.3.3), we perform TL by transferring the parameters (i.e., the weights) of all feature extraction stages obtained in training $A_S^i(\cdot)$ for T_S in D_S (referred to as *pre-training*; Goodfellow et al., 2016) to initialize the training of $A_T^i(\cdot)$ for T_T in D_T (referred to as *fine-tuning*; Yosinski et al., 2014).

3.3.3. DCNN architectures

Three predefined DCNN architectures are tested and compared considering their generalization capabilities, i.e., *InceptionResNetV2* (Szegedy et al., 2016b), *Xception* (Chollet, 2017), as well as *NASNet-A* (Zoph et al., 2018). We selected these architectures as they feature different complexities in terms of trainable parameters (Table 1) and recently achieved state-of-the art results on the ILSVRC2012 ImageNet

Height; multi-/ single-party	Material LLRS	Visual-structural criteria	Pre-SBST	SBST	Description GEM taxonomy string (short) ¹⁹
1 - 3 storeys; single-party	Unreinforced masonry, clay brick	≤ 4 storeys, colonial or neoclassical design (pre 1940, mostly 1920s); demarcation of masonry structure with boundaries where cracks could occur; thick walls: main load-bearing wall (MLBW) thickness ≥ 30 cm; wooden or reinforced concrete lintels above openings; sometimes reinforced concrete tie-beams on top of the walls ^{1,2,3}	MUR+CB/ HBET:1,3	01: MUR/HEX:1	Unreinforced masonry, clay brick or adobe; wall LLRS; 1 storey MUR+CB99/LWAL/HEX:1
	Unreinforced masonry, adobe	≤ 2 storeys; rustic self-constructed or colonial design (mostly pre 1940); demarcation of adobe structure with boundaries where cracks could occur; few small openings with wood lintels above; wooden roof system, often pitched with gable ends; shallow unreinforced foundations; thick walls: MLBW thickness ≥ 30cm ^{2,3,4}	MUR+ADO/ HBET:1,2	02: MUR/ HBET:2,3	Unreinforced masonry, clay brick or adobe; wall LLRS; 2 - 3 storeys MUR+CB99/LWAL/HBET:2,3
1 - 2 storeys; multi-party	Confined masonry	≤ 5 storeys; post 1930s design; demarcation of masonry structure with boundaries where cracks could occur; reinforced concrete cast-in-place tie-beams and tie-columns at regular intervals visible or indicated by specific vertical/horizontal boundaries; shallow reinforced concrete strip foundation; typical MLBW thickness 14 - 20 cm; single-family: houses often aligned in a row or semi-detached comprising pitched roofs with timber gable ends; apartments: often several building units of equal design, located close together in medium to low-income areas or social housing developments ^{2,5,6,7,8,9}	MCF/HBET:1,2	03: MCF/H:1,2	Confined masonry; wall LLRS; 1 - 2 storeys MCF/LWAL/HBET:1,2
	Reinforced masonry	≤ 5 storeys; post 1970 design; demarcation of masonry structure with boundaries where cracks could occur; shallow reinforced concrete strip foundation; MLBW thickness 14 - 20 cm; single-family: houses often aligned in a row or semi-detached comprising pitched roofs with timber gable ends; apartments: often several building units of equal design, located close together in medium to low-income areas or social housing developments ^{5,6,9,10}	MCF/HBET:3,4	04: MR/HBET:1,2	Reinforced masonry; wall LLRS; 1 - 2 storeys MR/LWAL/HBET:1,2
3 - 7 storeys; multi-party	Wooden	≤ 2 storeys; wooden wall/frame LLRS, thin walls; wooden window and door frames; wooden roof system with light covering ¹¹	MR/HBET:1,2	05: CR/HBET:1,2	Reinforced concrete; wall LLRS; 1 - 2 storeys CR/LWAL/HBET:1,2
	Unknown	Non-engineered structures mainly built with light wooden or unknown elements ¹²	MR/HBET:3,4	06: W_UNK/ HBET:1,2	Wooden; wall and/or frame LLRS, unknown or nonengineered; 1 - 2 storeys W+WLI/LWAL/HBET:1,2
≥ 8 storeys; multi-party	Reinforced concrete	1 - 30+ storeys; mainly recent to modern (post 1950 design); predominantly large reinforced concrete wall but also frame or dual frame-wall LLRS elements along entire height; medium to large opening spans; smooth continuous surfaces; typical Chilean „fish-bone“ plan medium to high rise residential buildings; vast majority of Santiago's buildings ≥ 5 storeys; some large and massive mainly non-residential neoclassical architecture buildings constructed in the first half of the 20th century ^{13,14,15,16,17,18}	W/HBET:1,2	07: MCF/ HBET:3,4	Confined masonry; wall LLRS; 3 - 4 storeys MCF/LWAL/HBET:3,4
	Commercial and industrial	Special warehouses, shopping malls and industrial buildings not unambiguously assignable	UNK/ HBET:1,2	08: MR/HBET:3,4	Reinforced masonry; wall LLRS; 3 - 4 storeys MR/LWAL/HBET:3,4
	Office buildings	Special office buildings; mostly glass cladded façades; not unambiguously assignable	CR/HBET:1,2	09: CR/HBET:3,4	Reinforced concrete; mainly wall LLRS; 3 - 4 storeys CR/LWAL/HBET:3,4
			CR/HBET:3,7	10: CR/HBET:5,7	Reinforced concrete; mainly wall LLRS; 5 - 7 storeys CR/LWAL/HBET:5,7
			CR/H:8+	11: CR/HBET:8,12	Reinforced concrete; mainly wall LLRS; 8 - 12 storeys CR/LWAL/HBET:8,12
			COM1_2_IND/ H:1+	12: CR/H:13+	Reinforced concrete; mainly wall LLRS; ≥ 13 storeys CR/LWAL/H:13+
			COM1_2_IND/ H:1+	13: COM1_2_IND/ H:1+	Warehouse, shoppingmalls, industrial not unambiguously assignable to SBST 1-12; ≥ 1 storey COM+COM1; COM+COM2; IND
			COM3/H:8+	14: COM3/H:8+	Office buildings; not unambiguously assignable to SBST 1 - 12; ≥ 8 storeys COM+COM3

Fig. 5. Scheme highlighting the hierachic image annotation procedure and addressed SBST taxonomy. The indication of (pre-) SBSTs follows the coding of the GEM taxonomy. With regard to the specification of heights, *HEX* refers to exact height, *HBET* to height between and *H:N+* to height of N storeys or higher. The GEM taxonomy strings in the description refer to the most frequently occurring instance of an SBST. Literature: ¹Monge, 1969; ²Astroza et al., 2012a; ³Jorquera et al., 2017; ⁴Alvarez et al., 2016a; ⁵Moroni et al., 2004; ⁶Astroza et al., 2012b; ⁷Moroni et al., 2014a; ⁸Alvarez et al., 2016b; ⁹MINVU, 2010; ¹⁰Moroni et al., 2014b; ¹¹Alvarez et al., 2016c; ¹²Villar-Vega et al., 2017; ¹³Alvarez et al., 2016d; ¹⁴Jüinemann et al., 2015; ¹⁵Santa María et al., 2017; ¹⁶Moroni et al., 2002a; ¹⁷Moroni et al., 2002b; ¹⁸Duarte et al., 2009; ¹⁹Brzev et al., 2013

(hereafter simply referred to as ImageNet) classification task. A 1.4 M natural images dataset labeled across 1000 classes (Russakovsky et al., 2015). In contrast to the first two, which were manually designed, NASNet-A is the result of an automated architecture engineering procedure, which is intended to adapt the model architecture in a beneficial way to the specific data.

The *InceptionResNet-v2* integrates the successively modified *Inception module* (Szegedy et al., 2016a) and *residual learning* (He et al., 2016). The central idea of the *Inception module* is to incorporate and capsule

variable receptive fields created by different kernel sizes to enhance the sparse capture of spatial and cross-channel information at different scales. Thereby, initial 1x1 convolutional filters perform a dimensionality reduction to regulate computational expenses. The latest Inception module was further enhanced by implementing factorized convolutions and kernel size reductions enabling an efficient up-scale of the DCNN to improve accuracy (Szegedy et al., 2016a). Residual learning addresses the vanishing gradient and degradation problem with short cut layer connections performing additive identity mapping (He et al., 2016). The

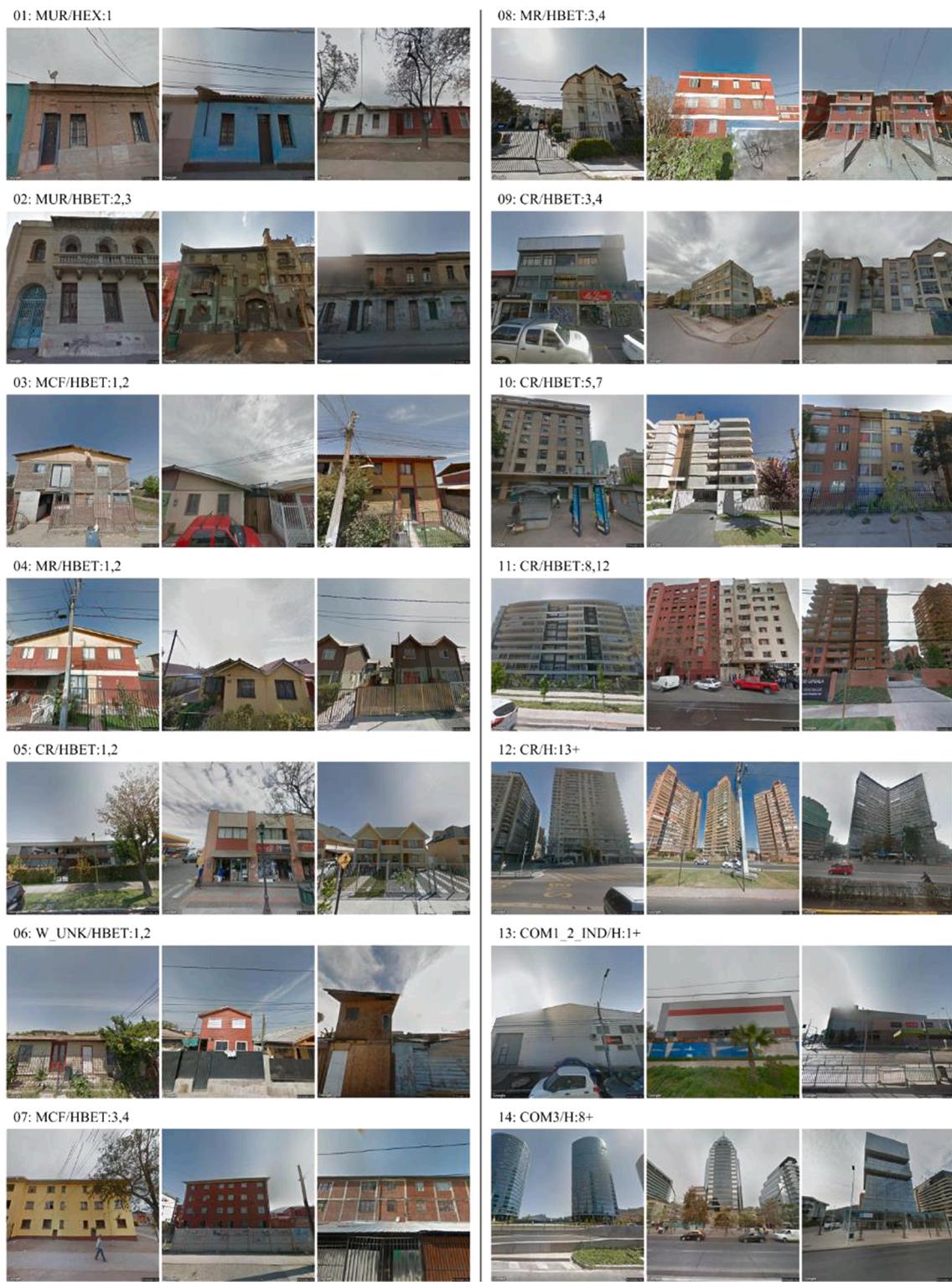


Fig. 6. Examples for GSV façade images annotated with one of 14 SBSTs according to the categorization scheme.

InceptionResNet-v2, is built-upon the revised inception modules equipped with short cut connections, leading to a faster training while showing a gain in recognition performance (Szegedy et al., 2016b).

The Xception architecture takes up the convolution factorization concept of the revised *inception* module, but restructures it to an instance of a *depth-wise separable convolutional layer*, i.e., an initial 1×1 *point wise convolution* to map cross-channel correlations followed by a convolutional layer performing a spatial convolution over each channel of an

input independently (*depth-wise convolution*). The whole Xception network is a linear stack of such depth-wise separable convolutional layers consecutively interlinked with linear residual short cut connections. Entirely decoupling spatial and feature channel correlation, this architecture showed to be particularly efficient in the use of parameters (Chollet, 2017).

While InceptionResNetV2 and Xception result from manual DCNN design, *NASNet-A* constitutes the outcome of an automated DCNN

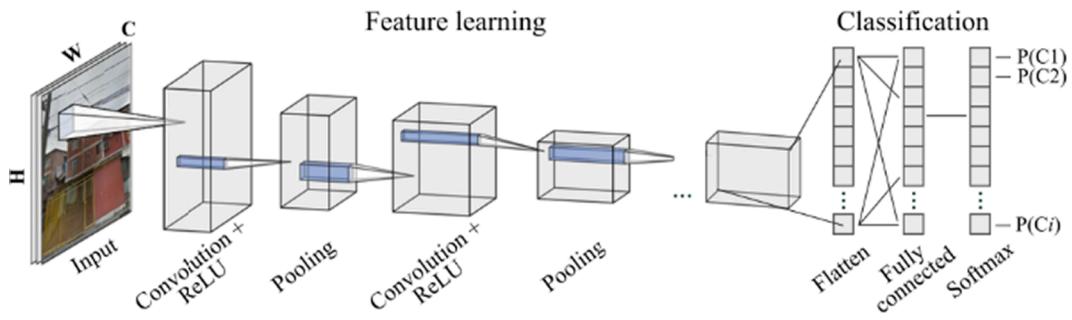


Fig. 7. Main parts of a DCNN architecture.

Table 1

Overview on deployed DCNN-Architectures. Order arranged according to the number of trainable parameters.

Architecture	Input image size [Pixels]	# Parameters [M]	ILSCVR2012 Top-1 acc. [%]
Xception	299 × 299	22.8	79.0
InceptionResNet-v2	299 × 299	55.8	80.1
NasNet-A	331 × 331	88.9	82.7

architecture engineering procedure (i.e., NAS). NAS renders the DCNN architecture search as a reinforcement learning problem, the configuration of a DCNN being the agent's action. A controller *recurrent neural network* (RNN) samples child networks with different architectures which are trained to convergence. Resulting recognition accuracies are used to consecutively update the controller and optimize architectures. Inspired by hand-crafted architectures composed of repeated modules, Zoph et al. (2018) search for the architecture of such motifs, rather than for whole DCNN architectures. Thereby, the search space is restricted to two different motifs, i.e., *Normal Cells* (convolutional cells preserving dimensionality) and *Reduction Cells* (convolutional cells reducing dimensionality). The NasNet architectures are subsequently constructed by sequentially concatenating both modules. Within the search, the controller RNN selects from a predefined set of hidden state operations (i.e., convolution, depth wise separable convolution, dilated convolution, average pooling and max pooling with several kernel sizes as well as identity mapping) to construct the convolutional cells. This modularity reduces the search space keeping computational costs in reasonable scales and facilitates scalability and transferability.

In this manner, Zoph et al. (2018) automatically optimized *Normal* and *Reduction Cells* to construct the NasNet-A architecture (Fig. 8).

4. Experimental setup

To assess the ability of DCNNs (Section 3.3.3) to predict a building's SBST from street-level-façade imagery, but also to extract individual structural attributes independently, the following experiments were carried out: *i*) training and evaluation of DCNNs to predict the 14 finally assigned SBSTS (Fig. 5); *ii*) training and evaluation of DCNNs to predict the 7 distinguished LLRS material classes (i.e., MUR, MCF, MR, CR, W_UNK, COM1_2_IND and COM3); *iii*) training and evaluation of DCNNs to predict 6 different building height classes (i.e., HEX:1, HEX:2, HBET:3,4, HBET:5,7, HBET:8,12 and H:13+). Thereby, prediction performances of the DCNNs were assessed when trained from scratch as well as when TL was applied. To take account for the class imbalance and to

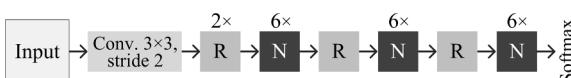


Fig. 8. Scheme of NasNet-A. R and N indicate Reduction and Normal cells.

ensure comparability across the experiments, we set up an appropriate sampling and balancing strategy.

4.1. Data balancing and partition

Considering the significant class imbalance in the reference data pool (Fig. 9), we carried out a balancing strategy that combines undersampling of the majority classes and data augmentation. With regard to classes with a limited amount of samples available ($n \leq 2,000$), we deployed all samples as *labeled in* data for learning and evaluating the DCNNs (Table 1). For larger populated classes ($n > 2,000$), a random sample of 2000 images was drawn (i.e., undersampling). The *labeled in* data was split into training, validation, and test data sets with respective shares of 65%, 15% and 20%. It was ensured that the training, validation, and test data set exclusively contain images representing buildings from different building blocks and thus are spatially disjoint. In this manner a potential positive bias in accuracy measures that can be attributed to overlapping image fields of views or images spotting at the same building from different perspectives was eliminated (Geiß et al., 2017b). Building blocks were represented by block level spatial entities from Chile's 2017 census (Section 2.3.3). In a final step, training data for each class was augmented to a balanced data set of 4000 images respectively by applying random combinations of the geometric image transformation operations *shearing*, *perspective skewing*, *cropping*, and *vertical flipping*. The data for all experiments, the SBST as well as the LLRS material and the height classification implies the described procedure. Thereby, to provide comparability of SBST prediction accuracies to the accuracies of the independent derivation of the LLRS material and height, training, validation, and test data sets for the determination of the latter two were created from the original training, validation, and test data set of the SBST experiments. An overview on the data setting for each experiment is given in Fig. 10.

4.2. Training of DCNNs

In order to prevent overfitting, we apply additional dropout regularization (Srivastava et al., 2014) with a dropout rate of 0.5 right at the fully connected softmax prediction layer of each of the deployed DCNN architectures. Considering the computational resources of the Nvidia GeForce RTX 2080 Ti GPU with 11 GB memory as well as the relatively large memory requirements with regard to the training of NasNet-A, we train all the networks uniformly with a relatively small mini-batch size of 10. The Adaptive Moment Estimation optimizer (ADAM; Kingma and Ba, 2014) is used to optimize categorical cross-entropy.

When conducting the training of the DCNNs from scratch trainable parameters were initialized randomly. In the TL experiments, training of all feature extraction stages started from the parameter set up resulting from a respective pre-training (Section 3.3.2) on the ImageNet data.

Different initial learning rates are set when the models are trained from scratch and when TL is applied. In the first case training is started with a larger learning rate of 0.001 in order to encourage a faster learning at the beginning. In the latter small initial learning rates of

0.0001 are set, allowing for a sensitive fine-tuning right from the start. To facilitate an efficient and exhaustive learning process, learning rates are reduced by a factor of 0.5 at validation accuracy plateaus, i.e., after 5 consecutive epochs without an increase. Early stopping is used in order to mitigate overfitting as well as an unnecessary extension of training time. The model status yielding the highest validation accuracy is kept. Apart from the initial learning rate set ups for the training from scratch and TL, training parametrization among all experiments remained unchanged to enable reasonable comparisons of the resulting models.

5. Results and discussion

5.1. Acquisition and labeling of street-level façade imagery

Following the workflow described in Section 3.1, 204,030 façade images were acquired, of which 29,597 were labeled according to the labeling procedure described in Section 3.2. Fig. 9 shows the numbers of available reference data categorized according to their assigned height, material LLRS, and SBST labels.

Particularly for the SBST and the material LLRS class frequencies exhibit substantial class imbalance. While the most populated SBST (i.e., 03:MCF/HBET:1,2) and LLRS (i.e., CR) classes respectively comprise 4361 and 13,236 assigned façade images, their least populated counterparts contain only 825 (08:MR/HBET:4,3) and 1017 (COM3) labeled samples. These imbalanced class-distributions are reasonable considering Santiago's history of construction along with its seismic events and spatio-temporal dynamics (Section 2.2).

5.2. Data sampling and balancing

To mitigate class imbalance as well as to enrich our training data, we conducted a proper sampling and data augmentation strategy (Section 4.1). Resulting frequency distributions across training, test, and validation data, as well as the amount of data augmentation for the addressed classification tasks are depicted in Fig. 10. An overview on aggregated overall numbers in terms of the partition of labeled data and data augmentation is given in Table 2.

5.3. Comparative accuracy and training times across resulting DCNNs

The test set classification accuracies of the DCNN models achieved through the training of the deployed architectures with the different training approaches are shown in Fig. 11. Respective training times in terms of GPU hours can be taken from Fig. 12.

Independently of the classification task and DCNN architecture, the levels of accuracy are fairly high with the Xception DCNN trained from scratch on SBST classification achieving the lowest Kappa value (κ), overall accuracy (OA), and weighted mean of F_1 scores (\bar{F}_1) with 0.757, 77.58%, and 0.776, respectively. The highest SBST classification accu-

racies attain a $\kappa = 0.813$, an OA = 82.7%, and a $\bar{F}_1 = 0.827$, and result from the transfer-learned NASNet-A. Overall, the highest accuracy levels are achieved for the LLRS material classification.

For all tasks highest accuracy levels are obtained when fine-tuning the NASNet-A architecture. In most of the experiments, TL also leads to a significant reduction in training time. Comparing deployed DCNN-architectures in terms of the training time required until convergence, it can be observed that, put into a sequence they are consistent with their number of parameters (Section 3.3.3). Xception converges fastest, followed by InceptionResnet-v2 and NASNet-A with markedly longer training times.

The attained accuracy levels demonstrate the automation potential in the retrieval of building structural attributes for seismic risk assessment on the basis of the presented data and methodology.

5.3.1. SBST classification

With regard to the classification of SBSTs, model training clearly improved from TL, which leads to a significant raise in accuracies across all three DCNN architectures but most evident for NASNet-A. Being nearly on par with InceptionResNet-v2 when trained from scratch, NASNet-A results in the highest accuracies on SBST classification ($\kappa = 0.813$, OA = 82.7%, and $\bar{F}_1 = 0.827$) injecting additional prior-knowledge via pretraining.

5.3.2. Material LLRS classification

The accuracies in the predictions of the LLRS material show a similar picture on a generally higher level, which reflects the reduced complexity of the classification task. Again, the learning of all applied DCNN architectures is clearly improved by TL. Whereas here, although the transfer-learned NASNet-A achieves the highest accuracies ($\kappa = 0.848$, OA = 87.1%, and $\bar{F}_1 = 0.870$), this time Xception profits the most and obtains a similar level ($\kappa = 0.845$, OA = 86.9%, and $\bar{F}_1 = 0.867$) with far less training effort (Fig. 12b). As with the prediction of SBSTs, InceptionResNet-v2 outperforms the other DCNNs when training the DCNNs from scratch.

5.3.3. Height classification

For the height predictions, only the NASNet-A benefits from TL and, as with the other tasks, shows the highest accuracy ($\kappa = 0.831$, OA = 85.9%, and $\bar{F}_1 = 0.858$). However, the Xception and InceptionResnet-v2 models trained from scratch slightly outperform their pretrained counterparts. The differences in accuracy between the models learned from scratch and the pretrained models are generally the lowest here. This could be attributed to the fact that the height classification task exhibits the most favorable ratio between the number of available training samples per class and the number of addressed classes, but also indicate that the features learned within the ImageNet-pretraining are less relevant for height classification.

We also assessed the accuracy of the results when combining the

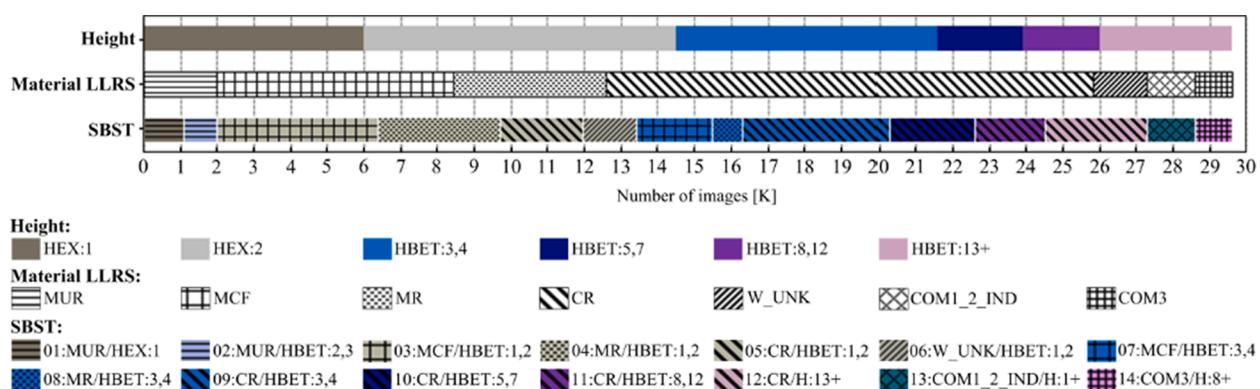


Fig. 9. Number and class distributions of labeled reference data pool.

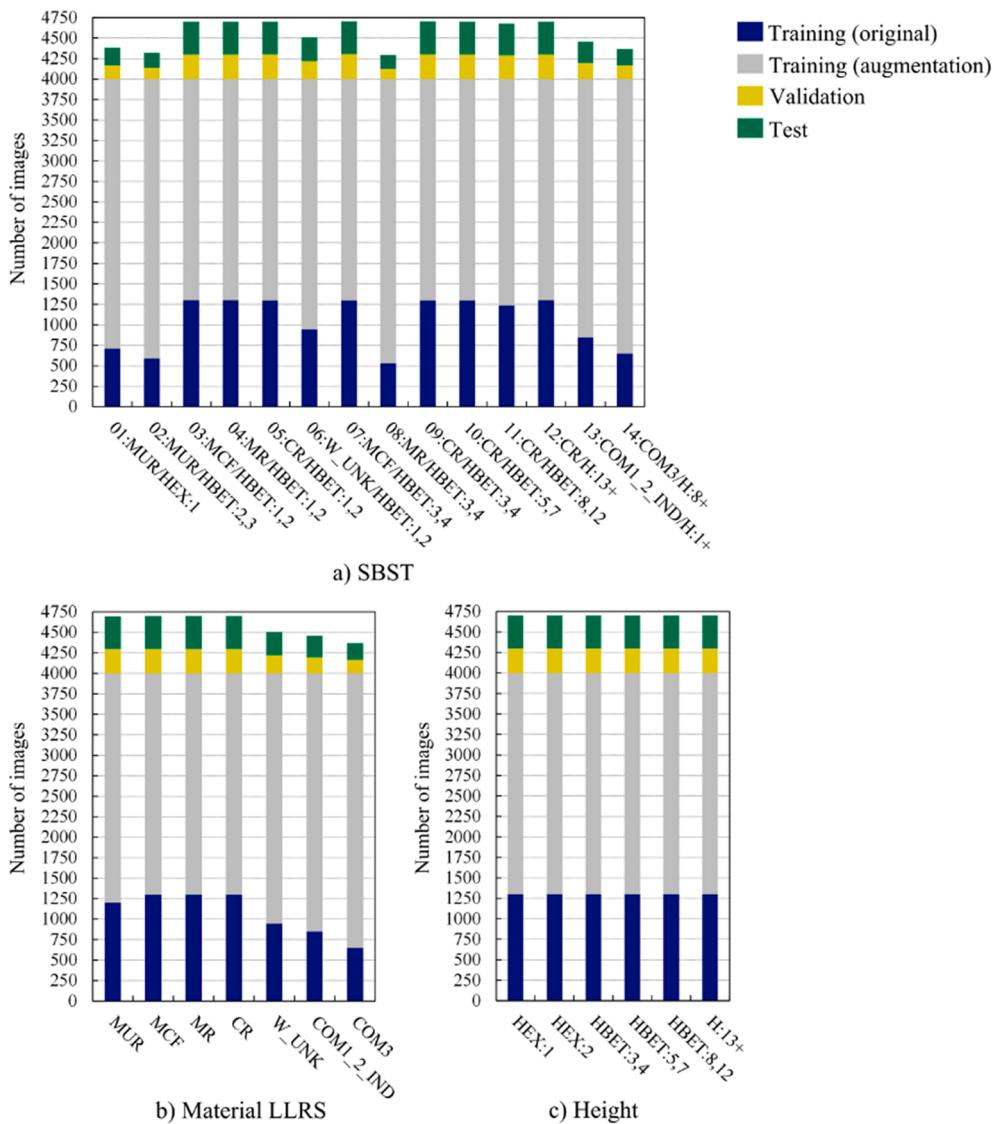


Fig. 10. Frequency distributions across training, test, and validation data as well as the amount of training data augmentation for the three classification tasks.

Table 2
Overview on labeled reference data used for training and testing of DCNNs.

Experiment	Labeled in	Training	Training augmented	Validation	Test
SBST-DCNN	22,515	14,599	56,000	3397	4519
Height-DCNN	12,000	7,800	24,000	1800	2400
Material LLRS-DCNN	16,000	10,400	32,000	2400	3200

predictions of the specifically trained LLRS material and height models to obtain the SBST and compared them to the direct SBST predictions. Thereby, the direct predictions of the SBSTS proved to be more accurate. The SBST accuracies, obtained combining the predictions of the best performing LLRS material and height DCNN (i.e., the NASNet-A respectively) were $\kappa = 0.785$, OA = 79.2%, and $\bar{F}_1 = 0.790$.

5.4. In detail classification accuracy of best DCNNs

The confusion matrices for the best performing models on the three addressed classification tasks (i.e., the NASNets) are shown in Fig. 13. The corresponding precision, recall and F_1 score values are illustrated in Fig. 14. Considering the SBST predictions, Fig. 13a reveals that

commission and omission errors generally occur among classes representing the same height but with a different LLRS material, or among classes of the same material LLRS but with a different height (among *adjoining classes* which are one height class above or below the considered one in particular). These difficulties in inference are confirmed by the confusion matrices of the LLRS material Fig. 13b and the height DCNN predictions Fig. 13c). While the errors in LLRS material prediction occur across several classes, in height prediction errors occur only among *adjoining classes*.

Three individual classes from the SBST predictions, 03:MCF/ HBET:1,2, 09:CR/HBET:3,4, and 11:CR/HBET:8,12 (F_1 scores of 0.778, 0.778, and 0.726, respectively) are particularly error prone. For 09:CR/ HBET:3,4 and 11:CR/HBET:8,12 this is due to the above-mentioned errors in regards to classes with *adjoining* heights, which are particularly pronounced here. In contrast, the errors in respect to 03:MCF/ HBET:1,2 can be clearly attributed to commission and omission errors across SBSTs of the same height but of different LLRS material. The difficulties in correctly predicting the LLRS material type MCF as well as the height class HBET:8,12 is also visible in the confusion matrices, precision, recall, and F_1 score values of the material LLRS and the height DCNNs, respectively. MCF construction (Section 2.2 and 3.2) has been built for a relatively long time period and, due to its good performance in earthquakes, is widespread, both in areas with a high and low

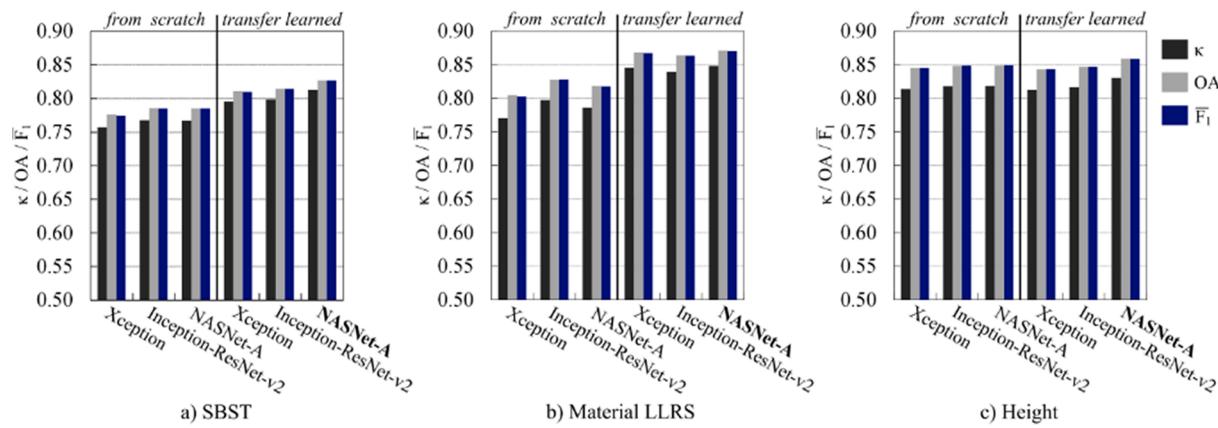
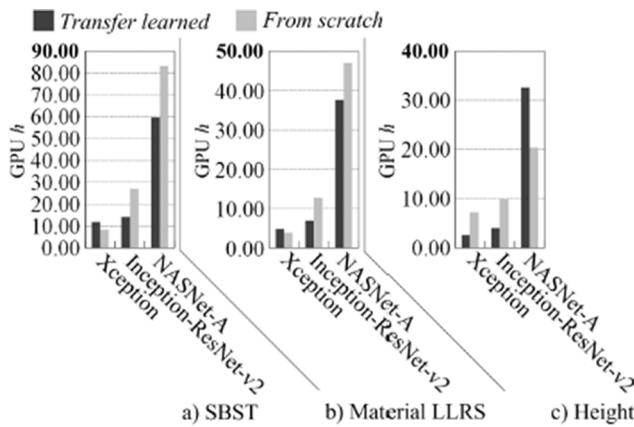
Fig. 11. κ , OA and \bar{F}_1 for the different classification approaches.

Fig. 12. Time (GPU hours) for training the DCNNs on a) SBST, b) material LLRS, and c) height classification.

socioeconomic level. Correspondingly, MCF comprises a large within-class variance. Especially when buildings are plastered or painted, the visual characteristics of MCF as well as the differences to other LLRS materials can be subtle and therefore difficult to recognize even when labeled manually. Older MCF buildings can bear a strong similarity to MUR buildings and newer MCF buildings to MR and CR promoting the susceptibility to prediction errors. Nevertheless, the accuracies for these comparatively error-prone classes are still high.

The deployed class balancing strategy (Section 4.1) is successful, since the classification results on the individual tasks show no deficiencies on the accuracies of the classes with less available reference data in comparison to the classes with a higher population of labeled images. These are particularly 01:MUR/HEX:1, 02:MUR/HBET:2,3, 06: W_UNK/HBET:1,2, 08:MR/HBET:3,4, 13:COM1_2_IND/H:1+, as well as 14:COM3/H:8+ considering the SBST class frequencies and W_UNK, COM1_2_IND as well as COM3 considering the LLRS material class frequencies.

It is worth stating that machine learning systems are subject to uncertainty (Hüllermeier and Waegeman, 2021). A contributing factor

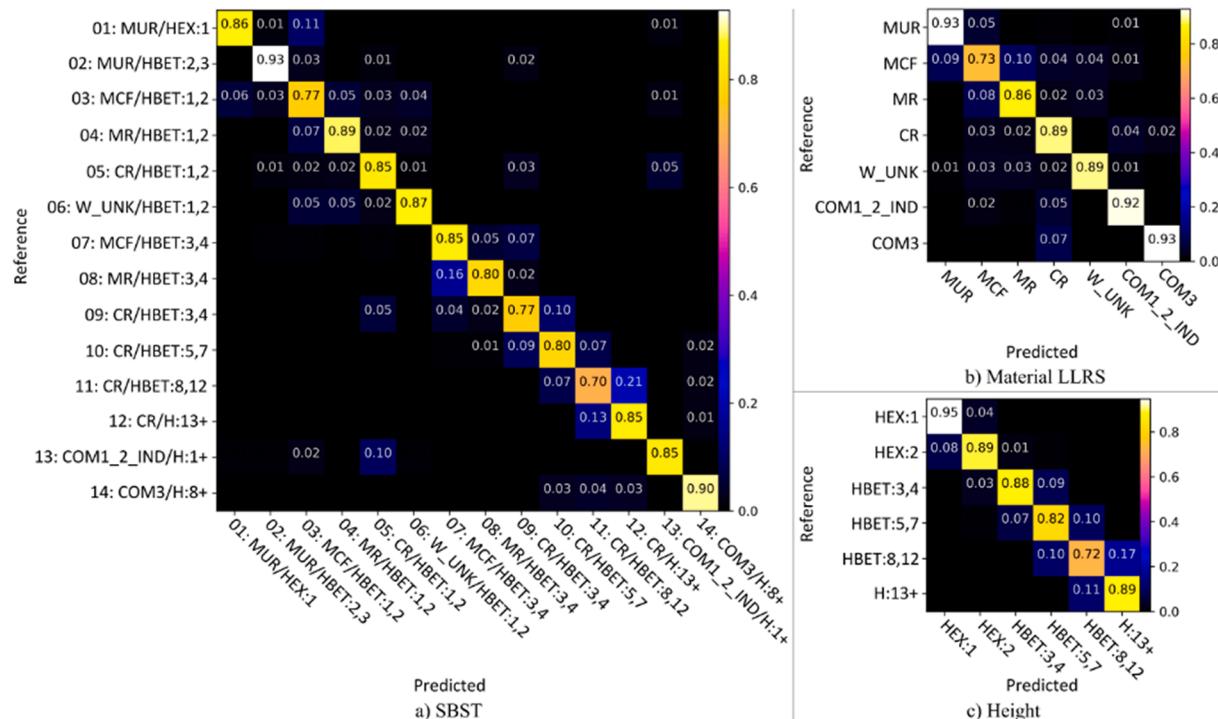


Fig. 13. Normalized confusion matrices of best performing DCNNs for the three classification tasks.

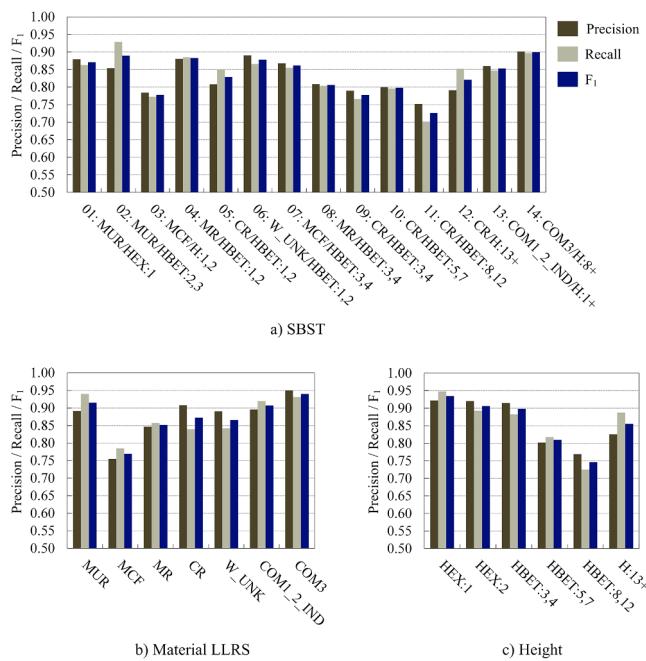


Fig. 14. Precision, Recall, and F₁ scores of best performing DCNNs for the three classification tasks.

here is that, despite our labeling strategy, the deployed reference data do not represent absolutely verified ground truth information (e.g., from construction plans). Instead, they result from human visual image interpretation. In less unambiguous cases, this is prone to induce label uncertainty that propagates to model learning and validation. With respect to the application of derived information in seismic risk modeling, such uncertainties should ideally be considered as epistemic uncertainties.

Overall, the experimental results clearly demonstrate the potential to automate the *in-situ* collection of vulnerability-related structural features from remote based on street-level images with extensive coverage and DCNNs. This also accounts for the retrieval of the often very subtly encoded and thus challengingly inferable LLRS material.

5.5. Application

To illustrate the application of the proposed procedure on an automated building characterization on a large-area, we applied the best performing DCNNs to all 204,030 filtered façade views. Resulting class frequency distributions are shown in Fig. 15.

The spatial distribution of each of the predicted SBSTs is depicted in Fig. 16 and reveals the distinct patterns of the different SBSTs across the area of Greater Santiago.

01:MUR/HEX:1 and 02:MUR/HBET:2,3 classifications are located in particular in the core of Santiago and its direct surroundings. Thereby, the higher 02:MUR/HBET:2,3 buildings concentrate almost exactly in the historic center, whereas the single-storey 01:MUR/HEX:1 buildings extend considerably beyond its boundaries and thus also include the historical outskirts of Santiago (e.g., Greene and Soler, 2005). 03:MCF/HBET:1,2 buildings spread nearly all over the City except for the high socio-economic status areas in the north-east. 05:CR/HBET1,2 is distributed throughout the conurbation, with increased amounts in areas of medium to high socio-economic status extending from the southwest to the northeast (a spatial consideration of Greater Santiago from a socio-economic perspective is provided, e.g., in Garreton, 2017). 04:MR/HBET1,2 as well as 06:W_UNK which can almost exclusively be attributed to low to medium economic status residential buildings, extend radially away from the boundaries of the city center and the municipalities with high socio-economic status in the north-east. 07:MCF/HBET:3,4 buildings are distributed dispersed throughout the city center and in addition, as with 08:MR/HBET:3,4, in a rather clustered manner across Santiago's residential areas. These clusters mainly refer to low to middle socio-economic status condominiums or social housing. 09:CR/HBET:3,4 CR is spread over the entire city, with more pronounced concentrations in the center as well as towards the north-east. The higher the CR SBST buildings are, the more clearly their occurrence is limited to residential areas of medium to high socio-economic status and the business and financial districts, which particularly expand northeastwards from the center. The presence of 14:COM3/H:8+ buildings is exclusively limited to these districts. Buildings classified as 13:COM1_2 are spread across the entire city area, with higher concentrations in the center and lower occurrences in the eastern parts. This shows that the quantity and spatial distribution of the different SBSTs is highly heterogeneous. Accordingly, the exposure to earthquakes follows a very distinct spatial variability.

For an application of the presented methodology to new geographic regions the target typologies as well as respective visual-structural criteria need to be revised and adjusted where necessary to capture the target domain properly. Furthermore, situation appropriate transfer learning techniques (Pan and Yang, 2010) could be considered to obtain a favorable trade-off between generalization capabilities and additional reference data collection efforts.

6. Conclusion

This study demonstrates the potential of the combined use of geocoded street-level imagery such as provided by GSV and state-of-the art DCNNs to automatize the classification of structural features of buildings for large-area seismic risk assessments. To this end, a properly tailored workflow for the targeted collection of street-level façade imagery, the compulsory labeling of reference data based on herein encoded *visual-structural criteria*, as well as information extraction based on DCNN image classification in a comparative experimental setting was

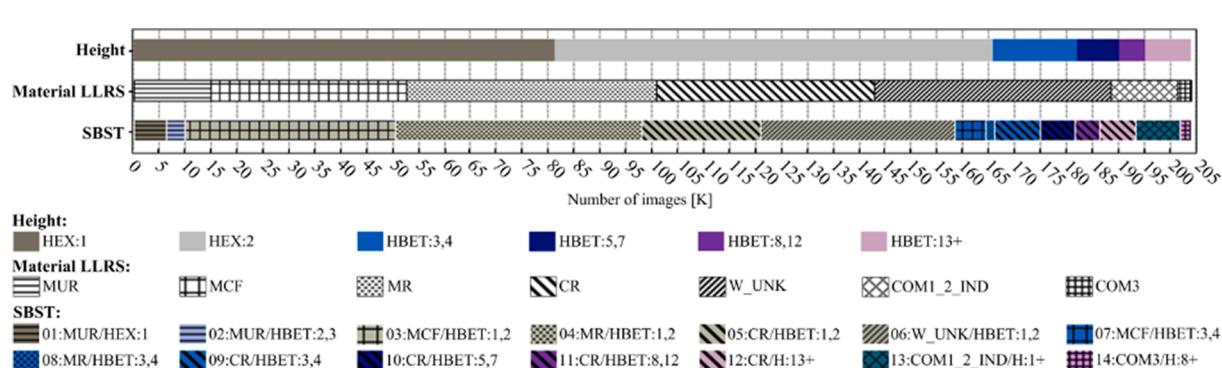


Fig. 15. Numbers and class distributions of classification outputs from best performing DCNNs.

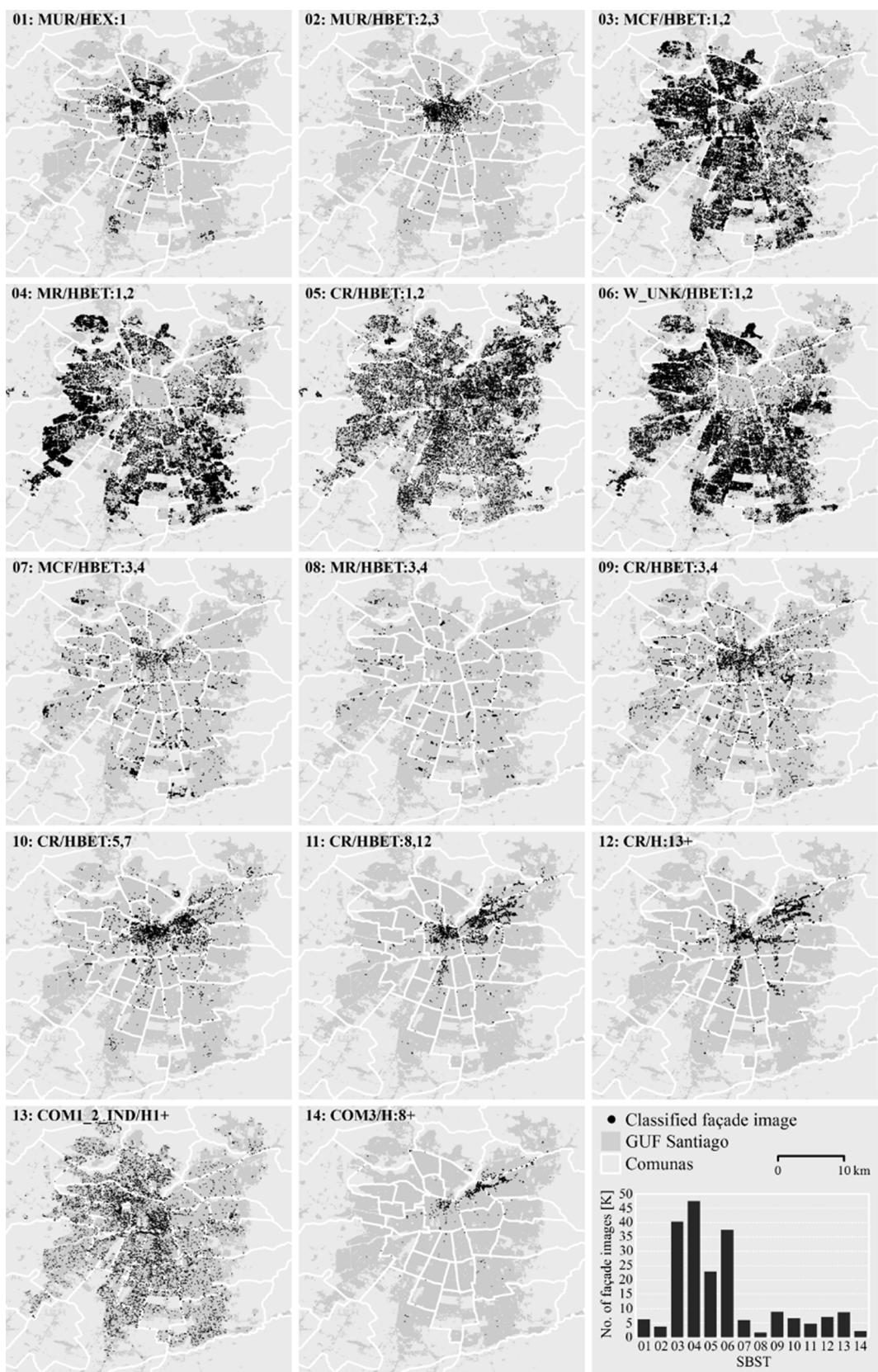


Fig. 16. Spatial distribution and class frequencies of SBST classification outputs from best performing DCNN.

elaborated. The experimental results for the study site Santiago de Chile prove that DCNNs allow for an accurate inference of key structural vulnerability-related building characteristics from street-level façade imagery, i.e., the SBST, the LLRS material, and the building height. Thereby, the transfer learned NasNet-A models have performed best. This enables an automated and thus efficient large-area *in-situ* data collection. The achievement of relatively high accuracy levels (>0.80 for SBST, >0.85 for LLRS material and height prediction) in spite of a limited amount of reference data indicates that the potential for extracting risk-related building structural information from street-level façade images in the natural hazard risk context can be further elaborated.

Previously, studies on the analysis of building vulnerability to earthquakes, but also to other natural hazards were limited with regard to their spatial coverage as well as their thematic depth with respect to meaningful structural building characteristics. These limitations can be primarily attributed to a lack of representative *in-situ* information. With the data and methods presented in this study, these limitations can now be overcome for many areas of the world while still remaining accurate in spatial and thematic depth. This level of detail is of central importance for building vulnerability analyses because, as shown in Fig. 16, the distribution of different building structural types can exhibit large variabilities across space.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Google for the access to the imagery and meta-data through their Street View Static API. This research received funding by the German Federal Ministry of Education and Research (BMBF) under grant no. 03G0876 (project RIESGOS). P. Aguirre and H. Santa María acknowledge funding from the National Research Center for Integrated Natural Disaster Management (CIGIDEN) CONICYT/FONDAP/15110017, and by the Regular Fondecyt Project CONICYT/FONDECYT/1191543.

References

- Aguirre, P., Vásquez, J., de la Llera, J.C., González, J., González, G., 2018. Earthquake damage assessment for deterministic scenarios in Iquique, Chile. *Nat. Hazards* 92 (3), 1433–1461.
- Alvarez, C., Hube, M., Rivera, F., Santa María, H., Hernandez, D., 2016a. Adobe House. World housing encyclopedia report 179 (Chile), EERI, IAEE; URL: <http://db.world-housing.net/building/179> (accessed on 29 May 2020).
- Alvarez, C., Hube, M., Rivera, F., Santa María, H., Hernandez, D., 2016b. Confined Masonry. World housing encyclopedia report 181 (Chile), EERI, IAEE; URL: <http://db.world-housing.net/building/181> (accessed on 29 May 2020).
- Alvarez, C., Hube, M., Rivera, F., Santa María, H., Labarca, M., 2016c. Timber Houses. World housing encyclopedia report 182 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/182> (accessed on 29 May 2020).
- Alvarez, C., Hube, M., Rivera, F., Santa María, H., Labarca, M., 2016d. Reinforced Concrete Shear Wall Houses. World housing encyclopedia report 180 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/180> (accessed on 29 May 2020).
- Ammirati, J.-B., Vargas, G., Rebollo, S., Abrahami, R., Potin, B., Leyton, F., Ruiz, S., 2019. The crustal seismicity of the Western Andean Thrust (Central Chile, 33°–34° S): implications for regional tectonics and seismic hazard in the Santiago Area. *Bull. Seismol. Soc. Am.* 109 (5), 1985–1999.
- Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J., 2010. Google street view: capturing the world at street level. *Computer* 43, 32–38.
- Aravena Pelizari, P., Spröhnle, K., Geiß, C., Schoepfer, E., Plank, S., Taubenböck, H., 2018. Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens. Environ.* 209, 793–807.
- Armijo, R., Rauld, R., Thiele, R., Vargas, G., Campos, J., Lacassin, R., Kausel, E., 2010. The West Andean thrust, the San Ramon fault, and the seismic hazard for Santiago, Chile. *Tectonics* 29, TC2007. <https://doi.org/10.1029/2008TC002427>.
- Astroza, M., Ruiz, S., Astroza, R., 2012a. Damage assessment and seismic intensity analysis of the 2010 (Mw 8.8) Maule earthquake. *Earthquake Spectra*. 28 (1_suppl1), 145–164.
- Astroza, M., Moroni, O., Brzve, S., Tanner, J., 2012b. Seismic performance of engineered masonry buildings in the 2010 Maule earthquake. *Earthquake Spectra*. 28 (1_suppl1), 385–406.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8), 1798–1828.
- Bilham, R., 2009. The seismic future of cities. *Bull. Earthq. Eng.* 7 (4), 839–887.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65 (1), 2–16.
- Borzi, B., Dell'Acqua, F., Faravelli, M., Gamba, P., Lisini, G., Onida, M., Polli, D., 2011. Vulnerability study on a large industrial area using satellite remotely sensed images. *B. Earthq. Eng.* 9 (2), 675–690.
- Branson, S., Wegner, J.D., Hall, D., Lang, N., Schindler, K., Perona, P., 2018. From Google Maps to a fine-grained catalog of street trees. *ISPRS J. Photogramm. Remote Sens.* 135, 13–30.
- Brzve, S., Scawthorn, C., Charleson, A.W., Allen, L., Greene, M., Jaiswal, K., Silva, V., 2013. GEM Building Taxonomy Version 2.0, GEM Technical Report 2013-02 V1.0.0. GEM Foundation, Pavia, Italy.
- Calvi, G.M., Pinho, R., Magenes, G., Bommer, J.J., Restrepo-Vélez, L.F., Crowley, H., 2006. Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET J. Earthquake Technol.* 43 (3), 75–104.
- Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Coburn, A., Spence, R., 2002. *Earthquake Protection*. John Wiley & Sons, New York.
- Duarte, P., 2009. Innovación constructiva a principios del siglo XX: preámbulo a la modernidad arquitectónica y arquitectura subestimada. *Revista de Arquitectura* 15 (20), 20–26.
- Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., Strano, E., 2017. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* 134, 30–42.
- FEMA, 2015. Rapid visual screening of buildings for potential seismic hazards: A handbook (FEMA P-154), third ed. Department of Homeland Security, Federal Emergency Management Agency, Washington, DC.
- Garretón, M., 2017. City profile: Actually existing neoliberalism in Greater Santiago. *Cities* 65 (2017), 32–50.
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E.L., Fei-Fei, L.i., 2017. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *PNAS* 114 (50), 13108–13113.
- Geiß, C., Taubenböck, H., 2013. Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Nat. Hazards* 68 (1), 7–48.
- Geiß, C., Taubenböck, H., Tyagunov, S., Tisch, A., Post, J., Lakes, T., 2014. Assessment of seismic building vulnerability from space. *Earthq. Spectra* 30 (4), 1553–1583. <https://doi.org/10.1193/121812EQS350M>.
- Geiß, C., Aravena Pelizari, P., Marconcini, M., Sengara, W., Edwards, M., Lakes, T., Taubenböck, H., 2015. Estimation of seismic building structural types using multisensor remote sensing and machine learning techniques. *ISPRS J. Photogramm. Remote Sens.* 104, 175–188.
- Geiß, C., Jilge, M., Lakes, T., Taubenbock, H., 2016. Estimation of seismic vulnerability levels of urban structures with multisensor remote sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9 (5), 1913–1936.
- Geiß, C., Thoma, M., Pittore, M., Wieland, M., Dech, S.W., Taubenbock, H., 2017a. Multitask active learning for characterization of built environments with multisensor earth observation data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (12), 5583–5597.
- Geiß, C., Aravena Pelizari, P., Schrade, H., Brenning, A., Taubenbock, H., 2017b. On the effect of spatially non-disjoint training and test samples on estimated model generalization capabilities in supervised classification with spatial features. *IEEE Geosci. Remote Sens. Lett.* 14 (11), 2008–2012.
- Geiß, C., Thoma, M., Taubenböck, H., 2018. Cost-sensitive multitask active learning for characterization of urban environments with remote sensing. *IEEE Geosci. Remote Sens. Lett.* 15 (6), 922–926.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press, Cambridge, MA. ISBN: 9780262035613.
- Greene, M., Soler, F., 2005. Santiago: de un proceso acelerado de crecimiento a uno de transformaciones. In: Santiago en la globalización: ¿una nueva ciudad?, Santiago: Ediciones SUR & IEU + T.
- Gu, D., Gerland, P., Pelletier, F., Cohen, B., 2015. Risks of Exposure and Vulnerability to Natural Disasters at the City Level: A Global Overview. United Nations Department of Economic and Social Affairs Population Division Technical Paper No. 2015/2.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Hoffmann, E.J., Wang, Y., Werner, M., Kang, J., Zhu, X.X., 2019. Model fusion for building type classification from aerial and street view images. *Remote Sens.* 11 (11), 1259. <https://doi.org/10.3390/rs11111259>.
- Hu, C.-B., Zhang, F., Gong, F.-Y., Ratti, C., Li, X., 2020. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Build. Environ.* 167, 106424. <https://doi.org/10.1016/j.buildenv.2019.106424>.
- Hüllermeier, E., Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learn.* 110 (3), 457–506. <https://doi.org/10.1007/s10994-021-05946-3>.

- Ilic, L., Sawada, M., Zarzelli, A., Ribeiro, H.V., 2019. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PLoS ONE* 14 (3), e0212814.
- INE, 2018. Base Cárptografica Censal. Alcances y consideraciones para el usuario. Departamento de Demografía y Censos, Instituto Nacional de Estadísticas. <http://www.censo2017.cl/servicio-de-mapas/descargas/mapas/alcances-base-cartografia-censo2017.pdf>.
- Jorquerá, N., Vargas, J., Lobos, M.L., Cortez, D., 2017. Revealing earthquake-resistant geometrical features in heritage masonry architecture in Santiago, Chile. *Int. J. Archit. Herit.* 11 (4), 519–538.
- Jünemann, R., de la Llera, J.C., Hube, M.A., Cifuentes, L.A., Kausel, E., 2015. A statistical analysis of reinforced concrete wall buildings damaged during the 2010, Chile earthquake. *Eng. Struct.* 82, 168–185.
- Kang, J., Körner, M., Wang, Y., Taubenböck, H., Zhu, X.X., 2018. Building instance classification using street view images. *ISPRS J. Photogramm. Remote Sens.* 145, 44–59.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Klotz, M., Kemper, T., Geiß, C., Esch, T., Taubenböck, H., 2016. How good is the map? A multi-scale cross-comparison framework for global settlement layers: evidence from Central Europe. *Remote Sens. Environ.* 178, 191–212.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2019. Deep learning for hyperspectral image classification: an overview. *IEEE Trans. Geosci. Remote Sens.* 57 (9), 6690–6709.
- Liuzzi, M., Aravena Pelizari, P., Geiß, C., Masi, A., Tramutoli, V., Taubenböck, H., 2019. A transferable remote sensing approach to classify building structural types for seismic risk analyses: the case of Val d’Agri area (Italy). *Bull. Earthquake Eng.* 17 (9), 4825–4853.
- MINVU, 2010. Política urbano – Habitacional de calidad e integración. URL: [http://minvuhistorico\[minvudominio\].cl/incejs/download.aspx?glb_cod_nodo=20100526104950&hdd_nom_archivo=Catalogo%20Minvu01.pdf](http://minvuhistorico[minvudominio].cl/incejs/download.aspx?glb_cod_nodo=20100526104950&hdd_nom_archivo=Catalogo%20Minvu01.pdf).
- Monge, J., 1969. Seismic behavior and design of small buildings in Chile. 4 WCEE, Santiago, Chile, Vol. VI, B6, pp. 1–9.
- Moroni, O., Gomez, C. 2002a. Concrete shear walls buildings. World housing encyclopedia report 4 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/4/> (accessed on 29 May 2020).
- Moroni, O., Gomez, C. 2002b. Concrete frame and shear wall building. World housing encyclopedia report 6 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/6/> (accessed on 29 May 2020).
- Moroni, M.O., Astroza, M., Acevedo, C., 2004. Performance and seismic vulnerability of masonry housing types used in Chile. *J. Perform. Constr. Facil.* 18 (3), 173–179. [https://doi.org/10.1061/\(ASCE\)0887-3828\(2004\)18:3\(173\)](https://doi.org/10.1061/(ASCE)0887-3828(2004)18:3(173)).
- Moroni, O., Gomez, C., Astroza, M., 2014a. Confined block masonry building. World housing encyclopedia report 7 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/7/> (accessed on 29 May 2020).
- Moroni, O., Gomez, C., Astroza, M., 2014b. Reinforced clay/concrete block masonry building. World housing encyclopedia report 5 (Chile), EERI, IAEE; URL <http://db.world-housing.net/building/5/> (accessed on 29 May 2020).
- Pan, S.J., Yang, Q., 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22 (10), 1345–1359.
- Panagiota, M., Chanussot, J., Erwan, P., Guéguen, P., 2012. A support vector regression approach for building seismic vulnerability assessment and evaluation from remote sensing and in situ data. In: Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium pp. 7533–7536, 22–27.
- Pittore, M., Haas, M., Megalooikonomou, K.G., 2018. Risk-oriented, bottom-up modelling of building portfolios with faceted taxonomies. *Front. Built Environ.* 4, 41.
- Rivera, F., Hube, M.A., Santa María, H., Alvarez, C., 2017. Use of remote digital surveys to generate exposure models of residential structures in Chile. In: 16th World Conference on Earthquake Engineering, Santiago de Chile, Paper No. 2414.
- Russakovskiy, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.i., 2015. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* 115 (3), 211–252.
- Santa María, H., Hube, M.A., Rivera, F., Yépez-Estrada, C., Valcárcel, J.A., 2017. Development of national and local exposure models of residential structures in Chile. *Nat. Hazards* 86 (S1), 55–79.
- Sarabandi, P., Kiremidjian, A., 2007. Development of algorithms or building inventory compilation through remote sensing and statistical inferencing. The John A. Blume Earthquake Engineering Center, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA, Report No. 158, p. 419.
- Silva, V., Yépez-Estrada, C., Dabbeek, J., Martins, L., Brzev, S., 2018. GED4ALL - Global Exposure Database for Multi-Hazard Risk Analysis – Multi-Hazard Exposure Taxonomy. GEM Technical Report 2018-01, GEM Foundation, Pavia, Italy.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv eprint 1409.1556.
- Srivastava, S., Vargas-Muñoz, J.E., Tuia, D., 2019. Understanding urban land use from the above and ground perspectives: a deep learning, multimodal solution. *Remote Sens. Environ.* 228, 129–143.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15 (2014), 1929–1958.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016a. Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016b. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv eprint 1602.07261.
- Taubenböck, H., Roth, A., Dech, S., Mehl, H., Münnich, J.C., Stempniewski, L., Zschau, J., 2009. Assessing building vulnerability using synergistically remote sensing and civil engineering. In: Kreck, A., Rumor, M., Zlatanova, S., Fendel, E. (Eds.), *Urban and Regional Data Management*. Taylor & Francis Group, London, pp. 287–300.
- Tucker, B.E., 2013. Reducing earthquake risk. *Science* 341 (6150), 1070–1072.
- Villar-Vega, M., Silva, V., Crowley, H., Yépez, C., Tarque, N., Acevedo, A.B., Hube, M.A., Gustavo, C.D., María, H.S., 2017. Development of a fragility model for the residential building stock in South America. *Earthquake Spectra* 33 (2), 581–604.
- Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B., Begalić, U., 2012. Estimating building inventory for rapid seismic vulnerability assessment: towards an integrated approach based on multi-source imaging. *Soil Dyn. Earthquake Eng.* 36, 70–83.
- Wyss, M., Rosset, P., 2013. Mapping seismic risk: the current crisis. *Nat. Hazards* 68 (1), 49–52.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* 27, 3320–3328.
- Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2018. Places: a 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (6), 1452–1464.
- Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2018. Learning Transferable Architectures for Scalable Image Recognition. arXiv eprint 1707.07012v4.