



Building Stock Classification Using Machine Learning: A Case Study for Oslo, Norway

Federica Ghione^{1,2*}, Steffen Mæland³, Abdelghani Meslem^{1,4} and Volker Oye^{1,2}

¹Department of Applied Seismology, NORSAR, Kjeller, Norway, ²Department of Geosciences, University of Oslo, Oslo, Norway, ³Department of Holistic Systems, SimulaMet, Oslo, Norway, ⁴Norwegian University of Life Sciences (NMBU), Ås, Norway

OPEN ACCESS

Edited by:

Roberto Paolucci,
Politecnico di Milano, Italy

Reviewed by:

Roberto Gentile,
University College London,
United Kingdom
Dennis Wagenaar,
Nanyang Technological University,
Singapore

*Correspondence:

Federica Ghione
federica.ghione@norsar.no

Specialty section:

This article was submitted to
Geohazards and Georisks,
a section of the journal
Frontiers in Earth Science

Received: 28 February 2022

Accepted: 26 May 2022

Published: 16 June 2022

Citation:

Ghione F, Mæland S, Meslem A and
Oye V (2022) Building Stock
Classification Using Machine Learning:
A Case Study for Oslo, Norway.
Front. Earth Sci. 10:886145.
doi: 10.3389/feart.2022.886145

This paper describes a new concept to automatically characterize building types in urban areas based on publicly available image databases, making parts of seismic risk assessment more time and cost-effective, and improving the reliability of seismic risk assessment, especially in regions where building stock information is currently not documented. One of the main steps in evaluating potential human and economic losses in a seismic risk assessment, is the development of inventory databases for existing building stocks in terms of load-resisting structural systems and material characteristics (building typologies classification). The common approach for building stock model classification is to perform extensive fieldwork and walk-down surveys in representative areas of a city, and in some cases using random sample surveys of geounits. This procedure is time and cost consuming, and subject to personal interpretation: to mitigate these costs, we have introduced a machine learning methodology to automate this classification based on publicly available image databases. We here use a Convolutional Neural Network (CNN) to automatically identify the different building typologies in the city of Oslo, Norway, based on facade images taken from *in-situ* fieldwork and from Google Street View. We use transfer learning of state-of-the-art pretrained CNNs to predict the Model Building Typology. The present article attempts to categorize Oslo's building stock in five main building typologies: timber, unreinforced masonry, reinforced concrete, composite (steel-reinforced concrete) and steel. This method results in 89% accuracy score for timber buildings, though only 35% success score for steel-reinforced concrete buildings. We here classify and define for the first time a relevant set of five typologies for the Norwegian building typologies as observed in Oslo and applicable at national level. In addition, this study shows that CNNs can significantly contribute in terms of developing a cost-effective building stock model.

Keywords: building stock model, convolutional neural network, machine learning, seismic risk assessment, Oslo (Norway)

1 INTRODUCTION

Within the last century, earthquakes, flooding and droughts have been the dominating phenomena responsible for causing fatalities and economic losses worldwide (see Data Availability Statement section). In the last few decades though, earthquakes are the dominating phenomena standing for most fatalities (Wallemaq and House, 2018), providing the motivation and requirement for more

detailed and effective seismic risk assessment studies. A seismic risk assessment estimates the probability of losses if an earthquake occurs, can assist through updated building regulations and to initiate other mitigation actions to avoid likelihood for casualties and economic losses. To do that, information regarding seismic hazard, vulnerability and exposure models of the area are needed (Silva et al., 2014). In particular, the exposure model contains information about all the buildings in a specific area, infrastructures, and population data. A seismic risk assessment for a building stock in a city needs a classification of the buildings in accordance with their structural typologies, an important parameter to define the building performance under seismic load.

The development of a building stock model is always challenging and time consuming, especially when the area of investigation is large. Until today it is a common practice to use both Google Earth and *in-situ* surveys to get information on building structural systems and material characteristics. However, machine learning methods to analyze visual imagery have recently been applied to classify building stocks, using online available façade images, obtained through Google Street View (GSV), as key information. To our knowledge, this automatic image analysis using Convolutional Neural Network (CNN) was previously only tested in two places to automatically detect building materials and types of lateral-load resistance: Gonzalez et al. (2020) for Medellin city, Colombia and Aravena Pelizari et al. (2021) for Santiago de Chile. Those two examples are similar to our study, in terms of applying CNNs to automatically detect and identify the different building typologies for the two cities using GSV images. In this study, however, we define for the first time a Model Building Typology (MBT) for Oslo, and we conduct a survey amongst experts within earthquake engineering. Based on the survey's results, the MBT identified in Oslo can further be applied at the national scale (Norway). Another example by GFDRR (2018) shows that using a combination of deep learning and GSV together with drone images may lead to remotely identify buildings that require further inspection and possible retrofitting/strengthening for Guatemala city, Guatemala.

This paper presents the building stock model for Oslo, capital of Norway. The general information of the buildings is obtained from the public cadastre from www.kartverket.no. This information contains the total number of buildings in Oslo with the corresponding coordinates, number of stories, number of housing units, usable area, and total area.

To perform seismic risk studies, the availability of detailed building stock typologies is a necessity. This is usually a very time-consuming process, and our innovative approach using transfer learning and CNNs for the classification may provide this required input in much shorter time and at lower cost. This could be a solution for most of the urbanized regions in the world to develop seismic risk assessments and incorporate earthquake preparedness actions.

The paper is organized as follow: in **Section 2** we describe the methodology followed and the CNN approach; **Section 3** describes the study area, the identified building typologies, and the data used for this work; **Section 4** introduces the results and **Section 5** presents the main conclusion.

2 METHODOLOGY

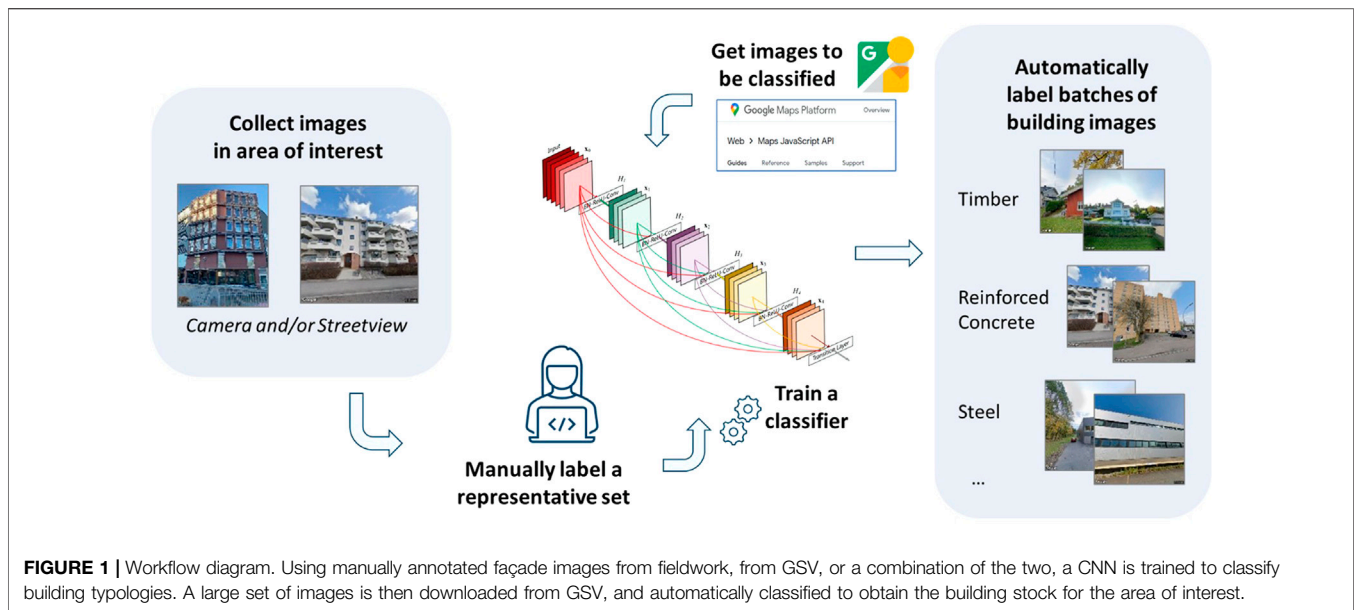
The methodology is based on the use of a CNN for automating the typology classification step. While the classification itself requires no manual involvement, the main workload goes into compiling a labelled image dataset (discussed in **Section 3**) and training the parameters of the CNN (described in this section). Once a CNN structure is selected and the CNN is suitably trained, classifying a set of unseen images is done quickly and at low computational cost. A diagram depicting the workflow is shown in **Figure 1**. First, a representative set of façade images must be collected and labelled, which entails the manual part of the process. Using these images to train a CNN, subsequent images of unknown typologies can be downloaded online and classified in bulk and used for seismic risk assessment. There are two challenges when using CNNs for this task: the need for a large set of labelled images, and the computational load of training them. We mitigate both by the use of *transfer learning*, which will be described in this section.

2.1 Image Recognition

Current state-of-the-art methods for recognizing objects in images are based on variants of CNNs, which are artificial neural networks that use sliding filters (or kernels) to process their inputs (LeCun et al., 2015). Unlike traditional image processing, these filters are learnt from examples during a training phase, and they are typically stacked so that the output from one convolutional layer is used as input to the next one. A benefit of the convolutional filters is that they are invariant under translations, meaning that the positioning of objects in an image does not matter (Goodfellow et al., 2016). Stacking convolutional layers allows for learning pattern hierarchies, where the first layers recognize simple shapes such as vertical or horizontal lines, while the last layers recognize compositions of these patterns, for instance the shape of a building. In most approaches, the convolutional layers are interspersed with pooling layers, which reduce the dimensions by downsampling. These reduce the required number of learnable parameters, and at the same time introduce invariance to rotation and scaling (Goodfellow et al., 2016). In addition, it is common to add special layers and mechanisms that facilitate the learning process, such as dropout (Srivastava et al., 2014) and residual (skip) connections (He et al., 2016a). The composition of layers is referred to as the *architecture* of the network, while we use *model* to indicate a network with a particular set of optimized parameters. For classification purposes, the output of the final convolution layer is input to one or several fully connected layers, which ultimately output a mutually exclusive prediction for which class an input image belongs to. The learnable parameters of a model include both the convolution filters and the weights of the classification layer.

2.2 Transfer Learning and Network Architectures

Modern CNNs typically have millions of free learnable parameters and optimizing them requires large corpora of



data. In image recognition, the standard dataset for training and performance evaluation is the *ImageNet (ILSVRC)* database (Russakovsky et al., 2015), which contains nearly 1.3 million labelled images of various common objects. When training a CNN on such a diverse dataset, the first layers will be sensitive to simple shapes like straight or curved lines, while the middle layers are sensitive to different compositions of these basic patterns. Since only the last layers are highly specific to the particular dataset the CNN is trained on, applications such as ours can benefit from *transfer learning*, where one first trains the CNN on a large, generic dataset such as ImageNet, and subsequently re-trains (or *fine-tunes*) the last layers on the application-specific data. This allows for re-using the knowledge contained in the first and middle layers and is particularly useful in cases like ours where the application-specific dataset is comparatively small.

To select the optimal CNN architecture for our case, we compute four performance metrics (described in next paragraph) for eleven candidates, all pre-trained on ImageNet data. These are available in the Keras framework (Chollet, 2015). For each candidate, we set all parameters to remain constant, but replace the final classification layer by a new layer with six output nodes, corresponding to the number of building typologies under consideration. As a measure against overfitting, we apply dropout (Srivastava et al., 2014) before the classification layer. This means temporarily removing nodes from the network during training, which improves robustness by reducing the nodes' reliance on each other. We randomly drop nodes at a 20% probability. We then train on our data, updating only the parameters of the classification layer. We perform 4-fold cross-validation, meaning we divide the training data into four equally sized parts, and use three parts for training and one for computing metrics. The parts are then rotated so that all parts are used for computing metrics. Since the weights of the final layer are randomly initialized for each training round, potentially affecting the final performance, we repeat the cross-validation process twice, and report the

average results for the in total eight trained models per architecture. Training is done for up to 100 epochs, where an epoch is a single pass over all the training data, followed by a metric evaluation. If learning fails to improve over five consecutive epochs, the learning rate is reduced by half, thereby taking shorter steps toward the optimal solution. If learning fails to improve over ten consecutive epochs, the training is stopped.

The four computed metrics are: accuracy, precision, recall, and area under the receiver operating characteristic curve. Accuracy is defined as the fraction of samples where the predicted class exactly matches the true class; in our case of 6 distinct classes, a random classifier would have an accuracy of $1/6 = 0.17$, assuming that the number of samples of each class are the same. Precision is defined as the number of true positives divided by the sum of true positives and false positives, indicating the quality of the positive predictions. Recall, on the other hand, is defined as the number of true positives divided by the sum of true positives and false negatives, indicating the completeness of the positive predictions. In addition, we compute the area under the receiver operating curve (ROC), which is obtained by plotting the true positive rate against the false positive rate. Since true and false predictions relate to a binary classification problem, we compute precision, recall and ROC individually for each class, in a one-vs-all fashion, and then report the average across all classes. The results are listed in **Table 1**. For all metrics, the best possible result is 1, while the worst possible result is 0.

2.3 Fine-Tuning

From **Table 1** we observe that most CNN architectures yield similar results, but the three members of the DenseNet (Huang et al., 2017) family stand out as the highest performers across all evaluation metrics. They are structurally similar but differ in the number of convolution layers, which is given by the number at the end of their names (121, 169 and 201, respectively). The DenseNet201 architecture excels on all metrics and is therefore

TABLE 1 | Metrics for different CNN architectures. Highlighted in bold are the best observed values. The last column indicates the total number of parameters for each architecture.

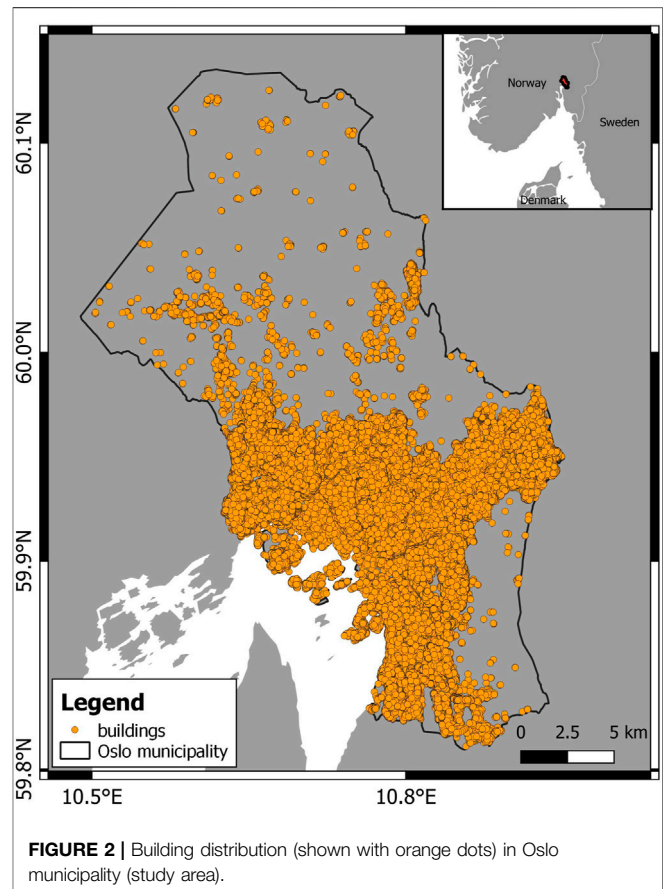
Architecture	Accuracy	Precision	Recall	ROC AUC	Parameters
Xception Chollet. (2017)	0.759	0.762	0.759	0.920	20,873,774
VGG16 Simonyan and Zisserman. (2015)	0.740	0.739	0.740	0.906	14,717,766
VGG19 Simonyan and Zisserman. (2015)	0.732	0.733	0.732	0.904	20,027,462
ResNet50V2 He et al. (2016b)	0.759	0.757	0.759	0.917	23,577,094
ResNet101V2 He et al. (2016b)	0.762	0.760	0.762	0.917	42,638,854
ResNet152V2 He et al. (2016b)	0.757	0.757	0.757	0.912	58,343,942
InceptionV3 Szegedy et al. (2016)	0.762	0.762	0.762	0.922	21,815,078
InceptionResNetV2 Szegedy et al. (2017)	0.750	0.752	0.750	0.912	54,345,958
DenseNet121 Huang et al. (2017)	0.764	0.764	0.764	0.922	7,043,654
DenseNet169 Huang et al. (2017)	0.770	0.772	0.770	0.926	12,652,870
DenseNet201 Huang et al. (2017)	0.778	0.780	0.778	0.932	18,333,510

selected for further optimization. While being largest in the DenseNet family, we note that this family of architectures is the smallest in terms of number of learnable parameters, compared to the other ones tested. This indicates that raw model size, as listed in the last column of **Table 1**, cannot be directly considered as a proxy for performance, and comparisons are required to find the best architecture for a given application. We find, however, that there is an approximately linear relationship between the model size and the training time. Compared to DenseNet201, the smallest architecture (DenseNet121) takes 12% shorter time to train a single epoch, while the largest architecture (ResNet151V2) takes 31% longer, when training on a Nvidia V100 (NVIDIA, 2020) graphics card.

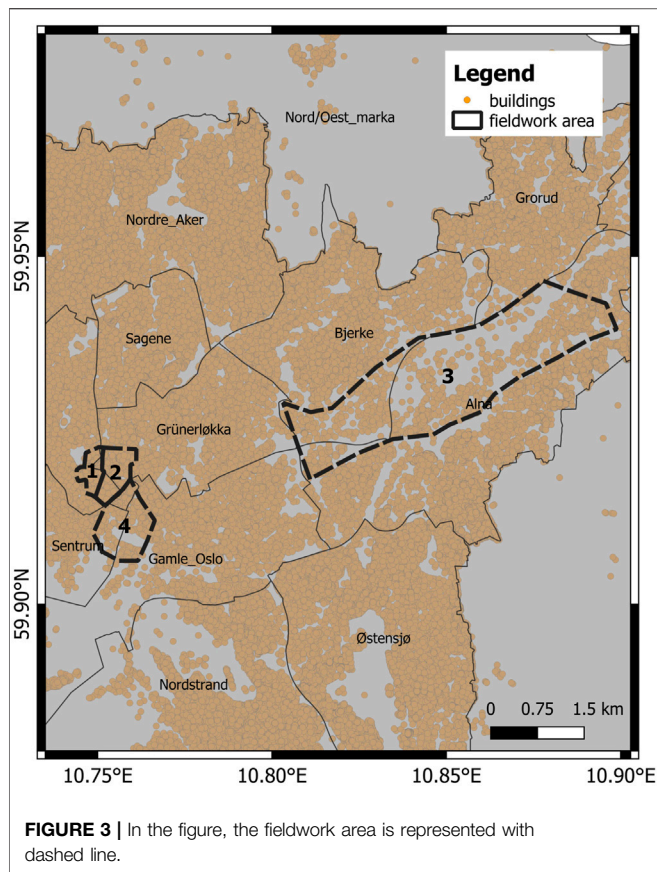
Proceeding with the DenseNet201 architecture, we fine-tune it to our data in a two-step procedure. Using the full training dataset, apart from 20% of images that are set aside as validation data to monitor the training progress, we first train only the final classification layer of the model. Like before, we halve the learning rate if training has stagnated for five epochs, and end training if it has still not improved after ten epochs. In the second step, we now additionally unfreeze the parameters of the last 64 convolution layers. This allows for adapting a larger part of the model specifically to our data, while still retaining the low-level pattern recognition of the first convolution layers. The DenseNet201 architecture is organized into five blocks of multiple convolution layers, where each block has a complicated internal structure, but has a single connection to the next block. Therefore, we only consider it useful to free the parameters of entire blocks at the time; freeing the last 64 convolution layers equals the entire last block. Doing so means we now train 38% of all parameters in the model (approximately 7 million out of 18 million). Experiments with freeing a larger number of blocks resulted in overfitting, where the accuracy on the validation data diverges greatly from that of the training data. As for the first step, training is run until improvement has failed to improve for ten epochs.

3 DATA

In order to conduct seismic risk assessment a building stock model needs to be available, and the key input data to develop



such building stock models are the building typologies as extracted in the relevant region. Building typologies are related to the ability of a building to resist lateral loads that mainly affects the structural system and depends on material and height. The lateral load-resisting system and its material can be identified only from the blueprints (two-dimensional set of technical/engineering drawings that specify a building's dimensions, construction materials, and the exact placement of all its components) or by direct expert observations. Unfortunately, structural blueprints are not always available. Therefore, expert



opinions seem to be the best option to assemble building inventories and fundamental information. However, in most of the cases, it is timewise and economically not possible to survey each asset. For this reason, many assumptions need to be made to establish a building stock model and interpolate data from surveyed areas to not surveyed neighborhoods. Recently, using GSV (Google Street View) and Google Earth has become an alternative way to carry out fieldwork for visualizing façade and material information remotely (GFDRR, 2018; Kang et al., 2018; Gonzalez et al., 2020; Aravena Pelizari et al., 2021). However, uncertainties are remaining in the building typology classification. In this paper, we use the EMS-98 building taxonomy (Grünthal, 1998) with some extension related to specific typology that are not present in the original classification [e.g. composite (steel-reinforced concrete) typology].

3.1 Study Area

Our study area is the city of Oslo, the capital of Norway (Figure 2). The oldest settlements in the area are from around 11000 BC, but Oslo was only founded in 1000 AC. The oldest part of the city is the eastern part, called Bjørvika in Gamle Oslo (see Figure 3). In 1624, a disastrous fire destroyed most of the city, and built on the ashes, the new town was called Christiania. After that devastating fire, a ban on wooden houses was introduced allowing only solid brick, and half-timbered brick houses (Eriksson et al., 2016). In 1769 Christiania had about 7500

inhabitants. During the 1600s the demand for wood rose, and Christiania became an important harbour for trading wood. Outside the city centre, many small wooden houses were built, and some of those still exist today. During the industrial period of the 1840s, many factories were built along the river Akerselva and the population increased significantly to about 113.000 inhabitants. In the late 1880s, many multi-stories brick tenements were constructed to fulfil the request for more living space. In 1925 the city changed its name back to Oslo and the Ring Road was introduced; in 1948 another important step for the city was the development of the new subway system (Eriksson et al., 2016). In 2020 Oslo had a population count of 697,549 and by 2040 the expected number is around 926,000 (Eriksson et al., 2016). This prediction has a direct consequence for the city, i.e. it will need to accommodate the new residents with new buildings. Oslo municipality covers an area of 480 km² and it is subdivided in 17 boroughs or *bydel* in Norwegian.

One key element to consider during the planning of a city expansion with respect to new infrastructures is an accurate evaluation of the seismic risk, which is strongly depending on the local seismic hazard conditions, in turn influenced by soil amplification. Norway shows low to medium seismicity that leads to a lower level of seismic hazard compared to other Southern European countries (Danciu et al., 2021). Compared on a national scale, the city of Oslo falls within a zone of intermediate seismic hazard. Oslo was hit by a significant earthquake (5.4 Mw) on the 23rd of October 1904 (Bungum et al., 2009). Although the epicenter was located 115 km south of Oslo, the event (known as 1904 Oslofjord earthquake) generated ground motions that propagated on both sides of the Oslo fjord from the south of Fredrikstad/Tønsberg to the north of Oslo. The earthquake was felt over an area of 800.000 km² from Namsos in the north to Poland, and across southern Norway to Helsinki in the east (Bungum et al., 2009). The maximum intensity on the Mercalli scale in Oslo was reported to VI and major damages were reported, mainly for wooden and unreinforced masonry buildings in the Oslo area. Although the severe damages to the buildings, no relevant and significant mitigation measures were put in place after this event. The first seismic standard introduced in Norway was by the end of the 1980s, and it was mostly destined for seismic design of offshore structures. Within the same period there was also another document destined for buildings, but it was introduced and used only as recommendation. In 2004 a new standard on the design of structures considering loads from seismic influence was adopted (NS 3491-12). In 2008, Norway adopted the Eurocode and since then it has been the only standard for seismic design for all types of structures and infrastructures.

3.2 Model Building Typology

Building stocks usually vary from locality to locality and certainly even more between different countries due to different construction practices, material availability and construction period.

The following steps are followed to recognize the different building typologies for Oslo and to develop a Model Building Typology (MBT):

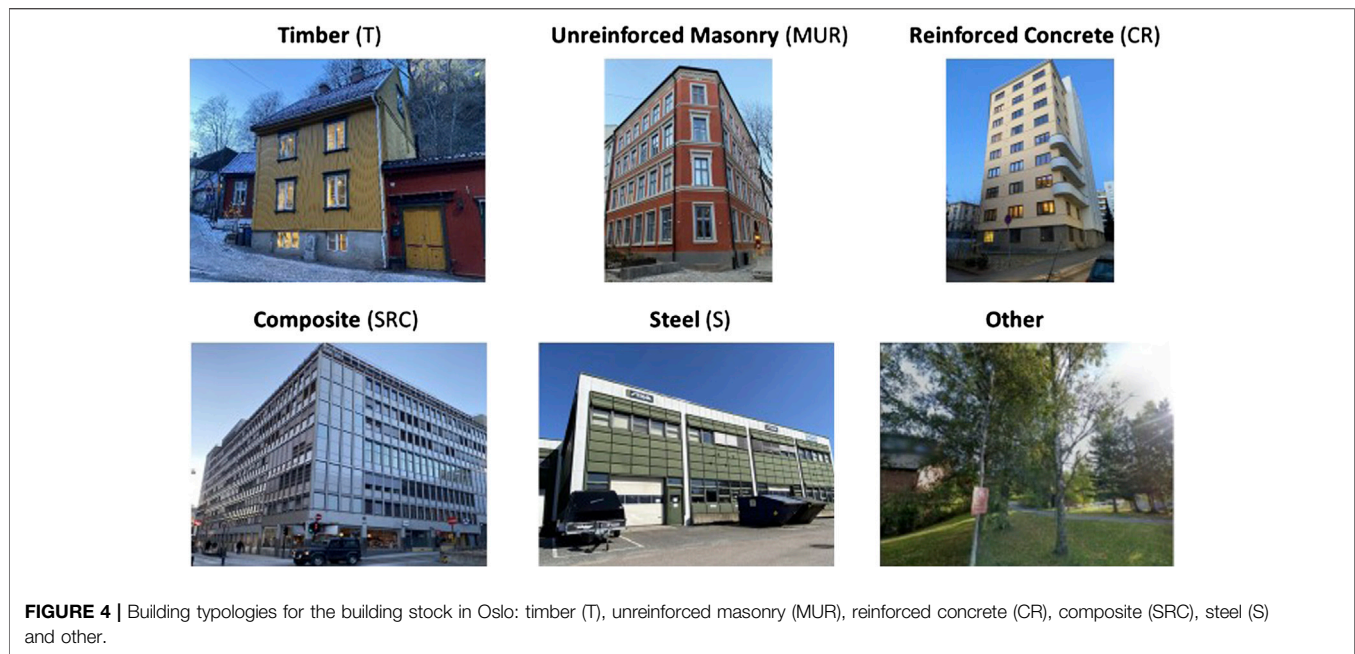


FIGURE 4 | Building typologies for the building stock in Oslo: timber (T), unreinforced masonry (MUR), reinforced concrete (CR), composite (SRC), steel (S) and other.

- 1) a first overview of the building typologies is obtained from Google Earth for the different neighborhoods in the city. The building stock is observed at different scales (from small to large):
 - the inbuilt 3D-building option allows us to identify the lateral load bearing systems, predominant work material for walls, façade decoration, flooring and roofing types and number of stories.
 - The Street View mode gives us a ground-level view of the different buildings and it helps us to identify similarity and heterogeneity in the same geounit.
 - 2) - With the plan view, we understand the general distribution of the building stock in the city looking from the top. It allows us to identify different roof materials and characteristics that can be potentially linked to different building typologies. Through the first evaluation using Google Earth, some area of the city center (Grünerløkka, Oslo central station, Bjørnvika) and Alna (see **Figure 3**) are chosen for detailed *in-situ* fieldwork. Those areas contain a good representation of all the typologies and are good candidates to test the machine learning methodology to automatically identify different building typologies.
 - 3) During 5 days of fieldwork during the winter of 2021, about 350 pictures of the facade buildings are taken manually, and information related to structural system and material related data are collected (e.g. lateral loadbearing systems observed, material type, flooring/roofing system, number of stories, usage/type of activity). These form key information to define a MBT.
 - 4) After steps 1 and 2, we define an initial building typology classification, and we divide the observations into five groups with the corresponding typology.
 - 5) To confirm and validate the initial evaluation of the typologies, a survey questionnaire regarding seismic vulnerability assessment of the existing building stock in Oslo is sent out to experts (mainly engineers working in Norway) in order to validate the building typologies that they have been identified. The survey was divided into eleven different sections, with questions related to the structural system and material characteristics of the building stock in Oslo, date of practice of a given typology and general practices. The expert's opinion confirmed the initial evaluation and they agreed that the building typologies recognized in Oslo are applicable at national scale. The questionnaire shows results compatible to the preliminary assessment.
 - 6) Combining results from both fieldwork and survey questionnaire, a final Model Building Typology (MBT) is defined, and it represents the existing building typologies in Oslo city and Norway.
 - 7) Images from GSV are downloaded for all the districts of the city. A total of 5074 pictures are manually labelled using the MBT previously defined. The pictures are used to train the Convolutional Neural Network (CNN). As Gonzalez et al. (2020), Aravena Pelizari et al. (2021), also here the CNN methodology is applied to automatically detect the different building typologies in Oslo.
- The results of the survey questionnaire combined with the field survey have shown that the existing building stock in Oslo can be divided into three main construction periods:
- 1) Buildings built before 1950, estimated to represent roughly 35% of the total existing building stock, and mostly made of timber and unreinforced masonry.
 - 2) Buildings built between 1950 and 1998, estimated to represent 25% of the total number of buildings, and they are mainly made of reinforced concrete.

TABLE 2 | Confusion matrix computed on the test data set. Rows show the true class, and the columns the CNN-identified class. Correct predictions follow the diagonal, highlighted in bold. As an example, 73% of CR buildings are correctly identified, but 3% are classified as MUR, 3% are classified as S, 2% SRC, 9% as T and 10% as other. In parentheses are the actual number of images.

		Predicted					
		CR	MUR	S	SRC	T	Other
True	CR	0.73 (86)	0.03 (4)	0.03 (3)	0.02 (2)	0.09 (11)	0.10 (12)
	MUR	0.05 (5)	0.71 (76)	0.00 (0)	0.00 (0)	0.20 (21)	0.05 (5)
	S	0.18 (5)	0.00 (0)	0.57 (16)	0.00 (0)	0.11 (3)	0.14 (4)
	SRC	0.60 (12)	0.00 (0)	0.05 (1)	0.35 (7)	0.00 (0)	0.00 (0)
	T	0.01 (3)	0.01 (6)	0.00 (0)	0.00 (0)	0.89 (424)	0.09 (42)
	Other	0.02 (4)	0.00 (1)	0.01 (2)	0.00 (0)	0.12 (32)	0.87 (232)

3) Buildings built between 1998 and up to today, estimated to cover 40% and they are found as reinforced concrete, composite (steel-reinforced concrete) and steel.

The following typologies are identified in Oslo and can be applied at national scale (see **Figure 4**):

- Timber (T): the construction with timber has been in practice for many years (before 1950), and the timber buildings mostly are frame structures consisting of wooden frames with solid or plate timber members. This typology is mainly used for housing and typically can be one to two stories, but there are also some 3 to 4 story-high buildings that can also be found as residential apartments and public buildings for various activities, e.g., schools. The timber construction is still in practice and starting from 2000 there have been new modern timber buildings that can also be found as high-rise buildings with more than 8 stories. As per the situation today, timber buildings represent around 20% of the existing building stock in Oslo.
- Unreinforced masonry (MUR): unreinforced masonry constructions are load-bearing wall system structures made of material that vary depending on the construction's age, and which can be burnt clay bricks, stone masonry blocks, concrete blocks or mixtures of other materials. Most of the existing buildings of this type were constructed before 1950. The majority can be found as low/mid-rise structures with 2–5 stories, and very few with more than 6 stories. They are used for residential, commercial and other general activities. In terms of location, this type of construction mostly can be found in the center of the city, and as a rough estimate, they represent 30% of the current building stock in Oslo.
- Reinforced concrete (RC): the construction of reinforced concrete buildings has been in practice since 1950, and the construction procedure of this typology consists of reinforced concrete frames (columns and beams), cast-in-place. But starting from 2000, the pre-cast practice has become more frequent in the construction of this type of structures. The existing RC buildings can cover different ranges of height classes (low, mid and high-rise) and can be found in most of the districts and zones of the city. It is estimated that this type of buildings represents 35% of the total existing buildings in the city.
- Composite (steel-reinforced concrete) (SRC): the composite construction has been in practice since 2000s, and it is estimated that it represents almost 10% of the total existing building stock of Oslo. Typically, the number of stories for this type of building can range between 2 to more than 15 stories for a few of the most recent buildings. The construction procedure of this typology consists of steel frames and cast-in-place concrete frames (columns and beams) and/or concrete shear walls. Buildings of this typology are mostly used as residential apartments, offices, commercial activities, and they are found in most of the districts and zones of the city.
- Steel (S): the existing steel structures are load-bearing steel moment frame constructions, and mostly are found outside the city center. This type of structures is, in general, used for industrial activities, also as big grocery stores, supermarkets, malls, parking lots, and hangars. Due to the nature of the utilization, story heights can be up to 6 m, and the number of stories can be up to 2 stories. The construction in steel has recently increased with the urban development, and currently it is estimated that this type of construction covers about 5% of the existing buildings in the city of Oslo.

In addition to the above typologies, one extra category “other” is added to account for the case that the algorithm could not recognize the typology or that no building could be identified in the picture at all.

3.3 Image Dataset

The dataset of all buildings in Oslo is obtained from the public cadastre (downloaded in December 2020) and it contains the number and the coordinates of all free-standing buildings of more than 50 m². In addition to façade photographs collected manually, images for each building position are downloaded automatically using the Google *Street View* API (GSV). The Street View service provides near-continuous street level imagery of most of the world's cities, and the service's API allows for direct download of images for a given set of coordinates. In order to accept an image to be related to a building, we set a threshold that the image needs to be taken within 40 m from the requested building coordinate. For the



FIGURE 5 | Examples of correctly (A) and falsely (B) classified images.

TABLE 3 | Distribution in terms of numbers and percentages of the predicted building typologies.

Typologies	Number of buildings	%
T	56,305	56.7
MUR	7,979	8.0
CR	7,536	7.6
SRC	399	0.4
S	515	0.5
other	26,605	26.8

134,432 buildings in Oslo, 74% of them fulfil this requirement and it has façade imagery available. Coverage is better in the city center where images for 94% of the buildings are available but can be as low as 50% in distant suburbs.

We do not manually check the image quality, such as how many of the requested buildings are included in the image, or whether it is occluded by trees, passing vehicles, scaffolding or similar. Such cases, where the building typology cannot be identified from the image, are labeled as “other”, and the occurrence in the fieldwork area is approximately 27%. An example of this category is shown in **Figure 4**. Labelling of these cases is often difficult and can be subjective, and is expected to be a considerable source of uncertainty for training the automatic labelling procedure.

A total of 5,074 images from fieldwork and from GSV are manually labelled, and they constitute the dataset used for training and validating the CNN model. We set aside 20% (1,019 images) as a test dataset for the final performance evaluation, leaving 4055 images to use for training. Of the latter, we will also set aside a fraction of them for monitoring the training progress that is described in **Section 2**.

For CNN training and prediction, all images are downsampled to 224 × 224 pixels, which is the input resolution of the CNN model. Images from GSV are downloaded in a square format, while rectangular images from fieldwork are center cropped before downsampling. In order to artificially increase the size of the training data, we augment the images by applying the following

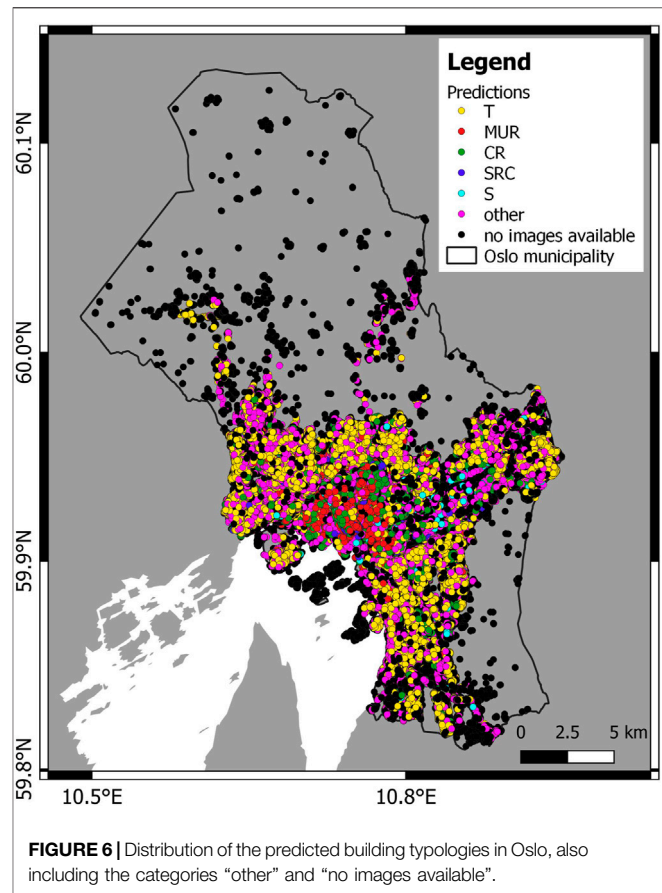
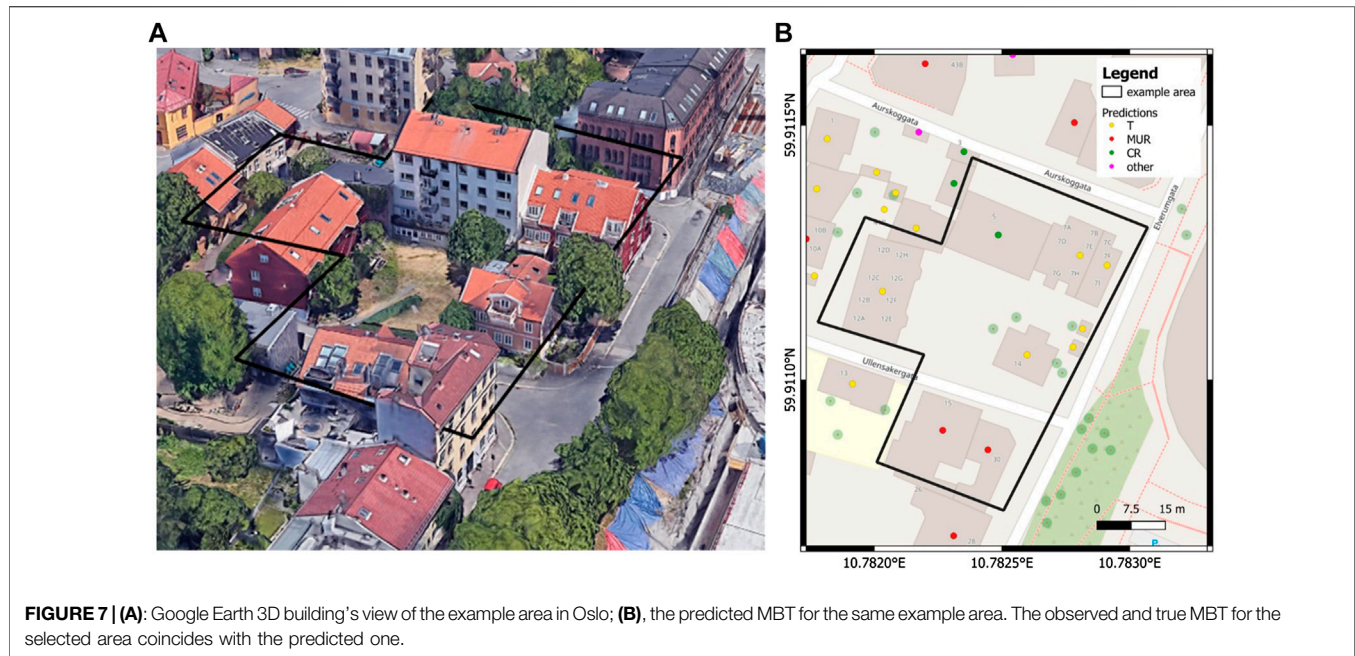


FIGURE 6 | Distribution of the predicted building typologies in Oslo, also including the categories “other” and “no images available”.

transformations: randomly zooming in by up to 20%, randomly rotating by up to 25°, and randomly mirroring along the vertical axis. This is done only during training, and it is a standard procedure to improve the model’s ability to generalize.



4 RESULTS

Having completed the fine-tuning, we compute the final performance metrics on the test dataset. The model achieves an accuracy of 0.825, a precision of 0.825, and a recall of 0.825. Without fine-tuning, we obtain an accuracy of 0.763, precision of 0.775 and recall of 0.769, showing that fine-tuning is greatly beneficial to performance. To investigate the classification performance per typology, we present a confusion matrix in **Table 2**. For each true typology (given by the rows), the confusion matrix shows the rate of test images that are assigned to each predicted label (given by the columns). The best performance is seen for timber (T), where 89% of images are classified correctly. This is expected, because timber is the most common typology in the training data, and therefore the category which the CNN has seen the largest variation of. It is also desirable, since we consider the training (and testing) data to be representative of the entire study area, and a high accuracy for the most common category will necessarily lead to a high overall accuracy.

The poorest performance is seen for steel-reinforced concrete (SRC), where 35% of the images are classified correctly, while 60% of them are classified as reinforced concrete (CR). These two typologies can be difficult to distinguish even for experts, as the façade characteristics may not unambiguously determine the typology. The classification of the “other” groups is in general quite successful with 87%. Within the 13% of misclassifications, almost all images in the “other” category are identified as timber buildings, which can be explained by the frequent presence of fences, rooftops and parked cars in these images, which are typical elements in residential areas common for timber buildings. Some examples of images and their predicted labels are shown in **Figure 5**. In **Figure 5** we can observe an example of a timber building correctly classified and a steel-reinforced concrete building misclassified as reinforced concrete.

Concerning the confusion matrix, we note that we did not apply class-specific weights to the images during training, i.e., all

images were given equal priority. It is an option, however, to weight the images by the inverse of the class prevalence, so that rare classes are given higher priority than common ones. This would lead to a more equal classification accuracy across the typologies, instead of having a very high accuracy for the most prominent class (timber) and low accuracy for the least prominent one (steel-reinforced concrete). At the same time, it would also reduce the overall accuracy, and thereby the quality of the final result, which is why we decided against it.

The predicted building typologies are shown in **Table 3**, in terms of number of buildings and percentages, and in **Figure 6** in terms of spatial distribution. It is important to mention that the total number of buildings of Oslo (134,432) also includes buildings where no GSV images are available to perform the classification (“no image available” category includes 35,093 buildings, shown in **Figure 6**). This category is not shown in **Table 3** because we want to present the percentages of the classified buildings that used GSV images for the classification using machine learning. The “no image available” category represents 26.1% of the total number of buildings, and together with the “other” typology, the two categories sum to 45.9% of the total number of buildings. This means that we were able to automatically attribute a model building typology to 54.1% of the buildings in Oslo.

The predominant typology identified in Oslo is timber, that represents 56.7% of the classified buildings. In **Figure 6** we can observe the distribution of the predicted building typologies: most of the classified buildings are localized in the city center and in the urbanized area. GSV images are not available in remote areas (as forest) in Nordmarka and Østmarka: based on our local knowledge and through supervised learning in the area with Google Earth, the main typology identified is timber, represented by private cabins.

Figure 7 shows the predicted building typologies for an example area in Oslo. This area is chosen because we have a

good representation of different typologies in a small area. On the left side of the figure, the Google Earth 3D building's view is allowing to identify the MBT of the buildings under investigation. The right side of the figure shows the predicted MBT, that coincide with the observed MBT.

5 DISCUSSION AND CONCLUSION

This work shows the potential for combining machine learning and publicly available street-level imagery to automate the process of classifying Model Building Typologies for large-scale seismic risk assessment, using Oslo city (Norway) as a case study. Using a state-of-the-art Convolutional Neural Network (CNN) pre-trained on the ImageNet database, we developed a model that classifies typology in unseen images with 83% accuracy, using only data sources available online: the public cadastre and Google Street View. Supplemental high-quality images taken during field work were also used but are not required. Our workflow shows how seismic risk assessment can be highly automated and performed quickly without or limiting time-consuming and costly on-site surveys.

We observe that the classification accuracy varies with building typology: from 89% for timber and down to 35% for steel-reinforced concrete. The reason for this is twofold:

1. first, the distribution of the building mass in Oslo is heavily skewed towards timber buildings, with few steel or steel-reinforced concrete (SRC) buildings to use in the CNN training.
2. Secondly, SRC buildings are typically misclassified as reinforced concrete (CR), which is unfortunately also often the case in manual classification analysis. These misclassifications were verified by breaking down the separate typology predictions for an unseen test dataset by the true labels.

Misclassifications also occur between typologies that share the surrounding environment; in particular, we observe timber buildings surrounded by trees and fences being classified as "other", and vice versa. This indicates that the CNN to some extent incorporates the environment into the prediction and does not rely on the building properties alone. There exist several techniques (Simonyan et al., 2013; Samek et al., 2017; Selvaraju et al., 2017) to investigate this on a per-image basis, which should be considered in future studies.

The performance of the classifier can be improved by adding more labelled images, which is a matter of additional analyst time. Still, using a pre-trained model reduces the need for training data drastically (Yosinski et al., 2014), as well as lowering the computational cost compared to training a model from scratch. Our method is limited by the availability and good quality of GSV images, which do not offer complete geographical coverage and may be blocked by trees or vehicles. Hence, supplemental field work may be necessary for certain areas. Other online image providers can also complement GSV.

Given the success of this transfer learning approach, also demonstrated by Gonzalez et al. (2020), future work should investigate how well the CNN methodology generalizes and

can hence be applied to other Nordic cities, both with and without additional re-training. We expect that this should be possible at least within Norway, likely also further to Sweden and Finland with generally similar building stocks.

Until recently, the identification of building typologies and the development of a building stock model for seismic risk assessment were limited regarding their spatial coverage as well as financial resources and the lack of representative *in-situ* information. With the data and methodology presented in this paper, these limitations can now be overcome for many areas of the world.

Future work within this topic could include a semi-automatic pre-classification of certain neighborhoods. For instance, we could feed the CNN with pre-conditioning data about the type of the current neighborhood, e.g. residential, commercial, industrial area, which will likely increase the success rate. In addition, one could think about merging information from google-maps roof aspect-ratio with the façade information as done in our current work.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found below: <https://github.com/NorwegianSeismicArray/ML-for-building-typologies>.

AUTHOR CONTRIBUTIONS

FG is the main author of the article. She focused in the abstract, introduction, data, results, discussion and conclusion sections. She prepared all the pictures in the article. Her main contribution to the research work is represented by fieldwork in Oslo, taking manually and labelling all the pictures, to develop the different building typologies and to prepare and to analyse the results of the survey sent to the experts. SM, as second author, focused on the methodology, results, discussion and conclusion. He implemented and trained the CNN model, compared the different architectures, and prepared the classification results. AM helped to recognize and identify the different building typologies, to write the description of the typologies and to review the paper. VO helped in the writing and review processes. He discussed with the other authors regarding the best approach to develop the research and the article itself.

FUNDING

The results and the research showed in this article are connected to the GEObyIT project, funded by the Research Council of Norway (grant number 311596).

ACKNOWLEDGMENTS

The authors are grateful for the economic support provided by the Research Council of Norway.

REFERENCES

- Aravena Pelizari, P., Geiß, C., Aguirre, P., Santa María, H., Merino Peña, Y., and Taubenböck, H. (2021). Automated Building Characterization for Seismic Risk Assessment Using Street-Level Imagery and Deep Learning. *ISPRS J. Photogrammetry Remote Sens.* 180, 370–386. doi:10.1016/j.isprsjprs.2021.07.004
- Bungum, H., Pettenati, F., Schweitzer, J., Sirovich, L., and Faleide, J. I. (2009). The 23 October 1904 MS 5.4 Oslofjord Earthquake: Reanalysis Based on Macroseismic and Instrumental Data. *Bull. Seismol. Soc. Am.* 99, 2836–2854. doi:10.1785/0120080357
- Chollet, F. (2015). Keras, GitHub. Available at: <https://github.com/fchollet/keras>.
- Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. doi:10.1109/CVPR.2017.195
- Danciu, L., Nandan, S., Reyes, C., Basili, R., Weatherill, G., Beauval, C., et al. (2021). The 2020 Update of the European Seismic Hazard Model - ESHM20: Model Overview. *EFEHR Tech. Rep. 001 v1* 0.0, 1–121. doi:10.12686/a15
- Eriksson, I., Borchgrevink, J., Sæther, M. M., Daviknes, H. K., Adamou, S., and Andresen, L. (2016). TU1206 COST Sub-urban WG1 Report. Available at: <https://static1.squarespace.com/static/542bc753e4b0a87901dd6258/t/5707869ae707eb820b4118a5/1460111042910/TU1206-WG1-012+Oslo+City+Case+Study.pdf>.
- GFDRR (2018). Machine Learning for Disaster Risk Management. Available at: <http://doi.wiley.com/51>, doi:10.1111/j.1468-0394.1988.tb00341.x
- Gonzalez, D., Rueda-Plata, D., Acevedo, A. B., Duque, J. C., Ramos-Pollán, R., Betancourt, A., et al. (2020). Automatic Detection of Building Typology Using Deep Learning Methods on Street Level Images. *Build. Environ.* 177, 106805–106812. doi:10.1016/j.buildenv.2020.106805
- Grünthal, G. (1998). *European Macroseismic Scale 1998*.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep Residual Learning for Image Recognition. in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016 770–778. doi:10.1109/CVPR.2016.90
- He, K., Zhang, X., Ren, S., and Sun, J. (2016b). “Identity Mappings in Deep Residual Networks,” in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 630–645. doi:10.1007/978-3-319-46493-0_38
- Heaton, J., Bengio, Y., and Courville, A. (2018). *Deep Learning*. MIT Press. Cambridge, Massachusetts, United States, doi:10.1007/s10710-017-9314-z
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. doi:10.1109/CVPR.2017.243
- Kang, J., Körner, M., Wang, Y., Taubenböck, H., and Zhu, X. X. (2018). Building Instance Classification Using Street View Images. *ISPRS J. Photogrammetry Remote Sens.* 145, 44–59. doi:10.1016/j.isprsjprs.2018.02.006
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- NVIDIA (2020). NVIDIA V100 TENSOR CORE GPU. Available at: <https://images.nvidia.com/content/technologies/volta/pdf/volta-v100-datasheet-update-us-1165301-r5.pdf>.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115, 211–252. doi:10.1007/s11263-015-0816-y
- Samek, W., Wiegand, T., and Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. Available at: <http://arxiv.org/abs/1708.08296>.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. in Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 Oct. 2017 doi:10.1109/ICCV.2017.74
- Silva, V., Crowley, H., Pagani, M., Monelli, D., and Pinho, R. (2014). Development of the OpenQuake Engine, the Global Earthquake Model’s Open-Source Software for Seismic Risk Assessment. *Nat. Hazards* 72, 1409–1427. doi:10.1007/s11069-013-0618-x
- Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. Preprint at <http://arxiv.org/abs/1312.6034>.arXiv.org.
- Simonyan, K., and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. in 31st AAAI Conference on Artificial Intelligence, AAAI 2017.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. doi:10.1109/CVPR.2016.308
- Wallemacq, P., and House, R. (2018). Economic Losses, Poverty & Disasters: 1998–2017. *Nat. Resour.* Available at: https://www.researchgate.net/profile/Pascaline-Wallemacq-3/publication/331642958_Economic_Losses_Poverty_and_Disasters_1998-2017/links/5c859a6d92851c69506b1f0f/Economic-Losses-Poverty-and-Disasters-1998-2017.pdf?origin=publication_detail.
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). “How Transferable Are Features in Deep Neural Networks?” in *Advances in Neural Information Processing Systems*.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ghione, Mæland, Meslem and Oye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.