

FIRST STEPS TO AUTOMATED INTERIOR RECONSTRUCTION FROM SEMANTICALLY ENRICHED POINT CLOUDS AND IMAGERY

Lars Sören Obrock, Eberhard Gülch

University of Applied Sciences Stuttgart, Schellingstr. 24, 70174 Stuttgart, Germany
lars.obrock@hft-stuttgart.de, eberhard.guelch@hft-stuttgart.de

Commission II, WG II/4

Key Words: Building Information Modeling, BIM, Photogrammetry, Deep Learning, Semantic Segmentation, Object Extraction, Interior Reconstruction

ABSTRACT:

The automated generation of a BIM-Model from sensor data is a huge challenge for the modeling of existing buildings. Currently the measurements and analyses are time consuming, allow little automation and require expensive equipment. We do lack an automated acquisition of semantical information of objects in a building.

We are presenting first results of our approach based on imagery and derived products aiming at a more automated modeling of interior for a BIM building model. We examine the building parts and objects visible in the collected images using Deep Learning Methods based on Convolutional Neural Networks. For localization and classification of building parts we apply the FCN8s-Model for pixel-wise Semantic Segmentation. We, so far, reach a Pixel Accuracy of 77.2 % and a mean Intersection over Union of 44.2 %. We finally use the network for further reasoning on the images of the interior room. We combine the segmented images with the original images and use photogrammetric methods to produce a three-dimensional point cloud. We code the extracted object types as colours of the 3D-points. We thus are able to uniquely classify the points in three-dimensional space. We preliminary investigate a simple extraction method for colour and material of building parts. It is shown, that the combined images are very well suited to further extract more semantic information for the BIM-Model. With the presented methods we see a sound basis for further automation of acquisition and modeling of semantic and geometric information of interior rooms for a BIM-Model.

1. INTRODUCTION

Building Information Modeling (BIM) is currently one of the major topics of research and development in building industries and will greatly affect it by moving from two-dimensional paper maps to computerized three-dimensional building models including a comprehensive monitoring over the whole life span of buildings from planning to demolition. BIM follows a holistic approach, which requests intensive collaboration of all stakeholders during the life span of a building. The central element is a three-dimensional building model, which is the basis for all tasks of the stakeholders. It includes all geometrical and relevant semantic information of all parts of a construction. Currently the main application of BIM is in planning and new constructions. An application of BIM for already existing buildings on a large scale for whole quarters or even cities is so far neither on the horizon nor really feasible, due to mainly the high costs of manual acquisition of geometric and semantic information.

Currently there are well established methods available to reconstruct the geometry of buildings and interiors (Borrmann et al., 2015). There are modern techniques of surveying, like tachymetry or terrestrial laser scanning but also digital close-range photogrammetric systems used daily world-wide in all sorts of construction, mapping and cadaster. Inherent to all those methods is the high degree of manual work either in the acquisition of the raw data or in generating the requested information of a building geometry in form of objects from point clouds to describe the detailed geometry of walls, windows or doors. The methods are thus very time consuming, require well trained personnel, and the high-level equipment used is rather expensive.

The major drawback with all those methods lies in the fact, that currently the important semantic information has to be collected manually and integrated manually into the BIM-model.

Photogrammetry in combination with image analysis tools offers a very high potential for automatic acquisition of geometric and semantic information of existing buildings, mainly because of the extremely high depth of information content of images.

In our work we develop first steps to a higher automation for the digitization of existing buildings in BIM-models using modern mobile imaging sensors, image analysis and deep learning approaches.

As a basic data source, we do not only have to provide geometric information, but especially semantic information of a building as central themes of a BIM-model so as to describe all major features and attributes of buildings and building parts.

The first and most important step is the classification of the building part or object, as all other information, both geometric and semantic, are dependent on it.

We decided not to focus on the building as a whole but on its interior and a room as the core element for the first steps to automated reconstruction from semantically enriched point clouds and imagery.

In our work we focus on the development of a strategy including photogrammetry and image analysis tools while using images as a basic input, which does allow generation of point clouds, but also the application of classification tools. For this purpose, our core development is based on Deep Learning Methods – already state-of-the-art in Computer Vision since with AlexNet (Krizhevsky et al., 2012) in the year 2012 a Convolutional Neural Network (CNN) has been presented for the purpose of image classification and have since been expanded for purposes of object detection and segmentation in imagery. Especially the Semantic Segmentation based on specially adapted CNNs, allocating object categories on a pixel level, is well suited for a detailed selection and classification of objects in images and allows a smart continuous usage for the generation of semantic and geometric information.

The combination of pixel level classification with point clouds allows us to distinguish between various classified building parts and objects in their three dimensional positions, laying the foundation for future research.

Based on the segmented and classified images and point clouds of real world data, we are developing an approach to further examine the semantic and geometric content. These investigations are important to integrate the developed method(s) in a more automated workflow.

2. RELATED WORK

Building Information Modeling is one of the major topics in building industries. There is an EU-BIM manual published by (EU BIM Task Group, 2017) with recommendations how to introduce BIM in public administration. In Germany several guidelines, concepts and investigations were performed on the requested conditions for the introduction, the expected benefit and the application, e.g. (Egger et al. 2013) (Eschenbruch et al., 2014), (Kaden et al., 2017), (Borrmann et al., 2015) and (Bramann et al., 2015b).

The Federal Minister of Transport and Digital Infrastructure has introduced an obligatory graduated scheme for introducing BIM in Infrastructure Projects of public administrations (Bramann et al., 2015a). The focus is, however, on planning and construction of new buildings. Possible procedures for a later BIM-conformal modeling of a building are addressed in (Borrmann et al., 2015), (Clemen and Ehrich, 2014) and (Kaden et al., 2017) but are based on geodetic methods used for two-dimensional building modeling. But no adaption to the increasing demands of three-dimensional geometry and semantics are envisaged. This in consequence means higher time efforts for the measurement and their analysis. This work starts at this point and investigates first approaches for an automated analysis of image data based on deep learning methods to attack this problem.

Since 2012 Convolutional Neural Networks, as described in (Krizhevsky et al., 2012), are building the basis for the state-of-the-art in image classification. By various adaptations of the first described architectures and methods, current networks reached even more promising classification accuracies and opened the road for more detailed image analyses, see e.g. (Szegedy et al., 2014), (Simonyan and Zisserman, 2015), (He et al., 2015), (Xie et al., 2017) and (Haung et al., 2018).

Instead of classifying a whole image, object detection as described and improved by (Gershick et al., 2014), (Gershick, 2015) and (Ren et al., 2016) allows to select, localize and classify certain areas within a bounding box. Such an integration of a local component allows the exact assignment of several classifications to different areas in an image.

An even more precise localization is possible with a segmentation and classification on pixel level, also called Semantic Segmentation, as it is applied by (Shelhamer et al., 2016) on the basis of CNNs and named Fully Convolutional Network (FCN). For this network, the well approved architectures for classification as contained in (Simonyan and Zisserman, 2015), are adapted with Fully Convolutional Layers, and extended by a decoder section. In this section categories for each pixel of the input image are extracted using skip connections with information of previous layers. This allows an exact match of each pixel to a category. Since the introduction of FCN it served as a basis for several continued developments, which further improved the accuracies by changed network architectures and methods, e.g. (Jégou et al., 2017), (Zhao et al., 2017), (Chen et al., 2016) and (Chen et al., 2018).

For this work, we still used the FCN-architecture as it offers an ideal starting point for our developments.

Most investigations using deep learning methods are focusing on a great variety of very different objects. We do, however, want to focus for BIM on building parts only, located in a room. Thus most generally accepted training- and validation data sets are related to categories of very different origin. There are but few research papers that focus on the segmentation of an interior room, and mostly using a combination of images and depth information as in (Song and Xiao, 2014) and (Song and Xiao, 2016). Mostly the intention is to detect things and objects in a room, but do not address building parts specifically. The main purpose is very often to assist robotics to detect obstacles for navigation or to detect target objects.

3. METHODS

3.1 Deep Learning for BIM

The core element of BIM is a three-dimensional and object based model. In this model the building parts of a building are not defined by lines and points like in a CAD model, but as complete objects with their geometric and semantic features. The dominant semantic feature is the related type of the building part, resp. object category, which is the basis for the geometric as well as the semantic information content.

To acquire both types of information, digital photogrammetry is a perfect tool. It allows to model objects in 3D, but the collected images can also be used to identify the type of objects. The state-of-the-art for the requested image analysis are deep learning methods on the basis of Convolutional Neural Networks (CNNs). Originally developed to classify whole images, they have advanced very quickly to analyse different local parts of images. The networks consist of different layers, in which figures, called neurons, are located in matrices of varying depth. The connection between the layers is realized by moving kernels, in which a weighted sum for a neuron of the next layer is derived. In the training phase of the network, the weights for the connections are adapted by backpropagation by comparing ground-truth with actual results, in such a way, that in most cases a good conformity is reached. After the training phase, the trained model can be applied to new data.

If we want to have a unique match of object types in a point cloud, photogrammetry as a connector requires, that the object types have been localized and classified accurately in the images. We thus evaluate two different CNN methods with substantially different results.

The object detection, which allows a localization and classification using bounding boxes in images, is only partially feasible in our use case. The final transfer of the classified bounding box of the two-dimensional image to the 3D point cloud is in principle possible with the photogrammetrically determined camera position and attitude. However, the bounding box of a localized object in an image is not exactly tailored for it. Thus also other objects can be included in the bounding box, which would then be incorrectly classified. A one to one match is thus hardly possible. Further error sources are several objects partly occluding each other, which in turn means, that the bounding boxes therefore would overlap. The object detection is thus not accurate enough, to satisfactorily localize and classify object types.

The other branch of the classical CNN method, which provides a one to one segmentation and classification of each pixel, is called Semantic Segmentation. By using the pixel wise matching, we achieve the same fine resolution as the image itself and it is thus perfectly suited for localizing and classifying

object types. In addition, as we do have as a result a segmented mapping of the original image, we can photogrammetrically produce a classified point cloud requesting only little preparatory work.

3.2 Training Data Set

For the training of the neural network an image data set was created for the interior reconstruction of building parts with a first selection of object categories. Those categories contain the major objects and object parts in an interior and model it in a comprehensive and representative way. Out of the big number of possible categories our focus here was on the room forming object parts and the corresponding connecting object parts as well as some additional essential elements and objects of interest as given in Table 1.

Room forming object parts	Connecting object parts	Objects of interest	Additional objects
Floor	Doors	Switches and Sockets	Picture frame
Wall	Windows	Lamps	Open door
Ceiling			

Table 1. Used categories of major object parts, connecting object parts, objects of interest and additional objects.

As additional category of objects we selected „Picture frame“, due to often appearing on our test site and another category of a changed state of the element „door“, namely „open door“, because this might play a major role in where the door is implemented in the BIM-Model, so we wanted to see how the neural network is handling this category. In addition, it can be very decisive for applications in which way a door opens.

In our test area, however, the walls and ceilings were rather homogeneous and very similar, so there was no clear distinction possible, as it came out by our first experiments. In addition this had a negative effect on the classification of some of the other object parts, our main focus. We thus decided to currently exclude wall and ceiling from our object categories and thus also from the training of the CNN.

To be able to fine tune the pre-trained model with different categories, image data of these new classes had to be generated. We used a combination of own and external images to be able to provide a broad portfolio of possible variations of objects. To each image, the ground truth was labelled, in a way that each object class was annotated an ID as shown in Figure 2.



Figure 2. From the original image (left) derived ground truth image (right) with marked building parts in assigned colours containing unique IDs.

The data set contains 166 images. Among these were 50 % that were produced by vertical mirroring of the original images. The images contain all selected object categories in their natural environment. Even if this is not a very huge data set, the amount

of images allowed the training of a neural network by fine tuning of a pre-trained model. For the Semantic Segmentation of images a FCN, according to (Shelhamer et al., 2016), based on the architecture of FCN-8s was chosen. To this architecture the corresponding pre-trained model fcn8s-heavy was selected, which was trained on the basis of the PascalVOC data set. In terms of transfer-learning, a fine-tuning of the model was performed during the training based on the previously generated data set. The gradient descent algorithm ADADELTA was applied. Based on it a Base Learning Rate with a value of 0.22 in connection with the Learning Rate Policy of „Exponential Decay“ and a Gamma of 0.9 provided the best results.

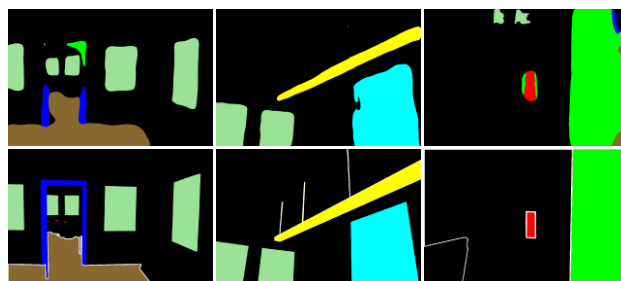


Figure 3. Selection of ground truth images at the bottom with the FCN segmented images on the top.

3.3 Classified Points in a Point Cloud

To generate a three-dimensional, object based building model in terms of BIM, the exact form, position, extension and orientation of an object is needed. Using photogrammetry we first need to generate a three-dimensional point cloud from the images. We focus in this research on an interior room as a core element of a building. Therefore we acquired a huge number of images for the later derivation of a photogrammetric point cloud. By using the above trained model, the objects contained in these images were segmented and classified (Figure 6). A direct derivation of the point cloud using the segmented images was not possible, as only the object type information is coded in the grey/colour values. The direct segmentation on pixel level, however, allows a matching of a classified pixel of the segmented image to the pixel at the same location in the original image. Beforehand index values are attached to the categories in the segmented images in such a way, that they have a large distance among them. For the seven categories of building parts a distance of 32 grey values was selected. The non-classified background areas were annotated with grey values of 0 and 255 each once for every image.



Figure 4. Combined images as a replacement of one of the RGB channels. Left: Replacement of the Blue channel. Middle: replacement of the Green channel. Right: Replacement of the Red channel. Background value on the top is 255 and 0 on the bottom.

These values are used for linking the object IDs in the segmented images with the original images and thus identifying the object type of a pixel. For this linking we selected two different variants. We on the one side used the segmented image as a replacement of one colour channel (2 colour channels and 1 Category-ID channel) and on the other side we added a 4th channel (3 colour channels and 1 Category-ID channel).

The replacement of one of the RGB channels of the original images produces varying combinations of the colour channels and thus, because of the two possible background IDs, six differently coloured variants of the images as seen in Figure 4. We might have lost the colour information of a channel, but due to the unique colour values of each category in the indexed channel we still can clearly distinguish the results visually. We can observe, that the bigger the difference to the other values, and especially to the non-classified areas, the higher is the contrast of the extracted category and the better its visibility.

If we use the segmented image as a 4th colour channel to the original image, it behaves visually like a transparency. The higher the value, the less the category is visible. On the other hand, we still have the full colour information of the original image available. As we do not focus on the visibility to a human user, but on the loss-free combination of all information, this variant offers the best basis for all further steps. However, the 4th channel could not be introduced to the Agisoft Photoscan (Agisoft, 2018) Software for the photogrammetric point cloud generation. We thus had to go back to variant 1 with the combined three channel images, thus losing colour and eventually also depth information. But, in the resulting point cloud the classified object types are three dimensionally localized in the points and their colours (Figure 8).

3.4 Additional Semantic Information

The produced 4 channel images can be used to additionally extract semantic information. These are essential parts of BIM besides the geometry and the building parts categories. Currently the additional semantic information is recorded manually during acquisition. The 4 channel images combine the original colour information and the detected object categories. We can use a simple assignment to derive colour and eventually even the material of single categories of an image. From the linked colour values to the categories for each image the median of a class was calculated and gross errors eliminated by using the standard deviation as decision criteria. From the median values we used the Euclidian distance in the three-dimensional colour space to determine the shortest distance to defined colour values and finally link each category to the related colour and material. We thus could with the first results show, that an extraction of semantic information of single object parts seems feasible using the categorized imagery.

4. EXPERIMENTAL RESULTS AND EVALUATION

4.1 Deep Learning Performance

The evaluation of the accuracy of the trained FCN-Model for Semantic Segmentations was performed using a randomly selected set of 50 images. It has to be noted, that a bigger part of these images had been used for the training, which means, that the degree of generalisation and an eventually occurring overfitting could not be determined. However, the results clearly can describe the quality of the segmentation and classification of building parts in our use case.

After optimization of the hyper parameters of the training the trained model produced good segmentation results.

By evaluating the Pixel Accuracy a value of 77.2 % could be reached. This shows a quite high consensus of the pixels. We have to note, that this value is weighted proportional to the area of a class. It can be stated, that categories covering big areas, are to a high degree classified correctly. To have a better comparison for the Semantic Segmentation, the method of "mean Intersection over Union" (mIoU) is generally used, with each class in an image having the same weight. For the mIoU a value of 44.2 % was obtained, which indicates a rather good consensus between segmented images and ground-truth images, despite the fine-tuning of the model using only so few images.

As all classes have the same weight, classes with a small area have a strong influence on the mIoU value. With small areas of wrongly classified or to be classified pixels strong variations of the mIoU are possible. Looking at the average IoU of the single classes, we can see, that especially those have low values which have a high percentage of mis-classifications or non-classified pixels. This is visible in Table 5, which shows the classes „Switches and Sockets“ with 54 % and especially the class „Open Door“ with 74 % are very often completely misclassified in the images. The average IoUs of these classes are very poor with 17.2 % and 11.4 % respectively. In the other classes we can see much less errors of that type and the class "Floor" shows with 72.9 % the best average IoU.

Category	average IoU (%)	complete mis-classifications (%)
Lamps	37.4	39.3
Switches and Sockets	17.2	54.3
Windows	55.3	35.0
Doors	49.4	34.4
Open door	11.4	73.9
Floor	72.9	12.5
Picture frame	50.6	35.1

Table 5. Average IoU of each class, compared with the percentage of complete mis-classifications of each class. A high percentage of complete mis-classifications at the same time means a worse average IoU for the class.



Figure 6. Comparison between a selection of the segmentation results and the related original images. Most of the segmentation results are very good. Some problems like incorrect classification on mostly the edges, wavy lines and rounded corners, missing objects and errors determining the state of the door can occur.

To be able to validate the approach a series of images of an interior was made. They were analyzed by the completely trained FCN to segment the given categories on pixel level.

By visually comparing the resulting images with the originals in Figure 6 it can be seen, that in most cases the detection of objects has been very well. But looking a bit closer it can be also stated, that for an object not always and exclusively the correct category was selected. Especially at the border between two objects or between an object and the background an incorrect selection of categories can occur. The extracted pixel areas are only coarsely matching the exact form of the objects. Especially at the edges of the objects we observe a wave form instead of a straight line, or we do see rounded corners. The differences along edges between categories might be rather low, which produces problems for the FCN-model to distinguish between categories.

Some of the objects were not recognized. This has happened for small areas and especially for the categories "Switches and Sockets", "Lamps" and "Windows", if they cover only a small area in the original images.

As an exception a category was recognized, even if there was no object in the original image visible. In these cases there was often a consensus in colour, shape or surface material with the respective category.

We could also observe problems to distinguish between open and closed door. In many cases an "open door" as well as closed "door" was recognized, but not always the correct state was observed. This was visible with open doors, as the wrongly classified "door" areas cover a big part of the area to be segmented. With closed doors, the wrongly classified "open door" areas cover often only a very small area. As a result, the distinction between the two conditions of a door produced mainly confusion and less profit.

The results are so far very promising, but definitely more training data and/or improved network architecture are needed to further reduce the amount of misclassifications and most of the observed problems. The segmentation itself produced rather good results, given the fact, that we used a fine-tuning of a previously trained model. Most of the objects, and especially the big ones, are recognized.

4.2 Image Processing and Photogrammetric Point Cloud Generation

In the combined images the bigger and well segmented areas that have a high ID are visually good to be distinguished, if the non-classified background has a value of 0 thus showing a high contrast as visible in Figure 7.



Figure 7. Combination of original images and the segmented images with a background value of 0. From left to right: As replacement of the Blue channel, as replacement of the Red channel, as replacement of the Green channel, as additional fourth channel acting as transparency

Using the interior images with the different replaced colour channels, the best functioning colour and category combination was selected to generate the point cloud. By using the combined images, where the category channel replaced the Blue channel, the best and most complete point cloud was produced.

No manual editing of the point cloud was applied to not influence the analysis. The point cloud modeled the room very good, except some few occlude areas. Important for the later extraction of geometries is the fact, that the by the neural network extracted categories in the images are still very well visible in the point cloud, and thus contain the category information as visible in Figure 8.

Especially the categories "Floor", "Picture frame" and "Door" are visually clearly recognizable. The other categories as "Window", "Switches and Sockets" and "Lamps" are less easy to recognize in the point cloud. They do, however, contain a low colour-value in the blue channel and thus a low contrast.

Looking at the colour information of the point cloud, we could observe, that all classes can be distinguished in the Blue channel. But also intermediate values are visible, probably caused by interpolation especially along borders.



Figure 8. Excerpts from the generated point cloud using the three channel images with replaced Blue channel. It becomes clear, that the extracted categories are very well visible in the point cloud. However, in the bottom right image there is an error with the floor

As a certain disadvantage it was also noted, that by missing the colour information of one channel minor deficiencies in the point cloud occurred. We do observe some gross errors sparsely scattered on not uniquely classified, larger, unstructured areas. They are, however, not decisive for the description of the structure of the interior room, and can be cleaned in a later processing stage.

We do also observe gaps at some places of the point cloud, even if these areas were well visible in the images. Very often large, homogenous areas hinder a good multi-image matching, which is even more affected by the missing colour channel. The biggest problem occurred on a small area of the floor which was imaged only by few images taken from similar directions.

To compensate the missing Blue channel, additional images can be taken, thus producing a more detailed point cloud. However, the point cloud provides extremely valuable information on which object is located in which area of a room.

4.3 Colour and Material as Additional Semantic Information

A simple assignment was applied to derive colour and material information of single categories of an image. A first investigation of the results was based on a manual assignment of ground-truth values of material for the classified part of an

image. The material can, however, also change within a category. Automatically assigned materials are also dependent from the assigned colour. Also here the manual annotation of ground-truth might influence the results.

For directly matching extracted colour values to a colour name a standardised look-up table is used, which is suitable, as long as the areas of the different categories are not consisting of different colour areas.

When assigning materials it becomes clear, that the linking via colour values is not as simple. This is mainly because there doesn't have to be a unique connection between colours and material, i.e. there might be materials with very similar colour or one material which can have many different colours.

Yet, even by using such a simple first approach, a decently good consensus of already 34.9 % has been reached. We do observe some problems especially caused by not excluding incorrectly classified categories. Also transitions of materials like from parquet to stone floor can produce problems which should be tackled in further developments, ranging from an improvement of the neural network for the Semantic Segmentation, an improvement and extension of the used algorithm or an alternative algorithm as well as the usage of deep learning methods also for the extraction of additional semantic information.

5. CONCLUSIONS

A new approach has been presented which shows a sound and promising basis to extract semantic and geometric information for BIM with a potential for increased automation. We can see that using images of digital cameras, tablets or even smartphones, i.e. rather low-cost devices can be suited for these tasks.

We so far reach a Pixel Accuracy of 77.2 % and a mIoU of 44.2 %. It is shown, that Deep Learning Methods in combination with photogrammetric procedures are well suited to recognize object types and their position in images and point clouds as major input source for a BIM-Model. We have so far applied rather simple approaches to evaluate the overall potential. We successfully started first developments to recognize color and material of building parts.

We have identified current problem areas and bottle necks and given hints on how to improve the performance. We can see the potential of an improved model-architecture to get a higher quality of the deep learning results. We do request more images in the process and have to increase the categories to a realistic high number for various building types.

With the presented methods we see a sound basis for further automation of acquisition and modelling of semantic and geometric information of interior rooms for a BIM-Model.

6. ACKNOWLEDGEMENTS

The project „i_city: BIM-konforme Gebäudeerfassung“ is sponsored by the Federal Ministry of Education and Research (BMBF) under the promotion code 13FH9E01IA and supervised by the project executing organization VDI Technologiezentrum GmbH for the BMBF. The responsibility for the content lies with the authors.

7. REFERENCES

Agisoft, 2018. Agisoft Photoscan. <http://www.agisoft.com/> (05 April 2018).

Borrmann, A., König, M. Koch, C., Beetz, J., 2015. *Building Information Modeling - Technologische Grundlagen und industrielle Praxis*. VDI-Buch, Springer Vieweg, Wiesbaden, pp. 343-361, ISBN 978-3-658-05603-3.

Bramann, H., May, I. and planen-bauen 4.0 – Gesellschaft zur Digitalisierung des Planens, Bauens und Betreibens mbH, 2015a. Stufenplan Digitales Planen und Bauen. https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufenplan-digitales-bauen.pdf?__blob=publicationFile (06 November 2017).

Bramann, H., May, I. and planen-bauen 4.0 – Gesellschaft zur Digitalisierung des Planens, Bauens und Betreibens mbH, 2015b. Konzept zur schrittweisen Einführung moderner, IT-gestützter Prozesse und Technologien bei Planung, Bau und Betrieb von Bauwerken – Stufenplan zur Einführung von BIM. https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-stufenplan-endbericht.pdf?__blob=publicationFile (06 November 2017).

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2016. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. <https://arxiv.org/pdf/1412.7062.pdf> (03 April 2018).

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. <https://arxiv.org/pdf/1802.02611v1.pdf> (03 April 2018).

Clemen, C., Ehrlich, R., 2014. Geodesy goes BIM. In: *allgemeine vermessungs-nachrichten. (avn) 121 (6)*. <https://gispoint.de/artikelarchiv/avn/2014/avn-ausgabe-62014/2552-geodesy-goes-bim.html> (20 December 2017).

Egger, M., Hausknecht, K., Liebich, T., Przybylo, J., 2013. BIM-Leitfaden für Deutschland. http://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-leitfaden-deu.pdf?__blob=publicationFile (20 December 2017).

Eschenbruch, K., Malkwitz, A., Grüner, J., Poloczek, A., Karl, C. K., 2014. Maßnahmenkatalog zur Nutzung von BIM in der öffentlichen Bauverwaltung unter Berücksichtigung der rechtlichen und ordnungspolitischen Rahmenbedingungen – Gutachten zur BIM-Umsetzung. https://www.bmvi.de/SharedDocs/DE/Anlage/Digitales/bim-massnahmenkatalog.pdf?__blob=publicationFile (06 November 2017).

EU BIM Task Group, 2017. Handbook for the introduction of Building Information Modelling by the European Public Sector – Strategic action for construction sector performance: driving value, innovation and growth. http://www.eubim.eu/wp-content/uploads/2017/07/EUBIM_Handbook_Web_Optimized-1.pdf (06 November 2017).

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. <https://arxiv.org/pdf/1311.2524v5.pdf> (19 December 2017).

Girshick, R., 2015. Fast R-CNN. <https://arxiv.org/pdf/1504.08083.pdf> (03 April 2018).

He, K., Zhang, X., Ren, S., Su, J., 2015. Deep Residual Learning for Image Recognition. <https://arxiv.org/pdf/1512.03385.pdf> (03 April 2018).

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2018. Densely Connected Convolutional Networks. <https://arxiv.org/pdf/1608.06993.pdf> (03 April 2018).

Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. <https://arxiv.org/pdf/1611.09326.pdf> (03 April 2018).

Kaden, R., Clemen, C., Seuß, R., Blankenbach, J., Becker, R., Eichhorn, A., Donaubaue, A., Kolbe, T. H., Guber, U., DVW – Gesellschaft für Geodäsie, Geoinformation und Landmanagement e. V., Runder Tisch GIS e.V., 2017. Leitfaden Geodäsie und BIM. https://rundertischgis.de/images/2_publicationen/leitfaeden/Geo_undBIM/Leitfaden%20Geod%C3%A4sie%20und%20BIM_Onl_ineversion.pdf (04 November 2017).

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (19 December 2017).

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. <https://arxiv.org/pdf/1506.01497.pdf> (03 April 2018).

Shelhamer, E., Long, J., Darrell, T., 2016. Fully Convolutional Networks for Semantic Segmentation. <https://arxiv.org/pdf/1605.06211.pdf> (21 December 2017).

Simonyan, K., Zisserman, A., 2015. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. <https://arxiv.org/pdf/1409.1556.pdf> (03 April 2018).

Song, S., Xiao, J., 2014. Sliding Shapes for 3D Object Detection in Depth Images. In: *Proceedings of the 13th European Conference on Computer Vision (ECCV2014)*. <http://slidingshapes.cs.princeton.edu/paper.pdf> (05 April 2018).

Song, S., Xiao, J., 2016: Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images. In: *Proceedings of 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016)*. <http://dss.cs.princeton.edu/paper.pdf> (05 April 2018).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going deeper with convolutions. <https://arxiv.org/pdf/1409.4842.pdf> (03 April 2018).

Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks. <https://arxiv.org/pdf/1611.05431.pdf> (03 April 2018).

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network. <https://arxiv.org/pdf/1612.01105.pdf> (03 April 2018).