Semantic Segmentation of Images of Building Facilities

Thomas Czerniawski, M.S., PhD student, Fernanda Leite, Ph.D., P.E., Associate Professor
The University of Texas at Austin, USA
tczernia@utexas.edu

Abstract. Scan-to-BIM is the process of converting a 3D reconstruction into a building information model (BIM). The process has two parts: (1) sorting subsets of the reconstruction into classes (semantic segmentation) defined by a BIM taxonomy and (2) identifying geometric parameters describing each class instance. Here we demonstrate the ability of deep learning artificial neural networks to semantically segment images of building facilities. We found this deep learning approach capable of simultaneously recognizing: ceiling, wall, plumbing, duct, door, floor, and stairs classes. This semantic scope surpasses state-of-the-art building system recognition methods and represents progress towards comprehensive BIM creation.

1. Introduction

Building information models (BIM) enable simulation, automation, and information sharing. However, building owners do not invest in creating and maintaining BIMs for most of their existing building facilities (Edirisinghe et al., 2016, Shen et al., 2016, Mayo et al., 2012, Giel and Issa, 2015). In order to make the adoption of BIM easier, researchers have been automating parts of BIM creation by applying computer vision (Lee and Lu, 2017, Fathi et al., 2015, Pătrăucean et al., 2015, Volk et al., 2014, Xiong et al., 2013, Musialski et al., 2013, Huber et al., 2011, Tang et al., 2010, Brenner, 2005). The process begins by digitizing existing facilities using reality capture technologies including laser scanners and range cameras. Information from the digital reconstructions is used to automatically generate objects in a BIM as defined by an object class label (e.g. wall, door, window, and column) and parameters (e.g. height, elevation, and relationships). This technical challenge has two parts.

(1) Sorting subsets of the spatial data into classes defined by a BIM taxonomy.
(2) Identifying geometric parameters describing each class instance.

Associating subsets of the input raw spatial data with a BIM taxonomy creates semantic interoperability between the raw spatial data and the computational/algorithmic design software generating the BIM. For this purpose, object recognition algorithms are used. Published work will typically focus on recognizing one type of object category such as plumbing (Ahmed et al., 2014), partition walls (Hamledari et al., 2017), building facades (Oskouie et al., 2017), or a limited number of categories together such as walls, floor, doors, and windows (Quijano and Prieto, 2016, Bassier et al., 2016, Ochmann et al., 2016, Anagnostopoulos et al., 2016, Mura et al., 2016). Researchers have yet to combine and scale existing methods to achieve comprehensive BIM creation because the diversity of objects and systems encountered in buildings far exceeds the scope of existing recognition models.

Emerging deep learning methods demonstrate a level of versatility that has the potential to greatly increase the scope of building system focused recognition. This paper presents a deep learning based object recognition method that successfully recognizes: ceiling, wall, plumbing, duct, door, floor, and stairs classes. This collection of building classes surpasses in size and diversity the state-of-the-art in the building systems literature.

2. Related Work

Several review articles covering the automated generation of BIMs have been published (Lee and Lu, 2017, Fathi et al., 2015, Pătrăucean et al., 2015, Volk et al., 2014, Xiong et al., 2013, Musialski et al., 2013, Huber et al., 2011, Tang et al., 2010, Brenner, 2005). Deep learning approaches are noticeably absent. Research in deep learning applied to computer vision has experienced success processing both 2D and 3D data (Krizhevsky et al., 2012, Qi et al., 2016a, Dai et al., 2017, Qi et al., 2016b). Although datasets used to train these emerging deep learning algorithms are not specifically building systems focused, there is some overlap. For example, of the 20 classes involved in the ScanNet Benchmark Challenge[1], the relevant to BIM include: door, floor, wall, and window. Despite this lack of BIM taxonomy coverage, these other datasets enable transfer learning (Pratt, 1993, Pan and Yang, 2010). Transfer learning is the process of reusing a model trained for one task as the starting point for a model on a second task. Deep learning requires algorithms to be trained on these large annotated datasets, but then they can be fine-tuned on building system focused specialty datasets such as 3DFacilities (Figure 1) (Czerniawski and Leite, 2018).

Figure 1: Example of data provided in the 3DFacilities dataset; 3D reconstruction with color texture (scan); 3D reconstruction annotated with instance-level category labels (scan); 2D image color (frame); 2D image depth (frame); 2D image annotated with instance- level category labels (frame)

[1] http://kaldir.vc.in.tum.de/scannet_benchmark/

3. Methodology

Here we implement a deep learning artificial neural network as a supervised machine learning algorithm. In addition to the network architecture, the implementation required training data and computational resources.

3.1 Training Data

Since its initial submission to EG-ICE in 2018 (Czerniawski and Leite, 2018), 3DFacilities (Figure 1) has grown to over 25,000 RGB-D frames and 110 3D reconstructions. Each individual RGB-D frame and each reconstruction have an associated annotation file where each pixel and vertex, respectively, have been categorized into one of 19 different building element categories. The instance count distribution in Figure 2 counts the number of physical building components represented in the dataset. 3DFacilities is a relatively small specialty dataset. Therefore, we make use of two additional datasets, MS-COCO (Lin et al., 2014) and VOC 2012 train_aug + trainval (image segmentation datasets) (Everingham et al., 2010) and transfer learning as described in Section 3.3.

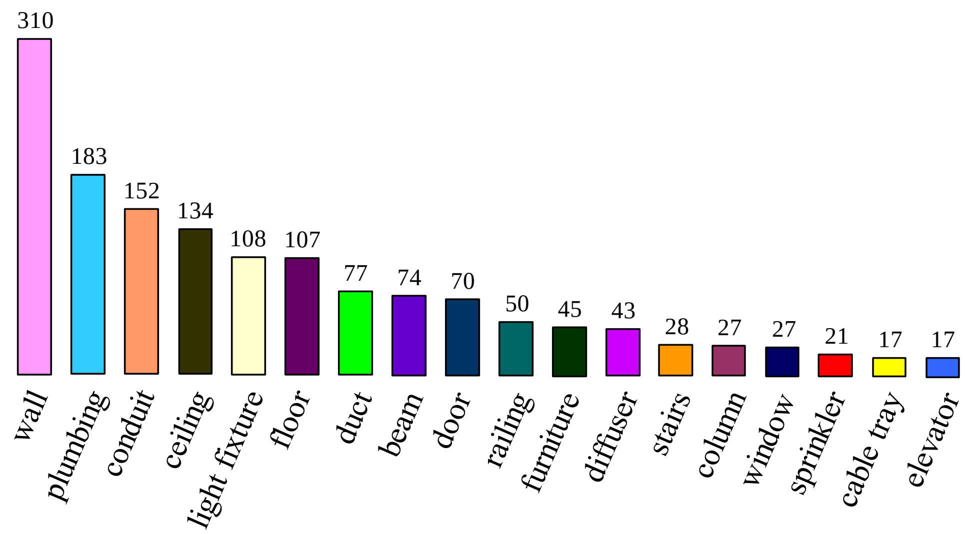

Figure 2: Instance count for each of the 18 categories in 3DFacilities

3.2 Programming Environment

The computing demand of training deep neural networks necessitates the use of high-performance computing. The Texas Advanced Computing Center (TACC) designs and operates some of the world's most powerful computing resources. Stampede2 is the flagship supercomputer at The University of Texas at Austin's TACC and provides high-performance computing capabilities to thousands of researchers across the U.S. It entered full production in Fall 2017 as an 18 petaflop system (TACC, 2018). The neural network is trained using compute nodes on Stampede2.

Although it is possible to code deep learning algorithms from scratch, there are several open-source libraries available. Tensorflow is an open source software library for high-performance numerical computation. It is written in Python, C++, and CUDA. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning (Abadi et al., 2016). Tensorflow was chosen as the development environment because of its supportive online community and its availability on TACC resources.

3.3 Semantic Image Segmentation

The specific neural network architecture we used is called DeepLab (Chen et al., 2014, Chen et al., 2018a, Chen et al., 2017, Chen et al., 2018b). DeepLab is a state-of-the-art deep learning model for semantic image segmentation, where the goal is to assign semantic labels to every pixel in the input image. DeepLab, along with its variations, currently performs amongst the top performing algorithms on the PASCAL VOC Challenge leaderboard (Everingham et al., 2018). PASCAL VOC is a standard recognition dataset and benchmark with detection and semantic segmentation challenges (Everingham et al., 2010). DeepLab is an especially attractive option for semantic segmentation because of its availability in the Tensorflow research model repository.

Data Preprocessing

Annotations from 3DFacilities must be converted into 8-bit images where the value of each pixel corresponds to the pixel's class. There is no input image size requirement as the network architecture is fully convolutional.

3DFacilities was split into a training set and a validation set using an algorithm that optimized for similarity between the pixel class distributions of both sets. This is so the neural network is validated on a dataset that has the same distribution as the dataset it is trained on. In an effort to reduce data dependence between the training and validation sets, sequences of 20 frames were used as inseparable units.

Neural Network Training

3DFacilities is a relatively small training dataset for deep learning. It is, nonetheless, possible to use the dataset to successfully train neural networks because of a technique called transfer learning. Transfer learning is a machine learning method where a model trained on one dataset is reused as the starting point for a model fine-tuned on a second dataset. This is a common approach given the vast compute and time resources required to develop datasets and neural network models.

The DeepLab implementation used by this research uses as its foundation MobilenetV2 (Sandler et al., 2018) that has been trained on ImageNet (Russakovsky et al., 2015) (an image classification dataset). The entire DeepLab implementation was then pre-trained on MS-COCO (Lin et al., 2014) and VOC 2012 train_aug + trainval (image segmentation datasets) (Everingham et al., 2010). The final layer of the neural network was replaced in order to account for the different semantic classes of 3DFacilities. The neural network was fine-tuned on 3DFacilities using the Tensorflow Momentum Optimizer for nearly 75 epochs at a learning rate of 0.0001 (Figure 3). For comparison, the learning rate used to pre-train on PASCAL VOC was 0.007. Total loss is a measure of classification error and is calculated using softmax cross-entropy.

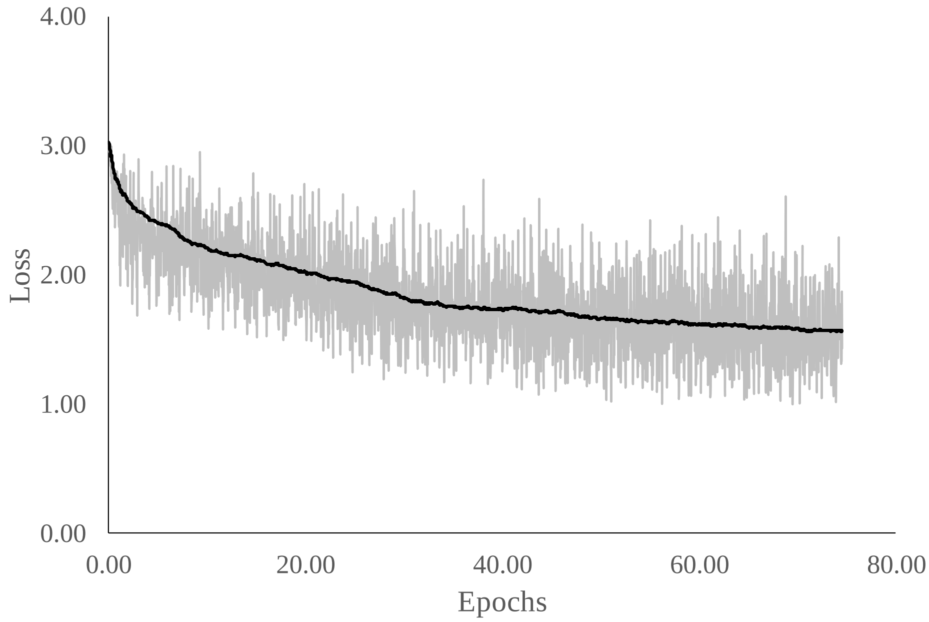

Figure 3: Deeplab fine-tuning on 3DFacilities, total loss vs. epochs

4. Results and Discussion

4.1 Semantic Image Segmentation

Example segmentations created by the DeepLab network can be seen in Figure 5. The neural network limits its pixel classifications to the following seven classes: ceiling, wall, plumbing, duct, door, floor, and stairs. It does not assign the following 12 classes: furniture, window, column, beam, railing, light fixture, elevator, diffuser, fire sprinkler, cable tray, conduit, and background. Comparing these results to the pixel class distribution in Figure 4, it is apparent the neural network has a bias for classes that occur more frequently in the training dataset and ignores those classes that occur less frequently.

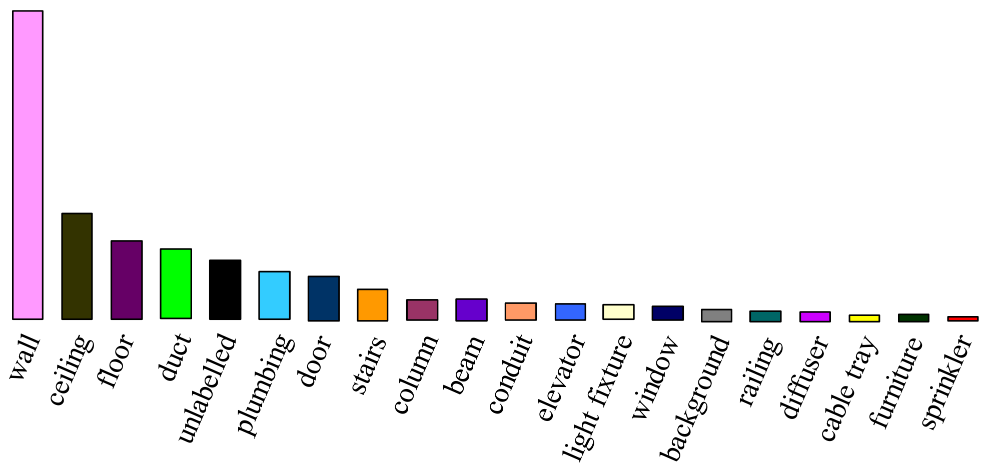

Figure 4: Pixel class distribution of 3DFacilities, horizontal axis is individual frames of 3D Facilities, and vertical axis is percentage of pixels in frame per class

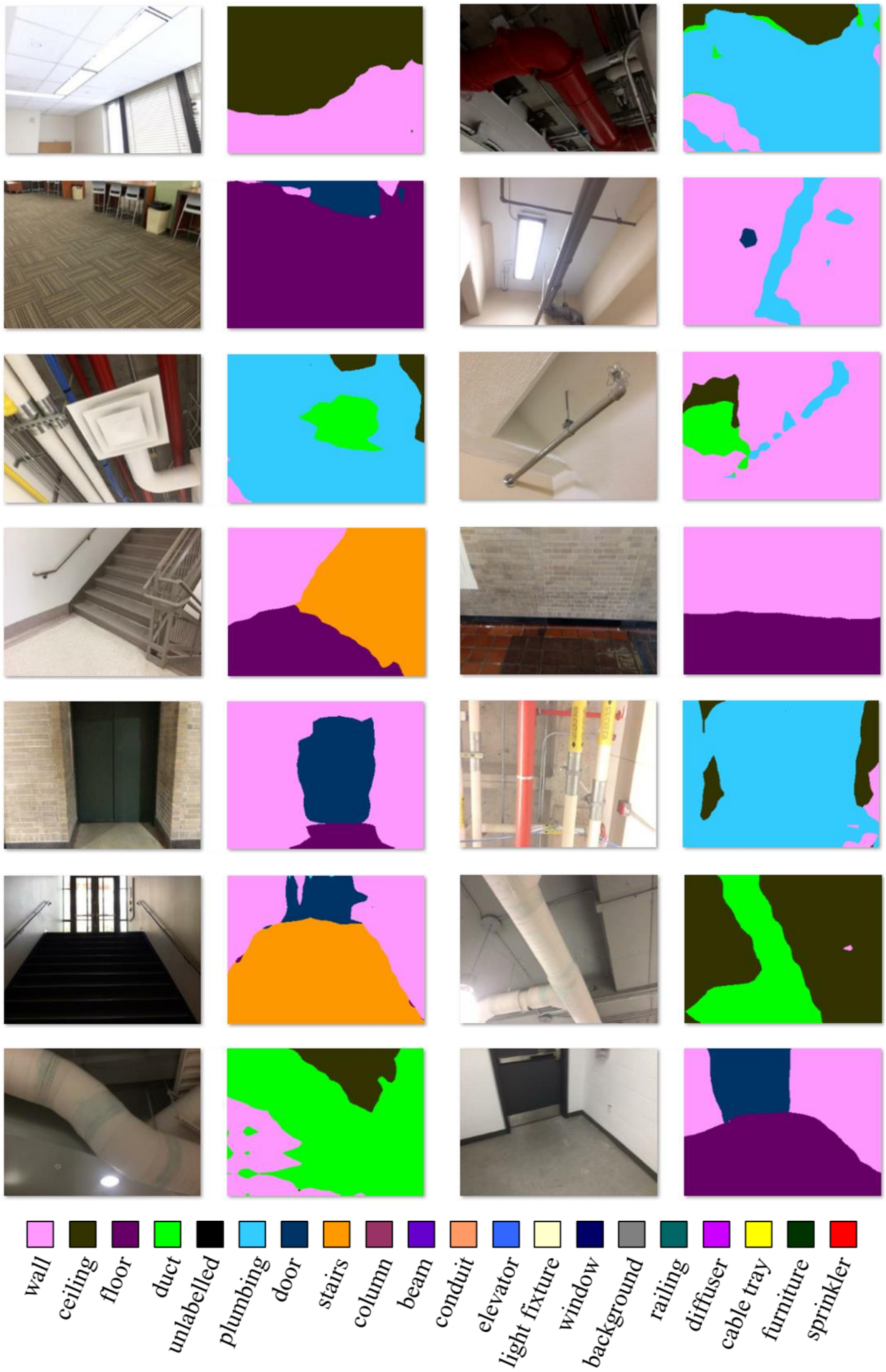

Figure 5: Cherry-picked results of semantically segmented images from the validation set

The seven classes provide partial semantic coverage of the BIM taxonomy as described in Tables 1 & 2 using Uniformat and Revit Families. Uniformat is a standard for classifying construction information in the U.S. and Canada endorsed by the American Society for Testing and Materials (ASTM). It classifies information based on functional elements or parts of a facility characterized by their function. Autodesk Revit is a prominent building information modeling software for people in the building industry. Components in Revit are categorized into families, which are groups of elements with a common set of properties.

Table 1: Top three levels of the UNIFORMAT classification system associated with Semantic Class Outputs from Neural Network

Level 1 **Major Group Elements**	**Level 2** **Group Elements**	**Level 3** **Individual Elements**
B SHELL	B10 Superstructure	B1010 Floor Construction
C INTERIOR	C10 Interior Construction	C1010 Indoor partition C1020 Interior Doors
	C20 Stairs	C20 Stairs
	C30 Interior Finishes	C3030 Ceiling Finishes
D SERVICES	D20 Plumbing	D20 Plumbing
	D30 HVAC	D3040 Distribution Systems

Table 2: Example Standard Families in Autodesk Revit Associated with Semantic Class Outputs from Neural Network

Included System Families		
• Ceilings • Ducts	• Floors • Pipes	• Stairs • Walls
Loadable Families		
Architectural Families	**Structural Families**	**MEP Systems Families**
• Doors • Furniture • Railings • Windows	• Columns • Framing • Trusses	• Conduit • Duct • Fire Protection • Lighting • Plumbing

4.2 Conclusions and Future Work

Scan-to-BIM involves segmenting 3D reconstructions into parts to conform to a BIM taxonomy. The set of seven classes successfully segmented by the neural network presented in this paper represents a substantial increase in semantic scope for recognition methods as compared to the state-of-the-art in the building systems literature.

Class Balancing

The neural network has a tendency to assign classes which are most dominant in the 3DFacilities dataset. In an effort to encourage the neural network to assign classes which appear less frequently in the 3DFacilities dataset, a modified loss function and a variable data feeder will be tested. Final image segmentation performance will be evaluated using standard metrics (mean IoU) (Shelhamer et al., 2017).

Incorporating Additional Data Channels

In an effort to improve semantic segmentation further, two additional inputs should be introduced to the segmentation system: a depth channel and inertial measurement unit (IMU) data. Since these two input types are relatively uncommon as compared to RGB input channels, incorporating them will be a challenge.

Uncertainties include: modifying the DeepLab architecture and cross-modality (Gupta et al., 2016) "warm-starting" with pre-trained parameters from standard DeepLab. The Slim Tensorflow interface provides functions that can be used to "warm start" training algorithms by using pieces of pre-existing model checkpoints[2].

Identifying Geometric Parameters for Class Instances

Combining segmentation results for the 2D data in order to segment the 3D reconstruction can be performed using a multi-view classification process (Pham et al., 2018, Qi et al., 2016b). Then work can begin on identifying geometric parameters describing each class instance.

References

ABADI, M. ı., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G., ISARD, M., KUDLUR, M., LEVENBERG, J., MONGA, R., MOORE, S., MURRAY, D. G., STEINER, B., TUCKER, P., VASUDEVAN, V., WARDEN, P., WICKE, M., YU, Y. & ZHENG, X. 2016. Tensorflow: a system for large-scale machine learning. *12th USENIX symposium on operating systems design and implmentation.* Savannah, GA.

AHMED, M. F., HAAS, C. T. & HAAS, R. 2014. Automatic detection of cylindrical objects in built facilities. *Journal of Computing in Civil Engineering,* 28**,** 04014009.

ANAGNOSTOPOULOS, I., PĂTRĂUCEAN, V., BRILAKIS, I. & VELA, P. Detection of walls, floors, and ceilings in point cloud data. Construction Research Congress 2016, 2016. 2302-2311.

BASSIER, M., VERGAUWEN, M. & VAN GENECHTEN, B. Automated semantic labelling of 3D vector models for scan-to-BIM. Proceedings of the 4th Annual International Conference on Architecture and Civil Engineering (ACE2016), 2016. 93-100.

BRENNER, C. 2005. Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation,* 6**,** 187-198.

CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K. & YUILLE, A. L. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. *International Conference on Learning Representations (ICLR).* San Diego, CA.

[2] https://github.com/tensorflow/tensorflow/tree/master/tensorflow/contrib/slim

CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K. & YUILLE, A. L. 2018a. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 40**,** 834-848.

CHEN, L.-C., PAPANDREOU, G., SCHROFF, F. & ADAM, H. 2017. Rethinking atrous convolution for semantic image segmentation.

CHEN, L.-C., ZHU, Y., PAPANDREOU, G., SCHROFF, F. & ADAM, H. 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation.

CZERNIAWSKI, T. & LEITE, F. 3DFacilities: annotated 3D reconstructions of building facilities. Workshop of the European Group for Intelligent Computing in Engineering, 2018. Springer, 186-200.

DAI, A., CHANG, A. X., SAVVA, M., HALBER, M., FUNKHOUSER, T. A. & NIEßNER, M. ScanNet: Richly-Annotated 3D Reconstructions of Indoor Scenes. CVPR, 2017. 10.

EDIRISINGHE, R., KALUTARA, P. & LONDON, K. An investigation of BIM adoption of owners and facility managers in Australia: institutional case study. COBRA 2016, 2016. Royal Institution of Chartered Surveyors, 1-10.

EVERINGHAM, M., GOOL, L. V., WILLIAMS, C. K., WINN, J. & ZISSERMAN, A. 2018. *Segmentation Results: VOC2012: Leaderboard* [Online]. Available: http://host.robots.ox.ac.uk:8080/leaderboard/displaylb.php?challengeid=11&compid=6 [Accessed 10/20/2018].

EVERINGHAM, M., VAN GOOL, L., WILLIAMS, C. K. I., WINN, J. & ZISSERMAN, A. 2010. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision,* 88**,** 303-338.

FATHI, H., DAI, F. & LOURAKIS, M. 2015. Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. *Advanced Engineering Informatics,* 29**,** 149-161.

GIEL, B. & ISSA, R. R. 2015. Framework for evaluating the BIM competencies of facility owners. *Journal of management in engineering,* 32**,** 04015024.

GUPTA, S., HOFFMAN, J. & MALIK, J. 2016. Cross modal distillation for supervision transfer. *Computer Vision and Pattern Recognition (CVPR).* Las Vegas, NV.

HAMLEDARI, H., MCCABE, B. & DAVARI, S. 2017. Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction,* 74**,** 78-94.

HUBER, D., AKINCI, B., ADAN, A., ANIL, E., OKORN, B. & XIONG, X. 2011. Methods for automatically modeling and representing as-built building information models. *NSF Engineering Research and Innovation Conference.* Atlanta, Georgia.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. 1097-1105.

LEE, S. & LU, Q. 2017. Image-Based Technologies for Constructing As-Is Building Information Models for Existing Buildings. *Journal of Computing in Civil Engineering,* 31**,** 4017005.

LIN, T.-Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., ROAMANAN, D., DOLLÁR, P. & ZITNICK, C. L. 2014. Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision (ECCV).* Zurich, Switzerland: Springer, Cham.

MAYO, G., GIEL, B. & ISSA, R. R. 2012. BIM use and requirements among building owners. *Computing in civil engineering (2012).*

MURA, C., MATTAUSCH, O. & PAJAROLA, R. Piecewise-planar Reconstruction of Multi-room Interiors with Arbitrary Wall Arrangements. Computer Graphics Forum, 2016. Wiley Online Library, 179-188.

MUSIALSKI, P., WONKA, P., ALIAGA, D. G., WIMMER, M., GOOL, L. & PURGATHOFER, W. 2013. A Survey of Urban Reconstruction. *Computer Graphics Forum,* 32**,** 146-177.

OCHMANN, S., VOCK, R., WESSEL, R. & KLEIN, R. 2016. Automatic reconstruction of parametric building models from indoor point clouds. *Computers & Graphics,* 54**,** 94-103.

OSKOUIE, P., BECERIK-GERBER, B. & SOIBELMAN, L. 2017. Automated recognition of building façades for creation of As-Is Mock-Up 3D models. *Journal of Computing in Civil Engineering,* 31**,** 04017059.

PAN, S. J. & YANG, Q. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering,* 22**,** 1345-1359.

PĂTRĂUCEAN, V., ARMENI, I., NAHANGI, M., YEUNG, J., BRILAKIS, I. & HAAS, C. 2015. State of research in automatic as-built modelling. *Advanced Engineering Informatics,* 29**,** 162-171.

PHAM, Q.-H., HUA, B.-S., NGUYEN, D. T. & YEUNG, S.-K. 2018. Real-time Progressive 3D Semantic Segmentation for Indoor Scene. arXiv.

PRATT, L. Y. Discriminability-based transfer between neural networks. Advances in neural information processing systems, 1993. 204-211.

QI, C. R., SU, H., MO, K. & GUIBAS, L. J. 2016a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv preprint arXiv:1612.00593*.

QI, C. R., SU, H., NIEßNER, M., DAI, A., YAN, M. & GUIBAS, L. J. Volumetric and multi-view cnns for object classification on 3d data. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016b. 5648-5656.

QUIJANO, A. & PRIETO, F. 2016. 3D Semantic modeling of indoor environments based on point clouds and contextual relationships. *Ingeniería,* 21**,** 305-323.

RUSSAKOVSKY, 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. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision,* 115**,** 211-252.

SANDLER, M., HOWARD, A., ZHU, M., ZHMOGINOV, A. & CHEN, L.-C. 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Computer Vision and Pattern Recognition (CVPR).* Salt Lake City, UT.

SHELHAMER, E., LONG, J. & DARRELL, T. 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 39**,** 640-651.

SHEN, L., EDIRISINGHE, R. & YANG, M. An investigation of BIM readiness of owners and facility managers in Singapore: institutional case study. CIB World Building Congress, 2016.

TACC, O. O. T. V. P. F. R. 2018. *STAMPEDE2: TACC'S flagship supercomputer* [Online]. Texas Advanced Computing Center (TACC). Available: https://www.tacc.utexas.edu/systems/stampede2 [Accessed 10/21/2018 2018].

TANG, P., HUBER, D., AKINCI, B., LIPMAN, R. & LYTLE, A. 2010. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction,* 19**,** 829-843.

VOLK, R., STENGEL, J. & SCHULTMANN, F. 2014. Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction,* 38**,** 109-127.

XIONG, X., ADAN, A., AKINCI, B. & HUBER, D. 2013. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction,* 31**,** 325-337.