

A Faster R-CNN Approach for Extracting Indoor Navigation Graph from Building Designs

L. Niu^{1,*}, Y.Q Song²

¹ School of Surveying and Urban Spatial Information, Henan University of Urban Construction, 467036 Pingdingshan, China - (l.niu, 064415104, zhmqz8)@hncj.edu.cn

² School of Geographic and Environmental Science, Normal University of Tianjin, West Bin Shui Avenue, 300387 Tianjin, China - syiq@hotmail.com

KEY WORDS: Faster R-CNN; indoor; extraction; navigation graph; building design

ABSTRACT:

The indoor navigation graph is crucial for emergency evacuation and route guidance. However, most of existing solutions are limited to the tedious manual solutions and inefficient automatic solutions of the indoor building designs. In this paper, we strive to combine the cutting-edge faster R-CNN deep learning models with spatial connection rules to provide fine quality indoor navigation graphs. The extraction experiment result is convincing for general navigation purpose. But there exist several shortages for faster R-CNN models to overcome, such as optimizations of the complex object detections and ability of handling irregular shape regions for indoor navigation graph extractions.

1. INTRODUCTION

Indoor graph is crucial for every demand related to navigation purpose, thus research of providing these data could mitigate the data shortage for relevant fields. One of the most promising research directions is the indoor navigation graph automatic extraction. One reason for using automatic solutions is the high volume of existing data and complex application scenes (Ali and Schmid, 2014, Schmid et al., 2013, Walton and Worboys, 2012, Siqueira et al., 2012).

Whilst the research topic is interesting, two major gaps have to be filled in order to achieve a sounding automation process for providing high quality indoor navigation graphs (Qian et al., 2015, Saberian et al., 2014, Ricker et al., 2014, Guex, 2014). One is that the data source quality cannot not be maintained due to inconsistent data formats and variant resolutions of building blueprints. The other one is to formalize extraction rules that could not be easily devised without difficulties.

Along this direction, the faster region proposal convolutional neural network (faster R-CNN in short) model is widely used for extracting graphic information from multiple visual data sources (Soltan et al., 2018, Ren et al., 2015). Therefore, its developing history should be reviewed to generate a clear concept for the indoor navigation graph extraction (Figure 1).

The history of faster R-CNN model originated from the object detection domain of machine learning. As for normal object detection scenes, early application models like supporting vector machines (SVM in short) strictly follow rigid rules, and struggle with low accuracy and efficiency. The first innovative deep learning model is the convolutional neural network (CNN in short) model (Soltan et al., 2018, Ren et al., 2015). This type of model can fill the convolutional neural network with sliding windows to brutally traverse all image pixels, and extract object boundaries and classes. However, this over-simplified scanning solution could result in high volume of data. Thus, it has to be improved on computational efficiency. Aiming for this need,

researchers proposed a new model: region proposal convolutional neural network (R-CNN in short) model. This model benefits from hypothesizing that many regions of interest (RoI in short) exist in the handling images. Therefore, the only subsequent work after image input is to refine these RoIs. By introducing the region proposal, R-CNN greatly reduces the computing burden troubling CNN.

Only introducing region proposal is not enough to totally relieve the solution from heavy computational burdens for searching and refining RoIs in large and complex images. Then, another important model based on R-CNN approach is invented: fast region proposal convolutional neural network (R-CNN in short) model (Soltan et al., 2018, Ren et al., 2015). This model uses region proposal with feature maps extracted for each RoI to parallel the region extraction task, and reduces the computation time to an almost-instant level such as video object tracking. But this is not the end for optimizing R-CNN.

Subsequently, the faster R-CNN model is introduced to accelerate the processing time to a millisecond level for the feature map extraction of the whole image, and subsequently fully benefit from using this holistic information to mark potential RoIs (Ren et al., 2015).

As mentioned above, the faster R-CNN is the most advanced solution for the fast object detection in images. We use it as the base frame for extracting navigation area from building designs, and this approach overcomes shortages of handling complex scenes of existing navigation space extraction methods for indoor navigation. These solutions could be classified into two main streams. The first class is the middle axis line extraction solutions, and the second class is the semantic information based spatial subdivision solutions.

The first category of space extraction method takes full usage of the existing geometric shells of the indoor space (Alattas et al., 2017, Zlatanova et al., 2013, Isikdag et al., 2013). A spatial topological correct set of boundaries for the research area is

* Corresponding author

formed, and then middle axis lines of this set are refined to establish the indoor navigation graph. This approach is purely computationally oriented and receives high reputation of time saving. But under complex shape for building space, it suffers from generating the geometric shells. For the second category, it concentrates on spatial decomposition, which is different from the first category. The spatial decomposition process produces several subdivided regions that could be used to generate graph nodes, and their spatial neighbouring relationships could be analysed to establish edges for generating indoor navigation graph (Alattas et al., 2017, Vanclooster et al., 2016, Scholz and Schabus, 2014, Krūminaitė and Zlatanova, 2014, Xu et al., 2013, Liu and Zlatanova, 2013).

These two categories of indoor navigation graph both favour the spatial topological correct situation, which is rare under real production situations. Thus, in order to mitigate existing flaws, we aim to combine the faster R-CNN method to extract indoor navigation regions and then utilize basic the spatial neighbouring analysis functions to generate indoor navigation graph. Thus, this paper is organized as follows: The second part explains the main workflow of the faster R-CNN indoor navigation graph extraction solution. The third part demonstrates the experiment area and raw test data. The fourth part elaborates the underlying phenomena and their potential meanings for experiment results. The final part concludes the pros and cons of our solution, and proposes several future research directions.

2. FASTER R-CNN SOLUTION FOR INDOOR NAVIGATION GRAPH EXTRACTION FROM BUILDING DESIGNS

The graph extraction workflow is straightforward (Figure 2). First, we provide the faster R-CNN model with training images, and then these images are marked with object boundaries and class names. Second, the faster R-CNN model learns from the supplied images with convolutional neural networks, and finally forms a tuned convolutional network corresponding to the original images. Next, this network is frozen to be used for extracting navigation space from untrained images. Finally, an indoor navigation graph is generated.

2.1 Faster R-CNN model

The introduced faster R-CNN models in this study are from the Tensorflow project. Existing Tensorflow models are most trained with the Microsoft COCO data set, which contains a large number of common objects in the collected images (Lin et al., 2014). Thus, the convolutional network setting is also optimized for the collection of images and their containing objects. We select `faster_rcnn_inception_resnet_v2_atrous_coco_2018_01_28` (resnet-v2 in short) model and `faster_rcnn_resnet101_coco_2018_01_28` (resnet101 in short) model from the Tensorflow model zoos. The reason is they receive balance from both accuracy and speed performance (tensorflow.org, 2018, Abadi et al., 2016). All the detail information could be accessed by the Tensorflow webpages in the reference section.

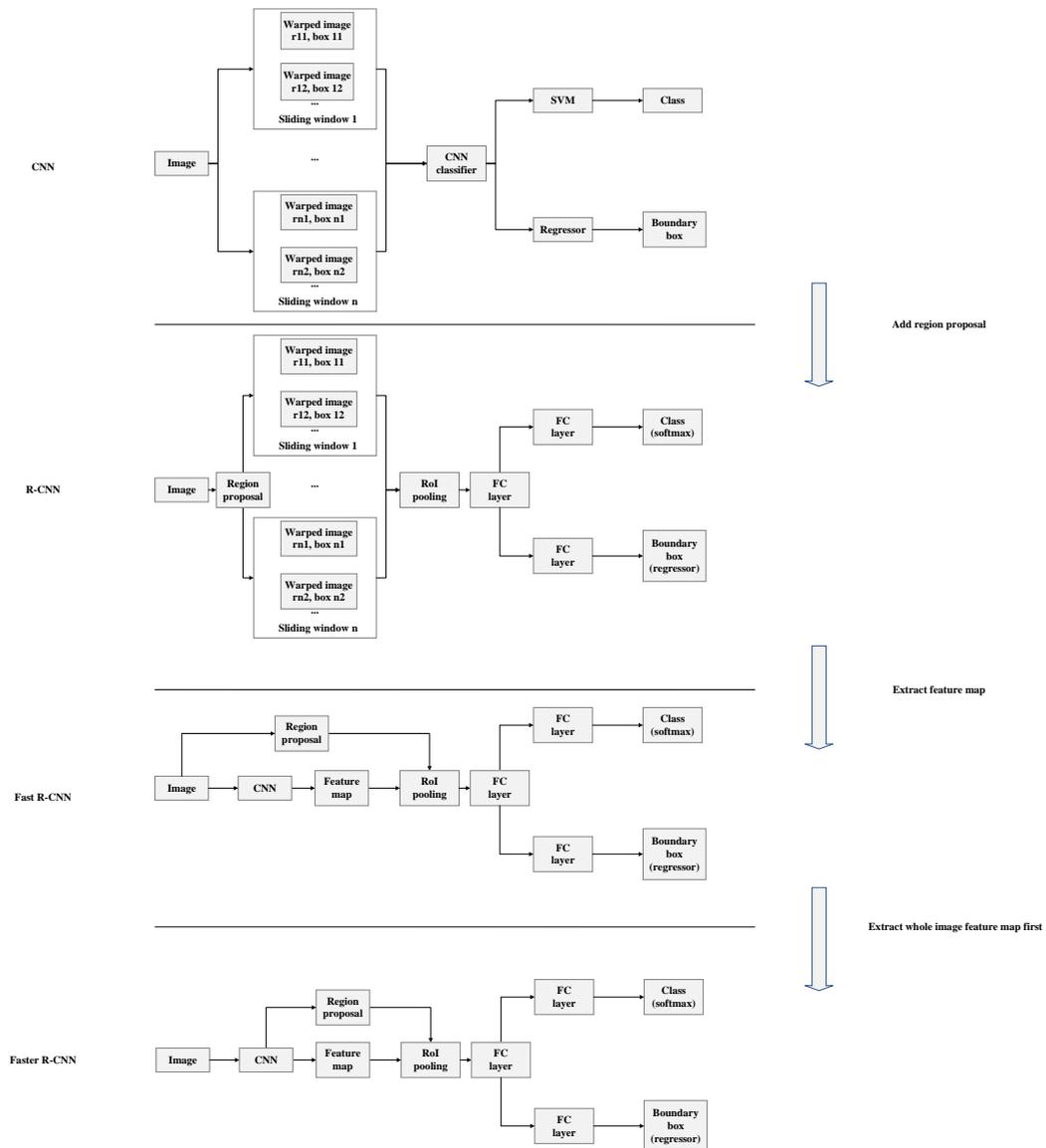


Figure 1. Faster R-CNN developing history and core mechanism explanation

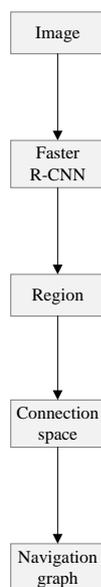


Figure 2. The whole solution workflow

2.2 Indoor navigation graph extraction principle

The indoor navigation graph extraction principle is also easy to follow. As our study is a preliminary step for applying the deep learning solution for indoor navigation graph extraction from building designs. We only aim for extracting rooms and doors between them to establish a basic navigation graph. We define ‘room’ is a full separate region for holding an indoor space, and also covers the scope of corridors, lobbies and any other functional areas surrounded by interior and outer walls. The door here is an opening on walls between rooms, and they can be two-direction doors, single doors or double doors.

3. EXPERIMENT

3.1 Experiment data

According to the requirement of applying deep learning models, the experiment data should be divided into two groups: training group and test group. The first one is used to tune parameters for embedded convolutional neural networks of the deep learning model, and the latter one is used to evaluate the learning quality of the trained model. Thus, we collect 255 images from the result of google image searching engine, which are filtered out from

around 1000 raw images generated by key words-building designs. The deep learning platform is composed of the hardware part and software part. The hardware part contains Intel I9 9900K, Nvidia GeForce RTX 2070, 1TB SSD hard disk and 32GB system memories; the software part contains the python 3.6 environment, Tensorflow 1.12, CUDA 9.2 and CUDNN 7.0.2. The experiment is designed to validate the performance of the proposed solutions for facing three challenges. The first challenge is that the foreground and background colour differences. After a preliminary study, we find that the building designs could be rendered with black background and white foreground, or vice versa, the deep learning model may be troubled by these colour setting differences. The second challenge is the inconsistent drawing of walls and doors, and this may lower the spatial classification result by deep learning models. The last and the greatest challenge is the enormous number of trivial objects existing in building space, such as decorations and furniture. Due to their existence, the region subdivision performance of deep learning model can be significantly hindered.

3.2 Raw experiment result

Because the raw results of two comparing models are huge in size and dimensions, we decide to upload them to an internet share disk folder to allow interested researchers to study. The English website file access link is: https://www.dropbox.com/s/61ruzma9vr4hxjv/20190103tensorflow_lowtest.zip?dl=0; Chinese website file access link is: <https://pan.baidu.com/s/1OV6sBECXYkgxWsjKoyOvVQ>. The file is in ZIP file format, and accessors need to install relevant unzip software, jpeg browser software and csv file parsing software to read the data. Besides, the frozen Tensorflow graph is also attached. The structure of the experiment data are as follows:

```
+ 20190103tensorflowtest
--
faster_rcnn_inception_resnet_v2_atrous_coco_2018_01_28_model
(frozen graph model for resnet-v2 model)
--
faster_rcnn_inception_resnet_v2_atrous_coco_2018_01_28_result
(test result for resnet-v2 model with marked images and
node/edge csv files)
--faster_rcnn_resnet101_coco_2018_01_28_model      (frozen
graph model for resnet101 model)
--
faster_rcnn_inception_resnet_v2_atrous_coco_2018_01_28_model
(test result for resnet101 model with marked images and
node/edge csv files)
--test (test raw images)
```

4. DISCUSSION

The analysis of experiment result could be classified into two categories. The first one is about the statistical analysis from the machine learning aspect, which contains the entropy of the learning result and step time cost. While the second one is about the spatial classification result of test images.

4.1 Statistical analysis

Here we explain the meaning for each statistical index. The box classification loss means the computational spending for correct class setting of each bounding box; the box localization loss means the computational spending for correct position setting of each bounding box; the global step time cost means the time duration for each global step; the regional proposal network

(RPN in short) localization loss means the computational spending for correct location setting of RoIs; the RPN objectness loss means the computational spending for RoIs successfully wrapping object; and the total loss is the sum of all losses.

The statistical result is shown from Figure 3 to Figure 14, which covers both the machine learning result generated by the resnet101 and resnet-v2 models. All models are trained by 200,000 steps. We can perceive that the box classification loss of resnet-v2 model oscillates more frequently than resnet101 model. When training step count approaches 120,000, the resnet-v2 model loss is just over 0.02. While the situation for resnet101 is more complex, its loss is unstable and finally around 0.02. The box localization loss for resnet101 shakes around 0.03 after 80,000 steps, and this loss for resnet-v2 finally converges to 0.02. The clone loss of resnet101 oscillates around 0.06 after 80,000 steps, and this loss finally converges to 0.04 for resnet-v2. The global time cost for each step shakes from 3.94 to 3.95 second for resnet101, and this figure is 1.95 second for resnet-v2 under most circumstances. The RPN localization loss for both models converges, but the beginning step for them to converge are different: 145,000 for resnet101 and 135,000 for resnet-v2. The RPN objectness losses for two models are quite different. The resnet101 model loss converges to 0 after 170,000 steps, but the resnet-v2 model loss does not converge after all 200,000 steps. As for the total loss of two models, the resnet101 loss is comparatively low for some steps but also high for the other steps, and the resnet-v2 loss converges to some value between 0.04 and 0.05.

4.2 Spatial classification analysis

As mentioned above, due to that the spatial classification result cannot be provided by one single illustration, we select 6 typical images to demonstrate the typical extraction results in Figure 15. These images are No. 1, No. 17, No. 24, No. 44, No. 45 and No. 46 images in our test. For each image, the left side is the resnet-v2 model extracted regions and navigation graphs, and the right side is the resnet101 model extracted regions and navigation graphs.

We can find that the doubt about negative influences caused by different foreground background colours could be eliminated, due to the proper handling for white background and black background by both models. Besides, the shades and shadows in the building designs also have very limitedly influenced the extracted results at all. As for the inconsistent drawing of walls and doors, this problem does hinder the extracted regions for image 44, 45 and 46. However, under these complex circumstances, even experienced human operators cannot draw the bounding box of each rooms without thinking for quite a while. The last challenge of overwhelming trivial objects in building space troubles both models in image 24, and resnet101 achieves better result than resnet-v2.

Furthermore, an interesting phenomenon could be found that for image No. 46, both models can extract several regions, but they are not certain about whether these regions are rooms, thus they just leave these regions unmarked. One more thing to complement is that due to the poor extracted result performance by both models on doors, we modify the connection relationship extracted algorithms to directly link the extracted rooms to establish a basic navigation graph.

5. CONCLUSION

In this study, we aim to combine deep learning models with basic indoor navigation space extracted principles to produce indoor navigation graphs. The experiment result shows this effort is promising for generating regions and navigation graph for indoor space from building designs.

Nevertheless, there are two major bottlenecks for applying deep learning solutions for indoor navigation graph extraction from mass building designs: training data preparing and extracted result accuracy. The former one is that a great amount of manual work have to be spend on the selection of good quality building designs and object marking of these designs. The latter one is that the extracted regions may not be a complete connected navigation graph.

In order to tackle these shortages, the future work have to be along two directions. First direction is to introduce the most advanced mask R-CNN model, which is optimized for extracting irregular shapes of objects like doors and irregular shape rooms. And more detailed classification of rooms and doors will also be discussed in the next step research. The second direction is to introduce more spatial semantic information to the extraction process, and a rigid indoor navigation graph could be expected with this improvement.

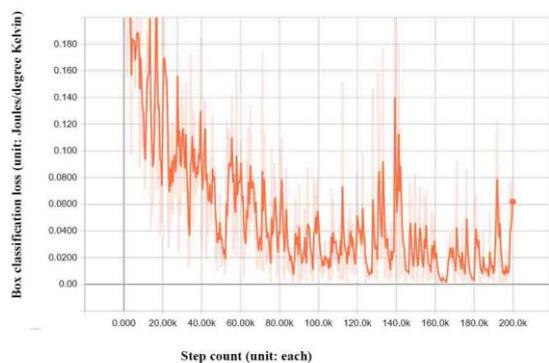


Figure 3. Box classification loss of resnet101 model

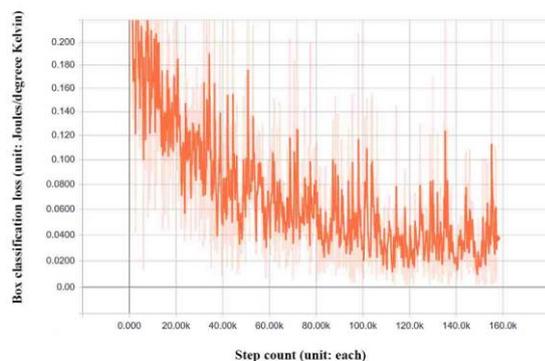


Figure 4. Box classification loss of resnet-v2 model

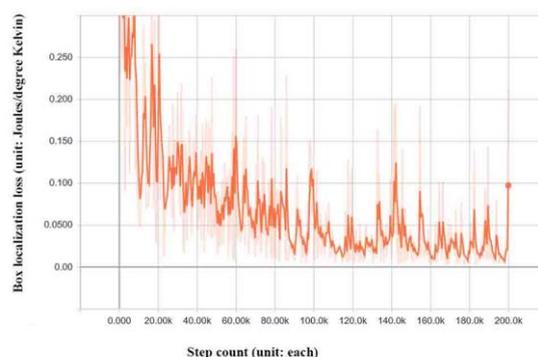


Figure 5. Box localization loss of resnet101 model

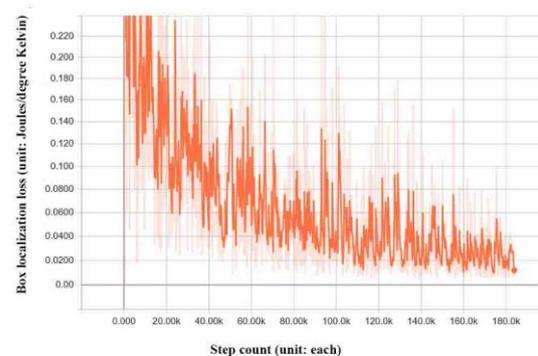


Figure 6. Box localization loss of resnet-v2 model

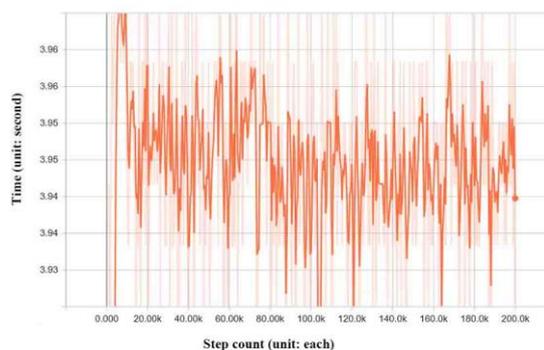


Figure 7. Global step time cost of resnet101 model

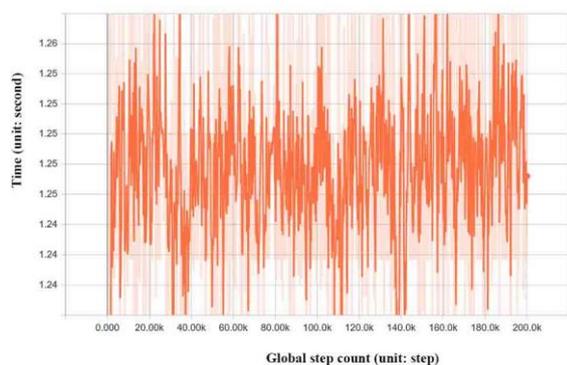


Figure 8. Global step time cost of resnet-v2 model

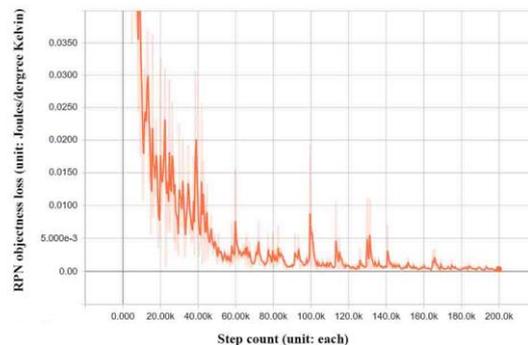


Figure 11. RPN objectness loss of resnet101 model

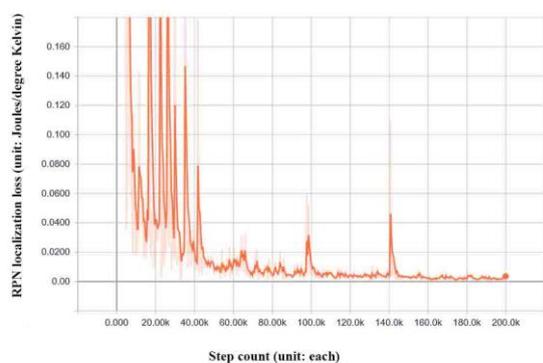


Figure 9. RPN localization loss of resnet101 model

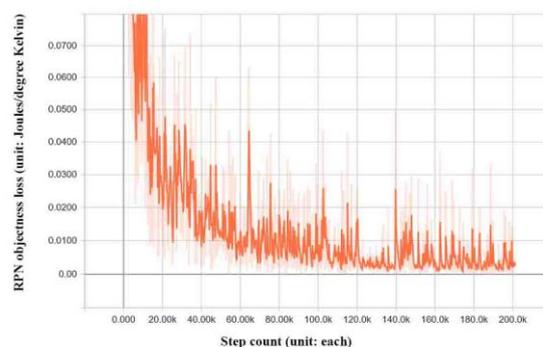


Figure 12. RPN objectness loss of resnet-v2 model

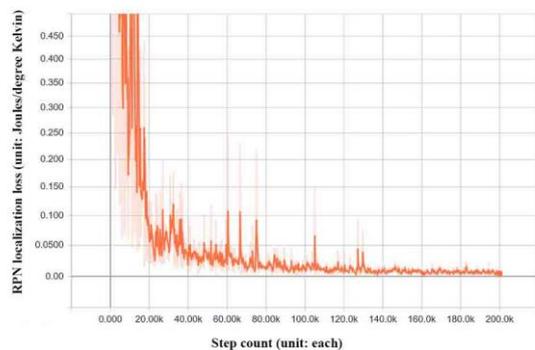


Figure 10. RPN localization loss of resnet-v2 model

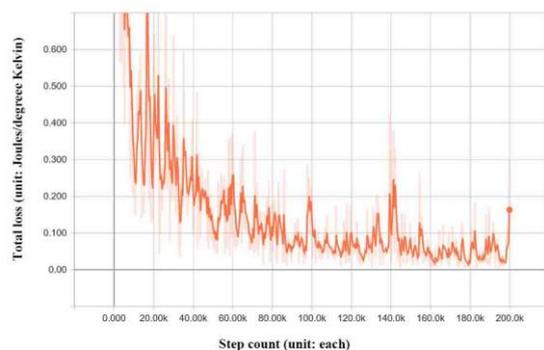


Figure 13. Total loss of resnet101 model

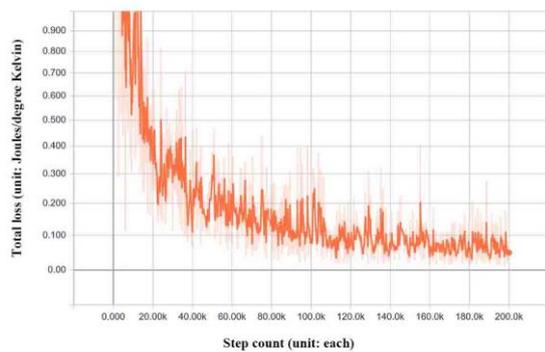


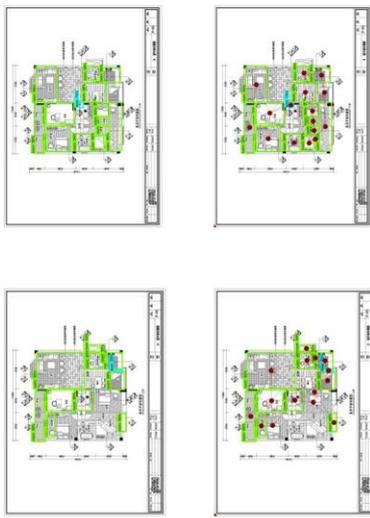
Figure 14. Total loss of resnet-v2 model

ACKNOWLEDGEMENTS

This paper is sponsored by the National Science Foundation of China (41771433, 41501440, 41571387, 41471341), 2015 Technology Leadership Foundation of Henan University of Urban Construction.

REFERENCES

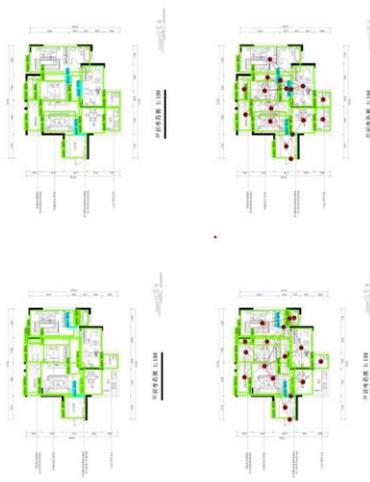
- ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G. & ISARD, M., TensorFlow: a system for large-scale machine learning. *OSDI*, 2016. 265-283.
- ALATTAS, A., ZLATANOVA, S., VAN OOSTEROM, P., CHATZINIKOLAOU, E., LEMMEN, C. & LI, K.-J., 2017. Supporting Indoor Navigation Using Access Rights to Spaces Based on Combined Use of IndoorGML and LADM Models. *ISPRS international journal of geo-information*, 6(12), 384.
- ALI, A. L. & SCHMID, F., 2014. Data quality assurance for volunteered geographic information. *Geographic Information Science*. Heidelberg, Germany: Springer.
- GUEX, G., 2014. Spatial Graphs Cost and Efficiency: Exploring Edges Competition by MCMC. *Geographic Information Science*. Heidelberg, Germany: Springer.
- ISIKDAG, U., ZLATANOVA, S. & UNDERWOOD, J., 2013. A BIM-Oriented Model for supporting indoor navigation requirements. *Computers, Environment and Urban Systems*, 41:112-123.
- KRŪMINAITĖ, M. & ZLATANOVA, S., Indoor space subdivision for indoor navigation. the Sixth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, 2014 Dallas, Texas, USA. ACM, 25-31.
- LIN, T.-Y., MAIRE, M., BELONGIE, S., HAYS, J., PERONA, P., RAMANAN, D., DOLLÁR, P. & ZITNICK, C. L., Microsoft coco: Common objects in context. European conference on computer vision, 2014. Springer, 740-755.
- LIU, L. & ZLATANOVA, S., 2013. Generating Navigation Models from Existing Building Data. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(4), 19-25.
- QIAN, J., PEI, L., MA, J., YING, R. & LIU, P., 2015. Vector graph assisted pedestrian dead reckoning using an unconstrained smartphone. *Sensors*, 15(3), 5032-5057.
- REN, S., HE, K., GIRSHICK, R. & SUN, J., Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 2015. 91-99.
- RICKER, B., DANIEL, S. & HEDLEY, N., 2014. Fuzzy Boundaries: Hybridizing Location-based Services, Volunteered Geographic Information and Geovisualization Literature. *Geography Compass*, 8(7), 490-504.
- SABERIAN, J., MALEK, M. R., WINTER, S. & HAMRAH, M., 2014. A New Framework for Solving the Spatial Network Problems Based on Line Graphs. *Transactions in GIS*, 18(5), 767-782.
- SCHMID, F., FROMMBERGER, L., CAI, C. & FREKSA, C., 2013. What you see is what you map: Geometry-preserving micro-mapping for smaller geographic objects with mapit. *Geographic Information Science at the Heart of Europe*. Springer.
- SCHOLZ, J. & SCHABUS, S., 2014. An indoor navigation ontology for production assets in a production environment. *Geographic Information Science*. Heidelberg, Germany: Springer.
- SIQUEIRA, T. L. L., DE AGUIAR CIFERRI, C. D., TIMES, V. C. & CIFERRI, R. R., 2012. Towards vague geographic data warehouses. *Geographic Information Science*. Heidelberg, Germany: Springer.
- SOLTAN, S., YANNAKAKIS, M. & ZUSSMAN, G. J. I. T. O. C. O. N. S., 2018. Power grid state estimation following a joint cyber and physical attack. 5(1), 499-512.
- TENSORFLOW.ORG. 2018. *TensorFlow home page* [Online]. Available: <https://www.tensorflow.org/> [Accessed March 3rd 2018].
- VANCLOOSTER, A., VAN DE WEGHE, N. & DE MAEYER, P., 2016. Integrating indoor and outdoor spaces for pedestrian navigation guidance: A review. *Transactions in GIS*, 20(4), 491-525.
- WALTON, L. A. & WORBOYS, M., 2012. A qualitative bigraph model for indoor space. *Geographic Information Science*. Heidelberg, Germany: Springer.
- XU, W., KRUMINAITE, M., ONRUST, B., LIU, H., XIONG, Q. & ZLATANOVA, S., 2013. A 3D Model Based Indoor Navigation System for Hubei Provincial Museum. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(4), 51-55.
- ZLATANOVA, S., LIU, L. & SITHOLE, G., A conceptual framework of space subdivision for indoor navigation. Proceedings of the Fifth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, 2013. ACM, 37-41.



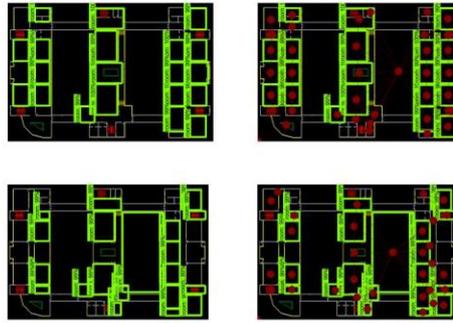
(c) image 24
 extracted result



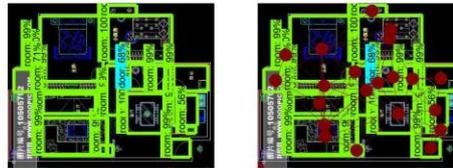
(b) image 17
 extracted result



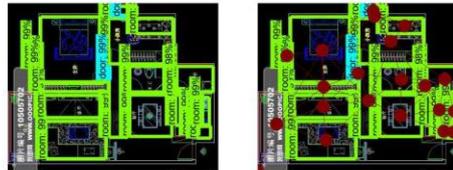
(a) image 1
 extracted result



(d) image 46
 extracted result



(d) image 45
 extracted result



(d) image 44
 extracted result

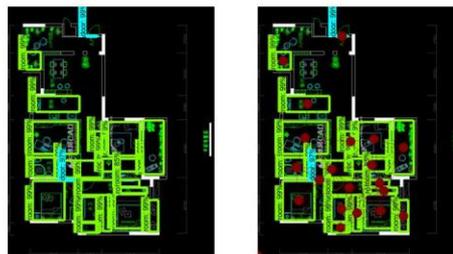


Figure 15. Selected extracted image result for two comparing models