Multi-layer Range Point Cloud Semantic Segmentation Based on U-Net

Delin Feng
2020233246
fengdl@shanghaitech.edu.cn

Longtian Qiu
Institution2
qiult@shanghaitech.edu.cn

Qianjing Shi
2018
shiqj@shanghaitech.edu.cn

Abstract

To date, as development in 5G network and AI or autonomous driving, there has been a surge of interest in 3D point cloud semantic segmentation. Generally speaking, there are three paradigms for point cloud semantic segmentation: projection-based, discretization(voxel) based, and point-based methods. Each of these three methods has its advantages, while, how to solve the lack of small object information in the scene is still a puzzle. In this project, we propose a multilayer range network for semantic segmention based on U-Net, to solve the problem of missing detailed structure and pose after projection or down-sampling and improve the efficiency and accuracy of the algorithm. Inspired by the voxel structure of Cylinder3D, multilayer processing of the original point cloud data is carried out on the basis of project-based method, so that it has the advantages similar to voxel. We verify our algorithm via SemanticKITTI dataset and compare results with state of the art methods to evaluate the performance of the algorithm. We also create our own(ShanghaiTech campus) dataset for experiments to verify the efficiency and accuracy of the algorithm.

1. Introduction

Since deep-learning, machine learning, and computer visions attracts more and more interest, point cloud segmentation (PCS) and 3D point cloud semantic segmentation (PCSS) are popular right now, which is a fundamental and essential capability for real-time intelligent systems like autonomous driving and augmented reality. In large scale 3D point clouds, more semantic information can be provided through efficient semantic segmentation.

People have found that deep learning's effectiveness for point cloud perception tasks. Conventionally, researchers change the point cloud into voxel grids ,then process them using 3D volumetric convolutions. It will produce information loss due to low resolutions during

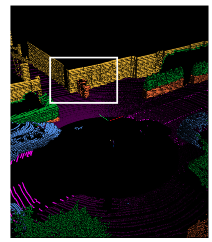

Figure 1. Ground truth segmentation

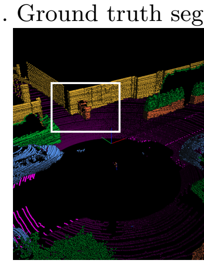

Figure 2. Ground truth segmentation

voxelization which means many points, if they lie in the same grid, will be merged together. So it is essential to find a high-resolution representation to preserve the fine details in the input data so as to keep it less blur.

But it is true that both memory requirement and computational cost increase cubically with voxel resolution. Recently, another models stream try to process the input point clouds directly. Due to the sparse representation, these point-based models need much lower GPU memory which is not as much as voxel-based models. But the fact is omitted that the random memory access is not as efficient other methods. For semantic segmentation, projection-based methods can perform more quickly, but the discretization errors and fuzzification of CNN output controls the limit of them.

Semantic segmentation is very basic in graph or imaging processing challenge, which is a method which associates pixels with semantic labels. As more and needs in 3D scenes, PCSS begins to take the responsibility, which plays the role in 3D form of it. Semantic segmentation can use 2D image's regular distributed pixels while through PCSS, it can use irregular or regular distributed points in a 3D place. The sensors get point cloud directly and it measures the distance. Point cloud can be also generated from multi- or stereo-imagery view.

We are recently witnessing an increasing availability of not interpreted point clouds and 3D models, often shared online using point-based rendering solutions (e.g. PoTree) of mesh-based portals (e.g. Sketchfab). When it comes to point clouds, innovative methods are needed more and more for the analysis and treatment of these data and for their classification, which we finally want to exploit in-depth the informative value of these surveys and representations. The simplest and powerful collection of elementary geometrical primitives are 3D point clouds, but at the same time able to represent shape, size, position and orientation of objects in space. This information may be augmented with additional contents obtained from other sensors or sources, such as colours, multispectral or thermal information, etc. For a successful exploitation of point clouds and to better understand them, we must first proceed with segmentation and classification procedures. The former refers to group points in subsets (normally called segments) characterized by having one or more characteristics in common (geometric, radiometric, etc.) whereas classification means the definition and assignment of points to specific classes ("labels") according to different criteria. Due to the complexity and variety of point clouds caused by irregular sampling, varying density, different types of objects, etc., point cloud classification and segmentation are very active research topics. There are multiple research studies related to these two topics, many driven by specific needs provided by the field of application (building modeling, Heritage documentation and preservation, robotics, etc.). Most of the segmentation algorithms are tailored to work with a 2.5D surface model assumption, coming for example from a LiDAR-based survey. Many algorithms require a fine-tuning of different parameters depending upon the nature of data and applications. Supervised methods are the majority with a training phase mandatory and fundamental to guide the successive machine learning classification solution. Some of the techniques developed for segmenting point clouds generated from airborne laser scanning can be applied or easily adapted to terrestrial point clouds. The results are generally affected by noise and density of the cloud as well as by the quality of the training data[10].

2. Related Work

For point-cloud segmentation, papers are based on two categories, one is about small-scale point clouds, the other is about large scale point clouds. In research of small scale point cloud segmentation like indoor scene understanding or object part parsing, researchers often use DGCNN[28] PointNet[21][22]. Another method, Deep-KdNet[14],groups neighbor point to extend the hierarchical architecture of PointNet++[22].

Previous approaches about point clouds recognition[8][11] mainly rely on complicated hand-crafted features, such as surface normal or generated descriptors, and hard thresh old decision rules based on a clustering algorithm. These approaches have two problems: (1) hand-crafted features cost much time and the results by hard threshold decision are not suitable for productions; (2) they can not recognize the pixel level object category as the same as the semantic segmentation, which makes it difficult to apply to some autonomous driving tasks[27]. Here are some recent approaches for deep learning in the 3D point cloud data, semantic segmentation network structure, semantic segmentation tasks and bounding box detection tasks[27].

Obtaining annotations, especially point-wise or pixel-wise annotations for computer vision tasks is usually very difficult. As a consequence, synthetic data sets have seen growing interest. In the autonomous driving community, the video game Grand Theft Auto has been used to retrieve data for object detection and segmentation[23][13][29].

CNN[26] approaches consider LiDAR point clouds in either two or three dimensions. Work with two-dimensional data considers raw images with projections of LiDAR point clouds top-down[2] or from a number of other views[4]. Other work considers three-dimensional data itself, discretizing the space into voxels and engineering features such as disparity, mean, and saturation[24]. Regardless of data preparation, deep learning methods consider end to end models that leverage 2D convolutional[17] or 3D convolutional[20] neural networks[29].

FCN[26] was the pioneering method for semantic segmentation based on deep learning. It replaced the last fullyconnected layers in the classification task with convolution layers. Recent approaches like DeeplabV3[3] used a dilated convolutional layer[30] and the conditional random field (CRF)[15] to improve the prediction accuracy. SegNet[1] used an encoder-decoder architecture to fuse the feature maps from the deep layer with spatial information from lower layers. Other approaches like ICnet[31], RefifineNet[18] took multi-scale features into consideration and combined image features from multiple refined paths. Among those methods, Networks like Deeplabv3 and FCN were compelled to increase the performance. SegNet and ICNet are able to achieve real-time performance.

Although, they have a big improvement in speed or accuracy, these methods still have some influence on the other side.

3D data has sufficient features and attracts much research attention. With the rapidly developing of deep learning, many methods apply convolutional neural networks (CNN) on the 3D point cloud data directly. 3DFCN[16] and VoxelNet[32] used a 3D-CNN[9] to extract features from width, height and depth simultaneously. MV3D[5] fused multi-perception from a bird's-eye view, a front view and a camera view to obtain a more robust feature representation. In addition, some works[6] considered the representation of three-dimensional data itself and divided it into voxels to undertake features such as intensity, distance, local mean and disparity. Although all of the above methods have achieved a good accuracy, they still cost too much time in computation which limited their applications in real-time tasks[27].

Previous works proposed several different algorithms for plane extractions from 3D point clouds, such as RANSAC based (random sample consensus)[25] methods and region grow-based methods[19]. However, RANSAC requires much computation on random plane model selection. Region grow-based methods, depending on the manually designed threshold, are not adaptive. Other traditional approaches based on clustering algorithms just realized the segmentation work but not pixel-wise region classifications.

Recently, researchers started focusing on the semantic segmentation of 3D Lidar point cloud data. PointNet[21] explored a deep learning architecture to do the 3D classification and segmentation on raw 3D data. However, it only works well in indoor. Also Dube[7] explored an incremental segmentation algorithm, based on region growing, to improve the 3D task performance. However, real-time performance is still challenging. SqueezeSeg[29] is similar with our task which used the SqueezeNet[12] as the backbone and performed compatible results. However, it only referred the CRF to improve the performance in the predicted 2D spherical masks, which could lose location information in the 3D space. Without considering the 3D constraints in the original point cloud, the results of SqueezeSeg is extremely limited by this CRF post-process [27].

2.1. Dual submission

Please refer to the author guidelines on the CVPR 2021 web page for a discussion of the policy on dual submissions.

2.2. Paper length

Papers, excluding the references section, must be no longer than eight pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of eight pages with two pages of references would have a total length of 10 pages. There will be no extra page charges for CVPR 2021.

Overlength papers will simply not be reviewed. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. Note that this LaTeX guide already sets figure captions and references in a smaller font. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

2.3. The ruler

The LaTeX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LaTeX users may use options of cvpr.cls to switch between different versions.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 095.5), although in most cases one would expect that the approximate location will be adequate.

2.4. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: http://www.pamitc.org/documents/mermin.pdf.

2.5. Blind review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one's own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words "my" or "our" when citing previous work. That is all. (But see below for techreports.)

Saying "this builds on the work of Lucy Smith [1]" does not say that you are Lucy Smith; it says that you are building on her work. If you are Smith and Jones, do not say "as we show in [7]", say "as Smith and Jones show in [7]" and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper just asking to be rejected:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Removed for blind review

An example of an acceptable paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of the paper of Smith et al. [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Smith, L and Jones, C. "The frobnicatable foo filter, a fundamental contribution to human knowledge". Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [?] as additional material and cite it as

> [1] Authors. "The frobnicatable foo filter", F&G 2014 Submission ID 324, Supplied as additional material fg324.pdf.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not require the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [?]". Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus et al.

You can handle this paper like any other. Don't write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus et al., but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ

Q: Are acknowledgements OK?
A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges? A: To conform with the double blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the

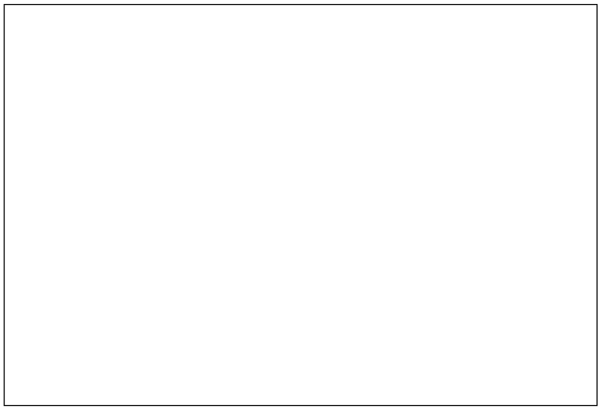

Figure 3. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

2.6. Miscellaneous

Compare the following:
$conf_a$ $\quad conf_a$
conf_a $\quad \mathit{conf}_a$
See The TEXbook, p165.

The space after e.g., meaning "for example", should not be a sentence-ending space. So e.g. is correct, e.g. is not. The provided \eg macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "et al." (not "et. al." as "et" is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: " Frobnication has been trendy lately. It was introduced by Alpher [?], and subsequently developed by Alpher and Fotheringham-Smythe [?], and Alpher et al. [?]."

This is incorrect: "... subsequently developed by Alpher et al. [?] ..." because reference [?] has just two authors. If you use the \etal macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher et al.

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [?, ?, ?] to [?, ?, ?].

3. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

3.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high. Page numbers should be in footer with page numbers, centered and .75 inches from the bottom of the page and make it start at the correct page number rather than the 4321 in the example. To do this fine the line (around line 20)

\setcounter{page}{4321}

where the number 4321 is your assigned starting page.

3.2. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 3 and 4. Short captions should be centred.
Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

Figure 4. Example of a short caption, which should be centered.

FIRST-ORDER HEADINGS. (For example, 1. Introduction) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, 1.1. Database elements) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

3.3. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

3.4. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [?]. Where appropriate, include the name(s) of editors of referenced books.

3.5. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

[1]This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

When placing figures in LaTeX, it's almost always best to use \includegraphics, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]
                {myfile.eps}
```

3.6. Color

Please refer to the author guidelines on the CVPR 2021 web page for a discussion of the use of color in your document.

4. Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

Please direct any questions to the production editor in charge of these proceedings at the IEEE Computer Society Press: https://www.computer.org/about/contact.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12):2481–2495, 2017. 2

[2] Luca Caltagirone, Samuel Scheidegger, Lennart Svensson, and Mattias Wahde. Fast lidar-based road detection using fully convolutional neural networks. 2017. 2

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017. 2

[4] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection network for autonomous driving. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 2

[5] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection network for autonomous driving. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 3

[6] Renaud Dubé, Andrei Cramariuc, Daniel Dugas, Juan Nieto, Roland Siegwart, and Cesar Cadena. Segmap: 3d segment mapping using data-driven descriptors. arXiv preprint arXiv:1804.09557, 2018. 3

[7] Renaud Dube, Mattia Guglielmo Gollub, Hannes Sommer, Igor Gilitschenski, and Juan Nieto. Incremental segment-based localization in 3d point clouds. IEEE Robotics and Automation Letters, PP(99):1–1, 2018. 3

[8] Chen Feng, Yuichi Taguchi, and Vineet R Kamat. Fast plane extraction in organized point clouds using agglomerative hierarchical clustering. In IEEE International Conference on Robotics & Automation, 2014. 2

[9] Ben Graham. Sparse 3d convolutional neural networks. arXiv preprint arXiv:1505.02890, 2015. 3

[10] E. Grilli, F. Menna, and F. Remondino. A review ofpoint clouds segmentation and classification algorithms. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W3:339–344, 2017. 2

[11] M. Himmelsbach, A Müller, T Lüttel, and H.-J Wünsche. Lidar-based 3d object perception. In Proceedings of 1st International Workshop on Cognition for Technical Systems, 2008. 2

[12] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. arXiv preprint arXiv:1602.07360, 2016. 3

[13] Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, Karl Rosaen, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? 2016. 2

[14] Roman Klokov and Victor Lempitsky. [ieee 2017 ieee international conference on computer vision (iccv) - venice (2017.10.22-2017.10.29)] 2017 ieee international conference on computer vision (iccv) - escape from cells: Deep kd-networks for the recognition of 3d point cloud models. 2017. 2

[15] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems, 24:109–117, 2011. 2

[16] Bo Li. 3d fully convolutional network for vehicle detection in point cloud. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1513–1518. IEEE, 2017. 3

[17] Bo Li, Tianlei Zhang, and Tian Xia. Vehicle detection from 3d lidar using fully convolutional network. arXiv preprint arXiv:1608.07916, 2016. 2

[18] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1925–1934, 2017. 2

[19] Zheng Lin, Jesse Jin, and Hugues Talbot. Unseeded region growing for 3d image segmentation. In ACM International Conference Proceeding Series, volume 9, pages 31–37, 2000. 3

[20] Daniel Maturana and Sebastian Scherer. 3d convolutional neural networks for landing zone detection from lidar. Proceedings IEEE International Conference on Robotics & Automation, 2015:3471–3478, 2015. 2

[21] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. 2017. 2, 3

[22] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30:5099–5108, 2017. 2

[23] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. 2016. 2

[24] Joel Schlosser, Christopher K. Chow, and Zsolt Kira. Fusing lidar and images for pedestrian detection using convolutional neural networks. In 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016. 2

[25] R. Schnabel, R. Wahl, and R. Klein. Efficient ransac for point-cloud shape detection. Computer Graphics Forum, 26(2):214–226, 2007. 3

[26] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. IEEE Computer Society, 2017. 2

[27] Yuan Wang, Tianyue Shi, Peng Yun, Lei Tai, and Ming Liu. Pointseg: Real-time semantic segmentation based on 3d lidar point cloud. 2018. 2, 3

[28] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. Acm Transactions on Graphics, 38(5), 2018. 2

[29] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In 2018 IEEE International Conference on Robotics and Automation (ICRA), pages 1887–1893. IEEE, 2018. 2, 3

[30] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. 2016. 2

[31] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In Proceedings of the European Conference on Computer Vision (ECCV), pages 405–420, 2018. 2

[32] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4490–4499, 2018. 3