Semantic Segmentation of 3D LiDAR Data at Dynamic Urban Scenes

Biao Gao*, Jilin Mei*, Donghao Xu*, Xijun Zhao†, Wen Yao†, Huijing Zhao*
*Peking University, Beijing, China
†201 Institution, Beijing, China

***Abstract*— This work studies semantic segmentation of 3D LiDAR data at dynamic urban scenes. LiDAR data plays an important role of perception in autonomous driving system. However, most semantic segmentation methods and datasets are designed for camera data nowadays. In this work, we propose a method which can generate semantic segmentation of LiDAR data and we evaluate its performance on a new 3D point cloud dataset collected in dynamic urban scenes by our driving platform. The experiments show that our method can recognize more kinds of labels and achieve an impressive result in dynamic urban scenes.**

I. INTRODUCTION

empty

II. RELATED WORKS

empty

III. METHODOLOGY

A. Data Preprocessing

The point cloud data for LiDAR is sparse and unorganized, so it is time-consuming to find neighboring relations between different points. In order to process these unorganized point cloud data with deep convolutional neural network, we convert the point cloud data into 2D range image by cylindrical projection. After that, it will be easier to implement deep convolution neural network on the LiDAR data.

After cylindrical projection, the point cloud data will be encoded as a dense matrix with shape of $[H, W, C]$. H means the number of lines for the specific LiDAR sensor (such as $H = 32$ for Velodyne HDL-32E). W equals to the number of points within each LiDAR scan line. C is the channels' number in the range image. Here, C is set to 3, which represents $[Range, Intensity, Height]$ three channels.

For each point $p_k = \langle x_k, y_k, z_k \rangle$ from the raw point cloud set P, the value of $Range$ in range image R is defined as r_k:

$$r_k = \sqrt{{x_k}^2 + {y_k}^2 + {z_k}^2}, r_k \subset [0, 255] \tag{1}$$

Similarly, the values of $Intensity$ and $Height$ are normalized into $[0, 255]$. These channels are very important properties of LiDAR data, which are enough to describe various objects in dynamic urban scenes.

B. Problem Definition

Let X denotes the range image extracted by cylindrical projection of 3D point cloud data S. In this kind of projection, there is a one-to-one correspondence between a 3D point in one frame point cloud data and a pixel in the range image X. As a result, the semantic segmentation task of 3D point cloud data is equal with giving each pixel x in the range image X a label y. The problem of this work is formulated as learning a semantic segmentation model f_θ which maps each pixel x to a label $y \in \{1, ..., K\}$, and subsequently associate y to the 3D points of S.

$$f_\theta : x \to y \in \{1, ..., K\} \tag{2}$$

The data samples are in the form of range images X. Given a set of supervised data samples $X_l = \{x_i, y_i\}$, where $\{x_i\}$ traverses each pixel of X and $\{y_i\}$ are labels for $\{x_i\}$, annotated manually by human annotators. In order to learning a semantic segmentation model f_θ, we need to find the best parameter set θ^* that minimize a loss function L as below.

$$\theta^* = \arg\max_\theta L(X_l; \theta) \tag{3}$$

C. Network Architecture and Loss Function

We use a FCN (Fully Convolutional Network) architecture for this semantic segmentation task. Compared with common deep convolutional networks, it removes last fully connected layers, and replaces them with the in-network up-sampled or deconvolutional predictions of convolutional layers as predicted feature maps. During training procedure, it generally computes cross-entropy like losses in pixel-wise, between the predicted labels and ground truths.

Our network is trained via end-to-end guided by the designed loss function. Because the number of pixels are imbalanced between different classes and a lot of invalid or unknown-class pixels, the following multi-class weighted cross entropy loss function is designed to regular the weights between imbalanced classes.

For labeled data X_l, we implement some changes on the widely-used definition of cross entropy, and define loss function L_l as below:

$$\Gamma_{i,j} = \begin{cases} \overrightarrow{\varphi_k}, & if \quad [y_{i,j} \neq k \quad and \quad y_{i,j} \neq 0] \\ \overrightarrow{0}, & otherwise \end{cases}$$

$$L_l(X_l, Y_l; \theta) = -\frac{1}{H * W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} \sum_{k=0}^{K} \Gamma_{i,j} \omega_k ln(P_\theta^k(x_{i,j})) \tag{4}$$

Where $\Gamma_{i,j}$ is a one-hot vector φ_k of label k, if $y_{i,j} \neq k$ and $y_{i,j} \neq 0$. Label 0 means invalid or unknown pixels, including many fine fragments belong to background or hard to be annotated, so we don't want to evaluate these pixels if they are predicted as non-zero labels. ω_k here is used to balance the sample numbers between different labels and $P_\theta^k(x_{i,j})$ is the probability that pixel $x_{i,j}$ be assigned a label k by our semantic segmentation model with the set of parameters θ.

IV. Experiment

A. Data Set

The performance of the proposed method is evaluated on a dynamic campus data set collected by an instrumented vehicle, which has a GPS/IMU suite and a Velodyne-HDL32, as shown in Fig. 1. The total route contains 1375 LiDAR frames. 880 frames for training, 220 frames for validation and 275 frames for testing.

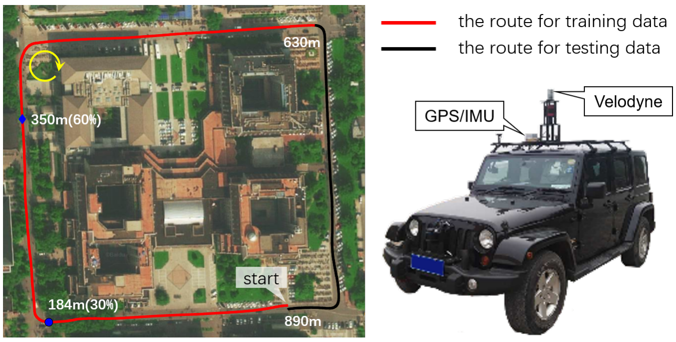

Fig. 1. The routes of data collection and the platform configuration.

TABLE I
CATEGORIES DISTRIBUTION IN DATASET

-	People	Car	Vegetation	Building	Road
Pixels	733,664	257,239	4,661,880	6,579,360	16,856,614

High quality pixel-level annotation is necessary for network training. Instead of working on the raw point cloud, human annotators work on the range image where object regions are associated with the ones in adjacent frames. Annotators only need to assign the category of some region in one frame, then a series of associated regions are marked with the same label. Although sometimes the data association brings errors, it largely reduce the annotation time.

The categories distribution in this dataset is shown in TABLE . I. Obviously, the data distribution is imbalanced between categories. So, we apply each label an unique weight based on data distribution to reduce the influence of data imbalance.

B. Setup

Our method is implemented with a FCN (Fully Convolutional Network). The range image size is 1080x32, which width is down-sampled for efficiency. A small batch size for training sets will be better. The network is implemented with TensorFlow in the environment with NVIDIA TITAN X GPU. We use the AdamOptimizer with 1e-5 learning rate.

It's important to aware that our data frames are captured sequentially. In order to avoid data correlation between adjacent frames, shuffling them before training process is necessary.

C. Preparing your Electronic Paper

Type sizes and typefaces: Follow the type sizes specified in Table II. As an aid in gauging type size, 1 point is about 0.35 mm. The size of the lowercase letter "j" will give the point size. Times New Roman is the preferred font.

1) US Letter Margins: top = 0.75 inches, bottom = 1 inch, side = 0.625 inches. Each column measures 3.5 inches wide, with a 0.25-inch measurement between columns.

2) A4 Margins: top = 19mm, bottom = 43mm, side = 13 mm. The A4 column width is 88mm (3.45 in). The space between the two columns is 4mm (0.17 in). Paragraph indentation is 3.5 mm (0.14 in).

TABLE II
TYPE SIZES FOR PAPERS

Type size (pts.)	Appearance		
	Regular	Bold	Italic
6	Table captions[1], table superscripts		
8	Section titles[1], references, tables, table names[1], first letters in table captions[1], figure captions, footnotes, text subscripts, and superscripts		
9		Abstract	
10	Authors' affiliations, main text, equations, first letters in section titles[1]		Subheading
11	Authors' names		
24	Paper title		

The column width is 82 mm (3.23 in). The space between the two columns is 6 mm (0.24 in). Paragraph indentation is 3.5 mm (0.14 in).

Left- and right-justify your columns. Use tables and figures to adjust column length. On the last page of your paper, adjust the lengths of the columns so that they are equal. Use automatic hyphenation and check spelling.

[1]Uppercase (**It is recommended that footnotes be avoided. Instead, try to integrate the footnote information into the text.**)

Fig. 2. Magnetization as a function of applied field.

V. Helpful Hints

A. Figures and Tables - Subsection Example

Position figures and tables at the tops and bottoms of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be centered below the figures; table captions should be centered above. Avoid placing figures and tables before their first mention in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

1) Sub-subsection example: Figure axis labels are often a source of confusion. Use words rather than symbols. For example, write "Magnetization", or "Magnetization, M", not just "M." Put units in parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization ($A \cdot m^{-1}$)." Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K."

Multipliers can be especially confusing. Write "Magnetization (kA/m)" or "Magnetization (10^3 A/m)". Figure labels should be legible, about 10-point type.

B. References

Number citations consecutively in square brackets [1]. Punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]. Use "Ref. [3]" or "Reference [3]" at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes (see Table 1). IEEE Transactions no longer use a journal prefix before the volume number. For example, use "IEEE Trans. Magn., vol. 25", not "vol. MAG-25".

Give all authors' names; use "et al." if there are six authors or more. Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. In a paper title, capitalize the first word and all other words except for conjunctions, prepositions less than seven letters, and prepositional phrases.

For papers published in translated journals, first give the English citation, then the original foreign-language citation [6].

C. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title unless they are unavoidable.

D. Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (5). To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use an en dash (-) rather than a hyphen for a minus sign. Use parentheses to avoid ambiguities in denominators. Punctuate equations with commas or periods when they are part of a sentence, as in

$$z = \sin^2 x + \cos^2 y. \tag{5}$$

Symbols in your equation should be defined before the equation appears or immediately following. Use "(5)," not "Eq. (5)" or "equation (5)," except at the beginning of a sentence: "Equation (5) is ..."

E. Other Recommendations

The Roman numerals used to number the section headings are optional. If you do use them, do not number ACKNOWLEDGMENTS and REFERENCES, and begin Subheadings with letters. Use two spaces after periods (full stops). Hyphenate complex modifiers: "zero-field-cooled magnetization." Avoid dangling participles, such as, "Using (5), the potential was calculated." Write instead, "The potential was calculated using (5)," or "Using (5), we calculated the potential."

Use a zero before decimal points: "0.25", not ".25". Use "cm^3" not "cc". Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter," not "webers/m^2". Spell units when they appear in text: "...a few henries", not "...a few H." If your native language is not English, try to get a native English-speaking colleague to proofread your paper. Do not add page numbers.

VI. Units

Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive."

Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally.

VII. Some Common Mistakes

The word "data" is plural, not singular. The subscript for the permeability of vacuum is zero, not a lowercase letter "o". In American English, periods and commas are within quotation marks, like "this period". A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you mean something that alternates). Do not use the word "essentially" to mean "approximately" or "effectively." Be aware of the different meanings of the homophones "affect" and "effect," "complement" and "compliment," "discreet" and "discrete," "principal" and "principle." Do not confuse "imply" and "infer." The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen. There is no period after the "et" in the Latin abbreviation "et al." The abbreviation "i.e." means "that is," and the abbreviation "e.g."

means “for example.” An excellent style manual for science writers is [7].

ACKNOWLEDGMENT

The preferred spelling of the word ”acknowledgment” in America is without an ”e” after the ”g”. Try to avoid the stilted expression, ”One of us (R.B.G.) thanks ...”. Instead, try ”R.B.G. thanks ...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized”, *J. Name Stand. Abbrev.*, in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.