

Métodos de *Deep Learning* aplicados à Segmentação Semântica de Imagens para Percepção de Veículos Autônomos

Gabriel Toffanetto França da Rocha

Laboratório de Mobilidade Autônoma – LMA

Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas

Campinas, Brasil

g289320@dac.unicamp.br

Abstract—

- Veículos autônomos
- Visão computacional
- Percepção do ambiente
- Segmentação Semântica de Imagem
- Métodos Vanilla
- Métodos Deep Learning
- Necessidades da aplicação
- Resultados
- Proximos passos (teste para obtenção do Perception grid)

Index Terms—Deep learning, Visão computacional, Segmentação Semântica de Imagem, Robótica móvel, Veículos autônomos

I. INTRODUÇÃO

Veículos com capacidade de se guiarem de forma autônoma estão cada vez mais presentes no dia a dia da sociedade contemporânea, possibilitando que o motorista possa realizar outras atividades durante a navegação, ou que o mesmo seja assistido em caso de alguma falha humana do condutor. Para que o automóvel seja capaz de se mover por conta própria, o mesmo deve ser capaz de perceber o ambiente, e sensores como sonares, radares, LiDARs e câmeras podem ser utilizados para tal. Porém, a câmera se faz como uma solução mais viável economicamente, e como visto na literatura, apresenta soluções que contemplam os desafios da navegação autônoma de veículos em ambientes urbanos, como visto nos trabalhos de [7] e [15].

Para que um veículo autônomo possa entender o ambiente à sua volta, é necessário que ele saiba reconhecer as entidades que o compõem, como por exemplo: estrada, veículos, calçadas, pedestres e vegetação, para que assim, o mesmo saiba diferenciar área navegável de obstáculos [11]. Para isso, o emprego da técnica de segmentação semântica de imagens, onde cada pixel da imagem é classificado de acordo com a entidade do ambiente da qual ele faz parte [9]. A Fig. 1 mostra a aplicação da técnica de segmentação semântica fundida à informação de profundidade dada por uma câmera *stereo*, permitindo a obtenção de um *grid* de percepção dinâmica local (DLP), que projeta no plano 2D o ambiente contendo a detecção de múltiplos objetos para que o veículo consiga planejar seu caminho [17].

Existem métodos de processamento de imagens que realizam o mascaramento de cada entidade da imagem, porém a

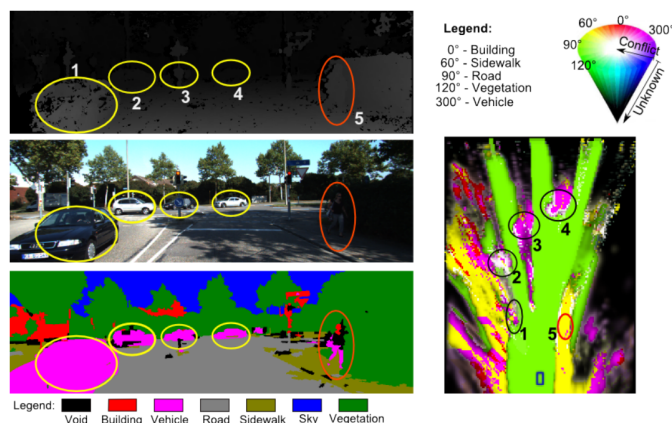


Fig. 1. DLP com ênfase na detecção múltipla de objetos móveis obtido com a fusão da imagem semanticamente segmentada e as informações de profundidade [17].

definição de qual é a classe de cada segmento se faz desafiadora, sendo anteriormente empregada a utilização de redes neurais artificiais (ANNs) para tal, como feito por [16]. Porém, devido à utilização de ANNs somente para a classificação final, era necessário muito pré-processamento para realização da segmentação semântica. Com o desenvolvimento das redes neurais profundas (DNNs), obteve-se métodos com poder suficiente para que, dada uma imagem bruta de entrada e uma imagem de referência (*ground truth*) segmentada para comparação, a rede profunda consegue aprender como realizar a segmentação da imagem do ambiente urbano, como nas várias arquiteturas mostradas por [13]. Com a popularização desses métodos, já existem diversos conjuntos de dados para treinamento das DNNs, como os [5] e [2], [3]. Existem também *datasets* que trazem cenas ainda mais desafiadoras, como o [12] que apresenta imagens urbanas durante noites chuvosas.

Dessa forma, a percepção do ambiente por meio de visão computacional se faz indispensável para o desenvolvimento dos veículos autônomos, e com isso, os métodos de *deep learning* se fazem uma grande ferramenta para conseguir-se reconhecer as entidades de uma cena urbana com robustez às variações de luz e reflexos, sendo assim uma solução a ser

explorada. Além do desempenho da segmentação semântica, o tempo demandado para tal operação também é vital, uma vez que durante a navegação, todos os módulos operam em tempo real, e a quantidade de *frames* segmentados por segundo é uma informação importante. Dessa forma, esse trabalho propõem a utilização de redes neurais profundas com diferentes arquiteturas para a segmentação semântica de imagens de cenas urbanas, utilizados *datasets* da literatura e para testes finais, utilizando imagens reais adquiridas pelo autor.

Este trabalho é dividido em seis partes, onde na Seção I é realizada a motivação e contextualização da pesquisa e na Seção II é apresentado o estado da arte, discorrendo sobre as soluções utilizadas atualmente. Com isso, a Seção III apresenta a metodologia a ser utilizada neste trabalho, seguida dos resultados obtidos na Seção IV e sua análise na Seção V. Por fim, são apresentadas as conclusões na Seção VI e as referências utilizadas.

II. ESTADO DA ARTE

- Artigos *Survey* [10]
- Tipos de redes utilizadas [1], [4], [6], [14], [18]–[20]
- Resultados estado-da-arte [13]
 - mIoU
 - FPS

III. METODOLOGIA

- Arquiteturas escolhidas
- Redes escolhidas
- *Datasets* escolhidos
 - Proposta de utilizar dados coletados no *campus*
- Método de treinamento
- Métricas utilizadas
- *Frameworks* utilizados
- *Hardware* utilizado

A. Arquitetura

B. Dataset de treinamento e dados de teste

C. Método de treinamento

D. Métricas de avaliação

E. Teste

IV. RESULTADOS

- Para cada rede:
 - Métricas
 - Segmentação
 - * Entrada
 - * *Ground truth*
 - * Saída
 - Tempo de treinamento

A. Métricas

B. Dados de teste

V. ANÁLISE DOS RESULTADOS

- Comparar as três redes:
 - mIoU
 - FPS
 - Dados segmentados do *dataset*
 - Dados segmentados coletados no *campus*
- Apontar relação custo *vs* desempenho de cada rede
- Considerar custos de treinamento

VI. CONCLUSÕES

- Retomar o problema inicial
- Destacar metodologia e os resultados que foram obtidos
- Comentar a análise dos resultados, mostrando que seria melhor para implementação
- Propor melhorias
- Propor validação de aplicação
- Listar proposta de aplicação dessa técnica
 - Perception grid

AGRADECIMENTOS

- Levy e Romis
- Giovani?

REFERÊNCIAS

- [1] BADRINARAYANAN, V., KENDALL, A., AND CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, Oct. 2016.
- [2] BROSTOW, G. J., FAUQUEUR, J., AND CIPOLLA, R. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters* 30, 2 (Jan. 2009), 88–97.
- [3] BROSTOW, G. J., SHOTTON, J., FAUQUEUR, J., AND CIPOLLA, R. Segmentation and Recognition Using Structure from Motion Point Clouds. In *Computer Vision – ECCV 2008*, D. Forsyth, P. Torr, and A. Zisserman, Eds., vol. 5302. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 44–57.
- [4] CHAO, P., KAO, C.-Y., RUAN, Y., HUANG, C.-H., AND LIN, Y.-L. HarDNet: A Low Memory Traffic Network. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul, Korea (South), Oct. 2019), IEEE, pp. 3551–3560.
- [5] CORDTS, M., OMRAN, M., RAMOS, S., REHFELD, T., ENZWEILER, M., BENENSON, R., FRANKE, U., ROTH, S., AND SCHIELE, B. The Cityscapes Dataset for Semantic Urban Scene Understanding, Apr. 2016.
- [6] FAN, M., LAI, S., HUANG, J., WEI, X., CHAI, Z., LUO, J., AND WEI, X. Rethinking BiSeNet For Real-time Semantic Segmentation, Apr. 2021.
- [7] GARCIA, O., VITOR, G. B., FERREIRA, J. V., MEIRELLES, P. S., AND DE MIRANDA NETO, A. The VILMA intelligent vehicle: An architectural design for cooperative control between driver and automated system. *Journal of Modern Transportation* 26, 3 (Sept. 2018), 220–229.
- [8] GÉRON, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, second ed. O'Reilly, 2020.
- [9] HE, H.-J., ZHENG, C., AND SUN, D.-W. Image Segmentation Techniques. In *Computer Vision Technology for Food Quality Evaluation*. Elsevier, 2016, pp. 45–63.
- [10] JANAI, J., GÜNEY, F., BEHL, A., AND GEIGER, A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. *Foundations and Trends® in Computer Graphics and Vision* 12, 1–3 (2020), 1–308.
- [11] JEBAMIKYOUS, H.-H., AND KASHEF, R. Autonomous vehicles perception (AVP) using deep learning: Modeling, assessment, and challenges. *IEEE Access* 10 (2022), 10523–10535.

- [12] JIN, J., FATEMI, A., LIRA, W., YU, F., LENG, B., MA, R., MAHDAVI-AMIRI, A., AND ZHANG, H. RaidaR: A Rich Annotated Image Dataset of Rainy Street Scenes, Oct. 2021.
- [13] PAPADEAS, I., TSOCHATZIDIS, L., AMANATIADIS, A., AND PRATIKAKIS, I. Real-Time Semantic Image Segmentation with Deep Learning for Autonomous Driving: A Survey. *Applied Sciences* 11, 19 (Sept. 2021), 8802.
- [14] POUDEL, R. P. K., BONDE, U., LIWICKI, S., AND ZACH, C. ContextNet: Exploring Context and Detail for Semantic Segmentation in Real-time, Nov. 2018.
- [15] VITOR, G. B. *Urban environment and navigation using robotic vision: conception and implementation applied to autonomous vehicle = Percepção do ambiente urbano e navegação usando visão robótica: concepção e implementação aplicado à veículo autônomo*. Doutor em Engenharia Mecânica, Universidade Estadual de Campinas, Campinas, SP, Sept. 2014.
- [16] VITOR, G. B., LIMA, D. A., VICTORINO, A. C., AND JANITO V. FERREIRA. A 2D/3D Vision Based Approach Applied to Road Detection in Urban Environments. In *2013 IEEE Intelligent Vehicles Symposium (IV)* (Gold Coast City, Australia, June 2013), IEEE, pp. 952–957.
- [17] VITOR, G. B., VICTORINO, A. C., AND FERREIRA, J. V. Modeling evidential grids using semantic context information for dynamic scene perception. *Knowledge-Based Systems* 215 (Mar. 2021), 106777.
- [18] WANG, Y., ZHOU, Q., AND WU, X. ESNet: An Efficient Symmetric Network for Real-time Semantic Segmentation, June 2019.
- [19] YU, C., GAO, C., WANG, J., YU, G., SHEN, C., AND SANG, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation, Apr. 2020.
- [20] YU, C., WANG, J., PENG, C., GAO, C., YU, G., AND SANG, N. BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation, Aug. 2018.