DESEMPENHO DE REDES CNN EM SEGMENTAÇÃO SEMÂNTICA PARA CANAVIAIS COM BASE EM DATASET DE PEQUENA ESCALA

Autor: Matheus Maquera Torres

Colaborador: Nícolas dos Santos Rosa

Orientador: Marco Henrique Terra

Escola de Engenharia de São Carlos - Universidade de São Paulo

matheusmaquera@gmail.com

Objetivos

Dentre as arquiteturas de aprendizado profundo recentes, as Redes Neurais Convolucionais (RNC) são destaque na visão computacional [1], seus pesos em formato de matriz aprendem a extrair padrões complexos analisando grupos de pixels. No contexto de veículos autônomos, esse tipo de rede neural é usado para auxiliar na percepção visual a partir da segmentação semântica, identificando áreas navegáveis, obstáculos e objetivos. Entretanto, grande parte dos dataset's usados nos treinos dessas redes são baseados em ambientes urbanos [2,3,4], o que não é representativo para contextos mais específicos como veículos autônomos de canaviais, por exemplo, que não possuem limites bem definidos entre suas classes e estão sujeitos a adversidades do clima e ambiente. Dessa forma, esta pesquisa visa a criação de um novo dataset representando a rotina de um caminhão transportador de cana em um canavial a fim de avaliar a performance de redes convolucionais do estado da arte considerando os requisitos de um veículo autônomo, técnicas de validação cruzada e metodologias para uma avaliação estatística robusta dado um dataset de pequeno porte.

Métodos e Procedimentos

Replicou-se a arquitetura DeepLabV3+ em PyTorch devido à sua eficiência inferencial, estrutura encoder-decoder e módulos ASPP para captura de contexto multiescala. Foram avaliados os backbones ResNet50, MobileNetV3-Large e Aligned Xception (esta última implementada *from scratch*). Utilizou-se um dataset próprio com 578 imagens anotadas (30 classes) de canaviais e vias não pavimentadas. Para evitar vazamento de dados, aplicou-se GroupKFold com grupos definidos por similaridade semântica: os *embeddings* das imagens foram reduzidos via PCA e clusterizados com DBSCAN (métrica cosseno), garantindo agrupamentos homogêneos por características visuais. Dada a limitação amostral, aplicou-se *data augmentation* com rotações aleatórias, cortes e inversões horizontais para aumentar a generalização. A avaliação de desempenho foi robustecida com *bootstrap* (1000 réplicas, IC 95%) sobre as métricas mIoU e *Dice Loss*, assegurando confiabilidade estatística aos resultados.

Resultados

A Tabela 1 apresenta os valores médios de mIoU e Dice Loss obtidos para cada backbone, juntamente com seus respectivos intervalos de confiança. A ResNet50 apresentou o maior valor médio de mIoU (23,7%), enquanto a MobileNetV3 alcançou desempenho intermediário e a Aligned Xception registrou o pior resultado médio, com maior Dice Loss (0,33). Apesar dessas diferenças, o desempenho geral permaneceu baixo para todas as arquiteturas, evidenciando as limitações impostas pelo tamanho reduzido do dataset e pela complexidade das 30 classes.

Tabela 1: Comparativo dos modelos (IC95%)

backbone	mIoU (%)	Dice Loss
Resnet50	23.7 (22.08 - 25.46)	0.24 (0.23 - 0.24)
Mobilenet V3	21.45 (20.20 - 22.82)	0.29 (0.29 - 0.30)
Xception	21.0 (19.55 - 22.25)	0.33 (0.32 - 0.33)

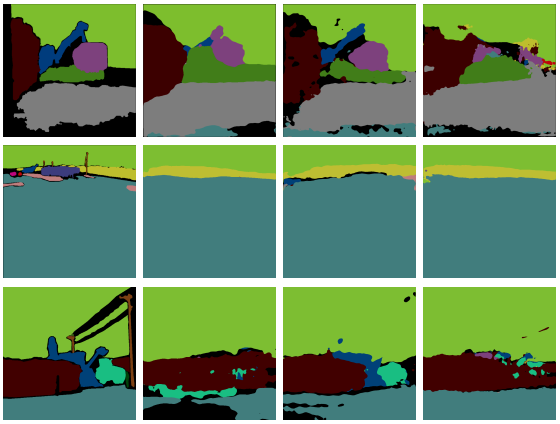

Figura 1: Comparação das máscaras de segmentação: referência (ground truth), MobileNetv3, ResNet50 e Xception.

A Figura 1 ilustra exemplos qualitativos de segmentação. Nota-se que as redes foram capazes de identificar regiões mais amplas e homogêneas, como céu e solo, mas tiveram dificuldade em discriminar objetos com contornos irregulares ou alta variabilidade, como folhas de cana e estruturas de maquinário.

Conclusões

No conjunto, os resultados indicam que, embora a DeepLabV3+ seja eficaz em ambientes urbanos, sua aplicação em canaviais requer cuidados adicionais, como datasets mais equilibrados e densos, de modo a reduzir o impacto do desbalanceamento entre classes e a melhorar a capacidade de generalização. Além disso, a análise comparativa sugere a necessidade de ponderar entre arquiteturas de *backbone* mais robustas, que oferecem maior acurácia, variantes otimizadas para tempo real, e ainda o uso de técnicas alternativas, como blocos de atenção ou redes baseadas em *transformers*, capazes de capturar dependências de longo alcance e auxiliar na classificação das diferentes classes em cenários agrícolas adversos.

Agradecimentos

Agradeço ao meu orientador pela oportunidade e confiança no desenvolvimento deste trabalho. Agradeço também ao colaborador pelo apoio técnico e pelas contribuições que enriqueceram este estudo.

Referências

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org

[2] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The Cityscapes Dataset for Semantic Urban Scene

Understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3213-3223).
[3] Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? The KITTI vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3354-3361).
[4] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. O., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 11621-11631).
[5] CHEN, Liang-Chieh et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2018, Munich. Proceedings... Cham: Springer, 2018. p. 801-818. (Lecture Notes in Computer Science, v. 11211).
[6] HE, Kaiming et al. Deep Residual Learning for Image Recognition. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016, Las Vegas. Proceedings... [S. l.]: IEEE, 2016. p. 770-778.
[7] HOWARD, Andrew et al. Searching for MobileNetV3. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2019, Seoul. Proceedings... [S. l.]: IEEE, 2019. p. 1314-1324.
[8] CHOLLET, François. Xception: Deep Learning with Depthwise Separable Convolutions. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2017, Honolulu. Proceedings... [S. l.]: IEEE, 2017. p. 1251-1258.