

Utilização de redes neurais e visão computacional para reconhecimento de características em campos agrícolas

Lucas Toschi de Oliveira

Vitor Akihiro Hisano Higuti

Marcelo Becker

Universidade de São Paulo (USP)

ltoschi@usp.br

Objetivos

O crescimento da população global demanda que os métodos e tecnologias aplicados à produção de alimentos sejam cada vez mais eficientes (Revichet al., 2016). Nesse contexto, o desenvolvimento de um sistema independente de navegação capaz de conduzir um robô dentro de uma plantação ainda é um desafio. Dentre as necessidades de um sistema de navegação autônomo, a criação de um mapa local com a tecnologia SLAM (“*Simultaneous Localization And Mapping*”) se destaca como um potencial objeto de estudo. Em locais dinâmicos, como o agrícola, esse método é mais suscetível a erros devido à típica suposição de ambientes estáticos adotada em abordagens tradicionais (Xiao et al., 2019).

Na busca de melhores resultados, o uso de *Deep Learning* na identificação dos objetos móveis da cena para eliminá-los do processo de mapeamento é uma alternativa promissora. Este projeto tem como **objetivo principal** a classificação e identificação das plantas com caixas utilizando *Deep Learning* como um passo fundamental para a futura aplicação de SLAM no ambiente agrícola.

Métodos e Procedimentos

Dentre os modelos existentes para realização da classificação e detecção de objetos, a rede YOLOv3 foi a escolhida para os primeiros testes do projeto. A principal razão foi a existência de implementações prontas para uso feitas pela comunidade online e os bons resultados já apresentados em aplicações com

SLAM em contextos diferentes (Qian et al., 2020). Foram adotadas as técnicas de *transfer learning* para a rede base (Darknet53), *image augmentation*, *batch-size training*, o otimizador de parâmetros Adam e um formato de imagem de entrada de 416x416 pixels.

Em relação ao processamento computacional, foram utilizados os serviços da plataforma *Amazon Web Services* (AWS). Para este projeto, adotou-se uma instância da máquina “p2.xlarge”. Ela possui 1 GPU (NVIDIA K80) com 12 GiB de memória e 4 CPU's com 61 GiB de memória.

As imagens utilizadas no projeto foram obtidas pelo robô agrícola móvel TerraSentia em campos de trigo. Ele foi desenvolvido por pesquisadores do Laboratório de Robótica Móvel (LabRoM – EESC/USP) e da Universidade de Illinois (Urbana-Champaign) e está retratado na Figura 1, juntamente com sua câmera monocular.



Figura 1: Robô TerraSentia e sua câmera.

Um novo *dataset* foi construído a partir desses dados utilizando a ferramenta *Computer Vision Annotation Tool* (CVAT). Ela possibilita a marcação das imagens com caixas padrões,

utilizadas no treinamento para otimizar os parâmetros da rede. O *dataset* final tem um total de mil imagens, sendo 750 delas utilizadas para treinamento e 250 para validação.

Resultados

Alguns exemplos de previsões da rede podem ser vistos na Figura 2. Nela, as imagens **a** e **b** representam exemplos onde a rede identificou plantas não previstas nas caixas padrão e as imagens **c** e **d** representam exemplos onde a rede falhou em identificar as caixas padrão.

Geralmente, em uma tarefa de detecção de objetos, utiliza-se uma métrica chamada de *Mean Average Precision* (**mAP**), responsável por medir quanto as previsões conseguiram se aproximar das caixas padrão. O valor obtido para o **mAP** foi de 65,70%. Essa métrica, porém, não é completamente adequada, já que, devido à subjetividade do processo de marcação do *dataset*, nem todas as plantas foram indicadas com caixas padrão em todas as imagens.

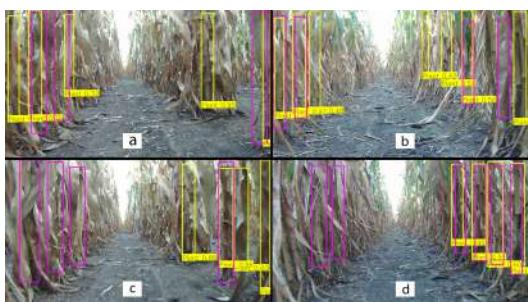


Figura 2: Algumas previsões da rede YOLOv3 (em amarelo) e caixas padrão (em rosa).

Para contornar esse problema, novas métricas foram definidas. Nas 250 imagens de validação, o número de caixas padrões (n_{gt}) e o número de previsões (n_{pred}) foram medidas. Além disso, foram contadas o número de boas caixas de previsão (n_{gtpred}) e o número delas que coincidem com as caixas padrão (n_{gtpred}). O termo “boa previsão” é definido pela inclusão da maioria dos pixels de uma planta. O número “total” de plantas identificáveis (n_{total}) pode ser definido pela Equação 1 e as novas métricas pelas Equações 2 a 6.

$$n_{total} = n_{gt} + (n_{pred} - n_{gtpred}) \quad (\text{Equação 1})$$

$$P = n_{gtpred}/n_{pred} \quad (\text{Equação 2})$$

$$C = n_{gt}/n_{pred} \quad (\text{Equação 3})$$

$$EB = (n_{gtpred} - n_{gt})/n_{pred} \quad (\text{Equação 4})$$

$$EBOT = (n_{gtpred} - n_{gt})/n_{total} \quad (\text{Equação 5})$$

$$NC = (n_{gt} - n_{gtpred})/n_{gt} \quad (\text{Equação 6})$$

A média dos cinco novos parâmetros em relação ao *dataset* de validação está retratada na Tabela 1.

mAP	mP	mC	mEB	mEBOT	mNC
65,70%	93,88%	62,61%	31,48%	26,18%	25,71%

Tabela 1: média das métricas do modelo YOLOv3

Observando a Tabela 1, é possível concluir que a precisão média (**mP**) do modelo é relativamente boa (93,88%). Apesar de apenas 62,61% (**mC**) coincidirem com as caixas padrão, 31,48% (**mEB**) das previsões eram boas e não coincidem com as caixas padrão. Considerando o total de plantas em cada imagem, 26,18% (**mEBOT**) das previsões não tem nenhuma referência anterior no *dataset*. No entanto, 25,71% (**mNC**) das caixas padrão não foram identificadas pelo modelo.

Conclusões

No desenvolvimento de algoritmos de SLAM com *Deep Learning*, este projeto propôs a identificação de características no meio agrícola. Os resultados indicam que o modelo YOLOv3 realizou a tarefa de detecção de objetos com sucesso. Para antigo melhor resultado, recomenda-se a expansão e diversificação do *dataset*, bem como a participação de mais pesquisadores em sua produção.

Este projeto foi desenvolvido nas dependências do Laboratório de Robótica Móvel (LabRoM, EESC-USP). Teve apoio da FAPESP (processos nº 2020/11262-0 e nº 2018/10894-2) e da empresa EarthSense pela doação de um exemplar do robô TerraSentia.

Referências Bibliográficas

- QIAN, Z. et al. Semantic SLAM with Autonomous Object-Level Data Association. 2020.
 REVICH, J. et al. Precision Farming - Cheating Malthus with Digital Agriculture. [S.I.], 2016. 4–5/18
 XIAO, L. et al. Dynamic-slam: Semantic monocular visual localization and mapping based on deep learning in dynamic environment. Robotics and Autonomous Systems, v. 117, p. 1 –16, 2019. ISSN 0921-8890.