

Projeto de Pesquisa

Iniciação Científica

Métodos de aprendizado profundo aplicados à segmentação de imagens

Aluno

Tibor Zequini

(Aluno do curso de Ciências Moleculares)

Orientadora

Nina S. T. Hirata

Departamento de Ciência da Computação

Instituto de Matemática e Estatística

Universidade de São Paulo

Resumo: A segmentação de imagens consiste em particionar o conjunto de pixels de tal forma que cada parte resultante corresponda a um elemento de interesse semântico presente na imagem. Trata-se de um dos principais processamentos em qualquer processo que envolve a análise de imagens digitais, típicos em problemas de visão computacional. Por outro lado, técnicas de aprendizado de máquina visam entre outras coisas o desenvolvimento de métodos que aprendem a transformação entrada-saída a partir de amostras dessa relação. Aprendizado profundo é o nome utilizado para técnicas sofisticadas de aprendizado que constituem o estado-da-arte na área. Neste projeto, o objetivo é a aquisição pelo estudante de conhecimentos sobre processamento de imagens, aprendizado de máquina, e técnicas de aprendizado profundo, por meio do estudo, implementação e avaliação de técnicas de aprendizado profundo aplicadas a problemas de segmentação de imagens.

São Paulo, 10 de maio de 2018

1 Introdução e justificativa

O processamento e a análise de imagens são fundamentais e ganham cada vez mais importância em problemas que envolvem visão computacional devido ao grande volume de imagens capturadas. As aplicações de visão computacional podem ser encontradas em diversas áreas, tais como em Astronomia, Oceanografia, diagnóstico médico, análise de documentos, entretenimento, entre outros.

Em qualquer problema que envolve uma análise de imagens há, tipicamente, a necessidade de se extrair das imagens as informações que são relevantes à aplicação em questão. Apesar dos grandes avanços em algoritmos e métodos para processamento e análise de imagens [Sze11, Pri12, Soi03], o processo de extração de informações ainda depende bastante da habilidade e experiência do usuário.

Uma das etapas de processamento importantes no processo de extração de informação é a segmentação. A segmentação de imagens consiste no particionamento do conjunto de pixels de uma imagem de tal forma que cada parte resultante corresponda a um elemento de interesse semântico presente na imagem. Um exemplo é mostrado na figura 1.

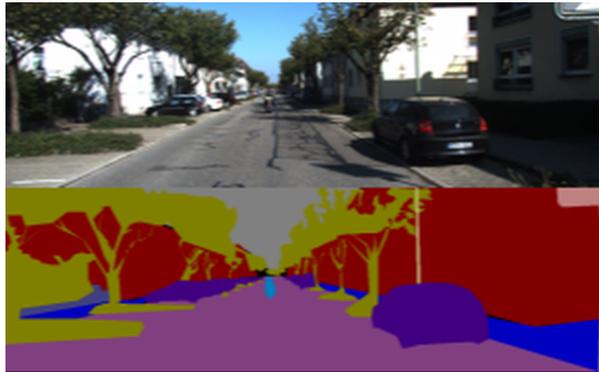


Figura 1: Um exemplo de segmentação de imagens. Acima, a imagem original e, abaixo, a imagem segmentada. Cada região na imagem segmentada corresponde a um elemento semântico de interesse na imagem original (imagem extraída de <http://adas.cvc.uab.es/elektra/enigma-portfolio/cvc10-semantic-segmentation-dataset/>).

O grande volume e variedade de imagens faz com que o desenvolvimento manual de algoritmos de segmentação específicos para cada caso seja uma solução extremamente ineficaz. Uma das abordagens utilizadas atualmente para se lidar com o volume e variedade

de dados é o aprendizado de máquina [HTF09, AMLMI12]. De forma geral, técnicas de aprendizado de máquina são capazes de “aprender” mapeamentos entre espaços de entrada e saída, a partir de amostras de pares entrada-saída do mapeamento esperado. No caso de segmentação de imagens, a formulação básica consiste em se utilizar algumas imagens originais e suas respectivas segmentações, preparadas manualmente, como exemplos a serem usados pelos algoritmos de aprendizado. Essa característica dos algoritmos de aprendizado, de serem guiados por dados, faz com que métodos baseados em aprendizado de máquina sejam uma abordagem ideal para se lidar com o grande volume e variedade de dados.

A área de aprendizado de máquina ganhou força nos últimos cinco a dez anos com a viabilização de técnicas que permitem o treinamento eficaz, isto é, o ajuste de parâmetros, de modelos altamente complexos de aprendizado de máquina. Especificamente, esses modelos complexos são variantes das redes neurais artificiais, envolvendo uma quantidade de camadas de processamento antes considerados impensáveis. O termo aprendizado profundo refere-se à grande quantidade de camadas dessas redes [GBC16].

2 Objetivos

O objetivo deste projeto de iniciação científica é o estudo de conceitos e algoritmos que fundamentam as áreas de processamento de imagens e de aprendizado de máquina.

Mais especificamente, no contexto de processamento de imagens, objetiva-se o estudo de segmentação de imagens. Para a execução da segmentação deverão ser estudados, no contexto de aprendizado de máquina, o problema de classificação e algumas técnicas usadas para a classificação. Essa combinação é pertinente uma vez que o problema de segmentação de imagens pode ser tratado como um problema de classificação individual dos pixels de uma imagem. Finalmente, o conhecimento de conceitos dessas duas áreas permitirá a compreensão de métodos de aprendizado profundo quando aplicados à segmentação de imagens, que será o principal objeto de investigação desse projeto.

3 Plano de trabalho

A seguintes atividades são consideradas para a execução deste projeto:

1. **Aquisição de conceitos de processamento de imagens:** poderá ser feito cursando-se uma disciplina, por meio de consulta às referências bibliográficas pertinentes, e principalmente por meio de participação e discussão no grupo de pesquisa em aprendizado de máquina do IME-USP.

Uma das tarefas concretas neste contexto será a busca e seleção, ou eventualmente a produção, de *datasets* de imagens para serem usadas no treinamento de algoritmos para segmentação de imagens.

2. **Aquisição de conceitos de aprendizado de máquina:** também poderá ser feito cursando-se uma disciplina, por meio de consulta às referências bibliográficas pertinentes, e participação e discussão no grupo de pesquisa em aprendizado de máquina do IME-USP.

Uma das tarefas concretas neste contexto será o estudo do livro *online* “Neural Networks and Deep Learning” de Michael Nielsen (<http://neuralnetworksanddeeplearning.com>), que permitirá ao aluno adquirir conhecimentos sobre redes neurais artificiais e alguma noção sobre redes neurais convolucionais, o modelo profundo usado em problemas de classificação.

Também serão trabalhados aqui a linguagem de programação Python, que será a principal linguagem a ser usada no desenvolvimento desse projeto.

3. **Estudo sobre segmentação de imagens:** será realizado um estudo por meio da aplicação prática do método previamente desenvolvido no grupo [MHH16, JAH17]. Para tanto deverá ser utilizada a biblioteca TRIOSlib (<https://trioslib.github.io>). Essa biblioteca foi desenvolvida pelo grupo de pesquisa ao qual a orientadora está associada, e pode ser aplicada em diversos problemas de processamento de imagens, incluindo a segmentação.

Com o uso dessa biblioteca será possível realizar a segmentação de imagens por meio de treinamento de classificadores de pixels. Especificamente, serão utilizados

classificadores profundos, baseados em redes neurais convolucionais.

4. **Segmentação de imagens usando técnicas profundas *end-to-end*:** embora técnicas profundas possam ser aplicadas para a classificação individual dos pixels, o interesse principal neste projeto de pesquisa são as abordagens *end-to-end* de técnicas de aprendizado profundo. Nessa abordagem, não se treinam classificadores de pixels individuais; treina-se uma rede neural profunda que mapeia uma imagem original em uma imagem que corresponde a sua segmentação. Tal mapeamento é realizado por meio de redes profundas conhecidas por *Fully convolutional networks* (FCN) e suas variantes [LSD15, RFB15, CPK⁺16, BKC17].

Do ponto de vista concreto, serão escolhidas algumas dessas redes para um estudo aprofundado. O estudo aprofundado consistirá de leitura de artigos relacionados, implementação da rede e sua aplicação (treinamento e avaliação) nos problemas de segmentação de imagens selecionados.

A parte de implementação deverá ser baseada na biblioteca TensorFlow (<https://www.tensorflow.org/>) e modelos dessas redes disponíveis na internet. A principal tarefa, do ponto de vista prático, será colocar esses modelos disponíveis para execução e, principalmente, aprender a realizar o ajuste dos parâmetros.

A avaliação poderá incluir a comparação de diferentes modelos, análise do efeito observado para diferentes parâmetros, entre outros.

Os conceitos necessários para a execução dessas tarefas serão estudados à medida que se mostrarem necessários.

5. **Escrita de textos técnicos e científicos:** uma parte das atividades será o treinamento de escrita de textos técnicos e científicos. Isto poderá ser praticado por meio da elaboração de projetos de pesquisa a serem submetidos para solicitação de bolsa de IC, elaboração de relatórios técnicos, e elaboração de artigos científicos.

O estudante terá oportunidades de vivenciar a comunicação técnica e científica não apenas na elaboração desses textos, mas também na exposição e discussões a serem realizadas no âmbito do grupo de pesquisa em aprendizado de máquina, que deverá se reunir semanalmente, e conta com vários participantes entre posdocs, alunos

de pós-graduação e de graduação, além dos professores responsáveis Nina Hirata e Roberto Hirata.

4 Cronograma

Este projeto de IC foi elaborado considerando-se uma vigência de 18 meses, a começar em fevereiro de 2018. Grosseiramente, as principais atividades e cronograma são conforme a seguir:

- (meses 1 a 3) Estudos iniciais sobre processamento de imagens e aprendizado de máquina
- (meses 4 a 6) Estudo da estrutura da biblioteca TRIOSlib e sua aplicação em algum problema de segmentação de imagens
- (meses 7 a 18) Estudo de técnicas *end-to-end* e sua aplicação em problemas de segmentação.

Referências

- [AMLM12] Yaser S. Abu-Mostafa, Hsuan-Tien Lin, and Malik Magdon-Ismael. *Learning From Data*. AMLBook, 2012. 2
- [BKC17] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, Dec 2017. 4
- [CPK⁺16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *arXiv preprint arXiv:1606.00915*, 2016. 4
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. 2
- [HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition, 2009. 2

- [JAH17] Frank D. Julca-Aguilar and Nina S. T. Hirata. Image operator learning coupled with cnn classification and its application to staff line removal. In *Submitted*, 2017. [3](#)
- [LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015. [4](#)
- [MHH16] Igor S. Montagner, Nina S. T. Hirata, and R. Hirata Jr. Image operator learning and applications. In *Conference on Graphics, Patterns and Images Tutoriais (SIBGRAPI-T)*, pages 38–50, 2016. [3](#)
- [Pri12] Simon D. J. Prince. *Computer Vision – Models, Learning and Inference*. Cambridge, 2012. [1](#)
- [RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. [4](#)
- [Soi03] P. Soille. *Morphological Image Analysis*. Springer-Verlag, Berlin, 2nd edition, 2003. [1](#)
- [Sze11] R. Szeliski. *Computer Vision – Algorithms and Applications*. Springer, 2011. [1](#)