

**Proposta de projeto de pesquisa  
para a disciplina MAC0215**  
Iniciação Científica

**Aprendizado de transformações de imagens via  
classificação de microrregiões**

Pedro Henrique Barbosa de Almeida  
**Estudante**  
N. USP 10258793

Nina S. T. Hirata  
**Orientadora**  
Contato: nina@ime.usp.br

Departamento de Ciência da Computação  
Instituto de Matemática e Estatística  
Universidade de São Paulo

São Paulo, 10 de março de 2019

# Introdução

Este documento descreve uma proposta de projeto de pesquisa a ser desenvolvido na disciplina MAC0215, no período de 18/02/2019 até 30/06/2019. Cabe observar que o estudante já realiza uma iniciação científica desde março de 2018, sendo com bolsa da FAPESP desde setembro de 2018. O projeto de pesquisa aqui proposto é parte do projeto de iniciação científica do estudante. Desta forma, é esperado que a maior parte das atividades de pesquisa a serem desenvolvidas pelo estudante sejam comuns a ambos os contextos (disciplina MAC0215 e iniciação científica com bolsa da FAPESP). Adequações com respeito à forma de apresentação dos resultados serão realizadas para cada um dos contextos.

Apresentamos inicialmente uma breve contextualização do trabalho sendo desenvolvido, com descrição das atividades já realizadas. Em seguida, descrevemos os objetivos assim como os métodos a serem explorados nos próximos quatro meses. Descrevemos também como o andamento do projeto poderá ser acompanhado pelos responsáveis pela disciplina. Por fim, detalhamos as atividades e o cronograma do plano de trabalho.

## 1 Contextualização do tema de pesquisa

O projeto de pesquisa apoia-se em dois conceitos importantes: a segmentação de imagens [Klavya, 2009, Gonzalez and Woods, 2002] e o aprendizado de operadores de imagens [Montagner et al., 2016].

A segmentação de imagens é um processamento comum em praticamente qualquer tarefa que envolve uma análise de imagens. Ela consiste em se obter um particionamento de uma imagem (isto é, de seus pixels) de tal forma que cada segmento (ou região) resultante corresponda a um componente de interesse. Por exemplo, em um software de reconhecimento de texto em imagens de documentos, separar componentes do tipo texto dos demais tipos é frequentemente um passo anterior ao reconhecimento propriamente dito.

Diversos outros tipos de processamento podem também ser necessários ou convenientes em uma análise de imagens. Os processamentos realizados em geral baseiam-se em uma combinação de vários tipos de transformações de imagens. A definição da combinação adequada dessas transformações, efeitos dos chamados operadores de imagens, é uma tarefa que demanda tempo e conhecimentos específicos [Montagner et al., 2016].

Vários desses operadores de imagens são transformações locais, caracterizados por uma função local. Por função local referimo-nos a uma função cuja entrada em geral é uma pequena região da imagem centrada num pixel. Essa função é aplicada pixel a pixel para se gerar a imagem transformada. Uma combinação finita dessas transformações locais é também uma transformação local. Desta forma, torna-se possível modelar o problema de projetar um operador (simples ou composto) como um problema de aprendizado dessas funções locais.

Usando-se pares entrada-saída de imagens que representam amostras da transformação desejada, temos um problema de classificação supervisionada no qual os exemplos do espaço de entrada são as regiões centradas em cada pixel da imagem e as classes do espaço de saída são os valores dos pixels correspondentes na imagem de saída.

Esta abordagem, formulada essencialmente como um problema de classificação de pixels, já vem sendo utilizada pelos membros do grupo de Visão Computacional do IME/USP a alguns anos [Montagner et al., 2016]. Um dos aspectos que precisam ser melhorados nessa abordagem é o tempo de processamento. Pelo fato do processamento consistir de aplicação pixel a pixel da função local (classificador), o custo computacional tende a ser elevado pois as imagens podem facilmente possuir mais de um milhão de pixels.

Uma abordagem comumente utilizada para evitar o tratamento individual de pixels é a representação de uma imagem por superpixels [Achanta et al., 2012]. De forma geral, superpixels são grupos de pixels adjacentes que compartilham características comuns (por exemplo, cores homogêneas). Superpixels são um tipo particular de segmentação na qual a imagem é supersegmentada em múltiplas pequenas regiões que são os superpixels. Mesmo sendo regiões pequenas, ao se passar da representação por pixels para a representação por superpixels, pode-se reduzir drasticamente o número desses componentes atômicos. Algoritmos projetados para tratar superpixels em vez de pixels tendem a ser bem mais eficientes. Note, porém, que o algoritmo para cálculo dos superpixels deve ser bastante eficiente para que haja um ganho de fato.

O projeto de pesquisa em andamento tem como proposta principal ampliar o escopo do problema de aprendizado de operadores de imagens para que possamos também executar o aprendizado no contexto dos superpixels. A principal diferença é que os classificadores locais passam a atuar sobre os superpixels (ou microrregiões).

## 2 Trabalhos realizados

As atividades realizadas desde o início da iniciação científica cobrem um espectro amplo incluindo estudos teóricos, estudo de algoritmos, implementação de algoritmos e aplicações visando teste das implementações.

**Formação básica:** Dado que o estudante começou a iniciação científica no início do seu segundo ano no curso, as atividades iniciais foram voltadas para o estudo de referências bibliográficas para a aquisição de conhecimentos e familiarização com a área de processamento de imagens.

O estudante realizou a leitura de diferentes materiais, e de forma concreta, em uma primeira etapa, focou no estudo dos operadores básicos da Morfologia Matemática [Soille, 2003] (abordagem utilizada para processamento de imagens). Além disso, realizou a implementação desses operadores, e experimentou os efeitos gerados ao se alterar seus parâmetros. Essas atividades foram suficientes para a aquisição de conceitos importantes tais como imagens, transformações de

imagens, operadores de imagens, e transformação local de imagens.

Em uma segunda etapa, o foco de estudos foram os algoritmos de segmentação de imagens [Klava, 2009, Achanta et al., 2012], visando gerar uma representação das imagens por superpixels.

Nos últimos meses o foco de estudo está localizado na área de aprendizado de máquina [Abu-Mostafa et al., 2012].

**Método para criação de superpixels:** O estudo de algoritmos de segmentação e sua experimentação sobre diferentes tipos de imagens mostrou que para as imagens consideradas os algoritmos testados são apenas parcialmente aceitáveis. A solução encontrada para gerar superpixels aceitáveis (com tamanhos regulares e aderência aos contornos naturais presentes na imagem) foi a combinação de dois algoritmos. Embora ainda não tenhamos obtido superpixels inquestionavelmente aceitáveis, a qualidade dos resultados obtidos até este momento foi avaliado como suficiente para prosseguir a pesquisa em direção ao desenvolvimento dos classificadores de superpixels.

### 3 Objetivos e resultados esperados

Considerando-se que o estudante já está bem familiarizado com processamento de imagens, tendo inclusive gerado a representação de imagens por superpixels, o próximo passo nesta pesquisa será o desenvolvimento de métodos para a classificação de superpixels.

Em particular, pretendemos utilizar um método que está sendo utilizado na tese de Ana Maia (aluna de doutorado) [Julca-Aguilar et al., 2017, Maia et al., 2018]. O desafio para o estudante será adquirir conhecimentos necessários para entender o método e adaptá-lo ao contexto considerado neste projeto.

Ao final, espera-se comparar o projeto de operadores que atuam no nível de pixels com o projeto de operadores que atuam no nível de superpixels. As comparações podem incluir tempo de processamento e também a qualidade dos resultados gerados pelos dois tipos de operadores. Pretendemos testar os métodos em tarefas de processamento de imagens de documentos.

### 4 Acompanhamento das atividades

O desenvolvimento desse projeto poderá ser acompanhado através da página <https://github.com/robonauta/IC/acompanhamento>, onde haverá a descrição das atividades relacionadas e, quando pertinentes, apontadores para os materiais gerados (código, *notebooks* e resumos).

A verificação dos estudos teóricos poderá ser feita através da leitura de resumos que serão produzidos.

Se os docentes responsáveis pela disciplina MAC0215 julgarem necessário, o estudante poderá elaborar relatórios que serão assinados pela orientadora e disponibilizados no mesmo link.

## 5 Plano de atividades e cronograma

1. Continuar o estudo sobre tópicos fundamentais relacionados à IC: segmentação e transformação de imagens, aprendizado de máquina, e aplicações. Em particular, o estudante está cursando neste semestre a disciplina MAC0460 - Introdução ao Aprendizado de Máquina e deverá ler parte do conteúdo da disciplina "Convolutional Neural Networks for Visual Recognition" da Universidade de Stanford<sup>1</sup>.
2. Desenvolvimento do método de treinamento sobre regiões. Esta atividade envolve desde a preparação de dados, o treinamento, e a geração de classificadores.
3. Aplicação e avaliação do método de treinamento sobre regiões e sobre *pixels*. O pipeline desenvolvido deverá ser aplicado sobre diferentes conjuntos de imagens.
4. Análise dos resultados. Os resultados obtidos da etapa anterior de experimentação deverão ser analisados.
5. Documentação. Ao final deverá ser gerado um relatório científico à FAPESP e à disciplina de MAC0215. Além disso, códigos gerados serão documentados. Caso pertinente, poderão ser elaborados artigos científicos descrevendo os métodos e os resultados.

Um cronograma proposto para essas atividades, com detalhamentos adicionais, é exibido a seguir.

1. Estudos sobre Redes Neurais Convolucionais			
Período coberto	Semana	Tarefas	Horas
25/02 a 03/03	1	Reunião semanal com o grupo de pesquisa	2
04/03 a 10/03	2	Reunião semanal com o grupo de pesquisa	2
		Leitura do Module 0 do material da disciplina CS231n <sup>1</sup> da Universidade de Stanford, incluindo os seguintes assuntos:  -Setup Instructions -Python / Numpy Tutorial -IPython Notebook Tutorial -Google Cloud Tutorial -AWS Tutorial	5

11/03 a 17/03	3	<p>Leitura do Module 1 do material da disciplina CS231n<sup>1</sup> da Universidade de Stanford, incluindo os seguintes assuntos:</p> <ul style="list-style-type: none"> <li>- Image Classification: Data-driven Approach, k-Nearest Neighbor, train/val/test splits: L1/L2 distances, hyperparameter search, cross-validation</li> <li>- Linear classification: Support Vector Machine, Softmax: parameteric approach, bias trick, hinge loss, cross-entropy loss, L2 regularization, web demo</li> <li>- Optimization: Stochastic Gradient Descent: optimization landscapes, local search, learning rate, analytic/numerical gradient</li> <li>- Backpropagation, Intuitions: chain rule interpretation, real-valued circuits, patterns in gradient flow</li> </ul>	5
		<p>Leitura do Module 1 do material da disciplina CS231n<sup>1</sup> da Universidade de Stanford, incluindo os seguintes assuntos:</p> <ul style="list-style-type: none"> <li>-Neural Networks Part 1: Setting up the Architecture: model of a biological neuron, activation functions, neural net architecture, representational power</li> <li>-Neural Networks Part 2: Setting up the Data and the Loss: preprocessing, weight initialization, batch normalization, regularization (L2/dropout), loss functions</li> <li>-Neural Networks Part 3: Learning and Evaluation: gradient checks, sanity checks, babysitting the learning process, momentum (+nesterov), second-order methods, Adagrad/RMSprop, hyperparameter optimization, model ensembles</li> <li>-Putting it together: Minimal Neural Network Case Study: minimal 2D toy data example</li> </ul>	5

18/03 a 24/03	4	<p>Leitura do Module 2 do material da disciplina CS231n<sup>1</sup> da Universidade de Stanford, incluindo os seguintes assuntos:</p> <ul style="list-style-type: none"> <li>- Convolutional Neural Networks: Architectures, Convolution / Pooling Layers: layers, spatial arrangement, layer patterns, layer sizing patterns, AlexNet/ZFNet/VGGNet case studies, computational considerations</li> <li>-Understanding and Visualizing Convolutional Neural Networks: tSNE embeddings, deconvnets, data gradients, fooling ConvNets, human comparisons</li> <li>-Transfer Learning and Fine-tuning Convolutional Neural Networks</li> </ul>	5
25/03 a 31/03	5	<p>Reunião semanal com o grupo de pesquisa</p> <p>Leitura do tutorial de Deep Learning com Keras<sup>2</sup>, feito François Chollet, incluindo os seguintes assuntos:</p> <ul style="list-style-type: none"> <li>- Chapter 2: <ul style="list-style-type: none"> <li>* 2.1: A first look at a neural network</li> </ul> </li> <li>Chapter 3: <ul style="list-style-type: none"> <li>* 3.5: Classifying movie reviews</li> <li>* 3.6: Classifying newswires</li> <li>* 3.7: Predicting house prices</li> </ul> </li> <li>-Chapter 4: <ul style="list-style-type: none"> <li>* 4.4: Underfitting and overfitting</li> </ul> </li> <li>-Chapter 5: <ul style="list-style-type: none"> <li>* 5.1: Introduction to convnets</li> <li>* 5.2: Using convnets with small datasets</li> <li>* 5.3: Using a pre-trained convnet</li> <li>* 5.4: Visualizing what convnets learn</li> </ul> </li> </ul> <p>Análise de códigos existentes produzidos pelo grupo de pesquisa para a classificação de elementos textuais em revistas</p>	<p>2</p> <p>5</p> <p>3</p>
2. Programação			
01/04 a 07/04	6	<p>Reunião semanal com o grupo de pesquisa</p> <p>Criação das redes neurais (ou reaproveitamento do código) para classificadores a nível de píxel</p>	<p>2</p> <p>5</p>
		Reunião semanal com o grupo de pesquisa	2

08/04 a 14/04	7	Criação das redes neurais para classificadores a nível de microrregiões Treinamento das CNNs	5 5
3. Experimentação			
15/04 a 21/04	8	Planejar e executar os experimentos a nível de pixel Planejar e executar os experimentos a nível de microrregiões	5 5
22/04 a 28/04	9	Reunião semanal com o grupo de pesquisa Definir métricas Ler artigos e comparar com algoritmos existentes	2 1 10
4. Análise dos resultados			
29/04 a 05/05	10	Reunião semanal com o grupo de pesquisa Discussão com a orientadora dos resultados obtidos Correção na experimentação de erros que possam ter comprometido os resultados	2 2 5
6. Documentação			
06/05 a 12/05	11	Reunião semanal com o grupo de pesquisa Elaboração de relatório final para a FAPESP	2 10
13/05 a 19/05	12	Reunião semanal com o grupo de pesquisa Elaboração de relatório final para a FAPESP	2 10
20/05 a 26/05	13	Elaboração de relatório final para a disciplina MAC0215	10
27/05 a 02/06	14	Reunião semanal com o grupo de pesquisa Elaboração de pôster para MAC0215	2 10
03/06 a 09/06	15	Reunião semanal com o grupo de pesquisa Elaboração de artigo científico	2 10
10/06 a 16/06	16	Reunião semanal com o grupo de pesquisa Elaboração de artigo científico	2 10
17/06 a 23/06	17	Reunião semanal com o grupo de pesquisa	2
24/06 a 30/06	18	Reunião semanal com o grupo de pesquisa	2
Total			150

<sup>1</sup><http://cs231n.github.io/>

<sup>2</sup><https://github.com/fchollet/deep-learning-with-python-notebooks>



## Referências

- [Abu-Mostafa et al., 2012] Abu-Mostafa, Y. S., Lin, H.-T., and Magdon-Ismail, M. (2012). *Learning From Data*. AMLBook.
- [Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282.
- [Gonzalez and Woods, 2002] Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Addison-Wesley Publishing Company, second edition.
- [Julca-Aguilar et al., 2017] Julca-Aguilar, F. D., Maia, A. L. M., and Hirata, N. S. T. (2017). Text/non-text classification of connected components in document images. In *30th Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 450–455.
- [Klava, 2009] Klava, B. (2009). Segmentação interativa de imagens via transformação watershed. Master’s thesis, Instituto de Matemática e Estatística - Universidade de São Paulo. <http://www.ime.usp.br/~klava/dissertacao/dissertacao.pdf>.
- [Maia et al., 2018] Maia, A. L. M., Julca-Aguilar, F. D., and Hirata, N. S. T. (2018). A machine learning approach for graph-based page segmentation. In *30th Conference on Graphics, Patterns and Images (SIBGRAPI)*.
- [Montagner et al., 2016] Montagner, I. S., Hirata, N. S. T., and Hirata Jr., R. (2016). Image operator learning and applications. In *Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 38–50.
- [Soille, 2003] Soille, P. (2003). *Morphological Image Analysis*. Springer-Verlag, Berlin, 2nd edition.