

APRENDIZADO DE TRANSFORMAÇÕES DE IMAGENS VIA CLASSIFICAÇÃO DE MICRORREGIÕES¹

Pedro H. B. de Almeida*²
Orientadora: Nina S. T. Hirata*

* Instituto de Matemática e Estatística da Universidade de São Paulo



INTRODUÇÃO

O tema de interesse é o problema de segmentação de textos em imagens de documentos e o uso técnicas de aprendizado profundo para solucioná-lo.



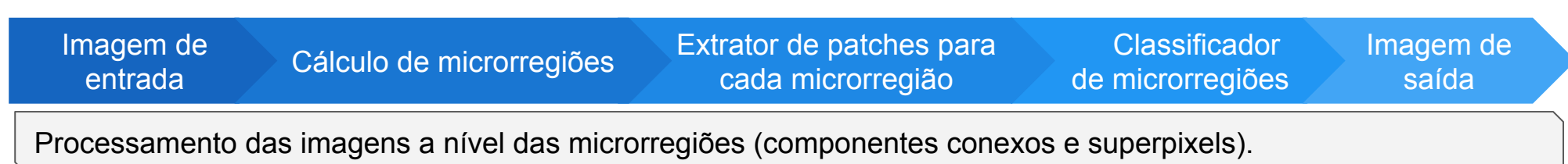
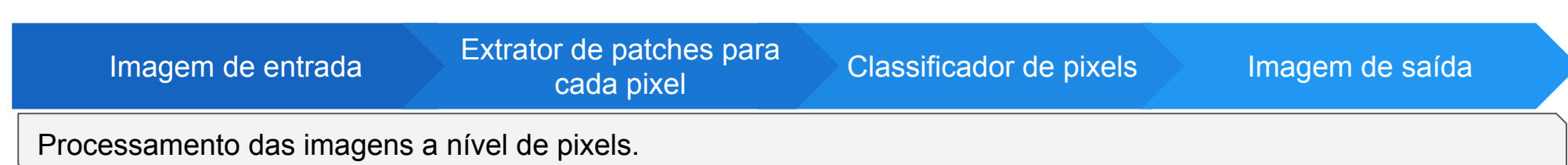
Exemplo de par de imagens entrada-saída usados para aprender a transformação Φ .

OBJETIVOS

Comparar a segmentação de imagens realizada via classificação pixel a pixel com a realizada via classificação de microrregiões.

METODOLOGIA

Pipelines propostos para classificação pixel a pixel e de microrregiões



Para o cálculo de microrregiões, foram considerados componentes conexos e superpixels calculados pelo algoritmo SLIC.

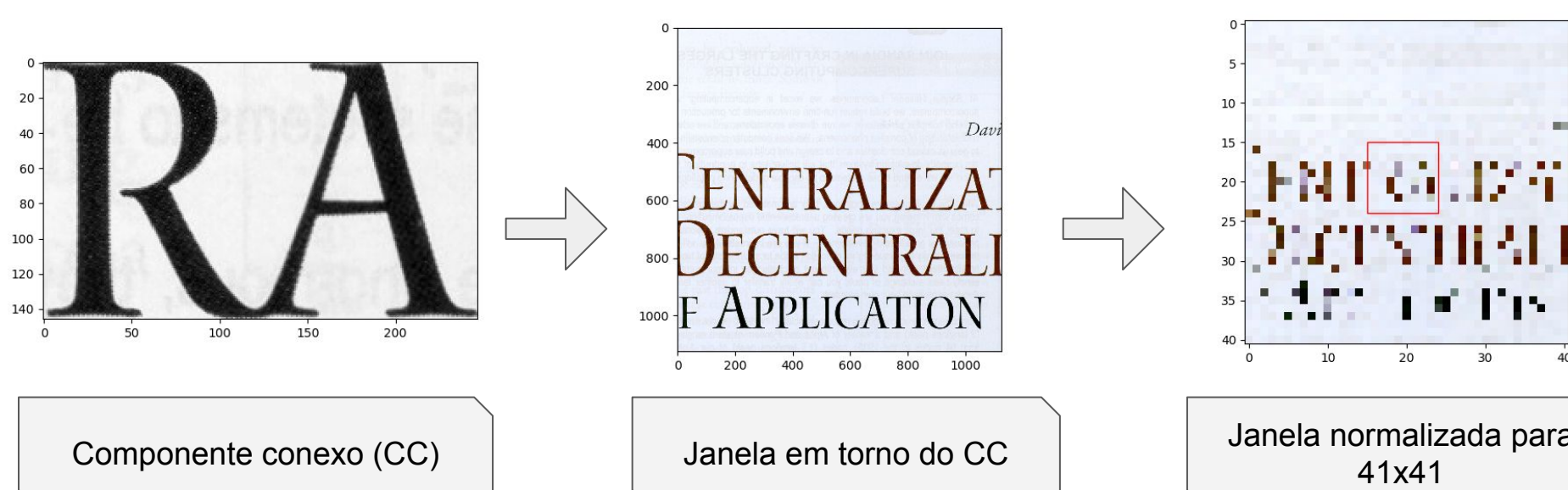


Componente conexo



Componente conexo segmentado em superpixels

Para a extração de patches, primeiro calcula-se a microrregião; em seguida, extrai-se uma janela em torno da microrregião; e por fim, normaliza-se todas as janelas para um tamanho fixo (41x41 para as microrregiões e 21x21 para o caso de pixels).



Por fim, para classificação, foi usada um CNN cuja entrada é um *patch* 41x41 (para as microrregiões) ou um *patch* 21x21 (para os pixels). A saída é 0 (não texto) ou 1 (texto). A implementação foi feita através da biblioteca Keras.

RESULTADOS

Primeiramente, o classificador foi treinado a partir de uma imagem de treinamento de dimensões 250x346x3. Em seguida, foi treinado em uma imagem de 2280x3257x3. Os resultados para as imagens de testes estão abaixo:

Conjunto de treinamento I			
	Pixels	Componentes conexos	Superpixels
Número de <i>patches</i>	86500	135	277
Tempo para extração de <i>patches</i> (segundos)	1,7740	0,2104	2,9262
Acurácia	0,9183	0,9832	0,9763
Precisão	0,8150	0,9857	0,9816
<i>Recall</i>	0,7692	0,9974	0,9940

Conjunto de treinamento II			
	Pixels	Componentes conexos	Superpixels
Número de <i>patches</i>	MemoryError	3322	16395
Tempo para extração de <i>patches</i> (segundos)	MemoryError	5,4418	62,4901
Acurácia	MemoryError	0,9959	0,9908
Precisão	MemoryError	0,9973	0,9969
<i>Recall</i>	MemoryError	0,9984	0,9889



DISCUSSÃO

- Número de patches reduzido em aproximadamente 99,84% no caso dos componentes conexos (CCs) e em 99,67% no caso dos superpixels (SPs).
- Desempenho da classificação de microrregiões superior ao da classificação pontual, sendo os CCs a granularidade que mais se destaca na acurácia.
- Poucos dados de treinamento e desbalanceamento de classes: apesar da alta acurácia, a rede falha em classificar corretamente os componentes da classe não-texto, cuja frequência é muito menor que componentes da classe texto.

REFERÊNCIAS

- 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.
- Karpathy, A. Class notes. CS231n: Convolutional Neural Networks for Visual Recognition.

¹ Projeto de Iniciação científica apresentando no contexto da disciplina MAC0215 - Atividade Curricular em Pesquisa.

² Pedro H. B. de Almeida foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) sob o processo nº 2018/11899-8.