

Título em Português: Elaboração de um Catálogo de Imagens de Vasos Sanguíneos para o Treinamento de Algoritmos de Aprendizado de Máquina

Título em Inglês: generation of a blood vessel image catalog for training machine learning algorithms

Autor: Natália de Carvalho Santos

Instituição: Universidade de São Paulo

Unidade: Instituto de Física de São Carlos

Orientador: César Henrique Comin

Área de Pesquisa / SubÁrea: Metodologia e Técnicas da Computação

Agência Financiadora: FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

Elaboração de um Catálogo de Imagens de Vasos Sanguíneos para o Treinamento de Algoritmos de Aprendizado de Máquina

Natália de Carvalho Santos

Cesar Henrique Comin

Universidade Federal de São Carlos

Universidade de São Paulo

naty_dcs@usp.br

Objetivos

O objetivo do presente projeto é *desenvolver uma base de imagens que possibilite a quantificação da performance de algoritmos de aprendizado de máquina em diferentes tipos de vasos sanguíneos*. A proposta é criar uma metodologia que possibilite a geração de imagens a partir da variação sistemática de diferentes características como comprimento, tortuosidade, contraste e nível de ruído na imagem. Tal metodologia possibilitará que a qualidade de qualquer algoritmo desenvolvido para a análise de vasos sanguíneos seja quantificada em diferentes situações. Especial foco será dado em imagens de microscopia confocal adquiridas a partir de amostras de cérebros de camundongos.

Métodos e Procedimentos

O projeto possui três principais etapas, sendo elas: definição e cálculo de propriedades relevantes para caracterizar imagens de vasos sanguíneos, seleção de imagens apropriadas para treinar algoritmos de segmentação e marcação manual das imagens. O enfoque são imagens de microscopia confocal, totalizando um total de 2800 imagens, que foram adquiridas sob diversas condições experimentais [1,2], sendo assim as imagens analisadas possuem grande variabilidade. Metodologias para a caracterização de vasos sanguíneos tipicamente envolvem quatro etapas principais. A partir de uma imagem original é realizada a segmentação gerando uma imagem binária, no qual é calculado o

esqueleto da vascularização, formado pelos eixos mediais dos vasos sanguíneos. Por último temos a representação em forma de grafo. Baseando nessas metodologias podemos analisar as propriedades dos vasos proposta pelo projeto. De maneira simplificada, a partir da imagem original e binária calculamos o nível de ruído e o contraste, e utilizando o esqueleto e o grafo extraímos o comprimento, a tortuosidade e a quantidade de bifurcações. As imagens a serem incluídas no catálogo serão selecionadas de forma manual, e o critério a ser utilizado será que elas possuam boa variação de contraste, nível de ruído, comprimento, tortuosidade e quantidade de bifurcações. Ademais, será feita uma análise cuidadosa nas segmentações realizadas e eventuais problemas encontrados serão corrigidos manualmente.

Resultados

O procedimento para selecionar as imagens foi extrair regiões de tamanho 256x256 das imagens originais juntamente com o grafo associado, sendo estas regiões em locais estratégicos das imagens. Após a obtenção de diversas imagens extraídas, nosso objetivo foi caracterizar as imagens de forma a criar uma base de dados possuindo imagens com variadas características. Investigamos diferentes definições das medidas consideradas de forma a selecionar as definições que fossem mais compatíveis com as imagens analisadas. O contraste foi calculado da seguinte forma :

$$C = \frac{I_v}{I_f} \quad (1)$$

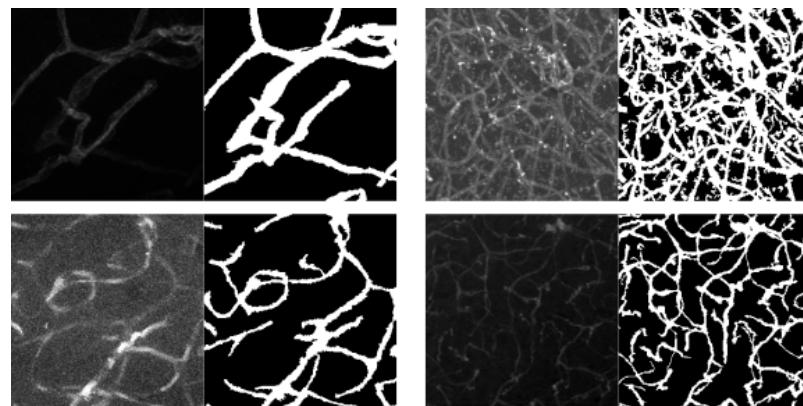


Figura 1: Exemplos de imagens extraídas até o momento

sendo \bar{I}_f e \bar{I}_v a intensidade média do fundo e do vaso respectivamente. O ruído foi calculado usando o método descrito em [3]. A estimativa da quantidade de bifurcações foi realizada utilizando os grafos das redes vasculares. O comprimento dos vasos foi calculado através do comprimento de arco do eixo medial dos vasos. A tortuosidade foi calculada através de estatísticas obtidas sobre a curvatura dos vasos, recentemente proposta e utilizada em [4]. Exemplos de imagens extraídas são mostradas na Figura 1. As imagens extraídas e caracterizadas foram escolhidas para cobrir da melhor forma possível as diferentes características que uma imagem de vasos sanguíneos de microscopia pode possuir. Isso possibilita o treinamento e teste de algoritmos de aprendizado de máquina em diferentes situações. Com isso, esperamos possibilitar que a qualidade de algoritmos de segmentação possam ser avaliados em diferentes situações de contraste, ruído e tipos de vasos, já que um algoritmo de segmentação deve possuir performance similar em qualquer região do espaço de atributos. Ademais, ao invés da performance ser calculada como uma média ao longo de todas as imagens, ele permite a identificação de possíveis situações nas quais um algoritmo de segmentação não gera resultados adequados.

Conclusões

Alterações e rupturas de vasos sanguíneos estão relacionadas com diversas doenças. Adicionalmente, a caracterização de vasos sanguíneos possui grande importância para o

entendimento do crescimento e atividade da vascularização de tecidos celulares e de neurônios. Portanto, atualmente há uma alta demanda por métodos computacionais que possibilitem identificar e caracterizar com precisão vasos sanguíneos contidos em imagens digitais. Tais métodos necessitam de bases de imagens para serem ajustados. Até o momento a base sendo desenvolvida possui 16 imagens organizadas de acordo com diferentes características. Essa organização possibilitará a verificação sistemática da performance de algoritmos de aprendizado de máquina em diversas situações experimentais. A base gerada será de grande utilidade para que pesquisadores desenvolvam novos algoritmos para o estudo de vasos sanguíneos.

Referências Bibliográficas

- [1] Naomi C Boisvert, et al. Hyperfiltration in ubiquitin c-terminal hydrolase l1-deleted mice. *Clinical Science*, 132(13):1453-1470, 2018.
- [2] Esther Kur, et al. Temporal modulation of collective cell behavior controls vascular network topology. *Elife*, 5:e13212, 2016.
- [3] David L Donoho , et al. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.
- [4] Matheus V da Silva, et al. An analysis of the influence of transfer learning when measuring the tortuosity of blood vessels. *Computer Methods and Programs in Biomedicine*, page 107021, 2022.

Generation of a Blood Vessel Image Catalog for Training Machine Learning Algorithms

Natália de Carvalho Santos

Cesar Henrique Comin

Federal University of São Carlos

University of São Paulo

naty_dcs@usp.br

Objectives

The objective of the project is to *develop an image database for the quantification of the performance of machine learning algorithms in different types of blood vessels*. The proposal is to develop a methodology that allows the generation of images while systematically varying different characteristics such as length, tortuosity, contrast and noise level of the images. Such a methodology will allow the quantification of the performance of algorithms developed for the analysis of blood vessels in different situations. Special focus will be given to confocal microscopy images acquired from mouse brain samples.

Materials and Methods

The project has three main steps, namely: definition and calculation of relevant properties to characterize blood vessel images, selection of appropriate images to train segmentation algorithms and manual image tracing. The main focus of the analysis is confocal microscopy images, totaling a total of 2800 images, which were acquired under different experimental conditions [1,2]. Methodologies for the characterization of blood vessels typically involve three main steps. From an original image, the segmentation is performed, generating a binary image for which the skeleton of the vascularization, formed by the medial axes of blood vessels, is calculated. Lastly, the skeleton is represented as a graph.

Based on these methodologies, the proposed properties of the vessels can be quantified. The noise level and the contrast are calculated from the original and binary image, while the skeleton and the graph are used for calculating the length, the tortuosity and the number of bifurcations. The images to be included in the catalog will be selected manually, and the criterion to be used will be that they show a good variation in contrast, noise level, length, tortuosity and number of bifurcations. In addition, the segmentations will be carefully analyzed and any problems found will be manually corrected.

Results

The procedure used to select the images was to extract regions of size 256x256 from the original images together with the associated graph, these regions being in relevant locations of the images. After obtaining several extracted images, our objective was to characterize the images in order to create a database containing images with different characteristics. We investigated different definitions of the measures considered in order to select the definitions that were most compatible with the analyzed images. The contrast was calculated as follows:

$$C = \frac{\bar{I_v}}{\bar{I_f}} \quad (1)$$

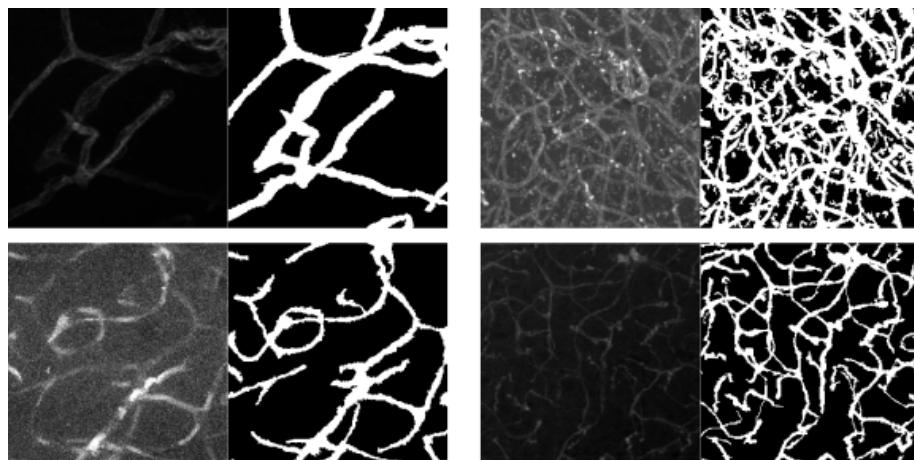


Figure 1: Examples of extracted images.

where $\bar{I_f}$ and $\bar{I_v}$ are the mean intensity of the background and the vessel, respectively. The noise was calculated using the method described in [3]. The length was calculated using the arc length of the medial axis of the vessels. The tortuosity was obtained through statistics of the curvature of the vessels, recently proposed and used in [4]. Examples of extracted images are shown in Figure 1. The extracted images were chosen to best cover the different characteristics that a microscopy image of blood vessels can possess. This makes it possible to train and test machine learning algorithms in different situations. Thus, we hope to enable the performance of segmentation algorithms to be evaluated in different situations of contrast, noise and types of vessels, since a segmentation algorithm must have similar performance in any region of a feature space. In addition, instead of the performance being calculated as an average over all the images, the proposed methodology allows the identification of possible situations in which a segmentation algorithm does not generate adequate results.

Conclusions

Changes and ruptures of blood vessels are related to several diseases. Additionally, the characterization of blood vessels is of great importance for understanding the growth and vascularization activity of cellular tissues and neurons. Therefore, there is currently a high demand for computational methods that make it

possible to accurately identify and characterize blood vessels contained in digital images. Such methods need image databases to be developed. So far, the database being created contains 16 images organized according to different characteristics. The complete database will enable the systematic verification of the performance of machine learning algorithms in several experimental situations. The generated base will be very useful for researchers to develop new algorithms for studying blood vessels.

References

- [1] Naomi C Boisvert, et al. Hyperfiltration in ubiquitin c-terminal hydrolase l1-deleted mice. *Clinical Science*, 132(13):1453-1470, 2018.
- [2] Esther Kur, et al. Temporal modulation of collective cell behavior controls vascular network topology. *Elife*, 5:e13212, 2016.
- [3] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.
- [4] Matheus V da Silva, Julie Ouellette, Baptiste Lacoste, and Cesar H Comin. An analysis of the influence of transfer learning when measuring the tortuosity of blood vessels. *Computer Methods and Programs in Biomedicine*, page 107021, 2022.