

SEGMENTAÇÃO DE BACIOS DE TUBERCULOSE EM IMAGENS DE MICROSCOPIA CONVENCIONAL USANDO REDES NEURAIAS CONVOLUCIONAIS

PHILIPPE RANGEL DEMUTH* PATRICK MARQUES CIARELLI† JORGE LEONID ACHING SAMATELO‡

*Programa de Pós Graduação em Engenharia Elétrica
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES, Brasil

Email: philiperd@aluno.ufes.br, patrick.ciarelli@ufes.br, jorge.samatelo@ufes.br

Abstract— Tuberculosis is one of the ten leading causes of death worldwide. Although it can be treated, the disease must first be diagnosed and for this, manual inspection processes are commonly used. Such processes are usually time consuming and tiring. In this work, we propose a method to achieve microscopy image segmentation, in which a convolutional neural network does a pre-segmentation of the image and the Graphcut algorithm does the post-processing in order to achieve finer results. The technique was evaluated over a dataset with 120 images obtained using conventional microscopy microscope in sputum smear sheets prepared according to the Ziehl-Neelsen technique. In the experiments carried out we obtained an accuracy of 96.70% and sensitivity of 96.75%. These results show that the method proposed has the potential to handle the given problem.

Keywords— Image Processing, Deep Learning, CNN, Optical Bacilloscopy, Ziehl-Neelsen, Conventional Microscopy, Tuberculosis Bacilli

Resumo— A tuberculose é uma das dez maiores causas de morte no mundo. Embora ela possa ser tratada, a doença primeiro deve ser diagnosticada, e para isso é comum o uso de processos de inspeção manual, em geral demorados e cansativos. Assim, neste trabalho é proposta uma técnica de segmentação de imagens de microscopia, na qual uma rede neural convolucional realiza uma espécie de pré-segmentação da imagem e o algoritmo *Graphcut* realiza o pós-processamento em busca de um maior refinamento dos resultados. A técnica foi avaliada em uma base de dados composta por 120 imagens obtidas por meio de um microscópio de microscopia convencional em lâminas de esfregaço de escarro preparadas segundo a técnica de Ziehl-Neelsen. Nos experimentos realizados foi obtida uma acurácia na segmentação de 96,70% e uma sensibilidade de 96,75%. Estes resultados mostram que a abordagem proposta tem potencial para tratar a tarefa indicada.

Palavras-chave— Processamento de imagens, *Deep learning*, CNN, Bacilloscopia óptica, Ziehl-Neelsen, Microscopia convencional, Bacilos de tuberculose

1 Introdução

Segundo o Centro de Controle e Prevenção de Doenças (CDC - *Centers for Disease Control and Prevention*), dos Estados Unidos, a tuberculose é uma doença infecciosa causada pelo bacilo *Mycobacterium tuberculosis*. Ela normalmente afeta o pulmão, mas também pode atacar qualquer parte do corpo, como o fígado, espinha e cérebro (CDC-Tuberculosis, 2018).

Os problemas de saúde causados pela tuberculose afetam aproximadamente 10 milhões de pessoas a cada ano e é uma das dez maiores causas de morte no mundo. Nos últimos 5 anos, ela foi a maior causa de morte no mundo por um único agente infeccioso, superando HIV/AIDS (WHO et al., 2017). Apesar desse fato, se o diagnóstico for feito no tempo correto, a maioria das pessoas que desenvolverem a tuberculose podem ser curadas (WHO et al., 2017).

Para ser tratada, a doença primeiro deve ser diagnosticada. O diagnóstico geralmente é feito colorindo uma lâmina de microscópio usando o método de coloração de Ziehl-Neelsen, que depois é analisada por um especialista, usando um microscópio óptico na busca de bacilos (Smart, 2017). Esta é uma técnica simples, rápida e barata, que é extremamente específica em áreas com

grande prevalência de tuberculose (ISTC, 2006). Dois métodos de microscopia podem ser utilizados para a detecção de bacilos de tuberculose: a microscopia convencional e por fluorescência. Embora a microscopia por fluorescência tenha sensibilidade em torno de 10% maior quando comparada à microscopia óptica, ela é menos comum em países em desenvolvimento, por ter um custo mais elevado (ISTC, 2006).

O processo de inspeção que inclui, além da verificação, a contagem dos bacilos, em geral é demorado e cansativo. Por isso, um sistema de reconhecimento automático de bacilos de tuberculose seria interessante, permitindo que o processo de diagnóstico se tornasse mais rápido. Para efetuar tal tarefa, um passo fundamental é a segmentação da imagem de microscopia. Entende-se por segmentação de uma imagem a tarefa de subdividir a imagem nas suas regiões constituintes. O modo que as subdivisões são feitas depende do problema a ser resolvido (Gonzalez and Woods, 2012). Para o caso da segmentação em duas classes (binária), uma imagem binária (que classifica cada pixel da imagem em uma de duas classes) é gerada a partir da imagem original. Neste caso, é entendido como objeto binário todo conjunto de pixels na imagem binária que correspondem a um bacilo ou agrupamento de bacilos.

Neste artigo é proposta uma técnica de segmentação que se divide, resumidamente, em duas etapas. Na primeira é utilizada uma rede neural convolucional (CNN – *Convolutional Neural Network*) que opera como um preditor de classes gerando uma imagem de probabilidades, denominado *heatmap*. Os valores dos pixels de um *heatmap* estão na faixa entre $[0, 1]$ e representam a probabilidade de um pixel pertencer a um bacilo. Na segunda etapa é usado o algoritmo *Graphcut* (Boykov et al., 2001). Tal algoritmo opera como segmentador binário, gerando a imagem segmentada final, além de eliminar numerosos artefatos presentes no *heatmap*. Resultados experimentais sobre uma base de dados indicaram que a técnica proposta obteve resultados com uma acurácia na segmentação de 96,70% e sensibilidade de 96,75%.

O restante deste artigo está estruturado como segue: na Seção 2 são mencionados os trabalhos relacionados, focando nas técnicas utilizadas em diferentes pesquisas e nos resultados obtidos; na Seção 3 é apresentada a proposta do trabalho explicando em detalhes a técnica desenvolvida; na Seção 4 são mostrados os resultados experimentais e finalmente na Seção 5 são descritas algumas conclusões e trabalhos futuros.

2 Trabalhos Relacionados

Embora os primeiros trabalhos de segmentação de imagens de microscopia tenham usado a técnica de fluorescência (Veropoulos et al., 1998; Forero et al., 2003), estes não serão mencionados por envolver um método de captura de imagens de custo mais elevado e consequentemente menos usado em países em desenvolvimento.

Sadaphal et al. (2008) propõem uma segmentação automática multiestágio bayesiana baseada em cor que identifica todos os possíveis objetos que poderiam ser bacilos de tuberculose. Os objetos segmentados são classificados pela sua forma, dividindo-os em três classes: “bacilos”, “possivelmente bacilos” e “não bacilos”. Para essa classificação é usada uma árvore de decisão, que decide a classe da estrutura segmentada baseada no comprimento de seu maior eixo e sua excentricidade. Entretanto, nenhuma avaliação quantitativa dos resultados é realizada.

CostaFilho et al. (2012) propõem usar a diferença entre os canais de cor R (vermelho) e G (verde) para segmentar a imagem usando uma técnica de segmentação global por limiar adaptativo. Os artefatos presentes na imagem segmentada são eliminados usando filtros morfológicos, filtros de cor e filtros de tamanho. Os objetos binários com menos de 200 pixels são eliminados. Os resultados alcançados foram de 76,65% e 12% para a sensibilidade e precisão, respectivamente. No entanto, não são fornecidos detalhes de como esses valores foram alcançados.

Siena et al. (2012) usam *decorrelation stretch* (Chitade and Katiyar, 2010) como etapa de pré-processamento seguido de uma técnica de segmentação baseada no algoritmo de clusterização *k-means*. Foram analisadas (testadas) nove imagens do CDC, enquanto que as imagens de treinamento foram as usadas por Forero et al. (2003). Nesse trabalho não são apresentados resultados quantitativos em relação à segmentação.

Chayadevi and Raju (2014) apresentam uma técnica de segmentação por cor usando o algoritmo de *Watershed* nos espaços de cores *YCbCr*, *HSI* e *Lab*. Em seguida, os objetos binários são analisados por características de forma, como área, perímetro, compacidade, excentricidade, eixos maior e menor, a razão entre os eixos e alguns de seus momentos invariantes. Entretanto, nenhuma avaliação quantitativa é realizada.

Em geral, os trabalhos citados têm as seguintes restrições: usam técnicas globais de segmentação dependentes de limiares (CostaFilho et al., 2012), precisam de etapas de pós-processamento fortemente parametrizadas (Sadaphal et al., 2008; CostaFilho et al., 2012; Chayadevi and Raju, 2014) e em alguns casos não são apresentadas análises quantitativas dos resultados (Sadaphal et al., 2008).

Diferentemente, a proposta deste trabalho está baseada no uso de uma rede CNN atuando como um preditor de classes binárias e o algoritmo *Graphcut* atuando como um segmentador binário. Ambos os algoritmos, por extraírem informações dos próprios dados, não requerem ajustes manuais dos parâmetros de segmentação: eles os “aprendem” na etapa de treinamento, gerando uma alta acurácia na segmentação.

Em relação ao uso de Redes CNN na tarefa de segmentação de imagens médicas, na literatura pode-se encontrar diferentes aplicações. Por exemplo, Chen and Ched'Hotel (2014) fazem uso de uma CNN para segmentar células do sistema imunológico presentes em lâminas de microscópio com amostras de tumores. Essas células são destacadas usando coloração imuno histoquímica (IHQ). Nesse trabalho é alcançado um coeficiente de correlação entre o número de células contadas pelo algoritmo e o número de células contadas manualmente de 99,49%.

Vivanti et al. (2015) propõem o uso de técnicas de segmentação na detecção de tumores no fígado usando imagens de tomografia computadorizada. Segundo os autores, a taxa de acerto é de 90,47%, superando o estado da arte até então.

A proposta de Xu and Huang (2015) também usa uma rede CNN para segmentar imagens médicas. Nesse caso, a segmentação é usada para identificar células cancerígenas. A base de dados consiste em imagens histopatológicas anotadas manualmente por um patologista treinado. O resultado desse trabalho superou o estado da arte

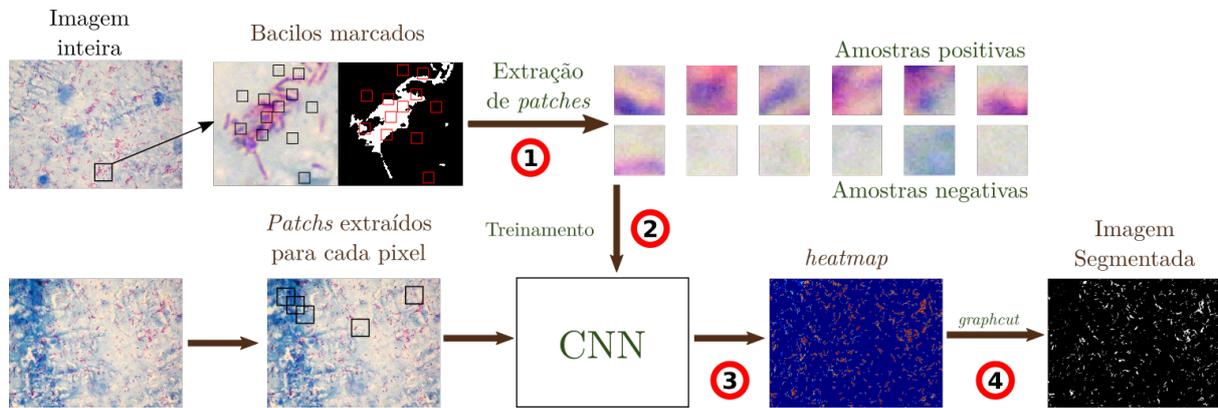


Figura 1: *Framework* do algoritmo implementado.

tanto em acurácia na detecção quanto em tempo de detecção.

O algoritmo *Graphcut* é um algoritmo de segmentação multiclasse que costuma ser usado em conjunto com arquiteturas de redes neurais profundas para o problema de segmentação semântica (Shakeri et al., 2016) (Lu et al., 2017). Em relação ao uso do algoritmo *Graphcut* como segmentador binário, existem diferentes aplicações na área médica. Por exemplo, Mahapatra (2017) usa essa técnica para obter consenso entre a marcação feita na base de dados por diferentes especialistas. Em casos em que a base de dados de segmentação de imagens médicas foi marcada por mais de um especialista, é necessário consenso entre as anotações.

3 Técnica Proposta

Nesta seção é descrita a técnica proposta para o problema de segmentação de imagens de microscopia. Tal proposta está baseada em duas etapas, especificamente:

- **Preditor de Classe.** Usando uma metodologia baseada em *patches*, uma rede CNN é treinada para prever a probabilidade de cada pixel da imagem de entrada pertencer a uma das duas possíveis classes: fundo ou objeto de interesse (bacilos ou conjunto de bacilos). A rede já treinada é aplicada a cada pixel da imagem de microscopia gerando como saída uma imagem de probabilidades, denominada *heatmap*.
- **Segmentador Binário.** Usando a imagem de microscopia e o *heatmap* como entradas para o algoritmo *Graphcut*, a imagem segmentada final é gerada.

Ambas as etapas serão explicadas em detalhes nas próximas subseções.

3.1 Preditor de Classe

As técnicas baseadas em *patches* têm sido cada vez mais usadas no campo de imagens médicas, com aplicações diversas em segmentação de imagens, *denoise* de imagens, registro de imagens, detecção de anormalidades e síntese de imagens (Wu et al., 2015). Soluções baseadas em redes neurais profundas, como CNN, demandam uma grande quantidade de amostras, neste caso imagens, para um treinamento adequado. Tal premissa é difícil de alcançar na área médica, devido as bases de dados de origem médica conterem normalmente poucas imagens, produto da necessidade de contar com um perito para efetuar a marcação manual das imagens.

Tal limitante é contornado usando *patches*, já que, a partir de cada imagem são extraídas várias regiões menores (os *patches*), aumentando assim a quantidade de amostras de treinamento para a rede CNN. Neste tipo de abordagem a segmentação da imagem é feita em cada pixel separadamente, ao invés da imagem toda por vez.

A metodologia proposta baseada em *patches* possui quatro passos principais resumidos na Figura 1 e explicados abaixo:

- ① Para cada uma das imagens de treinamento são extraídos aleatoriamente *patches* em que o pixel central faz parte de um bacilo (amostras **positivas**) ou pertence ao fundo (amostras **negativas**). As características desses *patches* serão explicadas em detalhes na Subseção 4.2.
- ② A rede CNN é treinada usando os *patches* do passo 1. A arquitetura da rede é descrita na Subseção 3.1.1.
- ③ Uma vez treinada, a CNN é aplicada sobre cada pixel da imagem de microscopia, onde a entrada da rede é um *patch* centrado no pixel a ser analisado e a saída é a probabilidade do pixel pertencer a um bacilo. O resultado final deste passo é o *heatmap* correspondente à imagem de microscopia de entrada.

- ④ Usando o algoritmo *Graphcut*, a partir da imagem de microscopia e seu *heatmap* correspondente, é gerada a imagem segmentada final. Uma descrição detalhada deste passo é feita na Subseção 3.2.

3.1.1 Arquitetura da Rede CNN

Uma rede CNN é um tipo de rede neural, contendo camadas de neurônios que realizam operações em cima dos resultados de camadas anteriores, sendo os parâmetros dessas operações ajustadas em uma etapa de treinamento. As CNN vêm sendo usadas em problemas de processamento de imagens como visão computacional, pois elas têm a capacidade de considerar características espaciais da imagem ao ser fundamentadas na operação de convolução. Um resumo atual sobre o tema pode ser encontrado em (Rawat and Wang, 2017) e (Gu et al., 2018).

A arquitetura da CNN usada é constituída por camadas convolucionais e de *max pooling*. As camadas convolucionais são formadas por máscaras que deslizam sobre os pixels da imagem realizando operações lineares, e suas saídas passam por uma função não linear (neste trabalho é usada a ReLU (*Rectified Linear Units*) (Nair and Hinton, 2010)), que adiciona uma maior capacidade para a rede em “aprender” uma tarefa. Por sua vez, as camadas de *max pooling* têm como objetivo reduzir a dimensão da matriz, selecionando o máximo valor de cada conjunto de valores definidos pela dimensão da máscara a ser aplicada.

Neste trabalho, a CNN foi definida com 2 camadas convolucionais intercaladas com mais 2 camadas de *max pooling*. A arquitetura da rede pode ser vista na Figura 2. Esta arquitetura foi baseada na rede LeNet-5 (LeCun et al., 1998), sendo modificada para o caso onde a entrada é um *patch* de dimensões $27 \times 27 \times 3$, extraído em torno de um pixel a ser classificado como fundo ou objeto. Tal pixel pertence a uma imagem em cores no espaço de cor RGB.

A saída da primeira camada convolucional tem dimensões 27×27 e profundidade 32. Cada um dos 32 *feature maps* (mapas de características) foi criado a partir da convolução do *patch* extraído com 32 filtros de dimensões $5 \times 5 \times 3$ e a aplicação da função de ativação ReLU. A camada seguinte é uma camada de *max pooling*, cuja saída é composta por 32 *feature maps* de dimensões 9×9 , calculadas a partir de vizinhanças de 3×3 pixels.

A saída da segunda camada convolucional tem dimensões 9×9 e profundidade 64. Os *feature maps* foram criados a partir da convolução da camada anterior com 64 filtros de dimensões 5×5 pixels e a aplicação da função de ativação ReLU. A camada seguinte é uma camada de *max pooling*, cuja saída é composta por 64 *feature maps* de dimensões 3×3 pixels, calculadas a partir de vizinhanças de 3×3 pixels.

À saída da última camada de *max pooling* é aplicada uma operação de *flatten*, que transforma os *feature maps* resultantes da última camada em um único vetor de dados. Tal vetor de dados é usado como entrada em uma rede neural do tipo perceptron multicamadas, com 512 neurônios na camada oculta e dois neurônios na camada de saída. Os neurônios da camada de saída possuem função de ativação *softmax*, onde cada um dos neurônios indica o valor de probabilidade do pixel central do *patch* ser um objeto ($P(B)$) ou fundo ($P(F)$). O mapa de probabilidades pode ser obtido a partir de qualquer uma das duas saídas já que $P(F) + P(B) = 1$. Neste trabalho, a saída $P(F)$ foi usada para gerar o *heatmap*.

3.2 Segmentador Binário

O algoritmo *Graphcut* faz uso da premissa Markoviana para efetuar a segmentação de uma imagem. Isso implica no uso tanto da informação de cada pixel como de sua vizinhança na definição do rótulo correspondente. Sendo assim, a segmentação é feita via a otimização de uma função objetivo que considera tanto a informação de cada pixel como da vizinhança. Na subseção seguinte é explicado o algoritmo *Graphcut* aplicado ao problema de segmentação binária e como são definidas as variáveis da função objetivo a otimizar para o problema em estudo.

3.2.1 Segmentação binária usando *Graphcut*

Seja $\mathbf{I}(\mathbf{x}) = \mathbf{z}$ uma imagem que mapeia cada pixel $\mathbf{x} \in \mathcal{X}$ para seu valor de cor \mathbf{z} . Seja $S(\mathbf{x}) = f_{\mathbf{x}}$ sua segmentação, onde $f_{\mathbf{x}} = \{0, 1\}$ são os rótulos de classe. O objetivo é resolver o problema de segmentação atribuindo para cada pixel \mathbf{x} um rótulo $f_{\mathbf{x}}$, onde $f_{\mathbf{x}}$ é localmente suave e consistente com a informação observada. Boykov et al. (2001) tratam o problema de segmentação de imagem como um problema de minimização de uma função objetivo definida como:

$$E(S) = E(S)_{smooth} + \lambda E(S)_{data}, \quad (1)$$

onde $E(S)$ mede a qualidade da segmentação S . $E_{data}(S)$, chamado de termo dos dados, mede quantos pixels em \mathcal{X} “se parecem” com os rótulos dados a eles. $E_{smooth}(S)$, chamado de termo de suavização, mede o quanto S não é suave. O parâmetro λ controla a influência dos pixels vizinhos e, portanto, a suavização do resultado da segmentação. E_{data} é normalmente escrito como

$$E_{data}(S) = \sum_{\mathbf{x} \in \mathcal{X}} D(\mathbf{x}, f_{\mathbf{x}}), \quad (2)$$

onde $D(\bullet)$ é a penalização de atribuir ao pixel \mathbf{x} o rótulo $f_{\mathbf{x}}$. Neste trabalho, $D(\bullet)$ é o *heatmap*

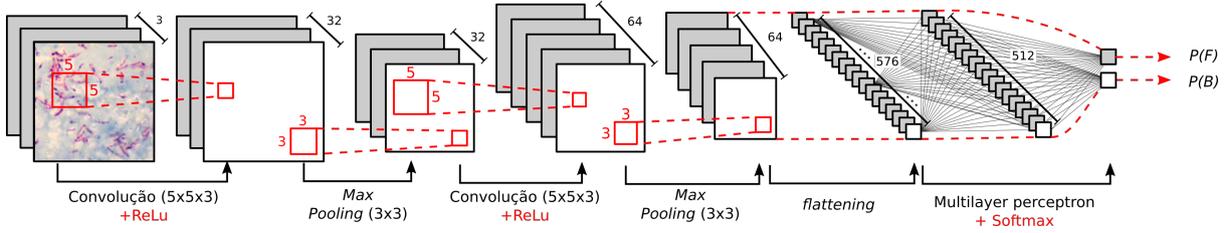


Figura 2: Arquitetura da CNN usada para a segmentação.

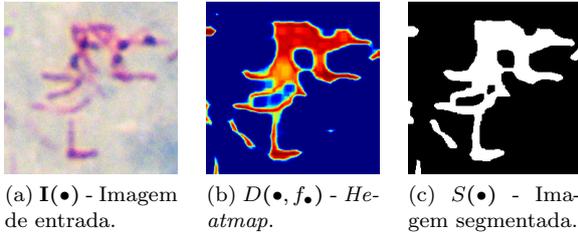
gerado pela rede CNN usada. $E_{smooth}(S)$ é normalmente formulado como

$$E_{smooth} = \sum_{\mathbf{x}, \mathbf{y} \in \aleph} V(\mathbf{x}, \mathbf{y}) \delta(f_{\mathbf{x}} \neq f_{\mathbf{y}}), \quad (3)$$

onde $V(\mathbf{x}, \mathbf{y})$ é a penalização imposta pela fronteira de segmentação entre \mathbf{x} e \mathbf{y} (também chamado de custo de suavização) e $\delta(S(\mathbf{x}) \neq S(\mathbf{y})) = 1$ se \mathbf{x} e \mathbf{y} possuem rótulos diferentes, do contrário ele é 0. A vizinhança \aleph é dada pelo conjunto de todos os pares $\{\mathbf{x}, \mathbf{y}\}$ presentes numa grade 4-conectado, correspondentes aos pixels da imagem no plano $2D$. Nesse trabalho V é dado por

$$V(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-[I_{R-G}(\mathbf{x}) - I_{R-G}(\mathbf{y})]^2}{2\sigma^2}\right), \quad (4)$$

onde $I_{R-G}(\mathbf{x}) = R(\mathbf{x}) - G(\mathbf{x})$. $R(\mathbf{x})$ e $G(\mathbf{x})$ são os canais de cor vermelho e verde da imagem $\mathbf{I}(\mathbf{x})$. Esses canais de cor foram escolhidos por mostrar bons resultados em (CostaFilho et al., 2012). A Figura 3 mostra um exemplo de segmentação usando a técnica de *Graphcut*.



(a) $\mathbf{I}(\bullet)$ - Imagem de entrada. (b) $D(\bullet, f_{\bullet})$ - Heatmap. (c) $S(\bullet)$ - Imagem segmentada.

Figura 3: Segmentação usando o algoritmo *Graphcut*.

4 Resultados Experimentais

Nesta seção é apresentada a base de dados usada nos experimentos e os resultados obtidos na avaliação da técnica proposta em função das métricas descritas na Subseção 4.3.

4.1 Base de Dados

A base de dados usada neste trabalho é composta por 120 imagens de 12 pacientes, sendo 10 imagens de uma mesma lâmina para cada paciente. As imagens foram obtidas usando uma câmera de 10 MegaPixels acoplada a um microscópio de microscopia convencional. As imagens das lâminas com

esfregaço de escarro foram preparadas segundo a técnica de Ziehl-Neelsen. As imagens do conjunto de treinamento e teste possuem resolução fixa de 2816×2112 pixels.

Em todas as imagens, os bacilos assinalados foram identificados e marcados pixel a pixel por um especialista. A Figura 4 mostra um recorte em uma das imagens da base de dados como exemplo.

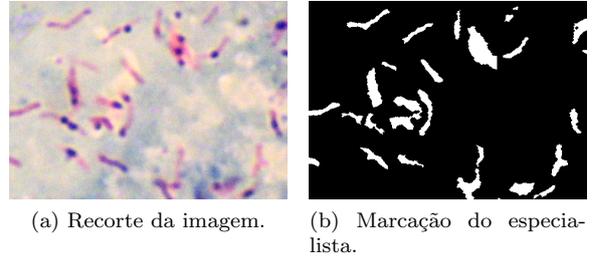


Figura 4: Exemplo de imagens da base de dados.

4.2 Características de Implementação

Para a implementação da CNN foi utilizada a biblioteca *TensorFlow*, versão 1.3.0, do *Python*, e para a implementação do algoritmo *Graphcut* foi usado o programa *MATLAB*, versão 2016a. Além disso, a configuração da máquina utilizada nos experimentos foi: (i) sistema operacional Linux, distribuição Ubuntu 16.04 LTS; (ii) processador Intel Core i5-7200U, 2.50GHz com 2 núcleos físicos; (iii) memória RAM de 8 GB DDR4; (iv) unidade de armazenamento de 1000 GB (disco rígido); (v) placa de vídeo Nvidia Geforce 940MX, com 2 GB de memória dedicada.

A partir da base de dados foram criados os conjuntos de treinamento e teste, cada um com 60 imagens. Um conjunto de validação foi separado dos *patches* de treinamento para poder validar o aprendizado da CNN. Este conjunto corresponde a 10% do número total de *patches* de treinamento. O conjunto de *patches* de treinamento foi criado segundo a seguinte metodologia: para cada uma das 60 imagens do conjunto de treinamento foram extraídas aleatoriamente 300 *patches* como amostras **positivas** e 300 *patches* como amostras **negativas**. Nos casos em que o *patch* tinha pixels fora da imagem, ele foi preenchido usando *padding* espelhado. Todos os *patches* possuem dimensões

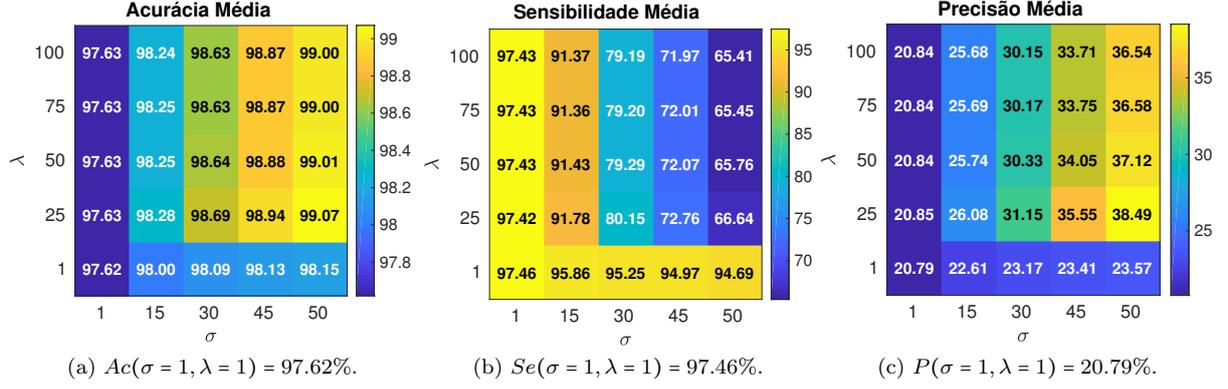


Figura 5: Acurácia, sensibilidade e precisão calculadas no conjunto de treinamento para vários valores de σ e λ .

de $27 \times 27 \times 3$ pixels.

São apresentados a seguir os hiperparâmetros usados para a rede CNN e o algoritmo *Graphcut*. (i) *Em relação à CNN*: a taxa de aprendizado usada tem valor 0,01, com decaimento exponencial de 0,95. A camada totalmente conectada tem seus pesos regularizados usando norma L2. A busca por esses parâmetros foi empírica. (ii) *Em relação ao algoritmo Graphcut*: considerando que são dois parâmetros λ e σ a serem estimados, a busca do valor desses parâmetros foi feita usando busca exaustiva. Os valores de σ e λ avaliados são: $\sigma \in \{1, 15, 30, 45, 50\}$ e $\lambda \in \{1, 25, 50, 75, 100\}$. Para cada valor de σ e λ usados, foram calculadas as métricas que serão descritas na Subseção 4.3.

4.3 Avaliação dos Resultados e Discussão

Considerando que a segmentação a nível de pixels foi modelada como um problema de classificação binária, a avaliação aqui utilizada será feita via as seguintes métricas: (i) *acurácia* (Ac), que avalia a capacidade da técnica de classificar corretamente os pixels como objeto ou fundo; (ii) *sensibilidade* (Se), que avalia a capacidade da técnica em classificar corretamente os pixels como objeto (bacilo); (iii) *Precisão* (Pr), que avalia a capacidade da técnica em classificar corretamente quais dos pixels do objeto segmentado pela técnica realmente correspondem a pixels de objeto; e finalmente a medida F_1 que é a média harmônica entre a sensibilidade e a precisão.

As relações que definem essas métricas são apresentadas nas Equações (5) – (8), onde: VP são os verdadeiros positivos (número de pixels corretamente classificados pela técnica como objeto), VN são os verdadeiros negativos (número de pixels corretamente classificados pela técnica como fundo), FP são os falsos positivos (número de pixels classificados erroneamente pela técnica como objeto), e FN são falsos negativos (número de pixels classificados erroneamente pela técnica como

fundo).

$$Ac = \frac{VP + VN}{VP + VN + FP + FN} \times 100\% \quad (5)$$

$$Se = \frac{VP}{VP + FN} \times 100\% \quad (6)$$

$$Pr = \frac{VP}{VP + FP} \times 100\% \quad (7)$$

$$F_1 = 2 \frac{Se \cdot Pr}{Se + Pr} \times 100\%. \quad (8)$$

Considerando as métricas apresentadas, a acurácia média, a sensibilidade média e a precisão média, calculadas a partir do conjunto de treinamento para diferentes valores de λ e σ , podem ser vistas na Figura 5. Os valores dos hiperparâmetros escolhidos para o algoritmo *Graphcut* foram $\sigma = 1$ e $\lambda = 1$, pois são os valores que retornaram alta acurácia e alta sensibilidade simultaneamente. Para fins médicos, é preferível uma alta sensibilidade ao invés de uma alta precisão, pois uma alta sensibilidade indica que a técnica foi capaz de encontrar (recuperar) boa parte dos pixels de bacilo. Embora isso seja feito com um grande número de falsos positivos (que implicam em uma baixa precisão), corre menos risco de um paciente com tuberculose ser diagnosticado como saudável (e não fazer o tratamento), do que o inverso.

Utilizando os hiperparâmetros determinados na etapa de treinamento, é efetuada a avaliação sobre o conjunto de teste, calculando-se as métricas de Ac , Se , Pr e F_1 , os resultados são mostrados na Tabela 1.

Tabela 1: Resultados da avaliação sobre o conjunto de teste. Os valores estão expressos em porcentagem.

	\overline{Ac}	\overline{Se}	\overline{Pr}	$\overline{F_1}$
CNN + <i>Graphcut</i>	96,70	96,75	28,62	39,64

Acerca dos valores obtidos, são feitos os seguintes comentários.

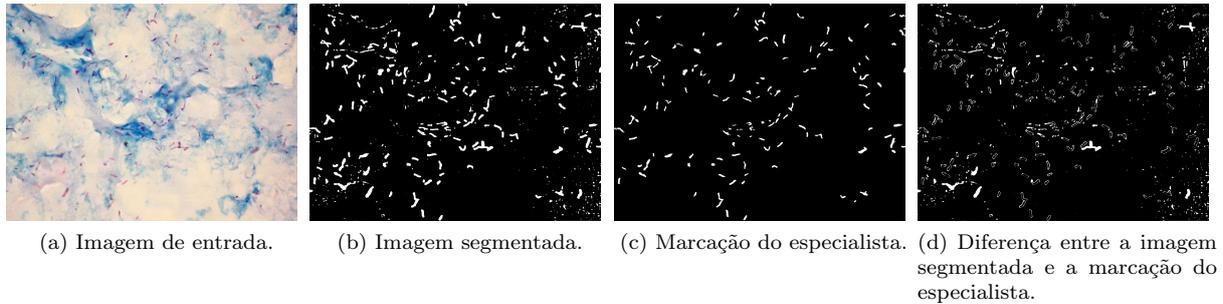


Figura 6: Exemplos de uma imagem segmentada pela técnica e marcação do especialista correspondente.

- O alto valor de acurácia é causado pela grande quantidade de verdadeiros negativos. Como o número de pixels pertencentes ao fundo é muito maior do que a quantidade de pixels pertencentes aos bacilos, a quantidade de verdadeiros negativos tem um peso muito maior no cálculo da acurácia do que os verdadeiros positivos.
- O valor alto de sensibilidade indica que a técnica é capaz de segmentar quase todos os pixels pertencentes aos bacilos. Em imagens médicas isso é preferível ao invés de uma alta precisão. Quanto mais estruturas forem identificadas como bacilo, maior a chance de se detectar a doença.
- O baixo valor de precisão indica que são detectados vários falsos positivos na segmentação, ou seja, alguns pixels pertencentes ao fundo são classificados como bacilos. Cabe indicar que o problema em si é fortemente desbalanceado, isto é, o número de pixels correspondentes ao fundo é superior ao número de pixels correspondentes aos bacilos (cada imagem tem em média 99,14% dos pixels pertencentes ao fundo e 0,86% pertencentes aos bacilos). Tal desbalanceamento das classes causa um forte impacto dos falsos positivos no cálculo da precisão.
- O valor baixo da medida F_1 é causado pela baixa precisão da técnica, já que o valor da sensibilidade é alto.

Um exemplo de imagem segmentada pode ser visto na Figura 6. Pode-se notar que: (i) devido à alta sensibilidade da técnica, ela consegue segmentar todos os bacilos marcados pelo especialista, além de algumas estruturas não marcadas (ver Figuras 6b e 6c); (ii) a segmentação das bordas não é precisa o que, juntamente com os objetos extras que não estão na marcação, diminui a precisão da técnica (ver Figura 6d).

5 Conclusão e Trabalhos Futuros

O objetivo principal deste trabalho foi propor uma técnica para a segmentação de imagens de microscopia de lâminas de escarros coloridos segundo o método de Ziehl-Neelsen. A técnica proposta é baseada fundamentalmente em uma Rede Neural Covolucional usando a metodologia de *patches* e o uso do algoritmo de *Graphcut* para geração da imagem segmentada. Como resultado, usando uma base de dados de imagens de microscopia, obteve-se acurácia de 96,70% e sensibilidade de 96,75%. A partir dos valores de precisão e sensibilidade foi observado que a técnica proposta apresenta uma alta acurácia devido ao grande número de pixels de fundo em relação a quantidade de pixels de bacilos, e uma alta sensibilidade, devido a que boa parte dos pixels pertencem aos bacilos foram identificados corretamente. Por outro lado, a precisão foi baixa (28,62%), produto do desbalanceamento das classes, já que em média cada imagem contém 99,14% dos pixels pertencentes ao fundo e 0,86% pertencentes aos bacilos. Finalmente, com o intuito de diminuir o número de falsos positivos, no futuro será implementada uma etapa de classificação dos objetos binários segmentados baseado em características morfológicas desses objetos.

Referências

- Boykov, Y., Veksler, O. and Zabih, R. (2001). Fast approximate energy minimization via graph cuts, *IEEE Transactions on pattern analysis and machine intelligence* **23**(11): 1222–1239.
- CDC-Tuberculosis (2018). <https://www.cdc.gov/tb/>. acessado em 01/03/2018.
- Chayadevi, M. and Raju, G. (2014). Automated colour segmentation of tuberculosis bacteria thru region growing: A novel approach, *Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the*, IEEE, pp. 154–159.

- Chen, T. and Ched'Hotel, C. (2014). Deep learning based automatic immune cell detection for immunohistochemistry images, *International Workshop on Machine Learning in Medical Imaging*, Springer, pp. 17–24.
- Chitade, A. Z. and Katiyar, S. (2010). Colour based image segmentation using k-means clustering, *International Journal of Engineering Science and Technology* **2**(10): 5319–5325.
- CostaFilho, C. F., Levy, P. C., Xavier, C. M., Costa, M. G., Fujimoto, L. B. and Salem, J. (2012). Mycobacterium tuberculosis recognition with conventional microscopy, *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE*, pp. 6263–6268.
- Forero, M., Cristobal, G. and Alvarez-Borrego, J. (2003). Automatic identification techniques of tuberculosis bacteria, *Applications of digital image processing XXVI*, Vol. 5203, International Society for Optics and Photonics, pp. 71–82.
- Gonzalez, R. C. and Woods, R. E. (2012). Digital image processing.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. et al. (2018). Recent advances in convolutional neural networks, *Pattern Recognition* **77**: 354–377.
- ISTC (2006). "international standards for tuberculosis care, http://www.who.int/tb/publications/2006/istc_report.pdf. acessado em 01/03/2018.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.
- Lu, F., Wu, F., Hu, P., Peng, Z. and Kong, D. (2017). Automatic 3d liver location and segmentation via convolutional neural network and graph cut, *International journal of computer assisted radiology and surgery* **12**(2): 171–182.
- Mahapatra, D. (2017). Semi-supervised learning and graph cuts for consensus based medical image segmentation, *Pattern Recognition* **63**: 700–709.
- Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.
- Rawat, W. and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review, *Neural Computation* **29**(9): 2352–2449.
- Sadaphal, P., Rao, J., Comstock, G. and Beg, M. (2008). Image processing techniques for identifying mycobacterium tuberculosis in ziehl-neelsen stains, *The International Journal of Tuberculosis and Lung Disease* **12**(5): 579–582.
- Shakeri, M., Tsogkas, S., Ferrante, E., Lippe, S., Kadoury, S., Paragios, N. and Kokkinos, I. (2016). Sub-cortical brain structure segmentation using f-cnn's, *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, IEEE*, pp. 269–272.
- Siena, I., Adi, K., Gernowo, R. and Miransari, N. (2012). Development of algorithm tuberculosis bacteria identification using color segmentation and neural networks, *International Journal of Video and Image Processing and Network Security* **12**(4): 9–13.
- Smart, T. (2017). Background on smear microscopy in tb diagnosis, <http://www.aidsmap.com/Background-on-smear-microscopy-in-TB-diagnosis/page/1426650/>. acessado em 01/03/2018.
- Veropoulos, K., Campbell, C., Learmonth, G., Knight, B. and Simpson, J. (1998). The automated identification of tubercle bacilli using image processing and neural computing techniques, *ICANN 98*, Springer, pp. 797–802.
- Vivanti, R., Ephrat, A., Joskowicz, L., Lev-Cohain, N., Karaaslan, O. A. and Sosna, J. (2015). Automatic liver tumor segmentation in follow-up ct scans: preliminary method and results, *International workshop on patch-based techniques in medical imaging*, Springer, pp. 54–61.
- WHO et al. (2017). Global tuberculosis report 2017, *Global tuberculosis report 2017*.
- Wu, G., Coupé, P., Zhan, Y., Munsell, B. and Rueckert, D. (2015). *Patch-Based Techniques in Medical Imaging*, Springer.
- Xu, Z. and Huang, J. (2015). Efficient lung cancer cell detection with deep convolution neural network, *International Workshop on Patch-based Techniques in Medical Imaging*, Springer, pp. 79–86.