

Processamento de Imagem

Trabalho prático – implementação da segmentação por volume utilizando rotulação.

Curso: Ciência da Computação

Professor: Gilmário Barbosa dos Santos

Alunos: Gustavo de Souza; José Augusto Laube.

Todo o projeto se encontra no GitHub e pode ser acesso pelo link abaixo:

<https://github.com/JoselLaube/Trabalho-1-processamento-de-imagens/tree/main>

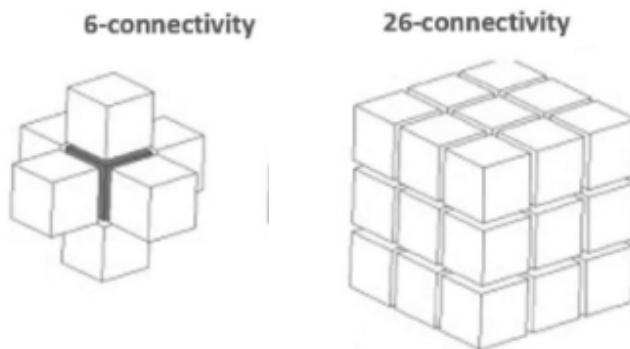
OBJETIVO DO TRABALHO

O relatório foi desenvolvido para documentar os resultados obtidos no a partir de uma análise de um volume tomográfico 3D de uma neoplasia utilizando a linguagem Python. O objetivo principal foi implementar, sem o uso de bibliotecas prontas, um algoritmo de rotulação de componentes conexos tridimensionais para identificar, quantificar e visualizar a distribuição espacial de três tipos distintos de células que fazem parte da composição da massa tumoral.

A respeito da classificação das células:

- Proliferativas: são as células doentes capazes de se reproduzir;
- Quiescentes: células vivas mas inativas, são capazes de se ativar na forma de proliferativas ou de morrer, neste caso elas se tornam células necróticas;
- Necróticas: são células mortas.

Para realizar uma análise completa, o trabalho investigou o impacto de duas abordagens distintas de vizinhança 3D: a conectividade-6 e a conectividade-26. Conforme os requisitos, para cada abordagem foram extraídos os seguintes dados: o número total de células de cada tipo, histogramas representando a distribuição dos tamanhos dos agrupamentos e uma visualização 3D para ilustrar a morfologia dos componentes celulares. A análise comparativa entre essas duas conectividades é, portanto, um ponto fundamental deste estudo.



METODOLOGIA UTILIZADA

O processo foi estruturado em um pipeline claro e sequencial, projetado para ser executado de forma independente para cada tipo de conectividade (6 e 26), garantindo a reprodutibilidade e a comparabilidade dos resultados. O pipeline de processamento de dados foi concebido para transformar o volume 3D bruto em insights e dados quantitativos e visuais, cada etapa alimenta a subsequente, conforme ilustrado no fluxograma abaixo:



Entrada de Dados e Pré-processamento:

A primeira etapa consiste no carregamento do arquivo `volume_TAC` utilizando a biblioteca `pickle`, que desserializa o arquivo diretamente em um objeto de matriz tridimensional `NumPy`. Para simplificar o algoritmo de busca de vizinhos, uma função de pré-processamento, `adicionar_borda_3d`, foi implementada. Esta função envolve a matriz original com uma camada de voxels de valor zero. O objetivo desta borda é pragmático: garantir que, ao verificar os vizinhos de um voxel localizado na extremidade do volume original, o algoritmo não tente acessar um índice fora dos limites da matriz, o que causaria um erro (`IndexError`). Isso elimina a necessidade de múltiplas verificações de borda dentro do loop principal do algoritmo, tornando o código mais limpo e eficiente.

Segmentação e Rotulação por Componentes Conexos:

Esta é a etapa central do trabalho, onde foi implementado um algoritmo autoral para rotulação de componentes conexos. Para evitar a duplicação de código e garantir a robustez, foi criada uma única função, `encontrar_componentes_conexos`, parametrizada pela lista de direções que define a vizinhança.

Esta é a etapa central do trabalho, onde foi implementado um algoritmo autoral para rotulação de componentes conexos. Como uma contribuição para a robustez e manutenção do código, optou-se por não criar funções redundantes. Em vez disso, foi criada uma única função, `encontrar_componentes_conexos`, parametrizada pela lista de direções que define a vizinhança. Esta decisão de design não apenas evitou a duplicação de código, mas também preveniu potenciais bugs de estado compartilhado entre chamadas sequenciais, garantindo a integridade de cada análise.

O algoritmo opera da seguinte forma:

- **Inicialização:** Uma matriz de controle visitado, do mesmo tamanho do volume e preenchida com `False`, é criada. Sua função é rastrear quais voxels já foram analisados e atribuídos a um componente, garantindo que cada voxel seja processado apenas uma vez.
- **Varredura do Volume:** O algoritmo percorre cada voxel do volume 3D (eixos X, Y e Z) de forma sistemática.

- **Descoberta de Novos Componentes:** Para cada voxel, ele verifica duas condições: (i) se o voxel pertence a uma das classes de interesse (valores 140, 200 ou 255) e (ii) se ele ainda não foi visitado (valor False na matriz visitado).
- **Busca em Largura (BFS):** Quando um voxel que satisfaz ambas as condições é encontrado, ele é identificado como o ponto de partida de um componente ainda não rotulado. A partir deste ponto, uma Busca em Largura (Breadth-First Search - BFS) é iniciada para encontrar todos os voxels que pertencem a este mesmo componente. A BFS utiliza uma estrutura de dados de fila (deque) para explorar os vizinhos de forma metódica, camada por camada, garantindo que todos os voxels conectados sejam encontrados.
- **Verificação de Vizinhança:** Durante a BFS, para cada voxel retirado da fila, o algoritmo verifica todos os seus vizinhos conforme a lista de direções fornecida (6 direções para conectividade-6; 26 para conectividade-26). Se um vizinho for válido (mesmo valor da célula, dentro dos limites da matriz e ainda não visitado), ele é marcado como visitado e adicionado à fila para exploração futura.
- **Armazenamento:** Ao final da BFS (quando a fila fica vazia), a lista de todos os voxels encontrados constitui um componente conexo completo. Este componente, junto com seu tipo (valor) e tamanho (número de voxels), é armazenado.

Este processo se repete até que todos os voxels do volume tenham sido visitados, garantindo que todos os componentes tenham sido identificados e rotulados.

Extração de Dados e Geração de Saídas:

Após a conclusão da etapa de rotulação, a função de análise retorna uma lista de todos os componentes encontrados. O código então processa esta lista para calcular as estatísticas detalhadas solicitadas, como o número total de componentes para cada tipo de célula, o volume total (soma dos voxels), e os tamanhos mínimo, máximo, médio e mediano dos componentes.

Finalmente, a biblioteca matplotlib é empregada para gerar as visualizações:

- **Histogramas:** Utilizando a função `plt.hist`, são gerados gráficos que mostram a frequência de componentes por faixa de tamanho para cada tipo de célula.
- **Gráficos 3D:** Utilizando o toolkit `mplot3d`, os voxels de cada componente são plotados como pontos em um gráfico de dispersão 3D, permitindo uma visualização intuitiva da distribuição espacial e da morfologia dos diferentes tipos de tecido dentro da neoplasia.

RESULTADOS OBTIDOS

Saida do terminal após executar o código "[main.py](#)" :

```
souza@machine:~/Documents/GitHub/Trabalho-1-processamento-de-imagens$ python3 main.py
Dimensões originais: (100, 100, 100)
Dimensões com borda: (102, 102, 102)
# ANÁLISE COM CONECTIVIDADE-6

Total de componentes encontrados: 247

=====
ESTATÍSTICAS DETALHADAS DOS COMPONENTES CONEXOS
=====

VALOR 140:
→ Quantidade de componentes: 58
→ Total de voxels: 6129
→ Tamanho mínimo: 2 voxels
→ Tamanho máximo: 5978 voxels
→ Tamanho médio: 105.7 voxels
→ Tamanho mediano: 2.0 voxels
→ Componentes pequenos (≤10): 57
→ Componentes médios (11-100): 0
→ Componentes grandes (>100): 1

VALOR 200:
→ Quantidade de componentes: 17
→ Total de voxels: 39
→ Tamanho mínimo: 2 voxels
→ Tamanho máximo: 5 voxels
→ Tamanho médio: 2.3 voxels
→ Tamanho mediano: 2.0 voxels
→ Componentes pequenos (≤10): 17
→ Componentes médios (11-100): 0
→ Componentes grandes (>100): 0

VALOR 255:
→ Quantidade de componentes: 172
→ Total de voxels: 1215
→ Tamanho mínimo: 2 voxels
→ Tamanho máximo: 380 voxels
→ Tamanho médio: 7.1 voxels
→ Tamanho mediano: 3.0 voxels
→ Componentes pequenos (≤10): 160
→ Componentes médios (11-100): 10
→ Componentes grandes (>100): 2

=====
RESUMO GERAL:
→ Total de componentes: 247
→ Total de voxels nos componentes: 7383
→ Média geral de tamanho: 29.9 voxels/componente
=====
```

```
# ANÁLISE COM CONECTIVIDADE-26

Total de componentes encontrados: 97

=====
ESTATÍSTICAS DETALHADAS DOS COMPONENTES CONEXOS
=====

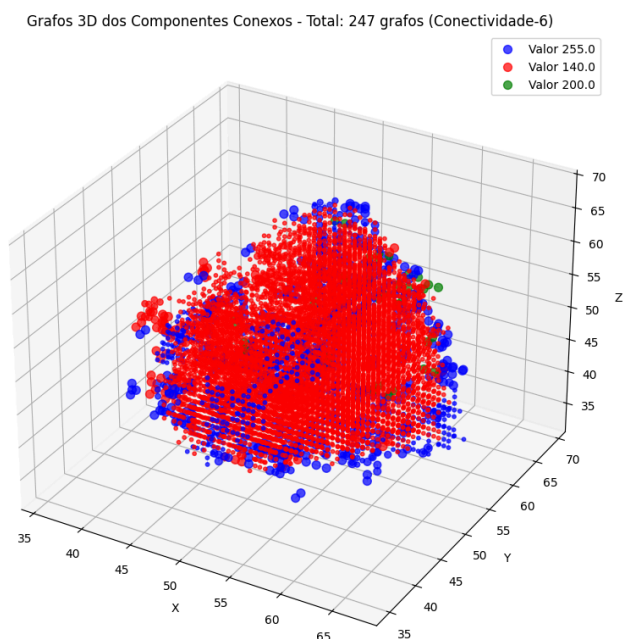
VALOR 140:
→ Quantidade de componentes: 1
→ Total de voxels: 6396
→ Tamanho mínimo: 6396 voxels
→ Tamanho máximo: 6396 voxels
→ Tamanho médio: 6396.0 voxels
→ Tamanho mediano: 6396.0 voxels
→ Componentes pequenos (≤10): 0
→ Componentes médios (11-100): 0
→ Componentes grandes (>100): 1

VALOR 200:
→ Quantidade de componentes: 29
→ Total de voxels: 110
→ Tamanho mínimo: 2 voxels
→ Tamanho máximo: 20 voxels
→ Tamanho médio: 3.8 voxels
→ Tamanho mediano: 3.0 voxels
→ Componentes pequenos (≤10): 28
→ Componentes médios (11-100): 1
→ Componentes grandes (>100): 0

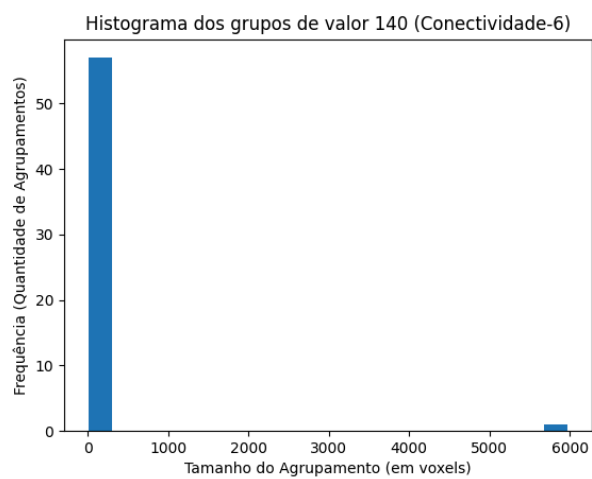
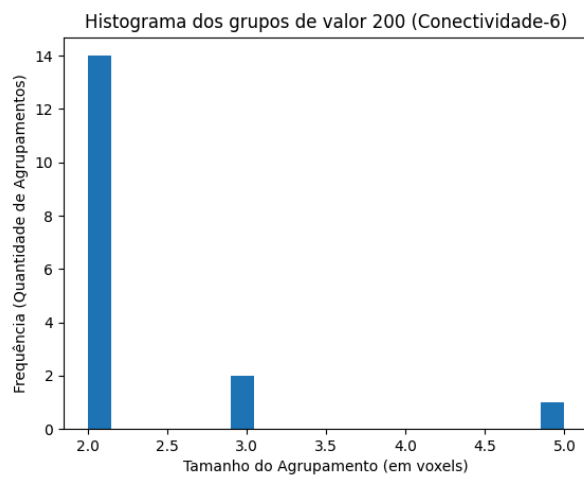
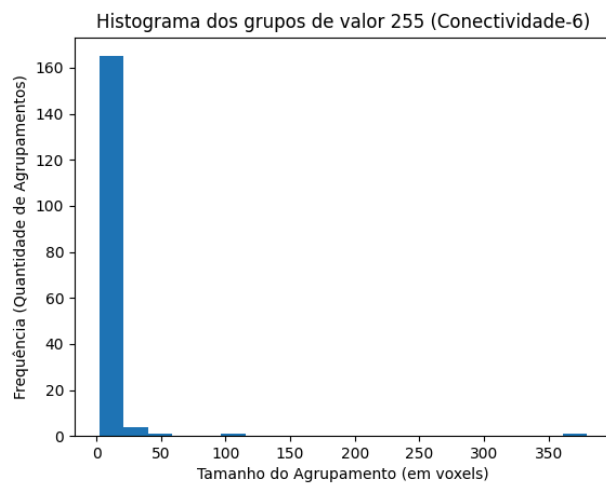
VALOR 255:
→ Quantidade de componentes: 67
→ Total de voxels: 1668
→ Tamanho mínimo: 2 voxels
→ Tamanho máximo: 1412 voxels
→ Tamanho médio: 24.9 voxels
→ Tamanho mediano: 2.0 voxels
→ Componentes pequenos (≤10): 61
→ Componentes médios (11-100): 5
→ Componentes grandes (>100): 1

=====
RESUMO GERAL:
→ Total de componentes: 97
→ Total de voxels nos componentes: 8174
→ Média geral de tamanho: 84.3 voxels/componente
=====
```

Análise com Conectividade-6:



Histogramas de Distribuição:



Análise por Tipo de Célula:

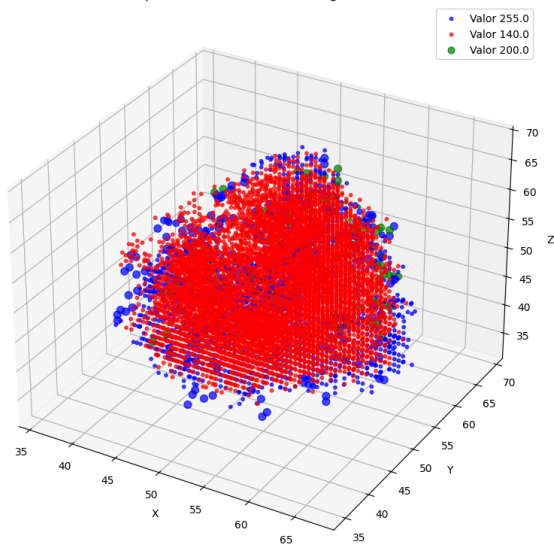
- Células Necróticas (140): Foram encontrados 58 componentes. O histograma correspondente aponta a seguinte conclusão: uma barra muito alta à esquerda mostra que a grande maioria dos componentes (57 deles) é extremamente pequena (mediana de 2.0 voxels). Em contrapartida, uma barra isolada à direita indica a presença de um único e massivo componente principal com 5.978 voxels. Isso sugere a existência de um núcleo necrótico principal acompanhado por uma pequena quantidade de pequenos fragmentos desconectados.
- Células Proliferativas (255): Um padrão similar foi observado, com 172 componentes. A grande maioria (160) é pequena, mas existem componentes maiores, sendo o principal de 380 voxels. A análise sugere uma camada proliferativa principal fragmentada em múltiplos agrupamentos.
- Células Quiescentes (200): Foram identificados 17 componentes, todos de tamanho reduzido (máximo de 5 voxels). O histograma confirma que estas células existem apenas em pequenos grupos isolados.

Dados Quantitativos em formato tabelar:

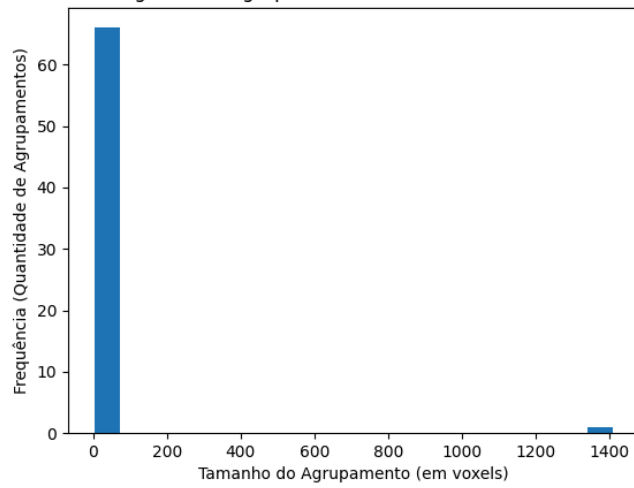
Métrica (Conectividade-6)	Células Necróticas (140)	Células Quiescentes (200)	Células Proliferativas (255)
Nº de Componentes	58	17	172
Total de Voxels	6129	39	1215
Tamanho Máximo	5978	5	380
Tamanho Médio	105.7	2.3	7.1

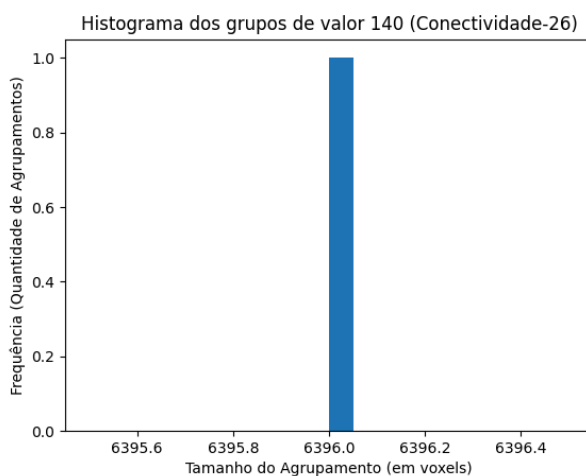
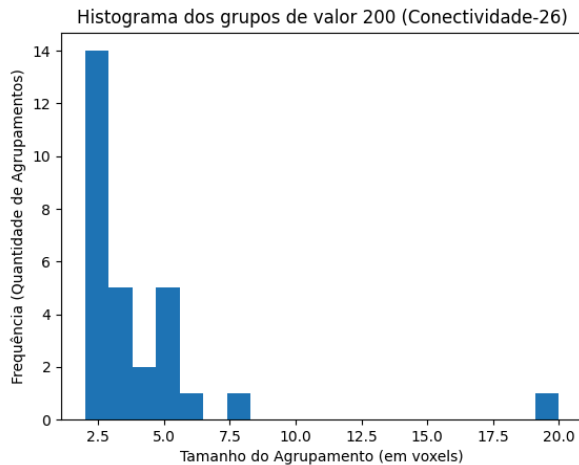
Análise com Conectividade-26:

Grafos 3D dos Componentes Conexos - Total: 97 grafos (Conectividade-26)



Histograma dos grupos de valor 255 (Conectividade-26)





Análise por Tipo de Célula:

- **Células Necróticas (140):** A mudança mais impactante ocorreu aqui em comparação com a conectividade 6. O número de componentes caiu de 58 para apenas 1. O histograma exibe uma única barra, mostrando que todos os 6.396 voxels necróticos pertencem a um único e contínuo componente. A poeira de fragmentos vista na conectividade 6 desapareceu, sendo integrada ao núcleo principal.
- **Células Proliferativas (255):** O número de componentes foi reduzido de 172 para 67. Mais importante, o tamanho do componente máximo saltou de 380 para 1.412 voxels. Isso indica que a camada proliferativa é, na verdade, uma estrutura muito maior e mais coesa do que a análise anterior sugeria.
- **Células Quiescentes (200):** O comportamento destas células não mudou drasticamente. Elas continuam a formar apenas pequenos agrupamentos (máximo de 20 voxels), confirmando que sua distribuição é de fato esparsa e isolada, não formando uma massa contínua.

Dados Quantitativos em formato tabelar:

Métrica	Células Necróticas	Células Quiescentes	Células
---------	--------------------	---------------------	---------

(Conectividade-26)	(140)	(200)	Proliferativas (255)
Nº de Componentes	1	29	67
Total de Voxels	6936	110	1668
Tamanho Máximo	6396	20	1414
Tamanho Médio	6396	3.8	24.9

DISCUSSÃO E ANÁLISE CRÍTICA

Comparativo de dados Conectividade-6 com Conectividade-26:

A verdadeira relevância deste estudo fica clara quando comparamos os resultados das duas abordagens. A análise com Conectividade-6 inicialmente sugeriu uma estrutura tumoral altamente fragmentada, com 247 componentes distintos. No entanto, ao aplicarmos a Conectividade-26, esse número caiu drasticamente para 97. Essa redução não é um erro visto que a Conectividade-26, por ser mais abrangente e conectar células que se tocam até mesmo pelas diagonais, fundiu dezenas de agrupamentos que a Conectividade-6 via como separados. Isso nos permitiu corrigir a visão fragmentada e entender a morfologia real do tumor: um núcleo necrótico massivo e único, envolvido por uma grande camada coesa de células proliferativas, com pequenos bolsões de células quiescentes verdadeiramente isolados. Fica evidente, portanto, que a Conectividade-26 oferece uma representação muito mais fiel da continuidade dos tecidos, sendo essencial para uma análise morfológica precisa.

Impacto da Conectividade: Custo Computacional vs. Acurácia:

(Resposta à Questão 4.3)

A escolha da conectividade representa um clássico trade-off entre acurácia e custo computacional. Conforme demonstrado, a acurácia da Conectividade-26 é superior para este problema, pois fornece uma representação mais fidedigna da estrutura tumoral. Ela evita a fragmentação artificial dos tecidos principais e revela a natureza contínua das massas necróticas e proliferativas, um insight que seria perdido com a Conectividade-6. Já o custo computacional é reflexo da precisão obtida. O algoritmo, ao operar com Conectividade-26, precisa verificar até 26 vizinhos para cada voxel, em contraste com apenas 6 na abordagem anterior. Isso resulta em um número significativamente maior de operações, tornando a análise mais lenta e computacionalmente mais onerosa. A escolha ideal dependerá, em um cenário real, da prioridade entre a velocidade de processamento e a precisão da segmentação morfológica.

Inadequação de Conectividades 2D em Análises Volumétricas:

(Resposta à Questão 4.2)

A aplicação de algoritmos de conectividade bidimensionais, como a conectividade-4 ou a conectividade-8, seria tecnicamente inadequada e metodologicamente incorreta para a análise de um volume 3D. Tal abordagem trataria o volume como uma simples pilha de imagens 2D independentes, processando cada "fatia" (plano XY) de forma isolada. A conectividade ao longo do eixo de profundidade (eixo Z) seria completamente ignorada.

Como consequência, um único componente tumoral que se estendesse por múltiplas fatias seria erroneamente segmentado em vários componentes distintos e menores, um para cada fatia que ele ocupa. Isso levaria a uma interpretação gravemente fragmentada e incorreta da estrutura tridimensional real do tumor.

CONCLUSÃO

O presente trabalho atingiu com sucesso seu objetivo de implementar e avaliar um algoritmo autoral de rotulação de componentes conexos para a análise de um volume tomográfico 3D. Por meio de um pipeline de processamento bem definido, foi possível segmentar, quantificar e visualizar três tipos de células tumorais utilizando duas abordagens distintas de vizinhança: conectividade-6 e conectividade-26. A análise comparativa dos resultados demonstrou de forma conclusiva que a escolha da conectividade impacta diretamente a interpretação morfológica. Enquanto a Conectividade-6 sugeria uma estrutura altamente fragmentada, a Conectividade-26 revelou uma organização mais coesa e realista, caracterizada por um núcleo necrótico único e massivo, envolto por uma camada proliferativa contínua. O estudo reforça, portanto, que a Conectividade-26, apesar de mais custosa computacionalmente, é essencial para obter uma representação fiel da continuidade dos tecidos em análises de imagem volumétrica.