

Segmentação de Imagens Médicas Hiperespectrais: Identificação de Próteses Odontológicas

Palavras-Chave: IMAGENS HIPERESPECTRAIS, SEGMENTAÇÃO, REDES NEURAIIS

Autores:

FELIPE HIROSHI KANO INAZUMI, FEEC, UNICAMP

Prof.^a Dra LETICIA RITTNER FEEC, UNICAMP

INTRODUÇÃO:

As imagens hiperespectrais (HSI, do inglês *Hyperspectral imaging*) são obtidas pela captura de imagens em múltiplos comprimentos de onda ao longo do espectro eletromagnético da luz, fornecendo a resposta espectral da substância amostrada em cada pixel da imagem [1]. Este tipo de imagem é frequentemente denominada de hipercubo, na qual cada imagem (banda espectral) corresponde a um comprimento de onda específico.

A grande vantagem da utilização das HSIs reside na combinação da informação espacial com a espectroscopia do material amostrado, sendo possível obter informações que estão no espectro visível da luz (400 - 700 nm) e estendendo para uma faixa do ultravioleta (350 - 400 nm) e do infravermelho próximo (700 - 2500 nm).

Além disso, como são obtidas informações em múltiplos comprimentos de onda, é possível construir a assinatura espectral detalhada da reflectância do objeto, superando a limitação das imagens RGB (*Red, Green, Blue*), que possuem informações em apenas três bandas.

Com a assinatura espectral é possível distinguir e identificar a presença de certos materiais na imagem HSI, uma vez que cada material interage de forma única com o espectro eletromagnético, resultando em uma assinatura espectral exclusiva.

As duas grandes áreas de aplicação de HSI na saúde são diagnóstico médico e auxílio intraoperatório. Isso ocorre porque esse tipo de imagem é não invasiva e não ionizante, revela detalhes além do espectro visível, não requer introdução de agentes no corpo e pode fornecer imagens em tempo real. Além disso, esta modalidade de imagem consegue detectar alguns biomarcadores importantes, como por exemplo, a concentração de hemoglobina e saturação da oxigenação dos tecidos [2], podendo assim diferenciar um tecido saudável de um tecido patológico.

Outra tarefa amplamente estudada em aplicações da saúde, que serve de suporte para a análise das imagens, é a segmentação, que consiste no agrupamento de regiões da imagem que pertencem à mesma classe, baseado nas suas propriedades físicas como por exemplo cor ou textura, no caso de imagens RGB.

A segmentação em imagens médicas identifica regiões de interesse na imagem, como por exemplo, órgãos, tecidos e lesões e é essencial para facilitar o diagnóstico, tratamento e o monitoramento da progressão de doenças [3]. Contudo, apesar da segmentação ser amplamente utilizada no estudo de doenças e no diagnóstico médico, ainda não há muitos trabalhos que estudam e utilizam a segmentação de imagens médicas hiperespectrais [4].

Portanto, o objetivo deste trabalho foi estudar se as HSIs são mais eficazes do que as imagens RGB para a segmentação de imagens

médicas. Em particular, foi utilizado um conjunto de dados composto por imagens odontológicas para treinar uma U-Net [5] capaz de diferenciar, através da segmentação, entre esmalte dentário (enamel) e próteses.

METODOLOGIA:

A) Banco de Dados

O Banco de Dados de Imagens Espectrais Oraís e Dentárias (ODSI-DB) [6] é um conjunto público de imagens hiperspectrais que consiste de 316 imagens de reflectância oral e dentária retirados de 30 pacientes humanos voluntários. As imagens do ODSI-DB incluem a vista frontal e as superfícies oclusais dos dentes inferiores, superiores, da mucosa oral e da face ao redor da boca dos pacientes.

O *dataset* foi construído utilizando duas câmeras espectrais diferentes: 145 imagens foram retiradas com a Nuance EX (CRI, PerkinElmer, Inc., Waltham, MA, USA) e 171 imagens com a Specim IQ (Specim, Spectral Imaging Ltd., Oulu, Finland).

A Nuance EX captura as imagens HSI por meio da aquisição global, na qual a câmera captura toda a informação espacial em cada um dos comprimentos de onda individualmente. As imagens possuem uma resolução espectral de 10nm na faixa do espectro entre 450-950 nm, resultando em um cubo espectral de 1392x1040x51 (Comprimento x Largura x Bandas Espectrais).

A câmera Specim IQ adquire as imagens utilizando a tecnologia de varredura eletrônica linear, na qual a câmera espectral captura toda a informação espacial e espectral de uma única vez, linha por linha. Esta câmera possui uma resolução espectral de 3 nm na faixa do espectro entre 400-1000 nm, dessa forma, as imagens capturas resultam em um cubo espectral de dimensão 512x512x201.

B) Máscaras de Segmentação

Juntamente com os dados originais, o ODSI-DB fornece também máscaras de segmentação geradas manualmente. Das 316 imagens disponíveis no conjunto de dados, 215

foram anotadas por médicos especialistas. No total, 35 classes foram anotadas, as quais incluem problemas técnicos nas imagens (região fora de foco, sombras e reflexão especular), tecidos duros (enamel, metal, plástico, prótese dentária e raiz), tecidos moles (gingiva marginal e inserida, vasos sanguíneos, palato duro e mole, lábios, pele, e língua) e problemas encontrados nos tecidos duros e moles (cálculo, cárie dentária, fibroma, gengivite, inflamação, úlcera entre outros).

C) Preparação dos Dados para Treinamento

Dentre as classes anotadas pelos especialistas, foram escolhidas as classes de prótese dentária e esmalte do dente para avaliar a segmentação das imagens RGB e HSI. Esta escolha baseia-se na semelhança visual entre o esmalte dentário e prótese, que apresentam características físicas como cor e textura muito similares, tornando difícil a distinção apenas por meio de uma imagem RGB

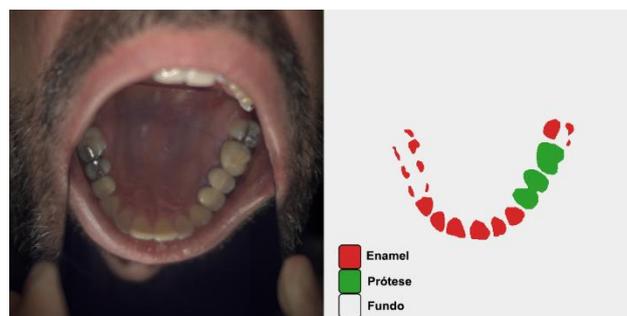


Figura 1 - Imagem RGB retirada do ODSI-DB (esq.) e máscara de segmentação correspondente anotada pelo especialista (dir.): enamel (vermelho); prótese dentária (verde); e fundo (cinza).

No entanto, enquanto o esmalte do dente é constituído majoritariamente de materiais inorgânicos como a hidroxiapatita (96%) e água (3%), a prótese dentária pode ser feita de diversos materiais diferentes, sendo os mais comuns a resina acrílica, cerâmica e a porcelana misturada com zircônio. Assim, como a composição dos materiais do esmalte dentário e da prótese são distintas, cada um desses materiais apresenta uma resposta espectral única, que pode ser captada pelas imagens HSIs, e portanto, uma rede de segmentação poderia aprender as respostas espectrais de cada uma das classes e diferenciar o esmalte do dente de prótese.

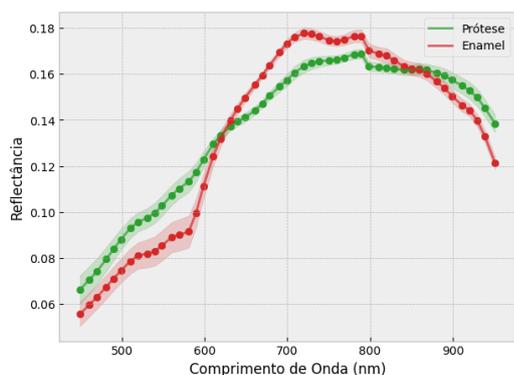


Figura 2 - Média da reflectância de todos os pixels das classes prótese dentária (pontos verdes) e esmalte dental (pontos vermelhos) para cada comprimento de onda. A linha contínua corresponde à extrapolação do espectro discreto do material amostrado. As regiões sombreadas correspondem ao desvio padrão das respectivas classes.

Para realizar o treinamento da rede neural, foi necessário garantir que todas as imagens possuíssem a mesma resolução espacial e espectral. Portanto, todas as imagens HSI e RGB foram redimensionadas para tamanho comum de 256x256 *pixels* e as imagens HSI foram ajustadas no espectro para possuírem 51 bandas espectrais, resultando em um hiper-cubo de 256x256x51 e um tamanho de 256x256x3 para todas as imagens RGB.

As imagens do ODSI-DB que continham anotações de esmalte dentário e prótese foram agrupadas por sujeitos e, em seguida, os sujeitos foram separados em diferentes conjuntos para evitar o vazamento de dados.

A divisão dos sujeitos não foi totalmente aleatória, pois uma inspeção visual das máscaras anotadas pelos especialistas revelou variabilidade entre os anotadores para a mesma máscara. Portanto, alguns pacientes foram propositalmente colocados no conjunto de teste para evitar que essas inconsistências prejudicassem o treinamento da rede. Dessa forma, a separação final resultou em 115 imagens para o conjunto de treino, 34 imagens para o conjunto de validação e 12 imagens para o conjunto de teste.

D) Arquitetura da Rede

A U-Net [5] foi a rede escolhida para fazer segmentação das imagens odontológicas RGB e HSI, pois tem apresentado bons resultados em segmentação de imagens

médicas e é uma das mais utilizadas atualmente para tarefas desse tipo. Na saída da rede foi aplicada uma função de ativação *softmax* com três canais de saída, correspondendo ao fundo, prótese dentária e esmalte do dente. O número de canais de entrada da rede foi alterado de acordo com o tamanho do número de bandas das imagens.

E) Métricas de Avaliação

O coeficiente Dice-Sørensen (DC) é uma métrica que avalia a sobreposição entre a máscara de segmentação prevista pela rede e a máscara real anotada pelos médicos [7]. O DC varia entre 0 e 1, sendo 0 nenhuma sobreposição entre as máscaras e 1 uma sobreposição perfeita.

F) Framework Computacional

Para realizar o treinamento da rede de segmentação, foram utilizados Jupyter Notebooks hospedados na plataforma do Google Colab e a biblioteca do Pytorch-Lighting. O treinamento foi realizado com uma Nvidia Tesla T4 GPU (2560 *Cuda cores*) Intel(R) Xeon(R) CPU @ 2.20GHz e 12.7 GB de memória RAM

RESULTADOS:

Os experimentos realizados tiveram como objetivo comparar o desempenho de segmentação de modelos treinados com imagens de entrada diferentes imagens: RGB; imagens com bandas no espectro infravermelho próximo (NIR); imagens com bandas no espectro visível (VIS); imagens com as 51 bandas (HSI)

Tabela 1 – Dados de Entrada da Rede

Conjunto	Bandas	Espectro
RGB	3	<i>Red, Green, Blue</i>
HSI	51	450 - 950 nm
VIS	36	450 - 800 nm
NIR	16	800 - 950 nm

A rede foi treinada por 35 épocas, usando o otimizador Adam, tamanho do lote

(*batch size*) igual a 20 e taxa de aprendizado (*learning rate*) de 10^{-4} . Algumas funções de perda como a *cross-entropy*, *dice-loss* e uma variação do *dice-loss* ponderado foram testados inicialmente, contudo, a função de perda *focal-loss* ponderada por classes foi a que apresentou melhores resultados e foi escolhida para o treinamento final.

Após o treinamento da rede com cada um dos conjuntos (Tab. 1), foi calculado o DC em comparação com as máscaras anotadas pelos especialistas. Foram computadas as médias de DC por classe e o DC médio total (Tab. 2).

DISCUSSÃO E RESULTADOS:

A partir da análise quantitativa dos resultados (Tab. 2) é possível notar que as redes treinadas com o conjunto de imagens RGB e NIR não conseguiram fazer a distinção entre esmalte do dente e prótese, apresentando ambas um DC de 0.00 para a classe prótese. Em ambos os casos a rede sempre previa que a imagem pertencia à classe esmalte dentário, demonstrando uma incapacidade de diferenciar corretamente entre essas duas classes. Esse resultado indica que, apesar do espectro na região NIR ser frequentemente utilizado na medicina por revelar informações como nível de oxigenação dos tecidos e fluorescência, neste caso, o espectro do infravermelho não traz informações relevantes para a distinção entre prótese e o esmalte do dente.

Já as redes treinadas com o conjunto HSI completo, contendo 51 bandas, e com o conjunto no espectro VIS, com 36 bandas,

apresentaram valores médios de DC maiores (78.38% e 80.20%, respectivamente), indicando que com a informação espectral a rede consegue aprender o espectro de cada amostra e distinguir prótese de dente. A redução do número de bandas de 51 para 36 não só reduziu o custo computacional do treinamento, mas também melhorou a segmentação da classe de prótese de 70,77% para 73,94%. Isso indica que nem toda a informação espectral é necessária para realizar a tarefa de segmentação de maneira eficaz.

A análise visual mostrou diversos casos de inconsistências das máscaras manuais geradas pelos especialistas. Em um paciente do conjunto de teste, por exemplo, em uma das imagens a parte inferior da arcada dentária é considerada como dente, enquanto na outra imagem, a mesma região é classificada como prótese (Fig. 3).

Contudo, a U-Net treinada com o conjunto de imagens VIS consegue identificar consistentemente que a parte inferior é, na verdade, constituída de dentes em ambas as fotos, indicando que a rede consegue distinguir bem entre prótese e dente utilizando a informação espectral. Além disso, a rede treinada com conjunto VIS conseguiu identificar ainda mais dentes do que a própria máscara anotada pelos especialistas. Todas essas inconsistências presentes nas máscaras indicam que os valores de DC resultantes da rede devem ser ainda maiores que os valores reportados (Tab. 2).

Tabela 2 – Coeficiente DC no conjunto de teste

Conjunto	DC Médio por Classe			DC médio
	Fundo	Prótese	Esmalte Dentário	
RGB	99.00 ± 0.02	0.00 ± 0.00	59.00 ± 2.78	52.67 ± 40.66
HSI	98.14 ± 1.08	0.00 ± 0.00	32.74 ± 25.81	43.63 ± 40.80
VIS	97.38 ± 1.26	70.77 ± 6.44	66.06 ± 10.78	78.38 ± 14.23
NIR	98.31 ± 0.91	73.94 ± 3.74	68.33 ± 15.49	80.20 ± 13.01

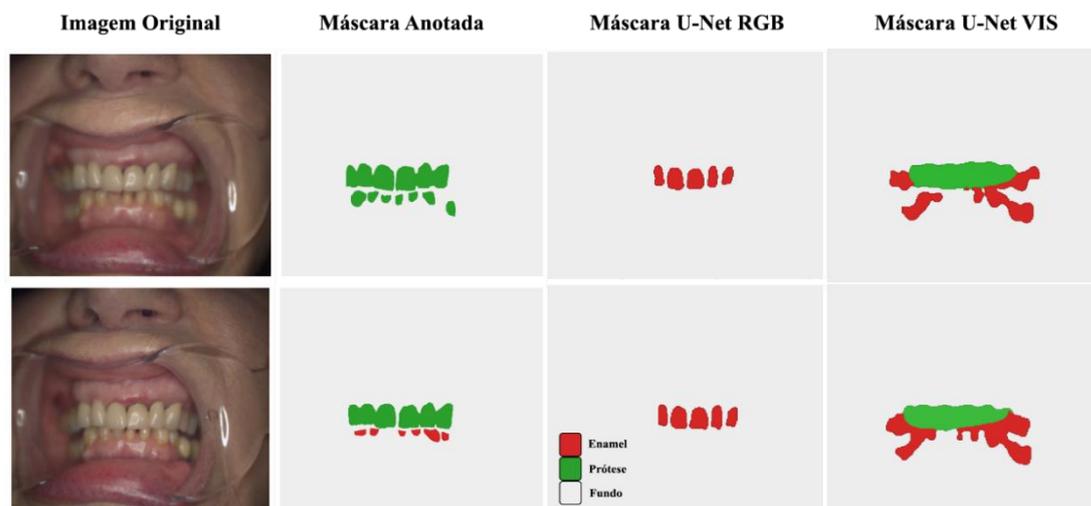


Figura 3 - Exemplos de duas imagens do mesmo paciente disponíveis no ODSI-DB, acompanhadas das respectivas máscaras de segmentação fornecidas por especialistas. Ao lado de cada máscara de referência, são apresentadas as previsões geradas pela rede treinada com imagens RGB e com o conjunto de imagens hiperespectrais no espectro VIS. Os pixels do fundo estão representados em cinza, os do esmalte dentário em vermelho e os da prótese em verde.

CONCLUSÃO:

As imagens hiperespectrais expandem as capacidades da visão computacional para além do espectro visível da luz, oferecendo um grande potencial de aplicação na área da saúde devido à sua natureza não invasiva e não ionizante. No contexto da segmentação, essas imagens demonstraram uma contribuição significativa em relação às imagens RGB, ao fornecerem informações espectrais mais detalhadas sobre os materiais que compõem as estruturas amostradas. Além disso, a segmentação utilizando apenas o espectro visível auxiliou na distinção entre esmalte dentário e prótese, indicando que não é necessário utilizar toda a informação espectral disponível para produzir uma segmentação eficaz. A análise direcionada do espectro e uma seleção adequada das bandas espectrais mais relevantes é suficiente para otimizar a performance da segmentação, sem a necessidade de dados espectrais completos.

INFORMAÇÕES ADICIONAIS:

Trabalho derivado de publicação dos mesmos autores nos anais do SIBGRAPI 2025, sob a licença Creative Commons Attribution 4.0 International (CC BY 4.0).

BIBLIOGRAFIA

- [1] A. Goetz, G. Vane, J. Solomon, and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, p. 1147–1153, Jun. 1985.
- [2] S. Panasyuk, S. Yang, D. Faller, D. Ngo, R. Lew, J. Freeman, and et al., "Medical hyperspectral imaging to facilitate residual tumor identification during surgery," *Cancer Biol Ther*, vol. 6, no. 3, p. 439–446, Mar. 2007.
- [3] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, Jan. 2024.
- [4] U. Khan, S. Paheding, C. P. Elkin, and V. K. Devabhaktuni, "Trends indeep learning for medical hyperspectral image analysis," *IEEE Access*, vol. 9, pp. 79 534–79 548, 2021
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional net-works for biomedical image segmentation," in *LCNS*. Cham: Springer International Publishing, 2015, pp. 234–241
- [6] J. Hyttinen, P. Fält, H. Jäsberg, A. Kullaa, and M. Hauta-Kasari, "Oral and dental spectral image database—odsi-db," *Applied Sciences*, vol. 10, no. 20, p. 7246, Oct. 2020.
- [7] K. Zou, S. Warfield, A. Bharatha, C. Tempny, M. Kaus, S. Haker, and et al., "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004.