

Identificação de próteses em imagens odontológicas: evitando imagens ionizantes com o uso de informações hiperespectrais

Felipe Hiroshi Kano Inazumi
Faculdade de Engenharia
Elétrica e de Computação
Universidade de Campinas (UNICAMP)
Campinas, SP, Brasil
Email: f215696@dac.unicamp.br

Leticia Rittner
Faculdade de Engenharia
Elétrica e de Computação
Universidade de Campinas (UNICAMP)
Campinas, SP, Brasil
Email: lrittner@unicamp.br

Resumo—As imagens hiperespectrais, imagens em múltiplos comprimentos de onda do espectro eletromagnético da luz que vão além do espectro visível, vêm sendo bastante utilizadas para o diagnóstico médico e voltadas principalmente para a classificação de tecidos saudáveis ou malignos, devido a sua capacidade de detectar alterações em alguns marcadores biológicos, como a saturação do oxigênio. Contudo, a segmentação utilizando imagens hiperespectrais, que consiste em agrupar regiões de interesse na imagem, como por exemplo órgãos, tecidos e lesões, ainda é pouco estudada na literatura, apesar do seu grande potencial. Com isso, o objetivo deste trabalho foi estudar se imagens hiperespectrais de reflectância oral e dentária auxiliam na diferenciação entre prótese e enamel (esmalte dentário). Experimentos de segmentação usando um *dataset* público de imagens orais e dentárias (ODSI-DB) mostraram um melhor desempenho da U-Net treinada com imagens hiperespectrais, quando comparada com uma treinada com imagens RGB. A avaliação qualitativa mostrou também que há um grande número de máscaras de segmentação manual da classe prótese inconsistentes (41%), indicando que possivelmente o desempenho da U-Net é muito melhor do que o Dice medido.

Abstract—Hyperspectral images, images in multiples wavelengths of the electromagnetic spectrum of light that go beyond the visible spectrum, have been widely used for medical diagnosis and focused mainly on classification of healthy or malignant tissues, due to its ability to detect changes in some biological markers, such as oxygen saturation. However, segmentation using hyperspectral images, which consists of grouping regions of interest in the image, such as organs, tissues and lesions, is still little studied in the literature, despite its great potential. Therefore, the objective of this present work was to study whether hyperspectral images of oral and dental reflectance help in the differentiation between prosthesis and enamel (tooth). Segmentation experiments using a public dataset of oral and dental images (ODSI-DB) showed better performance of a U-Net trained with hyperspectral images, when compared to the one trained with RGB images. A qualitative assessment also showed that there are a large number of inconsistent manual segmentation masks for the prosthesis class that are inconsistent (41%), possibly indicating that the performance of the U-Net is much better than the measured Dice.

I. 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) [2]. 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.

Originalmente, as HSIs têm sido exploradas em projetos de sensoriamento remoto, já que este tipo de imagem consegue, por exemplo, diferenciar tipos de terrenos, propriedades do solo e até mesmo identificar e revelar reservas minerais subterrâneas [3]. Mais recentemente, as HSIs começaram a ser exploradas na área da saúde, com um grande potencial de aplicação. Isso porque este é um tipo de aquisição 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 [4].

As duas grandes áreas de aplicação de HSI na saúde são diagnóstico médico e auxílio intraoperatório. Isso ocorre porque esse tipo de imagem consegue detectar alguns biomarcadores importantes, como por exemplo, a concentração de hemoglobina e saturação da oxigenação dos tecidos [5], podendo assim diferenciar um tecido saudável de um tecido patológico. Com isso, existe uma quantidade significativa de

trabalhos e pesquisas em HSI que envolvem a detecção e classificação de células cancerígenas, doenças retiniais, polineuropatia diabética e choque hemorrágico [4].

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 [6]. Contudo, apesar da segmentação ser amplamente utilizada no estudo de doenças e no diagnóstico médico [7], ainda não há muitos trabalhos que estudam e utilizam a segmentação de imagens médicas hiperspectrais [8].

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 [9] capaz de diferenciar, através da segmentação, entre esmalte dentário (enamel) e próteses. Foram testadas diferentes combinações de entrada na rede, bem como funções de perda e diferentes técnicas de aumento de dados para estabelecer uma boa forma de treinamento de uma U-Net utilizando HSIs.

II. CONJUNTO DE DADOS

A. Imagens Oraís e Dentárias

O Banco de Dados de Imagens Espectrais Oraís e Dentárias (ODSI-DB) [10] 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 na Universidade da Finlândia Oriental (UFI) após aprovação do comitê de ética do Distrito Hospitalar de Savônia do Norte (413/2016), 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 $1392 \times 1040 \times 51$ (Comprimento \times Largura \times 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 $512 \times 512 \times 201$.

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 (gengiva 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. Pré-processamento original

As HSIs adquiridas com ambas as câmeras foram pré-processadas pelo grupo que construiu o ODSI-DB antes de ser disponibilizado publicamente. O pré-processamento inicial incluiu a conversão de radiância para reflectância e normalização das imagens utilizando uma referência de cerâmica cinza ("Matt Diff Grey", Ceram Research, Ltd., Lucideon, Ltd., Stoke-on-Trent, UK). Além disso, imagens adquiridas com a câmera Nuance EX passaram por um algoritmo de registro [11] adicional para remover artefatos de movimento durante a aquisição das imagens e garantir que todas as bandas espectrais da mesma imagem estivessem devidamente alinhadas.

D. 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 (Fig. 1).

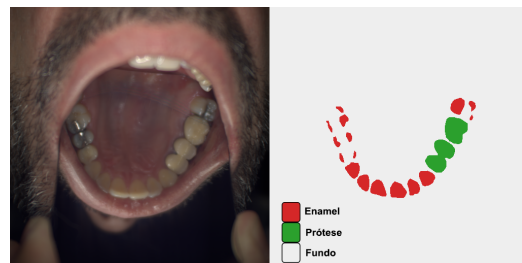


Fig. 1. Imagem RGB de um dos paciente do ODSI-DB (esq.) e máscara de segmentação correspondente (dir.): enamel (vermelho); prótese dentária (verde); e fundo (cinza).

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 (Fig. 2).

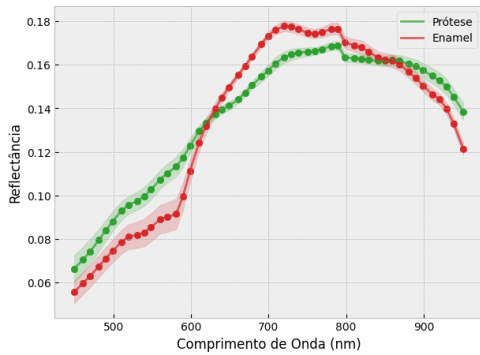


Fig. 2. Média da reflectância de todos os *pixels* das classes prótese dentária (pontos verdes) e esmalte dental (pontos vermelhos) para cada comprimento de onda. A linha contínua corresponde à extrapolação do espectro discreto do material amostrado. As regiões sombreadas correspondem ao desvio padrão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

Para realizar o treinamento da rede neural, foi necessário garantir que todas as imagens possuísem a mesma resolução espacial e espectral. Portanto, todas as imagens HSI e RGB foram redimensionadas para tamanho comum de 256×256 *pixels* e as imagens HSI foram ajustadas no espectro para possuírem 51 bandas espectrais, resultando em um hiper-cubo de $256 \times 256 \times 51$ e um tamanho de $256 \times 256 \times 3$ 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.

III. SEGMENTAÇÃO ATRAVÉS DA U-NET

A. Arquitetura da rede

A U-Net [9] 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 [12]. 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.

B. 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 [13]. O DC varia entre 0 e 1, sendo 0 nenhuma sobreposição entre as máscaras e 1 uma sobreposição perfeita.

IV. EXPERIMENTOS E RESULTADOS

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

TABELA I
DADOS DE ENTRADA DA REDE

Conjunto	Bandas	Espectro
RGB	3	Red, Green, Blue
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. I), 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. II).

TABELA II
COEFICIENTE DC NO CONJUNTO DE TESTE

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

V. DISCUSSÃO

A partir da análise quantitativa dos resultados (Tab. II) 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

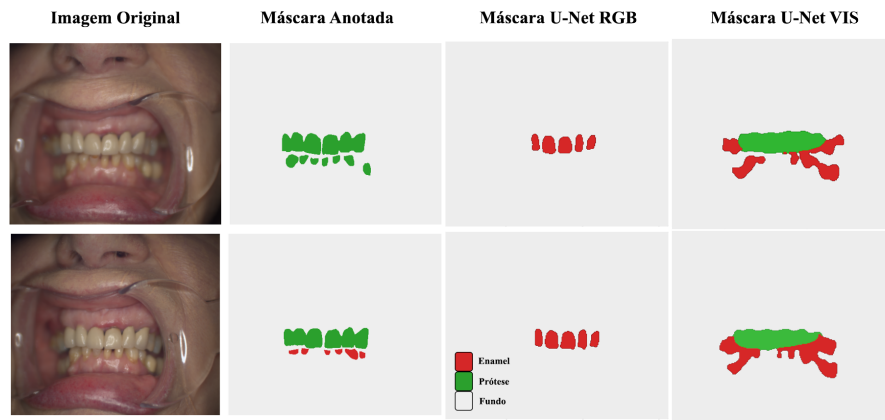


Fig. 3. Duas imagens retiradas do mesmo paciente presentes no ODSI-DB, juntamente com as máscaras anotadas pelos especialistas. Ao lado das máscaras anotadas, estão as previsões da rede treinada com imagens RGB e com o conjunto de imagens hiperespectrais no espectro VIS. Em cinza estão indicados os *pixels* pertencentes ao fundo, em vermelho os *pixels* do esmalte do dente e em verde os *pixels* da prótese.

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. II).

VI. CONCLUSÃO

As imagens hiperespectrais estendem a visão computacional para além do espectro visível da luz, apresentando um vasto potencial de aplicações na saúde devido à sua natureza não invasiva e não ionizante. As imagens hiperespectrais contribuíram significativamente para a melhoria dos resultados de segmentação em relação as imagens RGBs, ao proporcionar informações mais detalhadas dos materiais que constituem as estruturas amostradas na imagem. 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.

AGRADECIMENTOS

Os autores agradecem ao Instituto Eldorado e ao CNPq pelo apoio financeiro.

REFERÊNCIAS

- [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] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, p. 279–317, Dec. 2019.
- [3] M. Akewar and M. Chandak, "Hyperspectral imaging algorithms and applications: A review," Jan. 2024.
- [4] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, Jan. 2014.
- [5] 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.
- [6] 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.
- [7] Y. Shen, Z. Fang, Y. Gao, N. Xiong, C. Zhong, and X. Tang, "Coronary arteries segmentation based on 3d fcn with attention gate and level set function," *IEEE Access*, vol. 7, p. 42826–42835, 2019.
- [8] U. Khan, S. Paheding, C. P. Elkin, and V. K. Devabhaktuni, "Trends in deep learning for medical hyperspectral image analysis," *IEEE Access*, vol. 9, pp. 79 534–79 548, 2021.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *LCNS*. Cham: Springer International Publishing, 2015, pp. 234–241.
- [10] 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.
- [11] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, p. 1973–1989, Nov. 2007.
- [12] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2020.
- [13] K. Zou, S. Warfield, A. Bharatha, C. Tempany, 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.