Deep Learning Architectures for Spatial-Frequency 3D Fluorescence in Oral Cancer Surgery Models

Natalie J. Won[1], Anjolaoluwa Adewale[1], Jerry Wan[1], Mandolin Bartling[2], Jason Townson[1], Harley Chan[1], Esmat Najjar[2], Alon Pener-Tessler[2], Brian C. Wilson[1,3], Jonathan C. Irish[1,2], Michael J. Daly[1]

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, Canada
[2]Department of Otolaryngology-Head and Neck Surgery, University of Toronto, Toronto, Canada
[3]Department of Medical Biophysics, University of Toronto, Toronto, Canada

ABSTRACT

Fluorescence-guided surgery systems employed during oral cancer resection help detect the lateral margin yet fail to quantify the deep margins of the tumor prior to resection. Without comprehensive quantification of three-dimensional tumor margins, complete resection remains challenging. While interoperative techniques to assess the deep margin exist, they are limited in precision, leaving an unmet need for a system that can quantify depth. Our group is developing a deep learning (DL)-enabled fluorescence spatial frequency domain imaging (SFDI) system to address this limitation. The SFDI system captures fluorescence (F) and reflectance (R) images that contain information on tissue optical properties (OP) and depth sensitivity across spatial frequencies. Coupling DL with SFDI imaging allows for the near-real time construction of depth and concentration maps. Here, we compare three DL architectures that use SFDI images as inputs: i) F+OP, where OP (absorption and scattering) are obtained analytically from reflectance images; ii) F+R; iii) F/R. Training the three models required 10,000 tumor samples; synthetic tumors derived from composite spherical harmonics circumvented the need for patient data. The synthetic tumors were passed to a diffusion-theory light propagation model to generate a dataset of artificial SFDI images for DL training. Two oral cancer models derived from MRI of patient tongue tumors are used to evaluate DL performance in: i) in silico SFDI images ii) optical phantoms. These studies evaluate how system performance is affected by the SFDI input data and DL architectures. Future studies are required to assess system performance in vivo.

Keywords: Fluorescence-guided surgery, spatial frequency domain imaging, structured illumination, oral cancer surgery, deep learning, optical reconstruction

1. INTRODUCTION

Tumor delineation remains a challenge in oral cancer surgery.[1] As most oral cancer patients require surgical interventions for treatment, defining tumor margins is critical to achieving a complete resection while limiting the removal of healthy tissue.[2] To illustrate the borders of the tumor, surgeons rely on anatomical imaging, visual cues, pathology, and palpation; however, these traditional interoperative assessments are limited in their precision and contribute to poor patient outcomes.[3] Without precise technologies to illustrate tumors, both local recurrence and the proportion of positive margins are high.[4]

This need for better tumor delineation – in many surgical oncology applications – has motivated the integration of fluorescence-guided surgery systems in the operating room.[5] Fluorescence-guided surgery can provide improved delineation of target anatomy, leading to better patient outcomes.[6] This technique relies on fluorophore accumulation in the tumor to illustrate the boundary between tumor and healthy tissue.[7] When fluorophore-specific wavelengths of light are absorbed, the fluorophore becomes excited and re-emits light at a longer wavelength. This localized re-emission of light provides enhanced delineation of tumor boundaries. In oral cancer, fluorescence-guided surgery can help delineate the tumor at the mucosal surface, as these tumors tend to have "iceberg" configurations where the inclusion originates on the surface.[8] Although this technique can help localize tumors at the surface, the deep margin of the tumor remains unaccounted for prior to resection using traditional 2D fluorescence imaging.[9] Detecting the infiltration depth of these "iceberg" tumors is critical in achieving clear margins during resection.[10] To quantify in vivo tumor depth, we are developing a deep learning (DL)-enabled 3D fluorescence imaging system.

We are using spatial frequency domain imaging (SFDI) that can rapidly obtain depth information and tissue optical properties over a large field of view. SFDI systems project structured light at different spatial frequencies and capture reflectance or fluorescence images at each frequency.[11] Reflectance images are required to quantify tissue optical properties (absorption and scattering),[12] while fluorescence images contain depth-resolved information.[13] In this work, quantifying depth and fluorescence concentration involves passing these SFDI images to a custom DL architecture. Since this technology aims to provide intraoperative assessments, DL is advantageous over analytical methods as it offers near-real-time results.[14,15] Here, we compare three DL architectures that use SFDI images to predict depth and concentration.

2. METHODS

2.1 Spatial Frequency Domain Imaging

The benchtop SFDI system in Figure 1 is used to capture reflectance and fluorescence images across six spatial frequencies evenly spaced through $f_x=0\text{-}0.25\text{mm}^{-1}$.[16] The spatial light modulator (DLi6500 1080p Optics Bundle, DLi, Austin, TX) is based on a 1920×1080-pixel digital light projector (DLP) development kit (LightCrafter 6500, Texas Instruments, Dallas, TX). The system light source is a 6-LED light engine (Spectra X, Lumencor, Beaverton, OR) with 20nm bandwidth sources at 390, 438, 475, 512, 586, and 632nm. To couple the light engine and DLP, the system uses a 3mm liquid light guide (LGG0338, Thorlabs, Newton, NJ). A synchronization cable connects the projector to a 1392×1040-pixel, 14-bit, monochrome CCD camera (Pixelfly USB, PCO AG, Kelheim, Germany) coupled with a 25mm focal-length lens (Model 67715, Edmund Optics, Barrington, NJ). Crossed linear polarizers (LPVISE2X2, Thorlabs, Newton, NJ) are included to reduce specular reflections. To capture fluorescence images of the fluorophore protoporphyrin IX (PpIX), a 632nm source, and a 650nm long-pass filter (FELH0650, Thorlabs, Newton, NJ) are used. Before image acquisition, a polyurethane-based phantom containing carbon black and titanium dioxide (Biomimic PB0317, INO, Quebec City, QC) is imaged to calibrate the SFDI system.
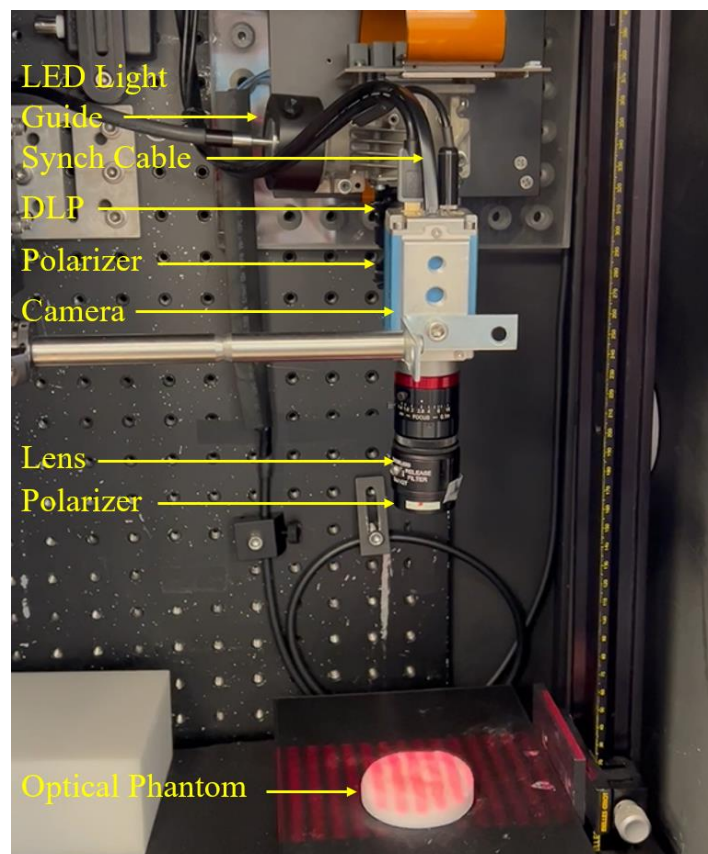

Figure 1. Benchtop spatial frequency domain imaging system collecting images of an agar tongue phantom.

2.2 Deep learning architectures

Three custom DL architectures quantify depth and fluorescence concentration from SFDI images, evolved from a previous architecture for time-domain fluorescence.[17] Each model architecture uses a variation of SFDI images (101×101 pixels at $0.5mm^2$/pixel) as inputs to make predictions: i) fluorescence and optical properties (F+OP); ii) fluorescence and reflectance (F+R); and iii) fluorescence divided by reflectance (F/R) (Born ratio[18]). Here, the DL algorithms only used the first four spatial frequency fluorescence and reflectance images (f_x=0-0.15mm^{-1}) as input. To obtain the optical properties, lookup tables are used to calculate scattering and absorption from SFDI reflectance images;[19] here, we use the pair f_x=0,0.2mm^{-1}. Scattering and absorption maps calculated at these frequencies are DL inputs.

Figure 2 illustrates the difference between the three convolutional neural network architectures. In Fig 2(a), the F + OP model has two distinct arms for each input. Fluorescence images go through a series of convolutions, one 3D convolution (kernel: $3\times3\times4$, activation function: ReLU) followed by two 2D convolutions (kernel: 3×3, activation function: ReLU). Optical properties are fed into a separate arm and go through three sequential 2D convolutions (kernel: 3×3, activation function: ReLU). The two arms concatenate, and a separable 2D convolution is applied. Obtaining a depth and concentration prediction required the downsampling of the feature maps via three sequential 2D convolutions (kernel: 3×3, activation function: ReLU). In Fig. 2(b), the F + R model has two symmetrical arms for the inputs. Both the fluorescence and reflectance images independently go through a series of convolutions, one 3D convolution (kernel: $3\times3\times4$, activation function: ReLU) followed by two 2D convolutions (kernel: 3×3, activation function: ReLU). The two arms concatenate, a separable 2D convolution is applied, and then the network splits into two arms. Feature maps are downsampled via three 2D convolutions (kernel: 3×3, activation function: ReLU) in each arm to produce either a depth or concentration prediction. In Fig. 2(c), the F/R model has a single input after dividing F by R for all spatial frequencies. The F/R images go through a series of convolutions, one 3D convolution (kernel: $3\times3\times4$, activation function: ReLU) followed by two 2D convolutions (kernel: 3×3, activation function: ReLU). A 2D separable convolution is applied, and the network splits. The feature maps are downsampled in parallel to produce depth and concentration predictions.
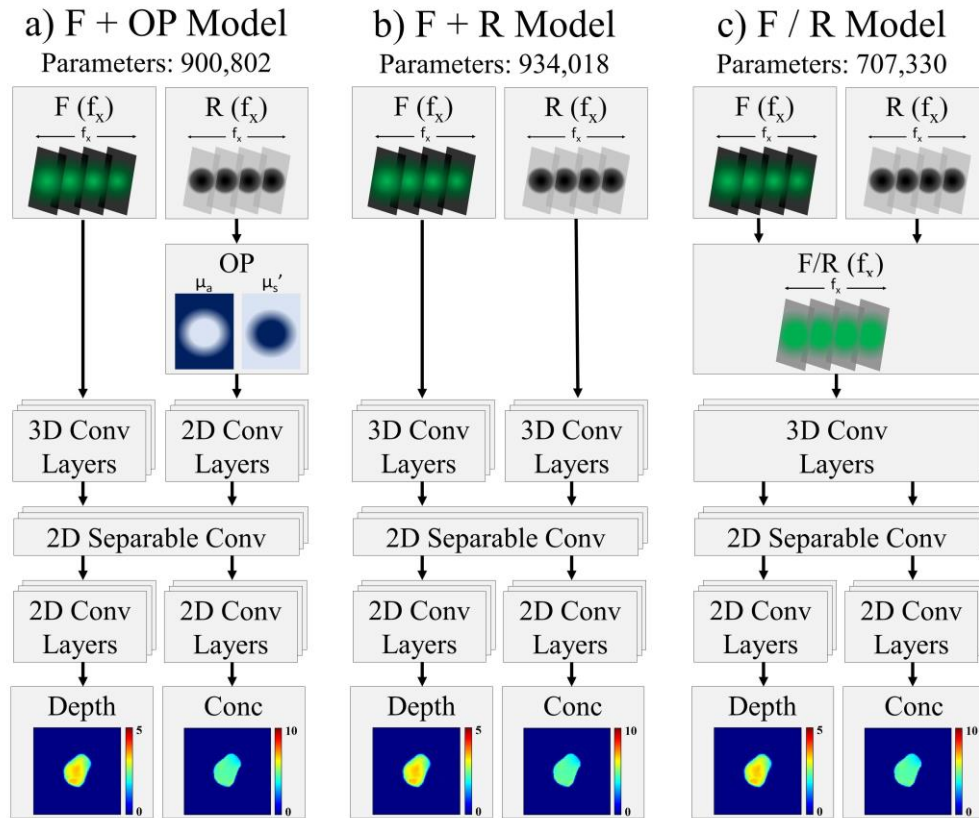

Figure 2. DL architectures for different variations of SFDI inputs: (a) F+OP as two separate inputs, (b) F+R as two separate inputs, (c) F/R as a single input.

2.3 Deep learning training

In silico SFDI data is required to train the DL architectures. To create a dataset that mimics the morphology of oral cancer tumors, synthetic tumors (Figure 3a) based on spherical harmonics were designed as a surrogate. Each tumor consists of four randomly generated spherical harmonics of varying order (2-20), degree (order-20), and radius. Merging the spherical harmonics creates a highly asymmetric composite spherical harmonic. Different transformations are applied to the composite spherical harmonics to create complex shapes: translations in the z-direction, scaling (widths=10-40mm, depths=1-10mm), and cutting at z=0 to create "iceberg" tumors. These parameters are all randomly chosen within a range of values to create a dataset that captures the complex geometric features of real-patient tumors. A comparison of the geometry of a synthetic tumor and a patient tumor is in Figure 3. MATLAB (MathWorks, Natick, MA) is used to generate 10,000 synthetic tumors. All synthetic tumors are passed to a numerical diffusion theory-based light propagation model to generate in silico fluorescence and reflectance SFDI images, optical property maps, depth maps, and concentration maps. Fluorescence, reflectance, and optical property images are inputs for DL training, while depth and concentration maps provide feedback during DL training.

Training the deep learning algorithms is done with Amazon Web Services (AWS) SageMaker (Amazon, Seattle, WA) using a ml.g4dn.2xlarge instance (1 NVIDIA T4 GPU, 8 vCPU). All DL training used the Adam optimizer (initial learning rate: $5e^{-5}$). Training continued for either 100 epochs or until the model met the early stopping criteria. The training process was repeated for all three DL architectures using the same in silico tumor dataset and DL parameters.
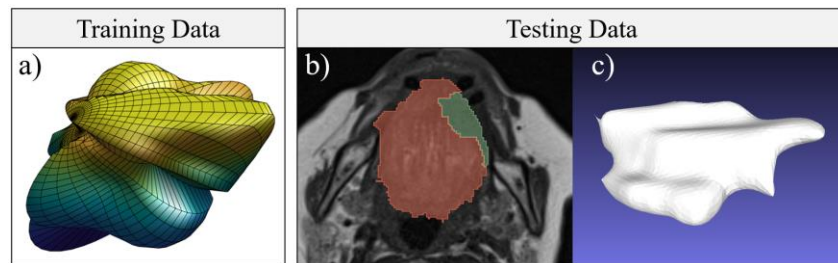

Figure 3. Comparison between synthetic tumors generated from composite spherical harmonics and patient tongue tumors contoured from MRI images. (a) Synthetic tumor generated from composite spherical harmonics. (b) Contoured MRI image of a tongue tumor. (c) 3D mesh of the contoured tongue tumor in the MRI image.

2.4 Deep learning implementation in silico

Twenty-one MRI contours of clinical tongue tumors create a base dataset for in silico validation (Fig. 3(b)). Twelve tumors were omitted as they fell outside the functional range of the DL-enabled system. Inclusion criteria required all tumors in the test set to have maximum depths within 10mm. Multiple transformations were applied to the tumors to increase diversity within the set; the tumors were rotated and scaled (depths=1-10mm) to create a dataset of 45 inclusions. The 3D meshes (Fig. 3(c)) were passed to the diffusion theory light propagation model to create synthetic SFDI images based on patient tumor geometries. These in silico fluorescence, reflectance, and optical properties were inputs for the DL models. Predictions were made in AWS SageMaker and downloaded for analysis in MATLAB. An error estimate was obtained by comparing the depth and concentration maps calculated from the mesh to the DL prediction. This process was repeated to compare the predictive performance across all three models.

2.5 Deep learning implementation in optical phantoms

The 21 contours of tongue tumors were also used to create meshes for 3D printing. Silicone molds of the 3D-printed tumors were produced so that an agar phantom of each tumor could be generated. Two solutions were required to make the optical phantom: i) agar for the fluorescent tumor and ii) agar for the background tissue. In both agar solutions, India Ink was the absorbing agent (Tumor μ_a=0.0093mm^{-1}, Background μ_a=0.0047mm^{-1}) and Intralipid was the scattering agent (Tumor μ_s'=2.3mm^{-1}, Background μ_s'=2.6mm^{-1}). PpIX (4.94 μg/mL) was added to the tumor solution to create fluorescence and iodine (10 μg/mL) was added to create contrast for a CT scan of the phantom. To create the fluorescent tumor, the fluorescent agar solution was poured into the silicone mold. Once the agar solution hardened, the tumor was removed, and a background agar solution was poured around it to represent the healthy tissue. To simplify the phantom for initial testing, any extruding tumor was removed to create a uniform surface. A cone-beam CT scan (Siemens Cios Spin) of the phantom was taken to act as the ground truth and provide information on the overall geometry of the tumor.

The experimentally collected SFDI images were processed to remove artifacts. A region of interest was drawn to highlight the tumor, and the exterior region was padded with the values that corresponded to the background tissue. Images were then resized to 101×101 pixels at a resolution of 0.5mm/pixel. The optical property maps and first four spatial frequencies (f_x=0, 0.05, 0.1, 0.15mm^{-1}) of both the fluorescence and reflectance images were exported and uploaded to a linked AWS S3 bucket.

The predictions from each DL model were passed to MATLAB for analysis. The CT scan of the phantom was contoured in 3D Slicer (Brigham and Women's Hospital) and exported as a mesh (.stl). To orient the mesh with the SFDI images, the mesh was manually rotated and translated in MeshLab (ISTI – CNR) to align with the experimentally collected images. The processed mesh was passed to the diffusion theory light propagation model to generate the true depth and concentration maps for assessing the DL predictions. The DL predictions were downloaded for analysis in MATLAB. This analysis was repeated for all three models and each DL performance is compared.

3. RESULTS

3.1 In silico

The respective average depth error and standard deviation for the F+OP, F+R, and F/R models are 0.510±0.845mm, 0.525±0.832mm, and 0.639±0.866mm when 45 patient-derived tumors are used as testing data. Concentration errors across all models are also comparable. The respective concentration error and standard deviation for the F+OP, F+R, and F/R models are 0.719±0.879μg/mL, 0.854±1.036μg/mL, and 0.902±1.055μg/mL. Figure 4 shows DL predictions for a representative patient-derived in silico case. The true depth and concentration maps represent the ground truth from the original patient-derived tumor meshes used to generate the in silico SFDI images. To evaluate DL performances, the ground truth images are compared to the DL predictions. The error map for both depth and concentration is the absolute difference between the true and predicted images calculated for each pixel. These error maps highlight the increased errors at the periphery of the tumor as demonstrated by the border in the error maps. All DL models demonstrate comparable depth predictive performances when dealing with patient-derived data.
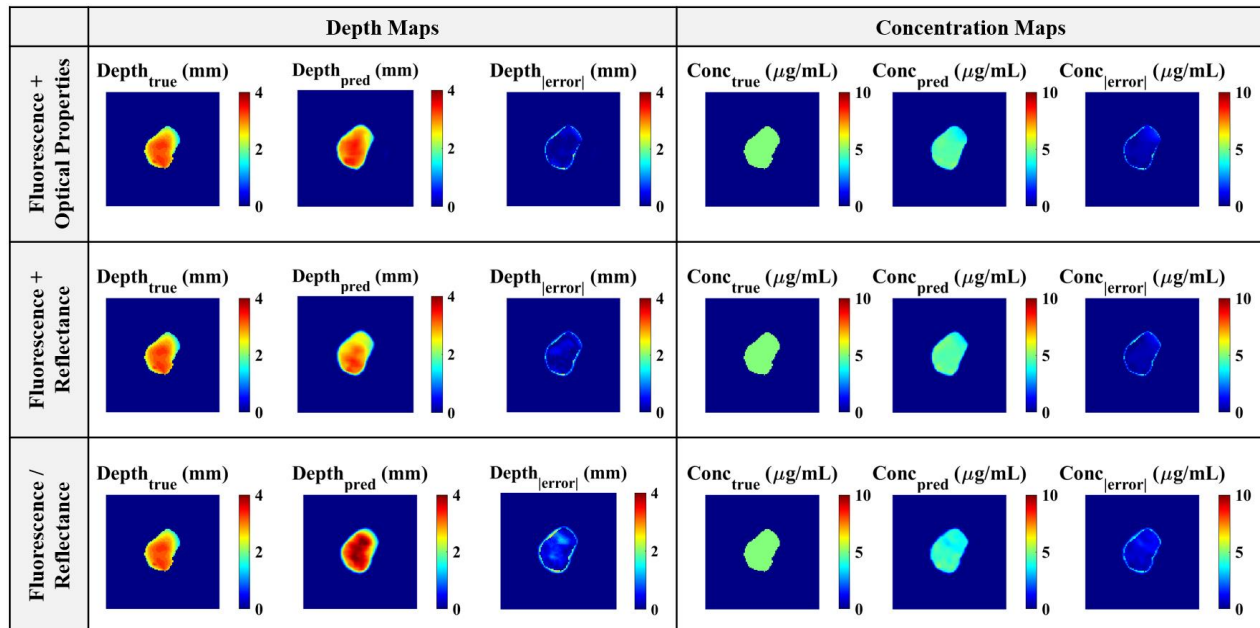

Figure 4. In silico SFDI images of patient-derived tumors are passed to each DL model: F+OP, F+R, F/R. Depth and concentration predictions are compared to the ground truth with error maps.

3.2 Phantom

Depth and concentration predictions for an optical phantom based on a representative patient tongue tumor are shown in Figure 5. The error map corresponds to the absolute difference between the true and predicted images. The respective depth errors for the F+OP, F+R, and F/R models are 2.178±0.999mm, 1.304±0.814mm, and 0.780±0.423mm. Concentration errors for the F+OP, F+R, and F/R models are 5.278±1.057μg/mL, 3.883±1.178μg/mL, and 0.435±0.423μg/mL. Both the F+OP and F+R models tend to underpredict depth and overpredict concentration. The F/R model demonstrates the most promise when dealing with experimental images of optical phantoms. The improved performance associated with the F/R model is likely due to any calibration issues or artifacts from the SFDI system being canceled through the division of fluorescence by reflectance.
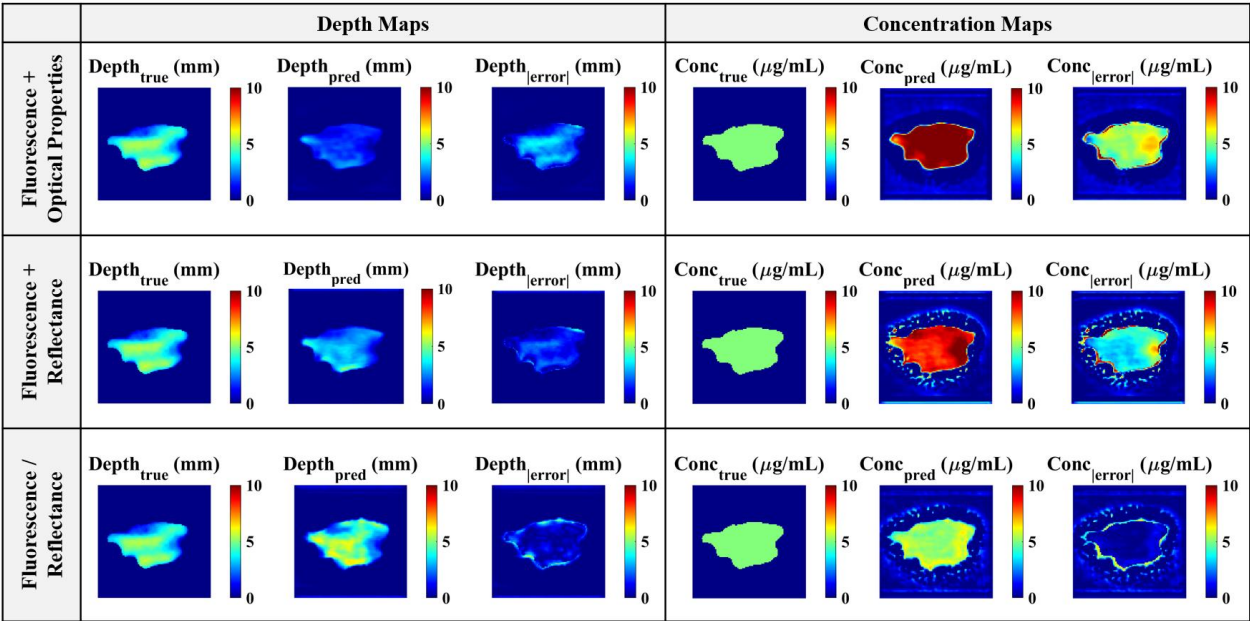

Figure 5. In silico SFDI images of patient-derived tumors are passed to each DL model: F+OP, F+R, F/R. Depth and concentration predictions are compared to the ground truth with error maps.

4. CONCLUSIONS AND FUTURE DIRECTIONS

The capabilities of a custom DL-enabled fluorescence SFDI system were investigated through simulations and experimental studies. The study aimed to determine what SFDI inputs coupled with distinct DL architectures produced the best depth and concentration predictions. A comparison between three distinct DL models was performed to explore which model had the best performance when coupled with the benchtop SFDI system. To train the DL models, synthetic tumors were generated to capture the complex features intrinsic to patient-derived oral cancer tumors. As validated during in silico experiments, these synthetic tumors demonstrate promise as training data for these initial simulations. All DL models demonstrated comparable performance during in silico simulations, but the F/R model showed the most promise in experimental studies, likely due to the reduced effects of SFDI system calibration errors.

Further study of the DL-enabled SFDI system is required for more complex cases. Future studies will focus on employing the F/R model in two pre-clinical models of oral cancer: i) phantoms with surface topography; and ii) in vivo and ex vivo mouse models. Phantoms with topography should be assessed to validate the ability of the F/R model to handle lighting variations on the surface. Animal studies are required to assess how the system performs with heterogeneous tissues. Validating the system in these complex cases is crucial for clinical translation.

ACKNOWLEDGMENTS

REFERENCES

[1] Pagedar, N. A., "Better Visualization of Oral Cancer Margins-A Struggle of Cancer and Technology," JAMA Otolaryngol Head Neck Surg **146**(12), 1156–1157 (2020).

[2] Mehanna, H., West, C. M., Nutting, C. and Paleri, V., "Head and neck cancer--Part 2: Treatment and prognostic factors," Bmj **341**, c4690 (2010).

[3] McMahon, J., O'Brien, C. J., Pathak, I., Hamill, R., McNeil, E., Hammersley, N., Gardiner, S. and Junor, E., "Influence of condition of surgical margins on local recurrence and disease-specific survival in oral and oropharyngeal cancer," Br J Oral Maxillofac Surg **41**(4), 224–231 (2003).

[4] Orosco, R. K., Tapia, V. J., Califano, J. A., Clary, B., Cohen, E. E. W., Kane, C., Lippman, S. M., Messer, K., Molinolo, A., Murphy, J. D., Pang, J., Sacco, A., Tringale, K. R., Wallace, A. and Nguyen, Q. T., "Positive Surgical Margins in the 10 Most Common Solid Cancers," Sci Rep **8**(1) (2018).

[5] Rosenthal, E. L., Warram, J. M., Bland, K. I. and Zinn, K. R., "The status of contemporary image-guided modalities in oncologic surgery," Ann Surg **261**(1), 46–55 (2015).

[6] Bou-Samra, P., Muhammad, N., Chang, A., Karsalia, R., Azari, F., Kennedy, G., Stummer, W., Tanyi, J., Martin, L., Vahrmeijer, A., Smith, B., Rosenthal, E., Wagner, P., Rice, D., Lee, A., Abdelhafeez, A., Malek, M. M., Kohanbash, G., Barry Edwards, W., et al., "Intraoperative molecular imaging: 3rd biennial clinical trials update," J Biomed Opt **28**(5), 050901 (2023).

[7] Pogue, B. W., Rosenthal, E. L., Achilefu, S. and van Dam, G. M., "Perspective review of what is needed for molecular-specific fluorescence-guided surgery," Journal of Biomedical Optics **23**(10), 1–9 (2018).

[8] Noorlag, R., de Bree, R. and Witjes, M. J. H., "Image-guided surgery in oral cancer: toward improved margin control," Curr Opin Oncol **34**(3), 170–176 (2022).

[9] DSouza, A. V., Lin, H., Henderson, E. R., Samkoe, K. S. and Pogue, B. W., "Review of fluorescence guided surgery systems: identification of key performance capabilities beyond indocyanine green imaging," J Biomed Opt **21**(8), 80901 (2016).

[10] Brouwer de Koning, S. G., Schaeffers, A., Schats, W., van den Brekel, M. W. M., Ruers, T. J. M. and Karakullukcu, M. B., "Assessment of the deep resection margin during oral cancer surgery: A systematic review," Eur J Surg Oncol **47**(9), 2220–2232 (2021).

[11] Gioux, S., Mazhar, A. and Cuccia, D. J., "Spatial frequency domain imaging in 2019: principles, applications, and perspectives," Journal of Biomedical Optics **24**(7), 1–18 (2019).

[12] Cuccia, D. J., Bevilacqua, F., Durkin, A. J., Ayers, F. R. and Tromberg, B. J., "Quantitation and mapping of tissue optical properties using modulated imaging," J Biomed Opt **14**(2), 024012 (2009).

[13] Hayakawa, C. K., Karrobi, K., Pera, V., Roblyer, D. and Venugopalan, V., "Optical sampling depth in the spatial frequency domain," Journal of Biomedical Optics **24**(7), 1–14 (2018).

[14] Smith, J. T., Ochoa, M., Faulkner, D., Haskins, G. and Intes, X., "Deep learning in macroscopic diffuse optical imaging," J Biomed Opt **27**(2) (2022).

[15] Tian, L., Hunt, B., Bell, M. A. L., Yi, J., Smith, J. T., Ochoa, M., Intes, X. and Durr, N. J., "Deep Learning in Biomedical Optics," Lasers Surg Med **53**(6), 748–775 (2021).

[16] Daly, M., Gilbank, A., Fleisig, J., Rajkotwala, M., Gupta, S., Sahovaler, A., Eu, D., Ferrari, M., Wilson, B. and Irish, J., [Fluorescence depth estimation in pre-clinical oral cancer models using spatial frequency domain imaging], SPIE (2021).

[17] Smith, J. T., Aguenounon, E., Gioux, S. and Intes, X., "Macroscopic fluorescence lifetime topography enhanced via spatial frequency domain imaging," Opt Lett **45**(15), 4232–4235 (2020).

[18] Ntziachristos, V. and Weissleder, R., "Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized Born approximation," Opt Lett **26**(12), 893–895 (2001).

[19] Angelo, J., Vargas, C. R., Lee, B. T., Bigio, I. J. and Gioux, S., "Ultrafast optical property map generation using lookup tables," J Biomed Opt **21**(11), 110501 (2016).