Real-time, wide-field and high-quality single snapshot imaging of optical properties with profile correction using deep learning

ENAGNON AGUÉNOUNON,[1,*] JASON T. SMITH,[2] MAHDI AL-TAHER,[3,4] MICHELE DIANA,[1,3,5] XAVIER INTES,[2] (iD) AND SYLVAIN GIOUX[1]

[1]*University of Strasbourg, ICube Laboratory, 300 Boulevard Sébastien Brant, 67412 Illkirch, France*
[2]*Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
[3]*Institute of Image-Guided Surgery, IHU Strasbourg, Strasbourg, France*
[4]*Maastricht University Medical Center, Maastricht, The Netherlands*
[5]*Research Institute against Digestive Cancer, IRCAD, Strasbourg, France*
[*]*faguenounon@unistra.fr*

Abstract: The development of real-time, wide-field and quantitative diffuse optical imaging methods to visualize functional and structural biomarkers of living tissues is a pressing need for numerous clinical applications including image-guided surgery. In this context, Spatial Frequency Domain Imaging (SFDI) is an attractive method allowing for the fast estimation of optical properties using the Single Snapshot of Optical Properties (SSOP) approach. Herein, we present a novel implementation of SSOP based on a combination of deep learning network at the filtering stage and Graphics Processing Units (GPU) capable of simultaneous high visual quality image reconstruction, surface profile correction and accurate optical property (OP) extraction in real-time across large fields of view. In the most optimal implementation, the presented methodology demonstrates megapixel profile-corrected OP imaging with results comparable to that of profile-corrected SFDI, with a processing time of 18 ms and errors relative to SFDI method less than 10% in both profilometry and profile-corrected OPs. This novel processing framework lays the foundation for real-time multispectral quantitative diffuse optical imaging for surgical guidance and healthcare applications. All code and data used for this work is publicly available at www.healthphotonics.org under the resources tab.

1. Introduction

Spatial Frequency Domain Imaging (SFDI) is a relatively inexpensive diffuse optical imaging method allowing quantitative, widefield and rapid measurement of optical properties [1–4]. The capabilities of SFDI have been demonstrated across a large number of biomedical applications including small animal imaging, burns wound assessment, skin flap monitoring, early stage diabetic foot ulcers diagnosis, among others [5–9]. However, because the standard implementation of SFDI requires the sequential acquisition of several sinusoidal patterns of light, its potential for clinical applications that require real-time feedback, such as intraoperative surgical guidance, is limited.

With the aim of deploying such technology in surgical settings, Single Snapshot imaging of Optical Properties (SSOP) was introduced as a new approach for performing SFDI in real-time. Conceptually, SSOP works by reducing the total number of necessary acquisitions to a single image – thereby increasing the speed of acquisition. First proofs of concept demonstrated the feasibility of real-time optical properties imaging (i.e. faster than 25 images per second) with high accuracy (i.e. with less than 10% error compared with SFDI). However, the approach initially suffered from both loss of image resolution and undesirable artifacts due to the single

image acquisition strategy [10,11]. Over the years, the SSOP method has been increasingly improved upon to enhance the technique's reconstructive visual quality and capabilities. Though at present current state-of-the-art SSOP implementations, while capable of acceptable visual quality retrieval in real-time, still suffer from edge artifacts, exhibit lower resolution than SFDI and do not include sample profile correction [12]. These limitations can mainly be attributed to a conventional processing workflow.

To overcome these limitations, deep learning (DL) and Graphics Processing Units (GPU) computation provide promising alternatives to conventional processing for diffuse optical imaging. Deep Neural Networks (DNNs), and more particularly Convolution Neural Networks (CNNs), have become ubiquitous in applications such as image classification, image segmentation, image translation, natural language understanding or speech recognition [13–15]. Over the last few years, they have also greatly impacted the field of biomedical optics [16–18]. Though, if these new methodologies have provided excellent performances, they are typically not amenable to real time processing due to the increased depth of the networks employed and software implementations. In this regard, the use of GPU-accelerated DL has demonstrated remarkable computational speed increases with regards to model inference in latency sensitives applications [19,20].

In this work, we present a real-time, wide-field and quantitative optical properties imaging implementation of SSOP that includes 3D profile correction and high visual quality reconstruction via GPU-accelerated CNNs. In particular, the workflow detailed herein is used to extract modulated images and the 3D profile of the sample concurrently from a single SSOP image input. A custom-made GPU implementation of the network, using NVIDIA's **cuDNN** library, optimizes the network's inference speed – exhibiting speeds comparable to that of the fastest existing approaches with enhanced accuracy. All together, we demonstrate profile corrected optical properties imaging in real-time with visual quality similar to SFDI and errors less than 10% in absorption and reduced scattering.

2. Materials and methods

2.1. Spatial frequency domain imaging

2.1.1. State of the art in SSOP

The standard SFDI method requires the sequential acquisition of several spatially-encoded images for the retrieval of optical property maps. The acquisition is typically performed using at least at two different spatial frequencies (e.g., $f_x = 0$ and $0.2\,\text{mm}^{-1}$) and 3 phase shifts or more for improved accuracy and visual quality (e.g. 7 phases) [4,21]. The patterns are generated and projected using a Digital Micromirror Device (DMD) and captured with a camera as depicted in Fig. 1. (A). Following the acquisition, the images are used to obtain the amplitude modulation at each spatial frequency, e.g. M_{DC} for $f_x = 0\,\text{mm}^{-1}$ and M_{AC} for $f_x = 0.2\,\text{mm}^{-1}$. From these measurements, the diffuse reflectance images $R_{d,DC}$ and $R_{d,AC}$ are obtained through calibration with a material of known optical properties. Finally, the measured reflectance allows for the pixel-wise estimation of the absorption and reduced scattering coefficients (μ_a, μ_s') across the image via the resolution of the inverse problem – such as with precomputed high density lookup tables (LUTs) generated using White Monte Carlo simulations [22].

In contrast to conventional SFDI, which necessitates a minimum of six images for quantification, SSOP is an alternative SFDI implementation capable of fast optical property extraction via acquisition of a single high spatial frequency image (e.g., $f_x = 0.2\,\text{mm}^{-1}$) – reducing the total acquisition time substantially by a few-fold, enabling real-time acquisition [4]. Amplitude modulations are extracted by filtering directly in the Fourier domain with both a low and high pass filter [10–12]. The general workflow is then similar to that of SFDI (illustrated in Fig. 1. (B)). Of importance, one should note that demodulation is the key step with which the resulting accuracies of both OP map retrieval and visual quality depend. Details of the acquisition,
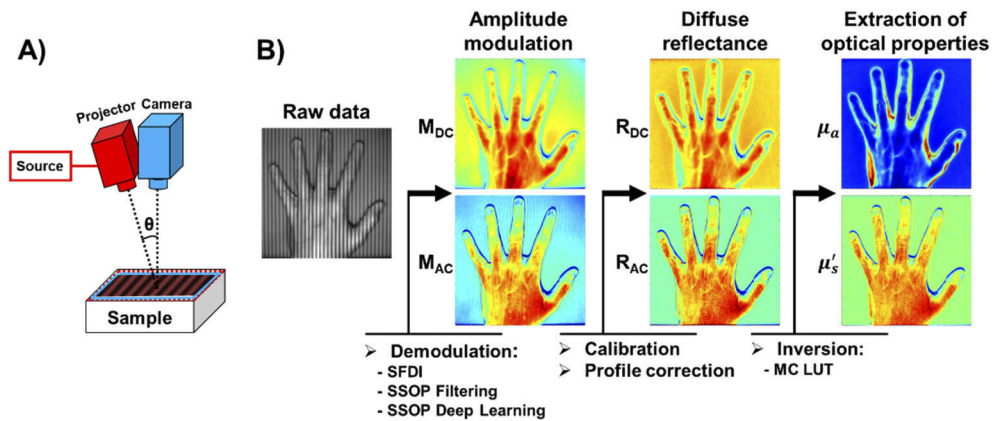
Fig. 1. (A) Schematics of a SFDI/SSOP imaging system. (B) SFDI/SSOP processing workflow.

demodulation and processing typically used in SFDI and SSOP are provided in [3,4], and details about creating a custom SFDI system and processing code are available on www.openSFDI.org and www.healthphotonics.org [23].

2.1.2.　SSOP demodulation

Typically, SSOP demodulation is performed either line by line [10], or in 2D spatial Fourier domain using rectangular low and high pass filters [11]. Recent developments in 2D spatial Fourier domain include using anisotropic low pass filtering based on sine windows and anisotropic high pass filtering based on Blackman windows [12]. This approach reported accuracies under 8.8% in absorption (μ_a) and 7.5% in reduced scattering (μ_s'). While this SSOP implementation was successfully used to perform real time oxygenation measurements [24], it is still affected by edge discontinuity, has a lower resolution than standard SFDI and does not include profile correction for the sample.

2.1.3.　3-dimensional profile correction applied to SFDI

Optical properties measurements of non-flat samples through SFDI conventionally suffer from inaccuracies due to intensity variations across the sample surface. To account for these errors, 3D profile correction methods for SFDI have been implemented by measuring the profile using structured-light profilometry [25,26] and correcting for its effect [27]. Several correction methods have been proposed for SFDI, all using either phase shifting profilometry or Fourier-based profilometry. Overall, they have illustrated high correction accuracy for large angles of tilt, from 40 up to 75 degrees depending on the method [11,27–29]. However, the use of a single image SSOP acquisition approach leads to lower profile accuracy that significantly affects the accuracy of retrieved OPs [11]. The main focus of the work herein is to leverage a novel deep learning approach to enable profile-corrected SSOP imaging in real-time and for complex non-flat samples.

2.2.　Deep learning approaches to SFDI

2.2.1.　State of the art

Several groups have focused their efforts on developing deep learning and machine learning approaches for SFDI or SSOP to improve accuracy, visual quality and/or speed. A random forest regression algorithm was developed for estimating optical properties in the spatial frequency domain, trained on Monte-Carlo simulations data [30]. The algorithm "learns" the nonlinear

mapping between diffuse reflectance at two spatial frequencies, and optical properties of the tissue, with the goal of reducing the computational complexity when a high density pre-computed LUT is used. Using this method, optical properties could be obtained over a 1-megapixel image in 450 ms with errors of 0.556% in μ_a and 0.126% in μ_s'.

Following, a DNN-based framework was proposed for multi-spatial frequency processing SFDI [31]. This network, a Multi-Layered Perceptron (MLP), outperformed other classic inversion methods (iterative method and nearest search) in terms of speed with computation times of 5ms for 100×100 pixels and 200ms for 696×520 pixels, using a 5 spatial frequencies SFDI acquisition and substantially reducing measurement uncertainties compared to the widely used 2 frequency SFDI processing method. More recently, CNNs have been introduced for diffuse optical imaging owing to their remarkable performance enhancements across imaging applications compared with MLPs [32]. In the case of SFDI, a Generative Adversarial Networks (GANs) framework (named "GANPOP") has been reported for the retrieval of optical properties directly from single input images with high accuracy [33]. In this work, SFDI was used to obtain ground-truth optical properties and the model developed took approximately 40 ms to process a 256×256 pixel image using an NVIDIA Tesla a P100 GPU Accelerator.

Both the random forest and MLP works have a commonality in that they both estimate optical properties from diffuse reflectance – though, GANPOP bypasses the demodulation step and retrieves optical properties directly from calibrated SSOP input. In addition, all these methods have been developed with a focus on providing a more accurate optical properties imaging ability, and though GANPOP illustrates partially profile-corrected OP retrieval, no DL-based workflow to date retrieves profilometry directly. Further, both the MLP and GANPOP's generator are comprised of a relatively large parametric size (on the order of 1×10^6 and 8×10^7 weights, respectively). In turn, these relatively large DL models can be both computationally burdensome and memory intensive, limiting their use for real time applications. In addition, though GANs have been increasing in popularity across applications of computer vision, they are notoriously problematic with regards to hyperparameter tuning, stable convergence and computational demand for proper model training [34]. Finally, no prior DL-based approach has actually been applied in real-time.

2.2.2. The proposed method

Inspired by these previous developments and results, we propose herein to combine GPU accelerated computing [24] with efficient DL architectures applied for a seamless and user-friendly SSOP processing methodology – allowing for real-time image reconstruction of mega-pixel images with high accuracy and visual quality along with direct retrieval of profilometry. The method consists of the following steps:

<u>U-NET 1: Demodulation</u>: A first small-size U-NET (less than 2×10^4 parameters) is used to get demodulation images (M_{DC} and M_{AC}), instead of directly map raw input data to optical properties. Since the network has dedicated weights according to the spatial frequency, both higher visual quality and more accurate optical properties can be obtained, in contrast with a Fourier filtering method with fixed spatial frequency parameters.

<u>U-NET 2: Profilometry</u>: A second small-size U-NET (less than 2×10^4 parameters), identical in structure to the first one, is used to get wrapped phase and obtain the profile, avoiding the use of Fourier domain 3D profilometry that leads to poor accuracy due to its dependency to spatial frequency. In our implementation, the real and imaginary part are extracted from the network to compute the wrapped phase by mean of atan2() function [35].

<u>Real-time profile-corrected optical properties imaging</u>: To reduce the inference time of both U-NETs, a custom-made GPU implementation of the networks was undertaken using the CUDA deep learning library **cuDNN**. Following the use of the two computationally inexpensive U-NETs described above, (used for demodulation and profilometry, respectively), the extraction of optical

properties can be performed along with profile correction at high speed using a previously developed real-time optical properties GPU processing code [24]. The use of the GPU-accelerated twin U-NET framework detailed above, followed by a GPU-optimized precomputed lookup table inversion method has a strong potential for delivering real-time and high visual quality SSOP imaging capabilities to many who would otherwise be unable to utilize more computationally burdensome approaches.

2.3. Deep learning network design

2.3.1. Structure

The general network architecture is illustrated in Fig. 2. and is inspired by the U-Net architecture [36]. It is a n-dimensional network ("Line U-NET" when $n = 1$ and "Image U-NET" when $n = 2$). It consists of a contracting path of 4 stages (left side) and an expansive path of 4 stages (right side). The expansive path is a fork and has 2 outputs. The contracting path consists of the repetition of a $7 \times 7^{n-1}$ convolution layer (padded convolution), each followed by a Tanh activation layer and a $2 \times 2^{n-1}$ max pooling layer (with stride 2). The number of feature channels is double at each level starting by 1 feature channel at stage 1. Every stage in the expansive path consists of an up-sampling $2 \times 2^{n-1}$ layer followed by a $3 \times 3^{n-1}$ deconvolution layer that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path, and a $7 \times 7^{n-1}$ convolution layer followed by a Tanh activation layer. At the final stage, a $7 \times 7^{n-1}$ convolution layer with linear activation is employed to retrieve the output. In total, the twin U-NET architecture is comprised of either 3064 or 18886 weights (according to n equal to 1 or 2, respectively) for 17 convolutional layers. The python code for each architecture is available at www.healthphotonics.org in the "resources" section.
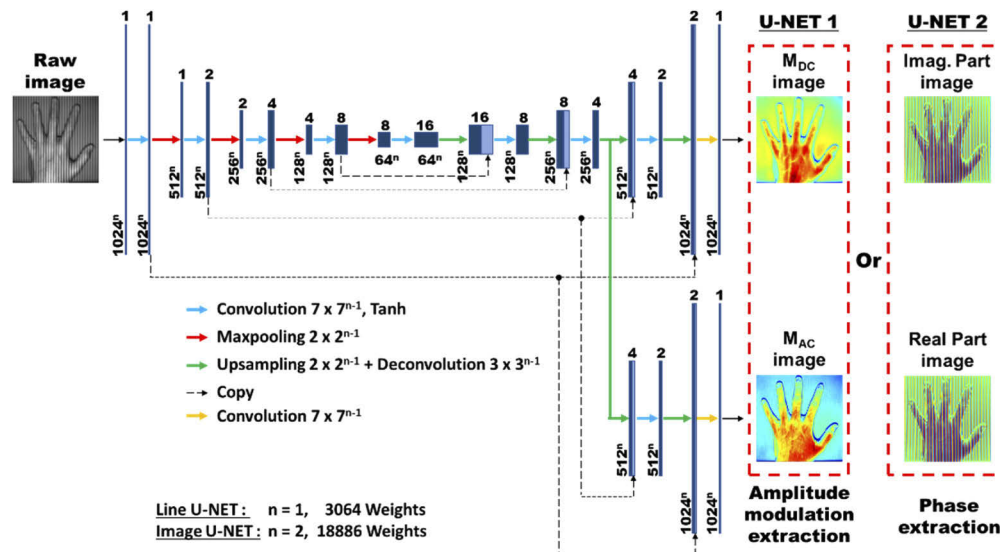

Fig. 2. Detailed architecture of the deep learning network architecture

2.3.2. Training

The network was implemented with *Tensorflow* using *Keras* backend and is compiled with mean squared error (MSE) as loss function and Adam as optimizer [42]. The training starts with a learning rate of 0.001 and employs a learning rate reduction (factor of 1.11) if there is no loss improvement after 15 epochs. The model is trained for 900 epochs using NVIDIA GTX 1080Ti

computation power in less than 2 hours. The training dataset consists of a total of 100 images that are 1024×1024 pixels in size composed of a combination of 7-phases SFDI acquisitions (for both accuracy and visual quality) of hands (N = 60 from 10 different hands placed in various orientations), silicon-based tissue mimicking phantoms (N = 20 from 18 different phantoms in various orientations) and ex vivo pig organs (N = 20 from 6 different organs). A 5-fold cross-validated was performed on the network and no signs of overfitting was observed.

2.4. GPU C CUDA implementation

Our custom SSOP processing code (using C CUDA) is made with Microsoft Visual Studio and NVIDIA CUDA toolkit environment. Based on the previous developed code [24], our work focused on the GPU implementation of the novel demodulation via the previously described twin U-NET workflow. To perform model inference, there are three options: using the training framework itself (convention), using TensorRT [39] or writing a custom code to execute the network using **cuDNN** low-level libraries and math operations [40,41]. We choose the latter for its high efficiency compared to other solutions. Since the network is a succession of many convolution, max pooling, up-sampling and concatenation, layers will be presented only once.

2.4.1. Convolution

Before one uses **cuDNN** functions to describe and execute a network, a handle must be created using **cudnnCreate**. A convolution consists of a description part and an execution part.

- Description part: (1) an input tensor is created with **cudnnCreateTensorDescriptor** and **cudnnSetTensor4dDescriptor**, this tensor memory space is then allocated; (2) the filter is described with **cudnnCreateFilterDescriptor** and **cudnnSetFilter4dDescriptor**, a memory space is allocated and the weights loaded; (3) the type of convolution is specified using **cudnnCreateConvolutionDescriptor** and **cudnnSetConvolution2dDescriptor**, to get the convolution output dimension **cudnnGetConvolution2dForwardOutputDim** is called; (4) a second tensor is created as previously for the convolution output; (5) the activation is set with **cudnnCreateActivationDescriptor** and **cudnnSetActivationDescriptor**; (6) a third tensor is created as previously to store the output after activation; (7) the optimized convolution algorithm is found with **cudnnGetConvolutionForwardAlgorithm** and the required work space for its execution with **cudnnGetConvolutionForwardWorkspaceSize**, this space is then allocated.

- Execution part: **cudnnConvolutionForward** is called first to execute the convolution, follow by a custom made **add_bias_gpu** kernel. Finally, to finish the execution, an activation function **cudnnActivationForward** is called.

2.4.2. Max pooling

The max pooling operation is also comprised of a description and an execution part: (1) it requires an input tensor (e.g., the previous activation tensor); (2) a description of the pooling is created via **cudnnCreatePoolingDescriptor** and **cudnnSetPooling2dDescriptor** and the pooling output dimension is defined with **cudnnGetPooling2dForwardOutputDim**; (3) a tensor is pre-allocated to store the output after pooling; (4) lastly, the max pooling is performed by calling **cudnnPoolingForward**.

2.4.3. Upsampling and concatenation

Upsampling and concatenation layers are custom made kernels since these two functions are not currently implemented on **cuDNN**'s library. With each of these bricks, the entire network is then built according to the structure described in the previous section.

2.4.4. GPU configuration

Processing times depending strongly on the GPU configuration. For each custom-made kernel, one thread was created for each pixel, broke down into blocks of 64 threads allowing to have a GPU occupancy of 100%.

2.5. Imaging system used for experiments

The instrumental setup (see Fig. 1. (A)) was custom-built using a DMD (Vialux, Germany) for the projection of custom patterns, fiber-coupled to a 665 nm laser diode (LDX Optronics, Maryville, Tennessee). The laser diode was mounted using a temperature-controlled mount (TCLDM9, Thorlabs, Newton, New Jersey). Diode intensity was controlled using current controllers (TDC240C, Thorlabs, Newton, New Jersey) and temperature controlled using thermoelectric cooler controllers (TED200C, Thorlabs, Newton, New Jersey). The projection system projects a sine wave pattern over a 175×175 mm^2 field of view at 45 cm working distance. Images of 1024 $\times$ 1024 pixels size are acquired using a scientific monochrome 16 bits sCMOS camera (PCO AG, pco.edge 5.5, Kelheim, Germany). Polarizers (PPL05C; Moxtek, Orem, Utah), arranged in a crossed configuration, were used to minimize the contribution from specular reflections at the surface of the sample. The projection system was set at an angle of $\theta = 4°$ to allow profilometry measurement [27]. A workstation with the following characteristics was used for controlling the acquisition and the processing: Intel i7-7800x 3.5GHz central processing unit, 16 GB of RAM, and an NVIDIA GeForce GTX 1080Ti GPU (3584 Cores CUDA, 11 GB of RAM).

2.6. Training set

20 images of the combination of 18 different phantoms (specifications provided below), 60 images of 10 different hands (Caucasian and Black men and women) in various orientations and 20 images from 6 different *ex vivo* pig organs in various configurations and orientations (intestine, colon, stomach, liver, esophagus and pancreas) were used to train the network. Silicone-based optical phantoms were built using titanium dioxide (TiO2) as a scattering agent and India ink as an absorbing agent [37,38]. A total of 18 phantoms were used with the following absorption and reduced scattering properties at 665nm: [(0.01,0.85); (0.01,1.13); (0.01,1.39); (0.01,1.58); (0.02,0.71); (0.02,0.80); (0.02,1.71); (0.05,0.92); (0.05,1.50); (0.06,1.31); (0.07,1.15); (0.08,0.78); (0.11,1.11); (0.12,0.80); (0.12,0.85); (0.14,0.83); (0.21,1.18); (0.23,0.70)] all expressed in mm^{-1}. A calibration phantom was made with a large size (210 mm $\times$ 210 mm $\times$ 20 mm) and with absorption of $\mu_a = 0.01$ mm^{-1} and reduced scattering $\mu_s' = 1.1$ mm^{-1} at 665 nm. All phantoms optical properties were referenced using a time domain photon counting system.

2.7. Validation and performance assessment

Our validation dataset consisted of the images of 4 different hands and 20 images from 6 different *ex vivo* pig organs in various configurations and orientations (intestine, colon, stomach, liver, esophagus and pancreas). None of the validation images were used for training. Images were acquired via 7-phase SFDI using 4 spatial frequencies (0.1 mm^{-1}, 0.2 mm^{-1}, 0.3 mm^{-1}, 0.4 mm^{-1}) to extract optical properties. These images were processed using four different processing modalities: SFDI, SSOP Filtering [12], SSOP Deep Learning (noted "SSOP DL conv") with convolutions of 7×1 and 7×7. For the quantitative performance assessment, 3D profile and optical properties maps obtained using SSOP methods at each spatial frequency were compared to that of 7-phase SFDI. These comparisons were undertaken by measuring the mean absolute percentage error given by the following Eq. (1) where N and M refer to image size

(number of pixels in x and y directions):

$$Error(\%) = \frac{1}{N \times M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left| 100 \times \frac{Result_{n,m}^{REFERENCE} - Result_{n,m}^{TEST}}{Result_{n,m}^{REFERENCE}} \right| \tag{1}$$

For the image visual quality assessment, structural similarity index (SSIM) of 3D profile and optical properties maps obtained using SSOP methods at each spatial frequency were compared to that of 7-phase SFDI.

For the GPU implementation performance assessment, NVIDIA Nsight environment was used to trace the C CUDA code processing time. This time does not account for the transfer time between the CPU and the GPU memories (which is 338 µs for a 1024×1024 pixels size image) on our system. The evaluation of the quality of the images was performed visually to consider degradations, such as ripples and edge artifacts. Finally, a movie of a moving hand was acquired at $f_x = 0.2\,\text{mm}^{-1}$ with rotation and translation to compare and contrast the three SSOP analytic method's real-time capabilities.

Note that the codes, for each demodulation method (SFDI 7 phases, SSOP Filtering, SSOP DL conv 7×1 and SSOP DL conv 7×7), and all data (calibration phantoms, hands, and *ex vivo* sample) used in this work are available at www.healthphotonics.org in the "resources" section.

3. Results

The 24 images of the test dataset were processed as described before with the four different modalities. To illustrate these results, we show in Fig. 3., Fig. 4. and Fig. 5. respectively along with their pixel-wise percentage error map the results of the measured profile, profile-corrected absorption coefficient and profile-corrected reduced scattering coefficient at each spatial frequency obtained on a hand. These images allow one to both qualitatively appreciate and quantitatively assess each SSOP processing method's reconstructive performance and against 7-phase SFDI. We also present in Fig. 6 the results obtained at spatial frequency $f_x = 0.2\,\text{mm}^{-1}$ on another hand image and 4 images of *ex vivo* porcine organs that illustrate how shape variety influence the accuracy of the optical properties extracted via each method.
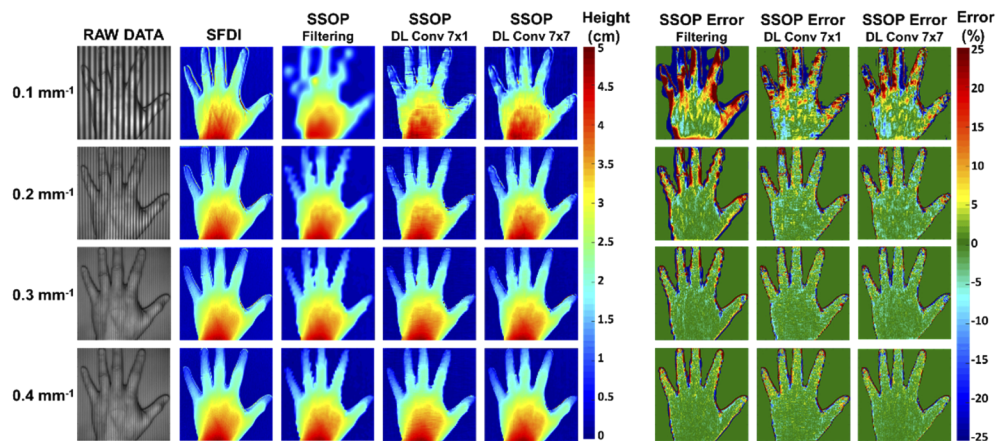

Fig. 3. Profile measurements: 3D profile maps obtained at four spatial frequencies with SFDI method and three SSOP processing methods along with their corresponding error maps. From left to right: Raw input data, SFDI reference, SSOP Filtering, SSOP Deep Learning Convolution (DL conv) 7×1, SSOP Deep Learning Convolution 7×7, SSOP Filtering percentage error, SSOP Deep Learning Convolution 7×1 percentage error, SSOP Deep Learning Convolution 7×7 percentage error.
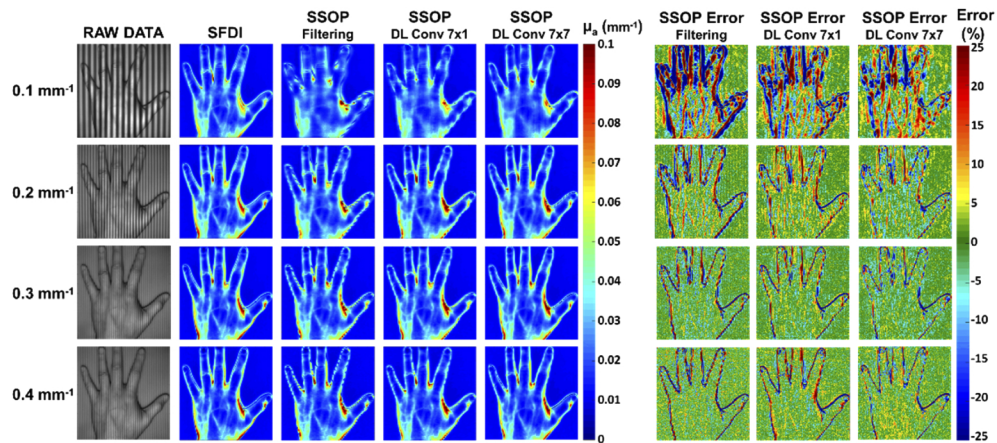
Fig. 4. Profile-corrected absorption measurements: profile-corrected absorption coefficient maps obtained at four spatial frequencies with SFDI method and three SSOP processing methods along with their corresponding error maps. From left to right: Raw input data, SFDI reference, SSOP Filtering, SSOP Deep Learning Convolution (DL conv) 7×1, SSOP Deep Learning Convolution 7×7, SSOP Filtering percentage error, SSOP Deep Learning Convolution 7×1 percentage error, SSOP Deep Learning Convolution 7×7 percentage error.

Fig. 5. Profile-corrected reduced scattering measurements: profile-corrected reduced scattering coefficient maps obtained at four spatial frequencies with SFDI method and three SSOP processing methods along with their corresponding error maps. From left to right: Raw input data, SFDI reference, SSOP Filtering, SSOP Deep Learning Convolution (DL conv) 7×1, SSOP Deep Learning Convolution 7×7, SSOP Filtering percentage error, SSOP Deep Learning Convolution 7×1 percentage error, SSOP Deep Learning Convolution 7×7 percentage error.
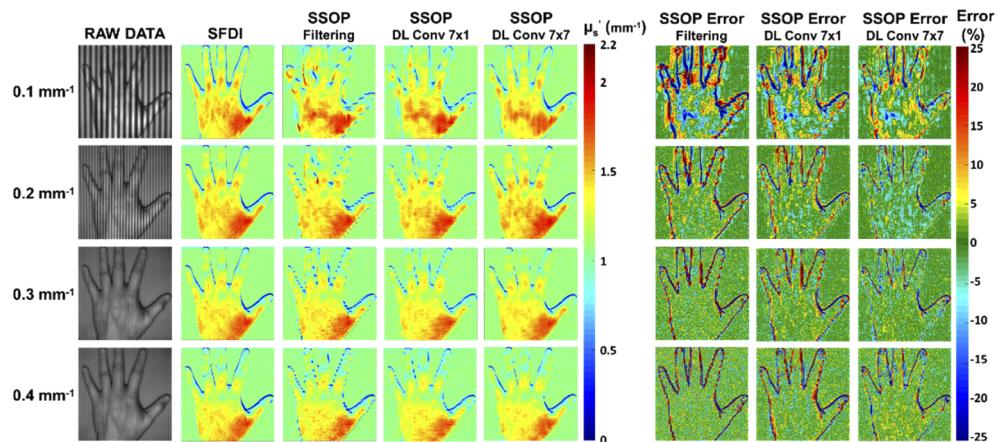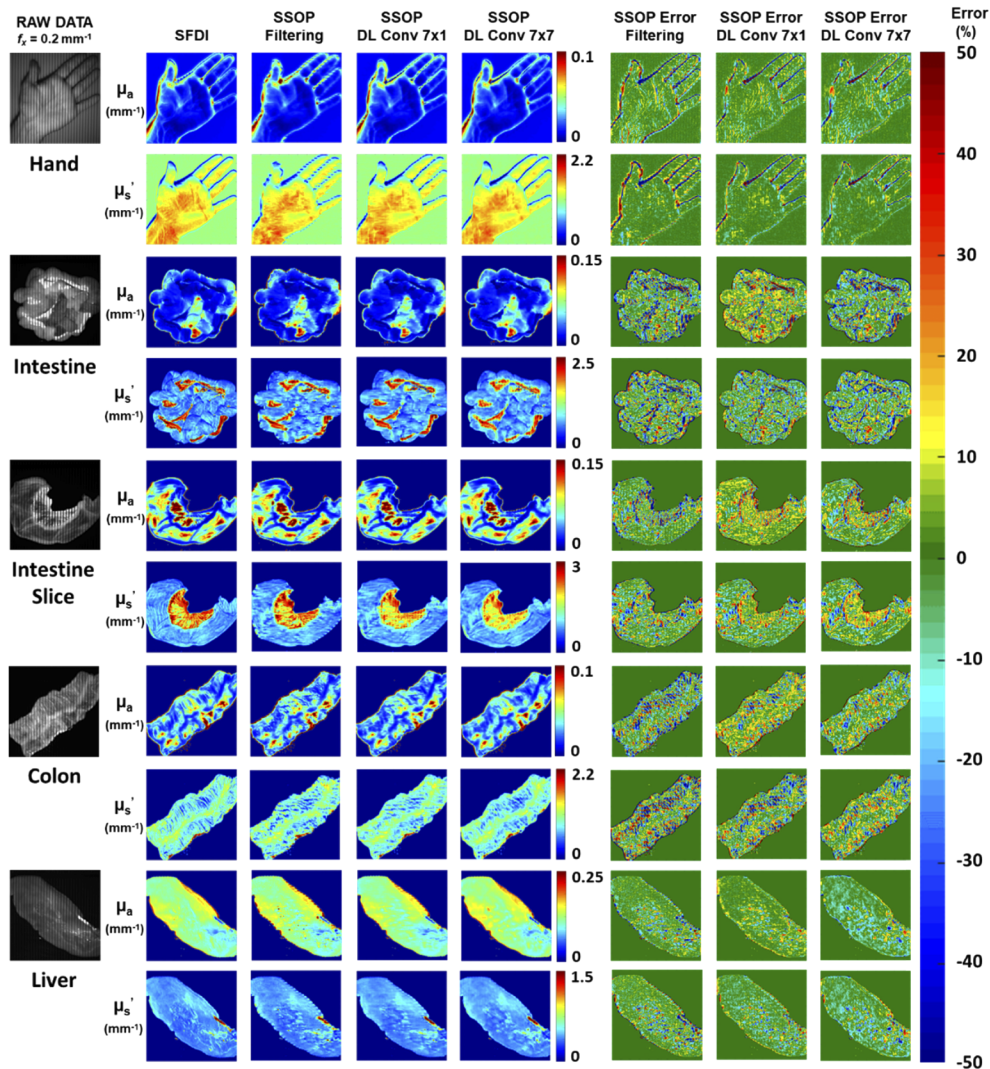
Fig. 6. Profile-corrected optical properties (absorption and reduced scattering) measurements obtained at $f_x = 0.2 \, \text{mm}^{-1}$ on a hand, an intestine, an intestine slice, a colon and a liver with SFDI method and three SSOP processing methods along with their corresponding error maps. From left to right: Raw input data, SFDI reference, SSOP Filtering, SSOP Deep Learning Convolution (DL conv) 7×1, SSOP Deep Learning Convolution 7×7, SSOP Filtering percentage error, SSOP Deep Learning Convolution 7×1 percentage error, SSOP Deep Learning Convolution 7×7 percentage error.

3.1. Image quality

<u>SFDI</u>: It is important for one to appreciate the reconstructive visual quality retrieved through the use of 7-phase SFDI. The 7-phase acquisition method improves visual quality compared to 3-phase and maintains better accuracy despite a signal to noise ratio decreasing when spatial frequency increases. It has been selected as the ground-truth for all error comparison herein.

<u>SSOP Filtering</u>: SSOP filtering results demonstrate an improvement in image resolution that corresponds with increased spatial frequency. This can be appreciated in the absorption coefficient maps by the vascular structure, which appears more clearly, and by the notable reduction of edge artifacts. Additionally, in the profile and reduced scattering coefficient maps, the improvement in resolution with increasing spatial frequency can be observed – with better definition of the hand shape and a reduction of the edge discontinuity effects. Both observations are particularly noticeable on percentage error maps.

<u>SSOP DL</u>: For the two SSOP DL methods, while still affected at low spatial frequency ($0.1 \, \text{mm}^{-1}$), the image degradation is significantly reduced compared to the SSOP Filtering approach. A much better definition of the hands and organs shape and a reduction of the edge discontinuity effect can be clearly noticed (supported by SSIM measurements, Table 1). Moreover, the SSOP DL Conv 7×7 method mimics SFDI image edges with near-perfection in stark contrast with SSOP DL Conv 7×1, which illustrates slightly degraded results.

Table 1. Structural similarity index of the 3D profile map and the optical properties map for all the validation dataset

Full image	Spatial frequency	Structural similarity index (SSIM)		
		SSOP Filtering	**SSOP DL Conv 7×1**	**SSOP DL Conv 7×7**
3D Profile	$0.1 \, \text{mm}^{-1}$	0.44 ± 0.16	0.49 ± 0.17	0.50 ± 0.16
	$0.2 \, \text{mm}^{-1}$	0.47 ± 0.15	0.50 ± 0.16	0.51 ± 0.16
	$0.3 \, \text{mm}^{-1}$	0.47 ± 0.15	0.51 ± 0.15	0.52 ± 0.15
	$0.4 \, \text{mm}^{-1}$	0.48 ± 0.15	0.51 ± 0.15	0.52 ± 0.15
Absorption	$0.1 \, \text{mm}^{-1}$	0.96 ± 0.02	0.98 ± 0.03	0.98 ± 0.01
	$0.2 \, \text{mm}^{-1}$	0.98 ± 0.01	0.98 ± 0.03	0.98 ± 0.01
	$0.3 \, \text{mm}^{-1}$	0.97 ± 0.02	0.97 ± 0.03	0.98 ± 0.02
	$0.4 \, \text{mm}^{-1}$	0.96 ± 0.02	0.96 ± 0.03	0.97 ± 0.02
Reduced scattering	$0.1 \, \text{mm}^{-1}$	0.75 ± 0.08	0.84 ± 0.07	0.86 ± 0.06
	$0.2 \, \text{mm}^{-1}$	0.75 ± 0.07	0.81 ± 0.08	0.83 ± 0.06
	$0.3 \, \text{mm}^{-1}$	0.71 ± 0.08	0.76 ± 0.09	0.78 ± 0.07
	$0.4 \, \text{mm}^{-1}$	0.67 ± 0.10	0.71 ± 0.10	0.74 ± 0.09

To quantify this analysis, we have calculated the structural similarity (SSIM) index versus 7-phase SFDI images (ground-truth) – results of which are provided in Table 1. The first observation is that, compared to the visual quality of the SFDI images at any spatial frequencies, we find a decreasing similarity index (hence a decrease in visual quality): (1) SSOP DL Conv 7×7, (2) SSOP DL Conv 7×1 and (3) SSOP Filtering. The similarity index varies according to the evaluated parameter but remains consistent with our visual analysis – with the exception of specific high spatial frequency conditions.

3.2. Quantitative analysis

Mean absolute percentage error of the full image for all the 24 samples SSOP error maps are computed verses 7-phase SFDI and reported in Table 2 (in mean ± standard deviation) for quantitative assessment.

Table 2. Mean absolute percentage error (%) of the 3D profile map and the optical properties map for all the validation dataset

Full image	Spatial frequency	Mean absolute percentage error (%)		
		SSOP Error Filtering	SSOP Error DL Conv 7×1	SSOP Error DL Conv 7×7
3D Profile	$0.1\,\text{mm}^{-1}$	17.6 ± 6.1	16.7 ± 6.3	16.6 ± 6.5
	$0.2\,\text{mm}^{-1}$	11.9 ± 5.1	12.3 ± 6.4	11.3 ± 5.8
	$0.3\,\text{mm}^{-1}$	10.1 ± 5.3	11.0 ± 6.6	10.1 ± 5.8
	$0.4\,\text{mm}^{-1}$	9.8 ± 5.8	10.2 ± 6.4	9.5 ± 5.9
Absorption	$0.1\,\text{mm}^{-1}$	12.6 ± 4.1	9.4 ± 6.9	9.4 ± 3.5
	$0.2\,\text{mm}^{-1}$	8.7 ± 2.6	8.6 ± 7.0	7.7 ± 2.8
	$0.3\,\text{mm}^{-1}$	9.9 ± 3.0	10.8 ± 7.6	9.5 ± 4.0
	$0.4\,\text{mm}^{-1}$	11.8 ± 3.8	12.4 ± 8.8	12.0 ± 5.6
Reduced scattering	$0.1\,\text{mm}^{-1}$	11.3 ± 3.2	8.3 ± 4.8	7.5 ± 2.7
	$0.2\,\text{mm}^{-1}$	9.3 ± 2.5	8.5 ± 6.6	7.6 ± 2.6
	$0.3\,\text{mm}^{-1}$	10.7 ± 3.1	11.1 ± 9.3	9.9 ± 3.8
	$0.4\,\text{mm}^{-1}$	12.5 ± 3.8	13.1 ± 10.6	12.3 ± 5.4

Profile: Generally, 3D profile errors decrease when spatial frequency increases: SSOP DL Conv 7×7 demonstrated the best results with errors as low as $\mathbf{9.5 \pm 5.9}\%$ followed by SSOP Filtering that gave errors as low as $\mathbf{9.8 \pm 5.8}\%$ and SSOP DL Conv 7×1 with errors as low as $\mathbf{10.2 \pm 6.4}\%$.

Profile-corrected absorption: Profile-corrected absorption coefficients illustrate a decreasing error trend along with spatial frequency until 0.2mm^{-1} and an increase at 0.3mm^{-1}: SSOP DL Conv 7×7 still performs best, with errors as low as $\mathbf{7.7 \pm 2.8}\%$, followed by SSOP DL Conv 7×1 with errors as low as $\mathbf{8.6 \pm 7.0}\%$ and SSOP Filtering with errors as low as $\mathbf{8.7 \pm 2.6}\%$.

Profile-corrected reduced scattering Profile-corrected reduced scattering coefficient errors increase with spatial frequency. SSOP DL Conv 7×7 still performs better with errors as low as $\mathbf{7.5 \pm 2.7}\%$, followed by SSOP DL Conv 7×1 with errors as low as $\mathbf{8.3 \pm 4.8}\%$ and SSOP Filtering with errors as low as $\mathbf{9.3 \pm 2.5}\%$.

Overall, qualitative and quantitative analyses using both image-to-image similarity and pixelwise error illustrate the edge improvements provided by the twin U-NET deep learning workflow as well as an enhanced accuracy in optical properties extraction.

3.3. Processing time

The analysis of the GPU processing times for the three different SSOP implementations is summarized in Table 3 on images having a size of 1024×1024 pixels.

Table 3. GPU processing time in milliseconds for the three different SSOP processing method

Image 1024 x1024	GPU Processing time (in ms)		
	SSOP Filtering	SSOP DL Conv 7×1	SSOP DL Conv 7×7
Get $\mathbf{M_{DC}}$ & $\mathbf{M_{AC}}$	0.51	6.7	8.8
Get Profile	0.41	6.9	9.0
Get $\mathbf{R_{DC}}$ & $\mathbf{R_{AC}}$	0.12	0.12	0.12
Get μ_a & μ'_s	0.09	0.09	0.09
TOTAL	**1.13**	**13.81**	**18.01**

On our workstation, the SSOP Filtering method was the fastest with 1.13 ms, followed by Deep Learning with 7×1 convolution with 13.81 ms and finally Deep Learning with 7×7 convolution with 18.01 ms. As expected, the computation time to obtain diffuse reflectance and to extract optical properties is the same for all method (using the same GPU-optimized LUT). Thus, the SSOP deep learning methods are only limited by the network inference time. Overall, since a video rate of 25 images per seconds is limited at 40 ms, all three processing methods demonstrate capabilities to be used in real-time environments. Note that processing time for the 7 phase SFDI method is not reported since the method is not suited for real-time imaging.

Finally, to compare all three SSOP approaches, a movie of a hand was acquired (**Visualization 1**, single-frame excerpt from the video recording shown in Fig. 7). In this movie, one can notice the edge discontinuity effect when using SSOP Filtering and how it depends on the hand position. In contrast, the SSOP Deep Learning workflow presented herein significantly reduces these effects and reconstructs OPs with high accuracy. Thus, for the first time, this work illustrates the potential of SSOP for use in challenging settings where motion, depth and real-time feedback are critical.
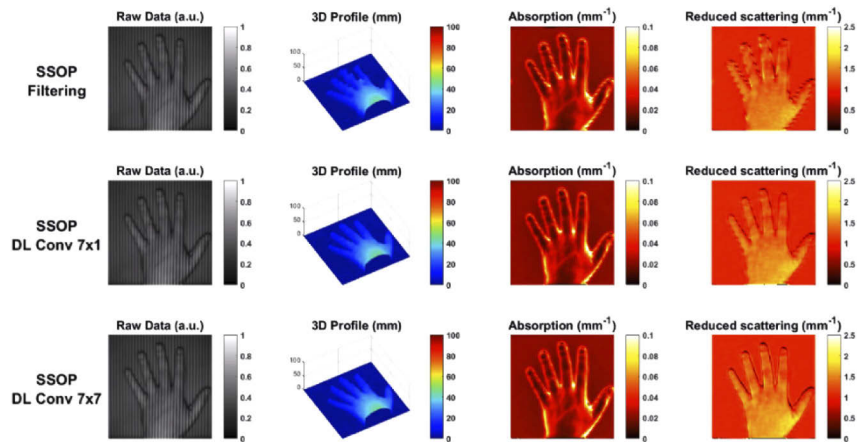

Fig. 7. Single-frame excerpt from video recording of a moving hand (Visualization 1): Raw input data, 3D profile, absorption coefficient and reduced scattering coefficient. From top to bottom: SSOP Filtering, SSOP Deep Learning Convolution 7×1, SSOP Deep Learning Convolution 7×7.

4. Discussion

Single snapshot imaging of optical properties was initially designed to provide fast acquisition and rapid processing but suffered from degraded visual quality compared with SFDI. Leveraging the recent developments in Deep Learning together with GPU-accelerated computation we demonstrated herein that SSOP is able to achieve reconstructive performances comparable to 7-phase SFDI with concurrent sample profile correction, leading to errors as low as **7.5** $\pm$ 2.7% for a complex object such as a hand and *ex vivo* pig organs. Such technology can readily be deployed in the intraoperative room to guide surgeons by offering real-time visualization feedback of functional and structural biomarkers.

Of note, it is important to highlight that real time capabilities were achieved thanks to the use of a relatively simplistic CNN architecture (comprising less than 2×10^4 parameters) coupled with low-level GPU implementation. Despite the use of a small size DL architecture, our approach allowed pushing the limits of SSOP to match data accuracy of SFDI. To obtain such result, the method combines the advantages of U-NET network demodulation and rapid GPU processing. It

is however interesting to note that simpler implementations such as SSOP Filtering demonstrate good results despite edge artifacts with reconstruction errors as low as **8.7** $\pm$ 2.6% and **9.3** $\pm$ 2.5% for absorption and reduced scattering, respectively. Though, such results using SSOP Filtering are possible only when state-of-the-art 2D anisotropic filtering is properly implemented, as described in prior work [12].

Another interesting observation is that using the SSOP Deep Learning methods described here, accurate and good visual quality of complex object can be performed at spatial frequencies as low as 0.2mm^{-1}. Although not demonstrated, the use of frequencies that are typically lower than 0.3mm^{-1} seems preferable to allow deeper penetration of the photons and better signal to noise ratio (amplitude modulation decreases significantly with spatial frequency). As such, the use of the presented SSOP DL methodology allows for a more reliable use of SFDI methods for healthcare application that necessitate real-time employability.

However, this work is not without limitations. The performance of a trained network always depends on the training dataset. In our case, we avoided the 3 phases SFDI acquisition signal to noise ratio limitation when spatial frequency increases by using 7 phases SFDI acquisition. Despite this precaution, SFDI ground truth at spatial frequencies 0.3 mm^{-1} and 0.4 mm^{-1} are still subject to some imperfections, which likely affect the network training and subsequent reconstruction upon inference (see Fig. 3–5). These artifacts could explain, in part, why the error values for absorption and reduced scattering are not decreasing when increasing the spatial frequency from 0.3 mm^{-1} to 0.4 mm^{-1}.

Another classical limitation regarding the datasets used for training in DL is the variety of configurations offered by the sample. Including more images of organs having different shapes and edge discontinuities would likely improve the network accuracy. Indeed, this was observed in previously published work [33] and noted as crucial for model generalizability.

This study highlights also the trade-off that exists between the choice of the spatial frequency used for acquisition and the capacity of SSOP to provide both accurate and high visual quality images. One can notice for all the SSOP Filtering images obtained at low frequency that the capability to retrieve the structure of the hand is limited by the half-period of the spatial frequency used.

Overall, this study demonstrates that with the combination of a simple (small-size, 2×10^4 parameters) DNN and low-level GPU implementation, accurate and high visual quality profile-corrected SSOP can be performed in real time for one wavelength. Since our goal is to provide interpretable information to healthcare practitioners, additional wavelengths will need to be acquired to obtain interpretable physiological parameters such as oxy-hemoglobin, deoxy-hemoglobin, lipids or water content. With a 18ms processing time using the best performing method (DL conv 7×7), at present only 2 wavelengths could reasonably be processed to meet real-time definition (i.e. frame rate faster than 25 frames per second). However, due to the rapid increase in GPU computing power and their low price, we expect that capabilities can be extended easily to process more than 2 wavelengths.

Finally, all code and data used for this work is available at www.healthphotonics.org under the resources tab. This is an open initiative for transparency and with the objective to share and accelerate developments in the field of SFDI. Sources codes, raw data as well as executable code readily usable can be downloaded. Other resources such as www.openSFDI.org can be used to fabricate a SFDI system that can be used with the codes provided here [23].

In summary, this work lays the foundation for real-time multispectral quantitative diffuse optical imaging for surgical guidance and healthcare applications.

5. Conclusion

In this work, using a simple (less than 2×10^4 parameters) convolutional deep learning architecture and low-level GPU implementation allowed to push the limits of SSOP to match data accuracy of

Biomedical Optics EXPRESS

SFDI with a real-time capability. In particular, we show that by designing a computationally inexpensive twin U-Net architecture for both the demodulation stage and the sample profile extraction stage, along with a custom-made GPU-accelerated implementation using C CUDA, megapixel profile-corrected optical properties (absorption and reduced scattering) maps were extracted in 18 ms with high visual quality and preserving a high accuracy. This work contributes to the development of a real-time, quantitative, and wide-field imaging for surgical guidance.

Funding

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. N. Dognitz and G. Wagnieres, "Determination of tissue optical properties by steady-state spatial frequency-domain reflectometry," Lasers Med. Sci. **13**(1), 55–65 (1998).
2. D. J. Cuccia, F. Bevilacqua, A. J. Durkin, F. R. Ayers, and B. J. Tromberg, "Quantitation and mapping of tissue optical properties using modulated imaging," J. Biomed. Opt. **14**(2), 024012 (2009).
3. J. P. Angelo, S. J. Chen, M. Ochoa, U. Sunar, S. Gioux, and X. Intes, "Review of structured light in diffuse optical imaging," J. Biomed. Opt. **24**(07), 1–20 (2018).
4. S. Gioux, A. Mazhar, and D. J. Cuccia, "Spatial Frequency Domain Imaging in 2019," J. Biomed. Opt. **24**(7), 07613 (2019).
5. A. Ponticorvo, D. M. Burmeister, B. Yang, B. Choi, R. J. Christy, and A. J. Durkin, "Quantitative assessment of graded burn wounds in a porcine model using spatial frequency domain imaging (SFDI) and laser speckle imaging (LSI)," Biomed. Opt. Express **5**(10), 3467–3481 (2014).
6. M. R. Pharaon, T. Scholz, S. Bogdanoff, D. Cuccia, A. J. Durkin, D. B. Hoyt, and G. R. Evans, "Early detection of complete vascular occlusion in a pedicle flap model using quantitative [corrected] spectral imaging," Plast. Reconstr. Surg. **126**(6), 1924–1935 (2010).
7. S. Gioux, A. Mazhar, B. T. Lee, S. J. Lin, A. M. Tobias, D. J. Cuccia, A. Stockdale, R. Oketokoun, Y. Ashitate, E. Kelly, M. Weinmann, N. J. Durr, L. A. Moffitt, A. J. Durkin, B. J. Tromberg, and J. V. Frangioni, "First-in-human pilot study of a spatial frequency domain oxygenation imaging system," J. Biomed. Opt. **16**(8), 086015 (2011).
8. A. Yafi, F. K. Muakkassa, T. Pasupneti, J. Fulton, D. J. Cuccia, A. Mazhar, K. N. Blasiole, and E. N. Mostow, "Quantitative skin assessment using spatial frequency domain imaging (SFDI) in patients with or at high risk for pressure ulcers," Lasers Surg. Med. **49**(9), 827–834 (2017).
9. D. J. Rohrbach, D. Muffoletto, J. Huihui, R. Saager, K. Keymel, A. Paquette, J. Morgan, N. Zeitouni, and U. Sunar, "Preoperative Mapping of Nonmelanoma Skin Cancer Using Spatial Frequency Domain and Ultrasound Imaging," Academic Radiology **21**(2), 263–270 (2014).
10. J. Vervandier and S. Gioux, "Single snapshot imaging of optical properties," Biomed. Opt. Express **4**(12), 2938–2944 (2013).
11. M. van de Giessen, J. P. Angelo, and S. Gioux, "Real-time, profile-corrected single snapshot imaging of optical properties," Biomed. Opt. Express **6**(10), 4051–4062 (2015).
12. E. Aguenounon, F. Dadouche, W. Uhring, and S. Gioux, "Single snapshot of optical properties image quality improvement using anisotropic 2D windows filtering," J. Biomed. Opt. **24**(7), 071611 (2019).
13. K. Suzuki, "Overview of deep learning in medical imaging," Radiol Phys Technol **10**(3), 257–273 (2017).
14. W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," Neural Computation **29**(9), 2352–2449 (2017).
15. G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning," IEEE Trans. Med. Imaging **37**(7), 1562–1573 (2018).
16. E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," Nat. Methods **16**(12), 1233–1246 (2019).
17. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**(8), 921–943 (2019).
18. J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, "Fast fit-free analysis of fluorescence lifetime imaging via deep learning," in *Proceedings of the National Academy of Sciences* (PNAS, 2019), pp. 24019–24030.

19. E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth, "Medical Image Segmentation on GPUs – A Comprehensive Review," Med. Image Anal. **20**(1), 1–18 (2015).

20. Loc N. Huynh, Youngki Lee, and Rajesh Krishna Balan, "DeepMon: Mobile GPU-based Deep Learning Framework for Continuous Vision Applications," In *Proceedings of the 15th Annual International Conference on Mobile Systems Applications and Services* (MobiSys, 2017), pp. 82–95.

21. J. Novák, P. Novák, and A. Mikš, "Multi-step phase-shifting algorithms insensitive to linear phase shift errors," Opt. Commun. **281**(21), 5302–5309 (2008).

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

23. Matthew B. Applegate, Kavon Karrobi, Joseph P. Angelo Jr., Wyatt M. Austin, Syeda M. Tabassum, Enagnon Aguénounon, Karissa Tilbury, Rolf B. Saager, Sylvain Gioux, and Darren M. Roblyer, "OpenSFDI: an open-source guide for constructing a spatial frequency domain imaging system," J. Biomed. Opt. **25**(1), 016002 (2020).

24. E. Aguénounon, F. Dadouche, W. Uhring, and S. Gioux, "Real-time optical properties and oxygenation imaging using custom parallel processing in the spatial frequency domain," Biomed. Opt. Express **10**(8), 3916–3928 (2019).

25. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," Appl. Opt. **22**(24), 3977–3982 (1983).

26. J. Geng, "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photonics **3**(2), 128–160 (2011).

27. S. Gioux, A. Mazhar, D. J. Cuccia, A. J. Durkin, B. J. Tromberg, and J. V. Frangioni, "Three-dimensional surface profile intensity correction for spatially modulated imaging," J. Biomed. Opt. **14**(3), 034045 (2009).

28. Y. Zhao, S. Tabassum, S. Piracha, M. S. Nandhu, M. Viapiano, and D. Roblyer, "Angle correction for small animal tumor imaging with spatial frequency domain imaging (SFDI)," Biomed. Opt. Express **7**(6), 2373–2384 (2016).

29. T. T. Nguyen, H. N. Le, M. Vo, Z. Wang, L. Luu, and J. C. Ramella-Roman, "Three-dimensional phantoms for curvature correction in spatial frequency domain imaging," Biomed. Opt. Express **3**(6), 1200–1214 (2012).

30. S. Panigrahi and S. Gioux, "Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging," J. Biomed. Opt. **24**(07), 1 (2019).

31. Y. Zhao, Y. Deng, F. Bao, H. Peterson, R. Istfan, and D. Roblyer, "Deep learning model for ultrafast multifrequency optical property extractions for spatial frequency domain imaging," Opt. Lett. **43**(22), 5669–5672 (2018).

32. Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (IEEE, 2010), pp. 253–256.

33. M. T. Chen, F. Mahmood, J. A. Sweer, and N. J. Durr, "GANPOP: Generative Adversarial Network Prediction of Optical Properties from Single Snapshot Wide-field Images," *IEEE Transactions on Medical Imaging*, (2019).

34. Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet, "Are gans created equal? a large-scale study," https://arxiv.org/abs/1711.10337v1.

35. S. Feng, Q. Chen, G. Gu, T. Tao, L. Zhang, Y. Hu, W. Yin, and C. Zuo, "Fringe pattern analysis using deep learning," Adv. Photonics **1**(02), 1 (2019).

36. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," https://arxiv.org/abs/1505.04597.

37. Frederick Ayers, Alex Grant, Danny Kuo, David J. Cuccia, and Anthony J. Durkin, "Fabrication and characterization of silicone-based tissue phantoms with tunable optical properties in the visible and near infrared domain," in *Proc. SPIE* 6870, 687007 (2008).

38. B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," J. Biomed. Opt. **11**(4), 041102 (2006).

39. NVIDIA Corporation "TensorRT Developer Guide," https://docs.nvidia.com/deeplearning/sdk/tensorrt-developer-guide/index.html

40. NVIDIA Corporation, "cuDNN Developer Guide," https://docs.nvidia.com/deeplearning/sdk/cudnn-developer-guide/index.html

41. NVIDIA Corporation, "cuDNN API," https://docs.nvidia.com/deeplearning/sdk/cudnn-api/index.html

42. Diederik P. Kingma and Jimmy Lei Ba, "Adam: A method for stochastic optimization," https://arxiv.org/abs/1412.6980.