https://doi.org/10.2352/ISSN.2470-1173.2020.15.COLOR-083

SPECTRANET: A DEEP MODEL FOR SKIN OXYGENATION MEASUREMENT FROM MULTI-SPECTRAL DATA

*Ahmed Kedir*** *Mohib Ullah*** *Jacob Renzo Bauer***

Department of Computer Science (IDI),
Norwegian University of Science and Technology, Norway.
***Equal contribution*

ABSTRACT

Skin oxygenation level is an important indicator for the anesthesiology and psychophysiology of a wide range of skin diseases. The non-contact patient monitoring approaches rely on traditional least square method which are not accurate and can't be deployed in clinical practices. In this paper, we exploited the power of deep learning to measure the skin oxygenation level from 16 channel spectral filter array cameras (SFA). Our architecture named SpectraNet consist of three important block i.e. a chain of Convolutional Neural Network (CNN) for feature extraction from the spectral data, an channel attention network for selecting the most informative channel selection and a bidirectional Long-Short Term Memory (LSTM) for incorporating the spatial and temporal information for estimating the final oxygenation curve from the input multi-spectral video. To show the validity of our proposed network, a clinically practiced oxygenation monitoring method (INVOS) is used as the reference. The subjective and objective evaluation shows that the techniques achieve promising results and can be deployed in the clinical practices. Moreover, due to a highly optimized nature of the proposed network, a fully trained model can be incorporated in a smartphone app for a real-time oxygenation measurement.

***Index Terms*—** Oxygenation, spectral filter array camera (SFA), CNN, channel attention, LSTM.

1. INTRODUCTION

Human skin provides the first layer of protection for the internal vital organs. Any internal or external fault has a direct influence on the condition of skin. Apart from different minerals and nutritions, oxygen is vital for different skin cells like the collagen and elastic tissues that provide the structural integrity to the skin. In this regard, skin oxygenation measurement is an important health feature [1] that indicates skin homeostasis and in general, gives key inside to a variety of diseases (skin cancer, internal organ damage, progression in chronic wound healing etc.). In a nutshell, a variety of clinical techniques are adopted (Oximetry [2], INVOS [3], conformal sensors [4]) that involve wearable electronic devices. However, with the advancement in imaging techniques, vision based skin assessment is becoming the next golden rule for the clinical applications. Especially, techniques based on deep learning has shown astonishing results in different computer vision tasks (tracking [5,6], behavior analysis [7,8], cyber security [9, 10], crowd analysis [11, 12], action recognition [13, 14], segmentation [15, 16]). Similarly, in the realm of medical imaging, deep learning has achieved human level performance in various applications like polyp detection [17] in colonoscopy images, thorax disease classification [18], interpolation for low resolution medical imaging [19], and assistance in laparoscopic surgery [20], to name a few.

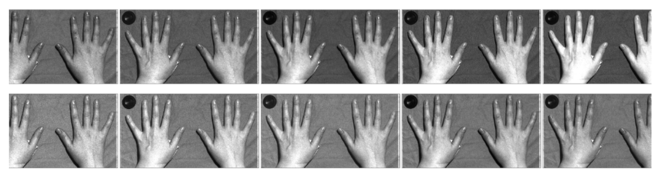

Fig. 1: Experimental setup

Inspired by the successes of deep learning in such applications, we come up with a sophisticated deep learning architecture for approximating the oxygenation curves of the human skin. The network is trained on multi-spectral skin images captured through spectral filter array (SFA) camera manufactured by XIMEA (XiSpec MQ022HG-IM-SM5X5-NIR) with an IMEC sensor. A visible range light is used as the illumination source. Technically, the skin is exposed to the light where it interact with the skin tissues. After interacting with the skin tissues, the reflected rays incarcerate useful information including but not limited to the blood perfusion [21], the oxygenation level [22] and the chromosphere concentration [23]. The information embedded in the reflected spectrum also depends on the light source. For the visible range light that is used in our experiment, the penetration depth is not very high but sufficient enough to give accurate estimation of the oxygenation level. The experiment is organized in such a way that initially, the normal blood flow is observed with much of oxygen. Then an occlusion is applied on the hand which virtually stop the blood flow and the skin tissues consumes most of the oxygen available in the blood stream. Later, the occlusion is removed and the blood is allowed to flow normally. Such a setup can be seen in Fig. 1. From clinical perspective, the important parameters in the oxygenation curve of skin are the the area under the curve, the de-saturation slope and the re-saturation slope as shown in Fig. 2.

The rest of the paper is organized in the following order. Section 2 briefly explained the related work. The proposed network and the optimization strategy with the chosen loss function is elaborated in section 3. The data acquisition and the quantitative results are discussed in section 4 and section 5 concludes the paper.

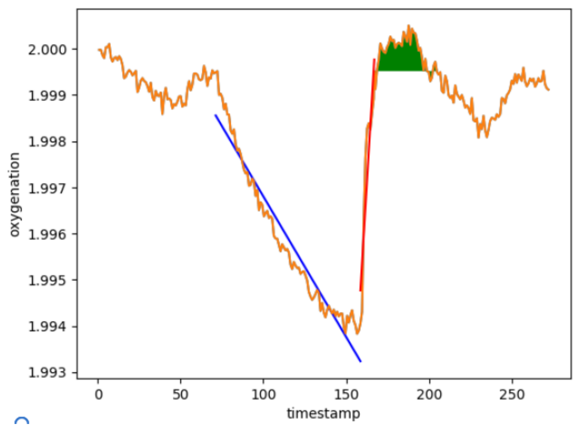

Fig. 2: A typical oxygenation curve obtained through INVOS system

2. BACKGROUND

With the advent of hardware technology, imaging and deep learning techniques made its way to a variety of medical and clinical applications. Non-context spatial resolved oxygenation monitoring is one of the applications that received lots of attention from the photonics, image processing and more recently, deep learning community. Earlier, many researchers tried to exploit RGB images and hand-crafted features to approximate the oxygenation curve. In the following, the RGB and multi-spectral imaging based approaches are briefly explained for oxygenation approximation.

2.1. RGB Imaging

The pioneer work of Wieringa et al. [24] illustrated the feasibility of using monochrome CMOS-camera with apochromatic lens for the skin oxygenation measurements. Primarily, the detection of two dimensional matrix of spatially resolved optical plethysmographic signals of the individual R, G, and B bands and their ratio is used as the criteria for oxygenation measurement (ratio-of-ratios rule). Nishidate et al. [25] used the FFT (fast Fourier transform) on each pixel of the RGB image for the total hemoglobin concentration. In addition to hemoglobin concentration, the two dimensional plethysmogram and vasomotion were also constructed for the activity evaluation of nervous system. Alessandro et al. [26] proposed an approach named Sophia for tracking the oxygen saturation changes in a controlled environment using RGB camera. One of major drawbacks of such approaches is the subject motion which introduces a high degrees of SNR in the measurement. Van Gastel et al. [27] come up with a solution to mitigate the subject motions effect. Technically, they used classical pulse Oximetry system but incorporated multi-site measurements and exploit camera spatial redundancy to reduce the motion artifacts in the measurements. In spite the success of RGB sensor and their potential use in variety if applications, they have implicit limitation of low spectral resolution with only three bands. With the continuously decreasing cost of multi-spectral cameras, it was natural to exploit it for the skin oxygenation measurements. In the following, a brief overview of multi-spectral imagining techniques is given.

2.2. Multi-spectral Imaging

Compared to RGB imaging, multi-spectral imaging is a technique to measure different narrow spectral bands. This allows a more accurate acquisition of the color or spectral changes in the reflectance of objects spatially. Multi-spectral imagining is usually based on a temporal decomposition of the spectral bands. In this line of work, Basiri et al. [28] used multi spectral imaging to access the progress of skin wound healing over extended period of time. Similarly, Schwarz et at. [23] used multispectral optoacoustic mesoscopy (MSOM) for skin dieases detection and generic dermotology. Essentially, they employed illumination at multiple wavelengths for enabling threedimensional multi-spectral optoacoustic mesoscopy (MSOM) of natural chromophores in human skin in vivo operating at 15-125 MHz. With such a setup, they disclosed insights of melanin, and blood oxygenation in human skin. Bernat et al. [22] designed a low cost multi-spectral imaging system for accessing changes in the oxygen concentration in human skin. The system consists of portable LEDs and an area scan camera all controlled by a Tablet computer. Bauer et al. [29] proposed an evaluation framework and tested three cameras including spectral filter array camera for the skin analysis. Additionally, an optimal model of skin is used to improve spectral reconstruction accuracy. Sowa et al. [30] designed a multi-spectral imaging system for skin tissues viability assessment. The system consists of a multi-spectral reflectance imaging device that measure the relative attenuation of reflected light form the oxygenated and de-oxygenated hemoglobin.

Similarly, most of the related approaches either rely on the design of camera or stick to the classical imaging approaches for skin oxygenation measurements. To the best of our knowledge, this is the first attempt where a sophisticated deep learn-

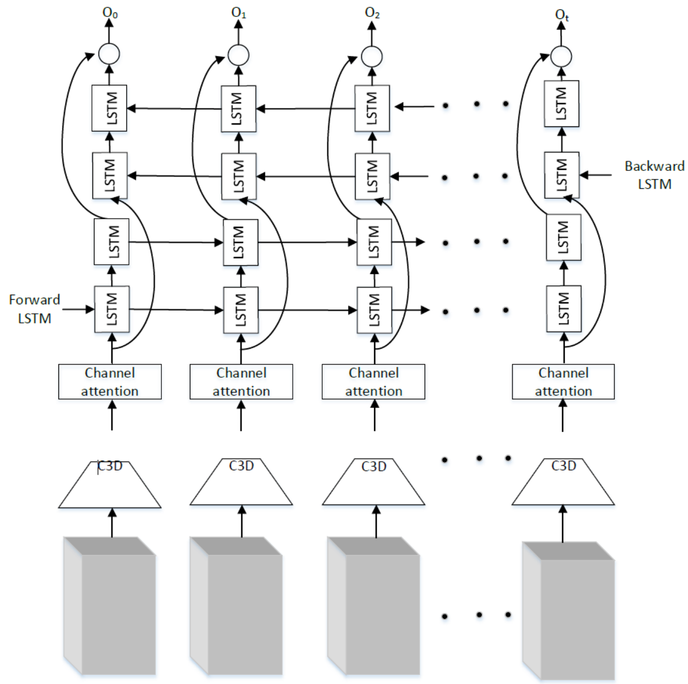

Fig. 3: Proposed method: The input to the network is space time volume of 16 channels. The Conv3D network takes the input and gives feature maps at the output. The feature maps are given the channel attention network for selecting the most informative channels which are consequently given to the bidirectional LSTM network. The bidirectional LSTM network encapsulate the spatial and temporal information for the final oxygen concentration predication in the input 16 channel image.

ing model is designed and trained on the multi-spectral data end-to-end to predict the oxygenation level in a non-context fashion.

3. PROPOSED METHOD

The block diagram of the proposed technique is given in Figure 3. The network can be divided into 3 main module i.e. a feature extraction module, a channel attention module and bidirectional long short term memory module [31]. For the feature extraction, we incorporated Conv3D network [32]. Conv3D network is originally trained on RGB video sequences for human action recongnition. Due to different nature of our problem, we followed similar architecture but trained the network from scratch on multi-spectral data and used it for the feature extraction. Hence, each 16 channel input is given to a Conv3D block in a temporal fashion as shown in Fig 3. After extracting the features from the multi-spectral data, the feature maps are passed through a channel attention module (spatial & temporal) [33]. The channel attention module select the most relevant and informative features from a bunch feature maps given by the Conv3D. Afterwards, the most informative feature maps are given to bidirectional LSTM that encapsulate the spatial and temporal information for the final prediction of the oxygenation curve. For the channel attention, the architectures is inspired my [33]. In the following section 3.1, the loss function and optimization strategy is explained.

3.1. Loss Function & Optimization

The aim of the proposed network is to predict the level of oxygenation from an input 16 channel multi-spectral image. Therefore, a classical mean square difference (MSD) is the most suitable loss as given in Eq. 1. Mathematically, the generic loss function could be written as:

$$L(y, \hat{y}) = \frac{1}{N}(y_i - \hat{y})^2 \quad (1)$$

In equation 1, y_i indicates the true value of the oxygenation of the given frame at time instance t_i and $\hat{y}$ is the predicted value given by our model. The network process the whole video and gives a point based prediction. As a post processing step, linear interpolation is used to get the final smooth curve of the oxygenation. Similarly, to train the network, we used conjugate gradient decent with ADAM optimizer. The details of hyperparameters and parameters will be given in the experiment section in the full version of the paper.

4. DATASET & RESULTS

4.1. Data Acquisition

The acquisition of spectral image from SFA cameras is rather complex. In order to obtain the image cube, the image has to be processed spatially and spectrally. For this research work, the data is acquired in Amsterdam hospital with a commercially available SFA camera based on the IMEC snapshot sensor, the XIMEA XiSpec SM4x4 VIS [34] operating in the visual range from 470nm to 630nm. In the wavelength between 470nm to 630nm, in total 16 channels are acquired.

4.2. Experiment

As obtaining labeled or groundtruth data for any deep learning algorithm is expensive and time consuming. It gets even more difficult in medical imaging because human subjects are involved in the data acquisition. In our case, in order to get the groundtruth data for testing our model, measurements taken with INVOS system is used as the benchmark. The detailed quantitative results will be provided in the full version of the paper. However, in order to support the claim that our model work,the predicted and the groundtruth oxygenation curve can be seen in Figure 4. It can be seen the model predicts the oxygen concentration in skin with very good accuracy.

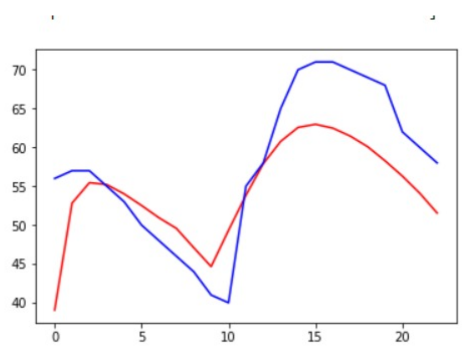

Fig. 4: The red curve corresponds to the predicted oxygenation and the blue correspond to the groundtruth obtained through INVOS system. The x-axis shows the temporal evolution of frames while the y-axis shows the oxygen concentration.

5. CONCLUSION

We proposed a deep model for predicting the skin oxygenation level from 16 channel spectral filter array cameras (SFA). The network consist of three main block i.e. a chain of Convolutional Neural Network (CNN) for feature extraction from the spectral data, an channel attention network for selecting the most informative channel selection and a bidirectional Long-Short Term Memory (LSTM) for encapsulating the spatial and temporal information for estimating the final oxygenation curve from the input multi-spectral video. The results of the proposed network is compared against clinically practiced oxygenation monitoring method (INVOS). The subjective and objective evaluation shows that the model achieve promising results and can be deployed in the clinical practices. Due to the light weight nature of the model, it could be deployed in resource limited devices like smartphones for real-time oxygenation measurements. In future, we are aiming to collaborate with hospital and test the model in real world scenarios.

6. REFERENCES

[1] Jacob Renzo Bauer, Arnoud A Bruins, Jon Yngve Hardeberg, and Rudolf M Verdaasdonk, “A spectral filter array camera for clinical monitoring and diagnosis: Proof of concept for skin oxygenation imaging,” *Journal of Imaging*, vol. 5, no. 8, pp. 66, 2019.

[2] CD Hanning and JM Alexander-Williams, “Fortnightly review: Pulse oximetry: a practical review,” *Bmj*, vol. 311, no. 7001, pp. 367–370, 1995.

[3] Jacob R Bauer, Karlijn van Bekuum, John Klaessens, Herke Jan Noordmans, Christa Boer, Jon Y Hardeberg, and Rudolf M Verdaasdonk, “Towards real-time non contact spatial resolved oxygenation monitoring using a multi spectral filter array camera in various light conditions,” in *Optical Biopsy XVI: Toward Real-Time Spectroscopic Imaging and Diagnosis*. International Society for Optics and Photonics, 2018, vol. 10489, p. 104890O.

[4] Zongxi Li, Emmanuel Roussakis, Emily Keeley, Gabriela Apiou-Sbirlea, Reginald Birngruber, Christene Huang, and Conor L Evans, “A wearable conformal bandage for non-invasive two-dimensional imaging of skin oxygenation (conference presentation),” in *Optical Diagnostics and Sensing XVI: Toward Point-of-Care Diagnostics*. International Society for Optics and Photonics, 2016, vol. 9715, p. 97150R.

[5] Mohib Ullah and Faouzi Alaya Cheikh, “Deep feature based end-to-end transportation network for multi-target tracking,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 3738–3742.

[6] Mohib Ullah, Habib Ullah, and Faouzi Alaya Cheikh, “Single shot appearance model (ssam) for multi-target tracking,” *Electronic Imaging*, vol. 2019, no. 7, pp. 466–1, 2019.

[7] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge, “Deeplabcut: markerless pose estimation of user-defined body parts with deep learning,” Tech. Rep., Nature Publishing Group, 2018.

[8] Saira Kanwal, Muhammad Uzair, Habib Ullah, Sultan Daud Khan, Mohib Ullah, and Faouzi Alaya Cheikh, “An image based prediction model for sleep stage identification,” in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1366–1370.

[9] Muhammad Mudassar Yamin, Basel Katt, and Vasileios Gkioulos, "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture," *Computers & Security*, p. 101636, 2019.

[10] Muhammad Mudassar Yamin and Basel Katt, "Modeling attack and defense scenarios for cyber security exercises," in *5th interdisciPlinary cyber research conference 2019*, 2019, p. 7.

[11] Habib Ullah and Nicola Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *ICPR workshop on pattern recognition and crowd analysis*, 2012, vol. 75.

[12] Habib Ullah, Ahmed B Altamimi, Muhammad Uzair, and Mohib Ullah, "Anomalous entities detection and localization in pedestrian flows," *Neurocomputing*, vol. 290, pp. 74–86, 2018.

[13] Jun Liu, Amir Shahroudy, Dong Xu, and Gang Wang, "Spatio-temporal lstm with trust gates for 3d human action recognition," in *European conference on computer vision*. Springer, 2016, pp. 816–833.

[14] Mohib Ullah, Habib Ullah, and Ibrahim M Alseadonn, "Human action recognition in videos using stable features," *Signal & Image Processing : An International Journal*, 2017.

[15] Ahmed Mohammed, Sule Yildirim, Ivar Farup, Marius Pedersen, and Øistein Hovde, "Streoscennet: surgical stereo robotic scene segmentation," in *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. International Society for Optics and Photonics, 2019, vol. 10951, p. 109510P.

[16] Habib Ullah, Mohib Ullah, and Muhammad Uzair, "A hybrid social influence model for pedestrian motion segmentation," *Neural Computing and Applications*, pp. 1–17, 2018.

[17] Ahmed Mohammed, Sule Yildirim, Ivar Farup, Marius Pedersen, and Øistein Hovde, "Y-net: A deep convolutional neural network for polyp detection," *arXiv preprint arXiv:1806.01907*, 2018.

[18] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng, and Yi Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv preprint arXiv:1801.09927*, 2018.

[19] Ahmed Mohammed, Ivar Farup, Sule Yildirim, Marius Pedersen, and Øistein Hovde, "Variational approach for capsule video frame interpolation," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 30, 2018.

[20] Congcong Wang, Ahmed Kedir Mohammed, Faouzi Alaya Cheikh, Azeddine Beghdadi, and Ole Jacob Elle, "Multiscale deep desmoking for laparoscopic surgery," in *Medical Imaging 2019: Image Processing*. International Society for Optics and Photonics, 2019, vol. 10949, p. 109491Y.

[21] Matija Milanič, Asgeir Bjorgan, Marcus Larsson, Paolo Marraccini, Tomas Strömberg, and Lise Lyngsnes Randeberg, "Hyperspectral imaging for detection of cholesterol in human skin," in *Optical Diagnostics and Sensing XV: Toward Point-of-Care Diagnostics*. International Society for Optics and Photonics, 2015, vol. 9332, p. 93320W.

[22] Amir S Bernat, Frank J Bolton, Kfir Bar-Am, Steven L Jacques, and David Levitz, "Assessing changes in oxygen saturation using a low cost multi-spectral imaging system," in *Optics and Biophotonics in Low-Resource Settings V*. International Society for Optics and Photonics, 2019, vol. 10869, p. 1086911.

[23] Mathias Schwarz, Andreas Buehler, Juan Aguirre, and Vasilis Ntziachristos, "Three-dimensional multispectral optoacoustic mesoscopy reveals melanin and blood oxygenation in human skin in vivo," *Journal of biophotonics*, vol. 9, no. 1-2, pp. 55–60, 2016.

[24] Fokko P Wieringa, Frits Mastik, and Antonius FW van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: a first step toward spo 2 camera technology," *Annals of biomedical engineering*, vol. 33, no. 8, pp. 1034–1041, 2005.

[25] Izumi Nishidate, Chihiro Tanabe, Daniel J McDuff, Kazuya Nakano, Kyuichi Niizeki, Yoshihisa Aizu, and Hideaki Haneishi, "Rgb camera-based noncontact imaging of plethysmogram and spontaneous low-frequency oscillation in skin perfusion before and during psychological stress," in *Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics*. International Society for Optics and Photonics, 2019, vol. 10885, p. 1088507.

[26] Alessandro R Guazzi, Mauricio Villarroel, Joao Jorge, Jonathan Daly, Matthew C Frise, Peter A Robbins, and Lionel Tarassenko, "Non-contact measurement of oxygen saturation with an rgb camera," *Biomedical optics express*, vol. 6, no. 9, pp. 3320–3338, 2015.

[27] Mark Van Gastel, Sander Stuijk, and Gerard De Haan, "New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring," *Scientific reports*, vol. 6, pp. 38609, 2016.

[28] Ali Basiri, Marjan Nabili, Scott Mathews, Alex Libin, Suzanne Groah, Herke J Noordmans, and Jessica C Ramella-Roman, "Use of a multi-spectral camera in the characterization of skin wounds," *Optics express*, vol. 18, no. 4, pp. 3244–3257, 2010.

[29] Jacob Renzo Bauer, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Rudolf M Verdaasdonk, "An evaluation framework for spectral filter array cameras to optimize skin diagnosis," *Sensors*, vol. 19, no. 21, pp. 4805, 2019.

[30] Michael G Sowa, "Snapshotnir: a handheld multispectral imaging system for tissue viability assessment," in *Photonics and Education in Measurement Science 2019*. International Society for Optics and Photonics, 2019, vol. 11144, p. 111440B.

[31] Zhiheng Huang, Wei Xu, and Kai Yu, "Bidirectional lstm-crf models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.

[32] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[33] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[34] Ximea, "Ximea hyperspectral cameras," 2018, https://www.ximea.com/.