Supplemental Document

Biomedical Optics EXPRESS

Rapid azoospermia classification by stimulated Raman scattering and second harmonic generation microscopy: supplement

JIE HUANG,[1,2,†] XIAOBIN TANG,[3,†] ZHICONG CHEN,[4,†] XIAOMIN LI,[4] YONGQING ZHANG,[3] XIANGJIE HUANG,[2] DELONG ZHANG,[3,5] GENG AN,[4,6] AND HYEON JEONG LEE[2,*]

[1] *Zhejiang Polytechnic Institute, Polytechnic Institute, Zhejiang University, Hangzhou 310058, China*
[2] *College of Biomedical Engineering & Instrument Science; Key Laboratory for Biomedical Engineering of Ministry of Education, Zhejiang University, Hangzhou 310058, China*
[3] *Interdisciplinary Centre for Quantum Information, Zhejiang Province Key Laboratory of Quantum Technology and Device, and Department of Physics, Zhejiang University; Hangzhou 310027, China*
[4] *Department of Obstetrics and Gynecology, Center for Reproductive Medicine; Guangdong Provincial Key Laboratory of Major Obstetric Diseases; Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology; Guangdong-Hong Kong-Macao Greater Bay Area Higher Education Joint Laboratory of Maternal-Fetal Medicine; The Third Affiliated Hospital of Guangzhou Medical University; Guangzhou 510150, China*
[5] *dlzhang@zju.edu.cn*
[6] *angeng0505@outlook.com*
[†] *These authors contributed equally*
[*] *hjlee@zju.edu.cn*

This supplement published with Optica Publishing Group on 5 October 2023 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.24205365

Parent Article DOI: https://doi.org/10.1364/BOE.501623

RAPID AZOOSPERMIA CLASSIFICATION BY STIMULATED RAMAN SCATTERING AND SECOND HARMONIC GENERATION MICROSCOPY: SUPPLEMENTAL DOCUMENT

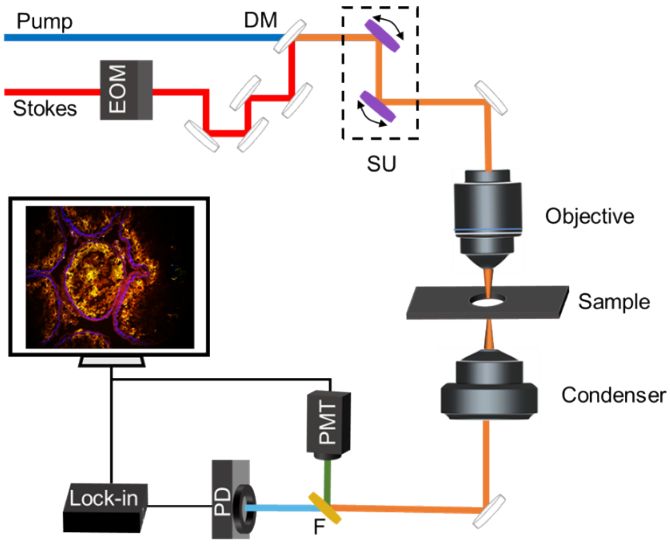

Fig. S1. Schematic representation of SRASH microscopy experimental setup. EOM: electro-optic modulator; DM: dichroic mirror; SU: scanning unit; F: filter; PD: photodiode; PMT: photomultiplier tube.

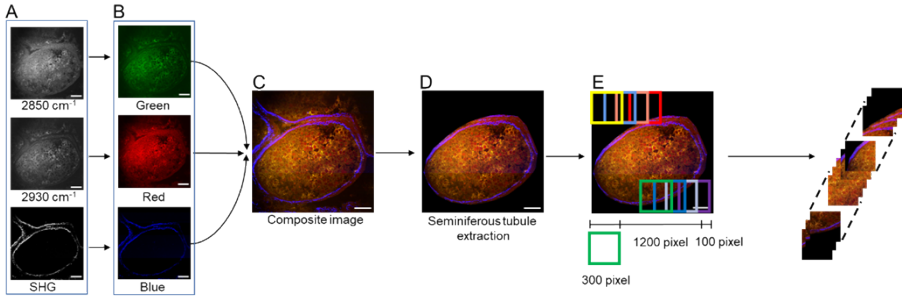

Fig. S2. Work flow of image processing. (A) Images obtained from SRS (2850 cm^{-1}, 2930 cm^{-1}) and SHG. (B) 2850 cm^{-1} channel, 2930 cm^{-1} channel, and SHG channel are colored in green, red, and blue, respectively. (C) Three-channel RGB image was generated. (D) Extract each seminiferous tubule. (E) The input image of our method is obtained by using dense sliding window algorithm and filtering invalid patches. Scale bar: 50 μm.

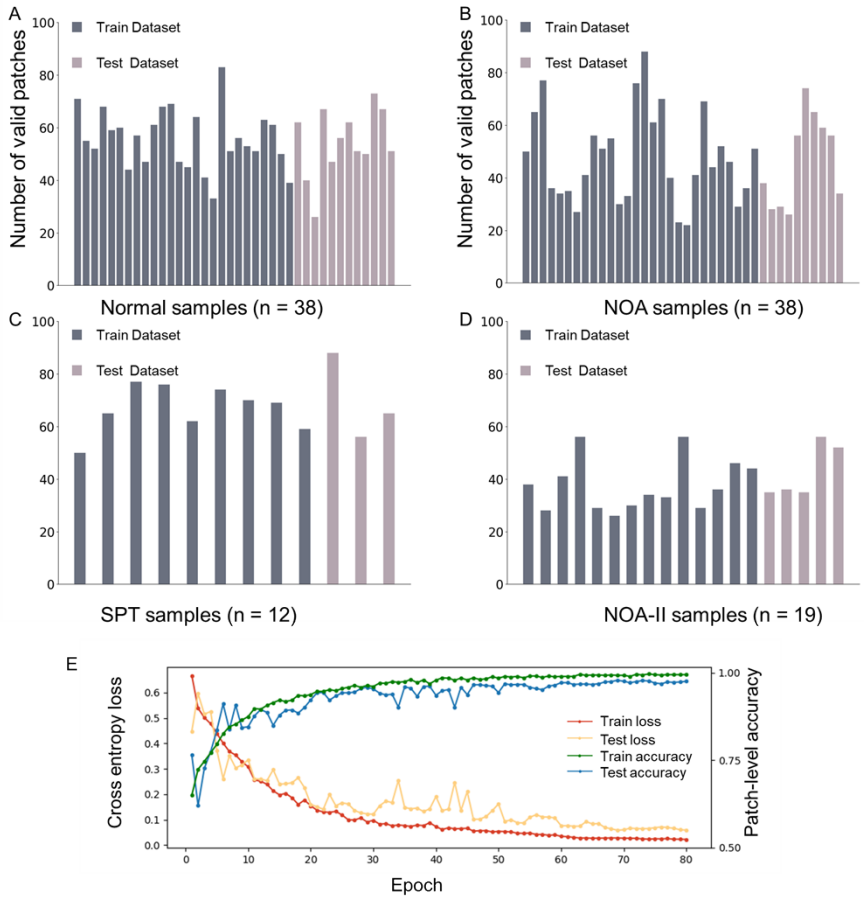

Fig. S3. SRASH image dataset. (A), (B) show the number of valid patches and the division of the training set validation set for normal samples and NOA samples, respectively. (C) Train and test dataset cross entropy loss and test patch-level accuracy is plotted for the training session.

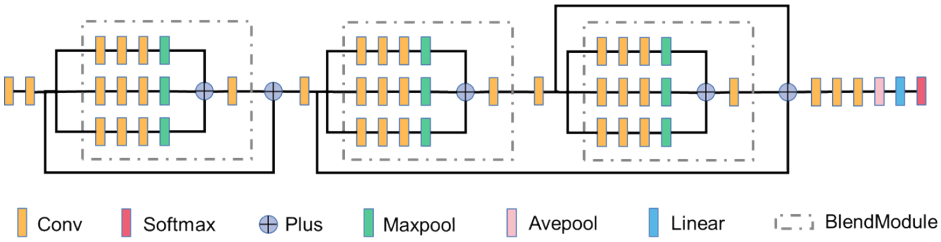

Fig. S4. Schematic of the LiteBlendNet architecture. LiteBlendNet (LBNet) is primarily composed of convolutional layers, BlendModule, and fully connected layers. BlendModule, as a key module, enables multi-scale feature fusion by integrating three branches, each utilizing a combination of standard convolutions and dilated convolutions to extract features from diverse receptive fields. This design promotes efficient feature integration across various scales.

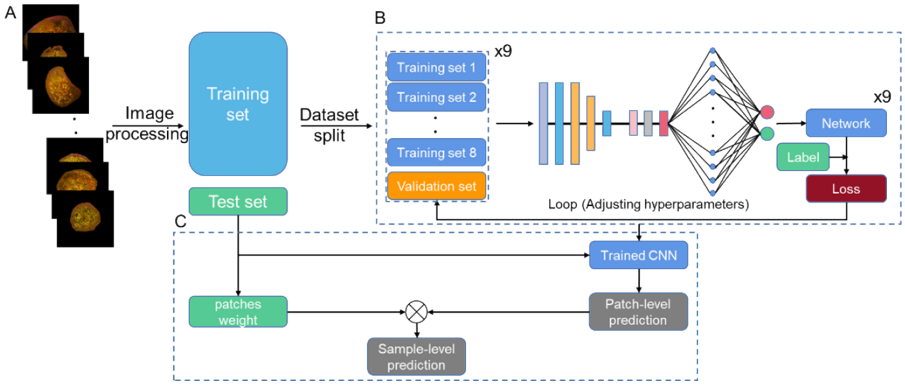

Fig. S5. Work flow of Deep Learning-based Assisted Diagnostic Inference Algorithm. (A) The composite image after extraction of seminiferous tubules was divided into training set and test set after image processing, and the image processing process is shown in Figure S2. (B) In the cross-validation process, the full training set is randomly divided into 9 parts, of which 8 are used as the training set and 1 as the validation set, for a total of 9 combinations. Utilize these nine combinations, each training round will generate 9 networks, and the loss of these networks on the validation set will be calculated to adjust the hyperparameters. The hyperparameters that minimize the average loss and variance are selected to train the LBNet on the full training set. (C) The weights of the patches are calculated, and the prediction results of all patches of the same sample are weighted and summed, and judged by a threshold to obtain the classification results at the sample level. See Figure S10 for details.

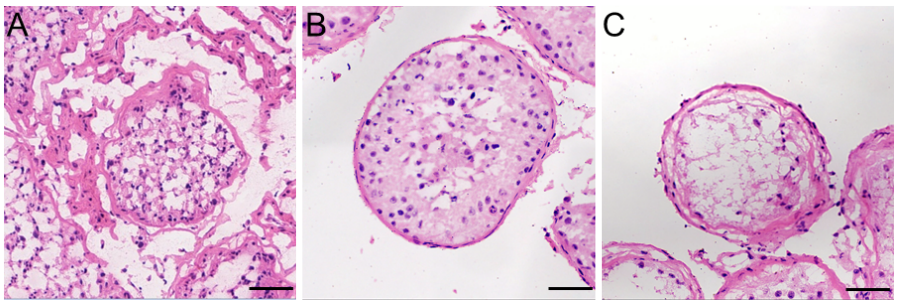

Fig. S6. Adjacent H&E staining for selected seminiferous tubular samples. (A) Normal. (B) SPT. (C) SPG. Scale bar: 50 μm.

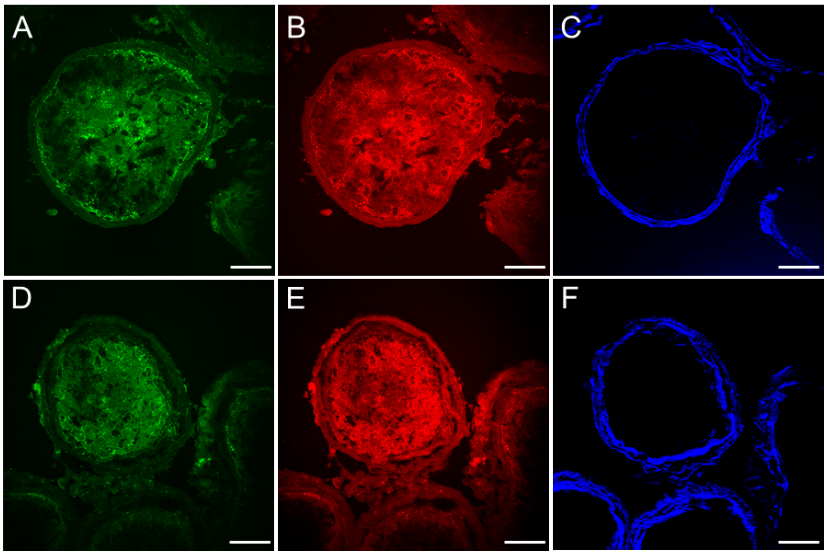

Fig. S7. Raw SRASH images of SPT and SPG seminiferous tubule. (A-C) Raw SRASH images of SPT seminiferous tubule, lipid channel (A), protein channel (B), collagen channel (C). (D-F) Raw SRASH images of SPG seminiferous tubule, lipid channel (D), protein channel (E), collagen channel (F). Scale bar: 50 µm.

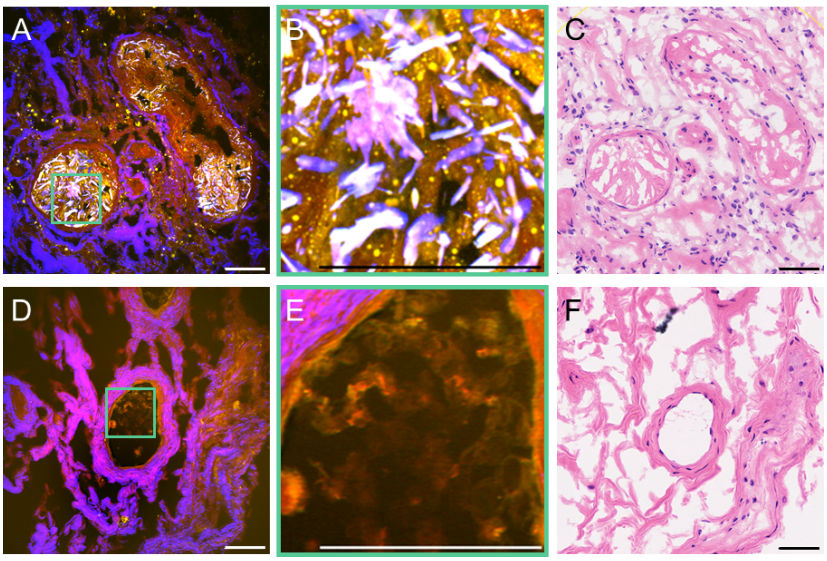

Fig. S8. SRASH and adjacent H&E images of SCO tubule and empty seminiferous tubule. (A) Composite SRASH image from SCO tubule. (B) Zoomed image from the rectangle in (A). (C) Adjacent H&E image of (A). (D) Composite SRASH image from empty seminiferous tubule. (E) Zoomed image from the rectangle in (D). (F) Adjacent H&E image of (D). Red: protein; Green: lipid; blue: collagen in (A, B, D, and E). Scale bar: 50 µm.

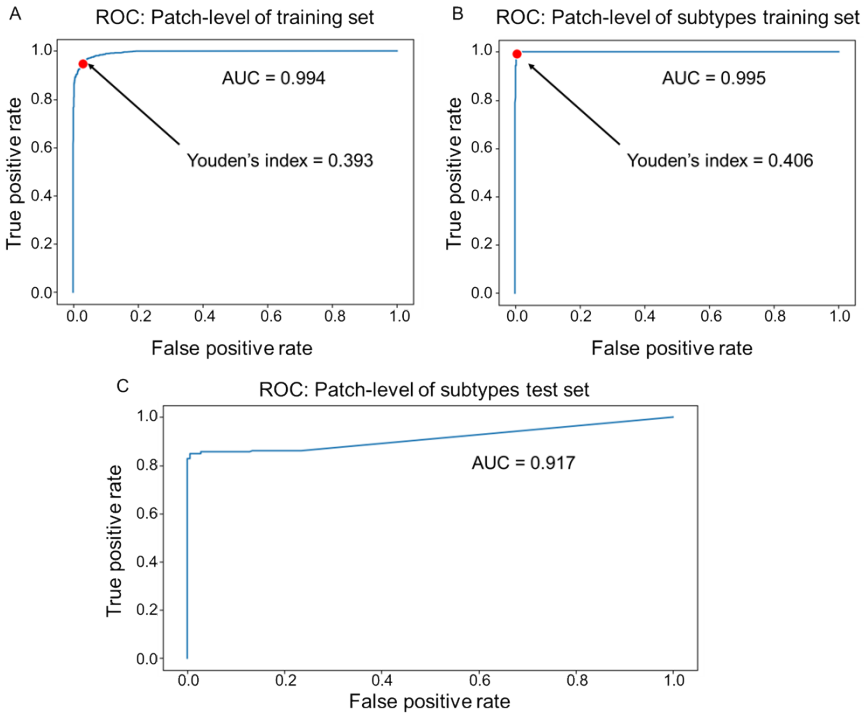

Fig. S9. ROC for the training dataset. (A) The Youden's index was found from the ROC curve as the cut-off point for predicting test dataset. (B) The Youden's index was found from the ROC curve as the cut-off point for predicting subtypes test dataset. (C) ROC curve for subtypes test set patch-level.

Input: Patches obtained after the sample passes through a dense sliding window; trained LBNet
Output: Category of sample

```
function PATCHES_WEIGHT(patches)
    weight_list ← []
    valid_patches ← []
    for patch in patches do
        temp_weight ← 0
        for pixel in patch do
            if pixel >0 then
                temp + = 1
            end if
        end for
        temp ← temp / (patch.height × patch.weight × patch.channel_nums)
        if temp >0.1 then
            append temp_weight to weight_list
            append patch to valid_patches
        end if
    end for
    return weight_list, valid_patches
end function

function PREDICTION(model, valid_patches)
    predition ← []
    for patch in valid_patches do
        output ← softmax(model(patch))
        if output[normal] >0.382 then
            append 1 to prediction
        else
            append 0 to prediction
        end if
    end for
    return prediction
end function

function DIAGNOSIS(patches, model)
    weight_list, valid_patches ← Patches_Weight(patches)
    prediction ← Prediction(model, valid_patches)
    summed_weight ← sum(weight_list)
    class_probability ← (weight_list × prediction) / summed_weight
    if class_probability >0.8 then
        class ← normal
    else if class_probability <0.2 then
        class ← idiopathic
    else
        class ← nondiagnostic
    end if
    return class
end function

return Diagnosis(patches, model)
```

Fig. S10. In classifying the high-resolution samples, we first cut the samples into patches using a dense sliding window. We then calculate the weight of each patch, which is determined by the number of valid pixels of the patch, and delete the patches with less than 10% of valid pixels. The remaining patches are fed into the trained LBNet to obtain the classification probability of each patch. Then, based on the patch classification threshold of 0.393 determined in the training set, we determine their labels for each patch. Finally, we obtain the category probability of the sample after weighting and normalizing all the valid patches of that sample. According to the classification threshold we set for the sample, greater than 0.8 is labeled as normal, less than 0.2 is labeled as NOA, otherwise it is non-diagnostic. The process of classifying NOA subtypes is similar.

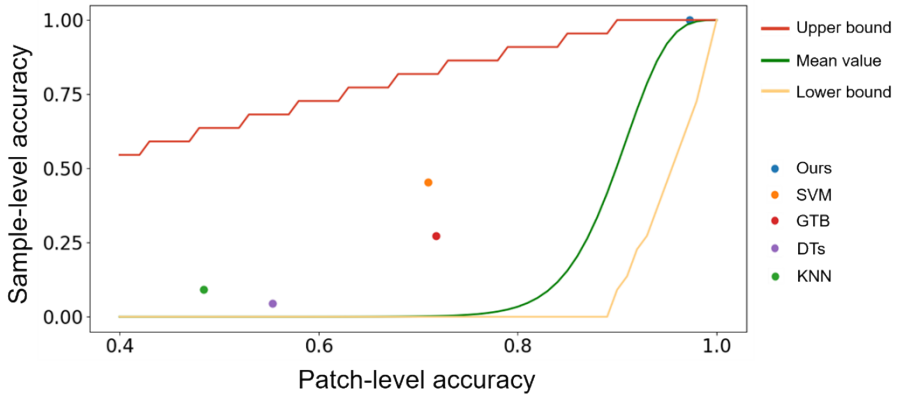

Fig. S11. The relationship between patch level accuracy and sample level accuracy. The results were obtained after one million simulations based on the test set and the designed inference algorithm.

Table S1. Description of sample categories during this study

Patient number	Age(year)	OA (Normal)	NOA (SPT)	NOA (SPG)	NOA (SCO, Empty tubule)
1	36	2	0	0	0
2	38	3	0	0	0
3	32	2	0	0	0
4	35	2	0	0	0
5	43	3	0	0	0
6	34	1	0	0	0
7	47	2	0	0	0
8	27	2	0	0	0
9	53	2	0	0	0
10	50	3	0	0	0
11	36	2	0	0	0
12	41	2	0	0	0
13	29	3	0	0	0
14	39	2	0	0	0
15	34	2	0	0	0
16	32	3	0	0	0
17	27	2	0	0	0
18	31	0	3	0	0
19	26	0	3	0	0
20	35	0	3	0	0
21	25	0	3	0	0
22	34	0	3	0	0
23	24	0	0	3	0
24	35	0	0	2	0
25	25	0	0	3	0
26	27	0	0	3	0
27	27	0	0	3	0
28	31	0	0	2	0
29	29	0	0	3	0
30	25	0	0	0	2
31	27	0	0	0	2

A total of 76 seminiferous tubules from 31 patient cases, including OA (n=38), SPT (n=15), SPG (n=19) and SCO, Empty tubule (n=4).

Abbreviations: obstructive azoospermia (OA), NOA blocked in sperm cells (SPT), NOA blocked in spermatogonia (SPG), Sertoli-cell-only (SCO).

Table S2. Comparison between LiteBlendNet and other deep learning methods

	Params	Patch-level accuracy	Patch-level AUC	Sample-level accuracy
ResNet-50	23.50M	0.920	0.965	0.909
Inception-ResNet-v2	54.31M	0.930	0.981	0.909
Swin-T	27.50M	0.793	0.889	0.727
LiteBlendNet	3.79M	0.962	0.991	1.000

Code implementing ResNet-50, Inception-ResNet-v2 and Swin-T all from: https://github.com/huggingface/pytorch-image-models. The four methods in the table use the same cross-validation process for tuning the hyperparameters.