Journal Pre-proof

Binocular Vision Fusion Enhanced 3D NIR-II in vivo Imaging of Bone and Vessel Networks

ZHICHAO YANG , SITONG WU , XUN ZHANG , MI CHAO ,
GUNGUN LIN , ZHIYONG GUO , DAYONG JIN

Please cite this article as: ZHICHAO YANG , SITONG WU , XUN ZHANG , MI CHAO ,
GUNGUN LIN , ZHIYONG GUO , DAYONG JIN , Binocular Vision Fusion Enhanced 3D NIR-
II in vivo Imaging of Bone and Vessel Networks, *Fundamental Research* (2024), doi:
https://doi.org/10.1016/j.fmre.2024.07.001

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition
of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of
record. This version will undergo additional copyediting, typesetting and review before it is published
in its final form, but we are providing this version to give early visibility of the article. Please note that,
during the production process, errors may be discovered which could affect the content, and all legal
disclaimers that apply to the journal pertain.

Binocular Vision Fusion Enhanced 3D NIR-II *in vivo* Imaging of Bone and Vessel Networks

ZHICHAO YANG[a, b, 1], **SITONG WU**[a, b, 1], **XUN ZHANG**[a, 1], **MI CHAO**[a,b], **GUNGUN LIN**[b,] , **ZHIYONG GUO**[a, *], **AND DAYONG JIN**[a,b, *]

a Institute for Biomedical Materials & Devices, Faculty of Science, University of Technology Sydney, Ultimo, New South Wales 2007, Australia

b UTS-SUSTech Joint Research Centre for Biomedical Materials & Devices, Department of Biomedical Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, China

1 These authors contributed equally to this work.

Abstract

Taking advantages of weak light scattering and minimal amount of auto-fluorescence background, optical imaging through the second near-infrared window (NIR-II,1000–1700nm) allows resolving the microscopic structures in deep tissues. However, current 2D wide-field imaging systems cannot provide the axial resolution to reveal depth information and the lateral resolution thereof has been limited by the pixel numbers of NIR-II cameras. We aim to improve 3D in vivo imaging at a large field of view (FOV) to achieve high-resolution 3D imaging of bone and vessel simultaneously. We developed a 3D NIR-II imaging technique utilizing iterative convolutional sub-pixel image reconstruction (ICSP-IR) and convolutional neural networks (CNN) in a binocular system. The instrumentation and CNN powered data analytics enable 3D *in vivo* volumetric imaging with a lateral resolution of 153 μm and an axial resolution of 480 μm at a FOV of 42.5 x 53 x 14 mm^3. High-resolution 3D imaging of both tibia structures and vessel networks can be simultaneously achieved by the 808-nm and 980-nm dual-beam excitations of NIR-IIa type nanoparticles that emit at 1060 nm and NIR-IIb type nanoparticles that emit at 1530 nm, respectively.

* Corresponding author**,** E-mail: guozy@sustech.edu.cn; dayong.jin@uts.edu.au

1   Introduction

Optical imaging through the second near-infrared window (i.e. NIR-Ⅱ, 1000–1700 nm) offers significant advantages for biomedical imaging due to its substantially reduced auto-fluorescence, minimal tissue absorption and decreased scattering attenuation. Compared with X-ray and MRI imaging techniques, NIR-II imaging allows regular visualization of live cells and tissue structures, making it a non-invasive method for disease diagnosis and prognosis in monitoring the efficacy of treatments [1-5]. Current NIR-II imaging techniques typically employ wide-field illumination over a large area of a small animal, capturing 2D information with a NIR camera [6-9]. To date, a series of imaging technologies have been developed to cover fields of view (FoV) from micro-scale to centimeter scale, such as detecting lesion areas [10-13]. Recently, 2D NIR-Ⅱ image-guided surgery has been demonstrated for excising metastatic ovarian tumors smaller than 1mm [14].

Despite the rapid achievements in NIR-II imaging instrumentation, achieving desirable axial resolution with depth information on tissue and organ structures remains a challenge. Laser confocal, two-photon, and light sheet scanning-based systems have been explored to resolve 3D structures. In 2018, Zhu et al. introduced a confocal near-infrared imaging system using NIR photomultiplier tubes, achieving a resolution of 10 μm at a depth of 1.3 mm in brain tissue [15]. In the same year, Qi et al. developed NIR-Ⅱ two-photon fluorescence microscopy, attaining a 5-μm resolution with a penetration depth of 1.1 mm in the mouse brain vascular system [16]. In 2019, Wang et al. developed a NIR-II light-sheet system, which achieved 3D volumetric imaging across a volume of 810 μm × 648 μm × 3,000 μm and attained a resolution surpassing 10 μm × 10 μm × 15 μm (FWHM) in glycerol-cleared brain tissue [17]. However, these systems are inherently limited by the well-known trade-offs among scanning speed, imaging resolution and the field of volume that can be imaged [18]. As a result, scanning-based imaging approaches typically cover a small volume at the sub-millimeter scale [19, 20]. In this context, the present work aims to achieve high-resolution (HR) and large-field-of-view imaging of live mice without the need for tissue clearing.

To date, ten-millimeter-scale FoV 3D HR imaging has not yet been achieved, primarily due to the resolution limitations of short-wave infrared (SWIR) cameras currently available on the market. The camera's design incorporates an array of back-illuminated InGaAs photodiodes linked to a readout circuit on a fixed-size silicon wafer. This configuration dictates the pixel size and diode spacing, resulting in a video graphics array (VGA) resolution, typically 640 x 512 pixels with a standard pixel size of 15 μm [21-24].

Here, we developed a binocular vision location system integrating a high-precision translation stage to obtain sub-pixel shifted image sequences and an algorithm based on CNN frameworks to improve the resolution over a large FoV. Moreover, we introduced two types of contrast agents, NIR-IIa (1,000-1,300) and NIR-IIb (1,500-1,700), for 3D imaging of the tibia and surrounding vessels using an integrated binocular vision system. This approach enabled high-resolution imaging of both bone structures and vascular networks in vivo within a FoV of 42.5 x 53 x 14 mm^3, which, to our best knowledge, is the first demonstration of such capability in the field of NIR-II imaging.

2 Materials and Methods

2.1 Binocular vision system for location and 3D reconstruction.

This binocular imaging system comprises two monocular vision units, a displacement stage (PI U-751), and dual laser sources (Fig.1a). Each monocular vision unit integrates a long-pass filter replacement module (Thorlabs FELH1000 nm or Semrock BLP02-1319R-50 nm) positioned between SWIR camera (Allied Vision Goldeye 640×512) and camera lens (Kowa 12.5mm). Pulsed lasers at either 980 nm or 808 nm (Changchun New Industries), with a beam expander, were employed to illuminate the mouse, delivering an average intensity of 2.2 mW/cm². Luminescent signals were collected by the camera lens and subsequently filtered by a 1319 nm or 1000 nm long-pass filter. Increasing the optical axis angle between the two views reduces the image overlap, resulting in a diminished portion of reconstructed 3D information. To optimize 3D structure reconstruction, the angle 2α of the two imaging axes was set to 24°, maximizing the reconstruction within a 42.5 x 53 mm² FoV. The imaging was focused on the lower limbs of anesthetized mice. The master control software was used to uniformly trigger both binocular cameras to capture fluorescence images simultaneously upon excitation by either the 980 nm or 808 nm laser. All the animal experiments are approved by Southern University of Science and Technology Committee and Laboratory Animal Center.

Camera calibration is critical in the realm of binocular vision. The prevailing standard procedure involved computing the internal and external reference matrices of the camera matrix, as well as the distortion coefficient of the imaging system, through a series of well-defined steps [25-26]. After correcting image distortions, spatial points were aligned along the same row of the epipolar line in both images. This alignment streamlines the process, necessitating only a single one-dimensional search within that row to establish correspondences, thereby significantly reducing the matching time.
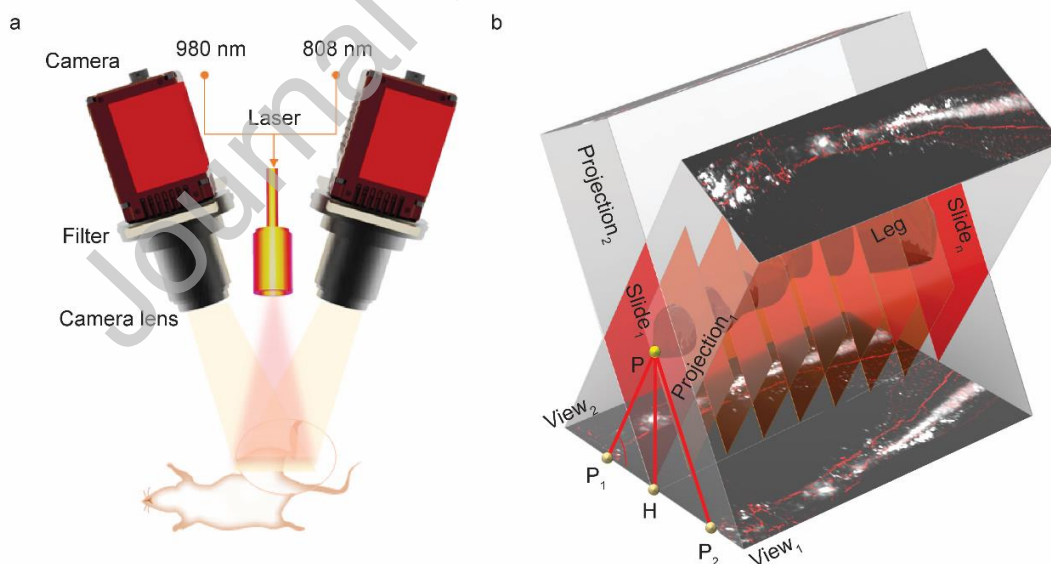
Fig. 1. Schematic of the binocular vision system for NIR-Ⅱ 3D imaging of tibia and vessels. (a) Image acquisition. The pulsed 808 nm and 980 nm laser beams excite NdNPs-PEG as the NIR-Ⅱa' contrast agent that emits at 1060 nm and ErNPs-PEG as the NIR-Ⅱb contrast agent that emits at 1530 nm to target

vessels and bone (tibia), respectively. (b) Volumetric reconstruction. Views 1 and 2 show two views of the same fluorescent object, here a leg including tibia and vessel. Through the two view images, the spatial points of the object can be geometrically deduced. The space point P is projected onto views 1 and 2. The red diamond slide$_1$ is obtained by projecting views 1 and 2 according to the imaging angle α, which contained whole spatial points of first row of two images.

Precision in point matching is pivotal for achieving accurate reconstruction. In this context, the cosine similarity is employed as the evaluation metric [27]. In the cosine similarity Eq. 1, L_i and R_i represent corresponding row vectors in views 1 and 2, respectively.

$$S\left(L_i, R_i\right) = \frac{\sum_{i=1}^{n}(L_i \times R_i)}{\sqrt{\sum_{i=1}^{n}(L_i{}^2)} \times \sqrt{\sum_{i=1}^{n}(R_i{}^2)}}$$
(1)

The calculation basis is established by capturing two images with an angle of α in binocular vision. To pinpoint the exact spatial location of point P within the two images, namely P1 and P2, the cosine similarity matching method is employed. The inter-point distances between P1 and P2 are determined through coordinate values. In this case, the same row in both views is utilized to generate a red diamond-shaped cross-section representing the 3D object (Fig.1b). Subsequently, all these cross-sectional slides (from slide$_1$ to slide$_n$) are consolidated in a single direction to reconstruct the ultimate 3D volume. The spatial height value (PH) of point P can be computed using Eq. 2.

$$PH = \frac{P_1 P_2}{2} \times \tan\alpha$$
(2)

2.2   Resolution enhancement by sub-pixel shifting.

The Nyquist criterion requires the sampling frequency to be at least twice the highest frequency contained in the signal; otherwise, the information about the signal will be lost. A camera's resolution can be expressed as follows:

$$Re_c = P_s/M \tag{3}$$

where Re_c represents the camera resolution, P_s stands for the pixel size, and M denotes the magnification of the optical system. For an image taken by a SWIR camera with a 640 x 512-pixel array, with each pixel measuring 15 μm in physical size. The theoretical resolution of the image is 83 μm for a FoV of 42.5 mm x 53 mm. Image restoration is typically accomplished by convolving the raw image with the point spread function (PSF) in addition to system noise. In the spatial domain, the restored image [28] is expressed as below:

$$g(x,y) = h(x,y) * f(x,y) + \eta(x,y) \tag{4}$$

where (x, y) is the coordinate value, $g(x, y)$ is the restored image, $f(x, y)$ is the raw image, $h(x, y)$ is the system PSF, and $\eta(x, y)$ is the additive noise.

To enhance imaging resolution, a technique that involves the fusion of half-pixel-shifted images is utilized. This approach has been effectively used to improve resolution when pixel size and number are constrained [29-31]. As illustrated with a one-dimensional pixel array in Fig.2a1, array H represents the desired HR image. Array LA corresponds to the image acquired after applying a 1/2-pixel shift to the raw image array LB. By initializing the HR G_{LA1} as 0, each pixel's value in the HR image can be sequentially computed using Eqs. 5- 6 as

$$G_{LA1} = (G_{H1} + G_{H2}) * \frac{1}{K} \tag{5}$$

$$G_{LB1} = (G_{H2} + G_{H3}) * \frac{1}{K} \tag{6}$$

where G_{LAn}, G_{LBn}, G_{Hn} are the pixel values and K is the resolution enhancement ratio.

In the case of a 2D array, this process yields four raw images following half-pixel shifts along four directions, resulting in the creation of four consecutive positions denoted as ① to ④ (Fig.2a2). A displacement stage was employed to achieve the precise half-pixel shift of the sample. The critical requirement is to maintain a shift step length that exactly corresponds to half a pixel when projected onto the camera. Given a 42.5 x 53 mm FoV to be projected onto a 640 × 512-pixel array, each pixel is responsible for collecting signals from 83 µm. Consequently, each movement on the stage was precisely 41.5 µm.

Rather than the traditional method (TR) of measuring or estimating the PSF of the imaging system [32], CNNs were chosen to enhance image resolution, renowned for their feature extraction prowess [33-35]. Each set of four sub-pixel shifted images was restored to their original positions, doubling their size via nearest interpolation. Subsequently, Subsequently, each image was shifted by one pixel in the opposite direction of the sample's movement, effectively aligning them half a pixel back to their original locations. Three convolution layers were then utilized to merge the four input images and reconstruct the HR image⑤ (Fig.2a2). Shifting the HR image by one pixel in four directions, mirroring the image acquisition process, provided a critical test. If the reconstructed HR image was accurate, the four shifted images closely matched the four original images (Fig.2b). Through iterative processing without regularization, we leveraged the convolutional layer's inherent preservation of adjacent pixel continuity. Employing Mean Square Error (MSE) as the loss function (Eq.7), discrepancies between the network's outputs and the four obtained images were minimized. The MSE calculation is as follows:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - f(x_i))^2$$
(7)

where x_i is pixel index, y_i is true intensity, $f(x_i)$ is the predictive intensity, m is the sample number.

In the process of image fusion using convolutional layers, sub-pixel reconstruction techniques are employed to convert low-resolution images into HR ones. In sub-pixel reconstruction, the consistency of

pixel reconstruction is maintained, as all pixels in the image move together. This ensures that the convolution kernel size precisely covers the surrounding pixels, preventing information loss or insufficiency. The weight sharing mechanism in convolution layers of CNN maintains the consistency of pixel reconstruction as all image pixels move together. This mechanism effectively reduces the number of network parameters, enhancing network efficiency. Consistency in the reconstruction modality during sub-pixel reconstruction ensures the smoothness and continuity of the output image, resulting in a clearer and more natural HR image. We refer to the application of CNN networks based on such image sequences as iterative convolutional sub-pixel image reconstruction (ICSP-IR).
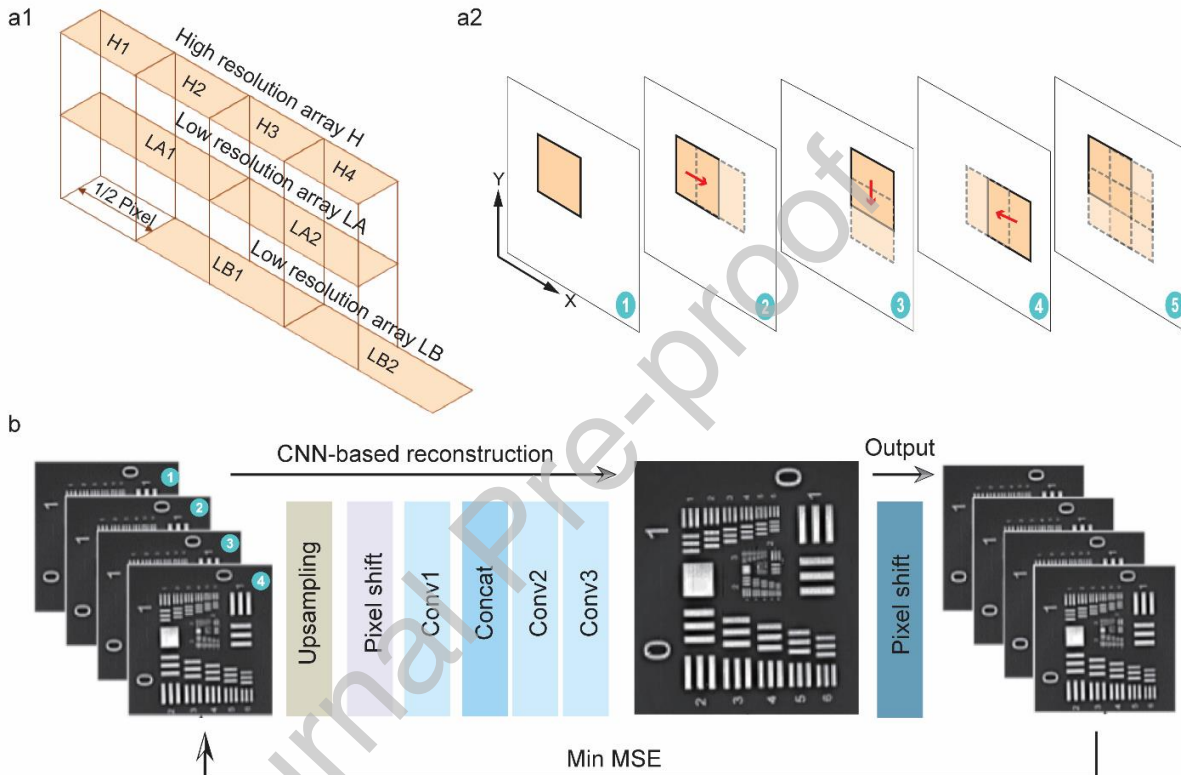
Fig. 2. Schematic illustration showing the principle for enhancing the resolution by using a sub-pixel shifting method and an ICSP-IR-framework imaging processing algorithm. (a1) ½ pixel shift of a line pixel array LB to form a line pixel array LA. Both pixel arrays were used to construct a HR image. (a2) Four ½ pixel shifted image views can be obtained by moving the sample stage from Position ① to ④ by half pixel values respectively along the directions, as indicated by red arrows. (b) Workflow of imaging processing with the ICSP-IR framework which integrated 4 raw images into 3 convolution layers (conv1, conv2, conv3) to output a HR image.

2.3   Sample Preparation

We utilized two biocompatible NIR-II down-conversion nanoparticle nanoparticles, including $NaYF_4$: 3% Nd@ $NaYF_4$ (core-shell) nanoparticles (NdNPs-PEG) for vessels imaging (excitation at 808 nm and emission at 1060 nm) and $NaYbF_4$: Ce, Er@$NaYF_4$ (core-shell) nanoparticles (ErNPs-PEG) for tibia imaging (excitation at 980 nm and emission at 1530 nm) [36-38]. The size of NdNPs-PEG with rectangular cross-section and ErNPs-PEG with hexagonal dimensions are 75 nm and 72 nm, respectively (Fig.3a). This strategy optimizes the performance of both contrast agents in imaging vessels and bones (situated deeper within the leg), allowing for the differentiation, and merging of the respective images acquired for bone and blood vessels.

3   Results

The imaging results of the two contrast agents revealed distinct scattering effects at different infrared bands in biological tissues, particularly in bone and blood vessels (Fig.3b1/b2). Vascular imaging at 1060 nm displayed strong scattering effects, illuminating adjacent skin and bone structures. In contrast, bone imaging at 1530 nm exhibited weaker scattering effects [39]. The Y-shaped vascular structure within the yellow box (Fig.3b1), featuring the strongest fluorescent signal, confirmed its origin from blood vessels. However, not all blood vessels displayed clear signals, with only vague outlines visible in the paw skin. The bone image provided precise details of bone joints, with certain local structures exhibiting high intensity, possibly due to individual differences among mice. Furthermore, the relative distributions of bone and blood vessels can be distinguished in the fused images (Fig.3b3). Notably, all figures presented in Fig.3 are 2D images.
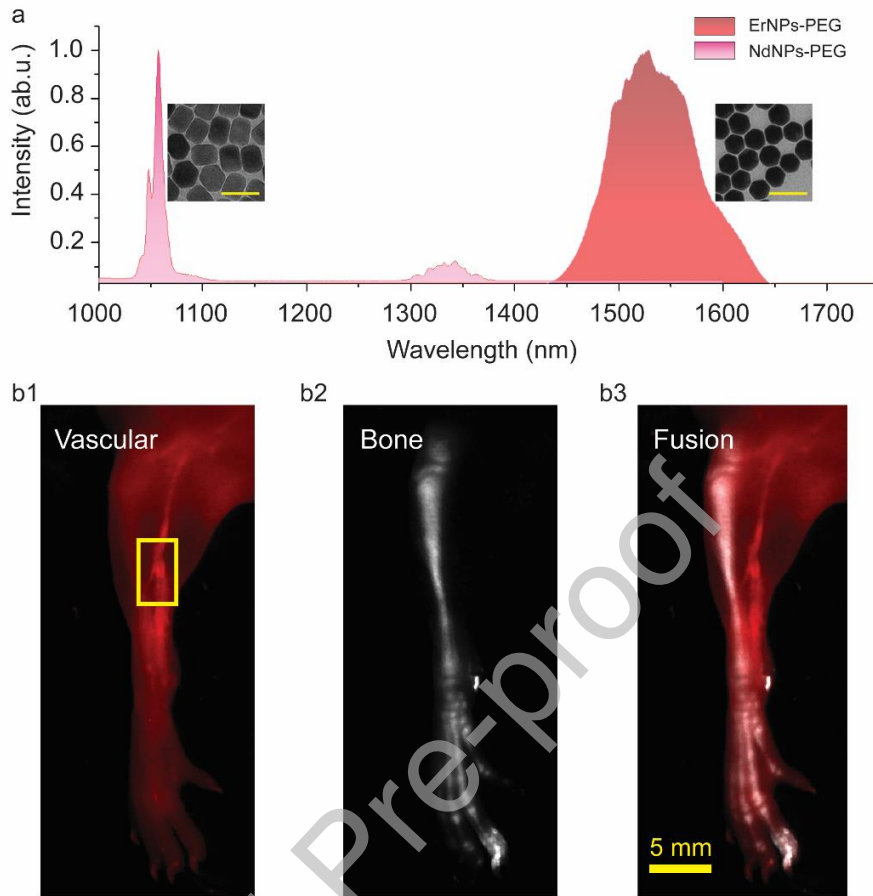
Fig. 3. Emission spectrum of two contract agents and the merged image of vessels and bone. (a) The NdNPs-PEG probe has two emission peaks, 1060 and 1330 nm, respectively. The ErNPs-PEG has broader emission peak (size of FWHM 60nm). Scale bar of TEM images: 100 nm; (b1) vessel image revealed by NdNPs-PEG. (b2) bone image revealed by ErNPs-PEG. (b3) The single view fusion of 2D image bone (white) and blood vessels (red).

To evaluate the resolution, a USAF 1951 resolution test board was positioned above the 1530 nm fluorescence sample. As indicated, the resolution in Fig.4b2, processed using the TR method, is 1.4 times higher than that in Fig.4b1. The reported enhancement of 1.4-1.8 times was attributed to precise PSF measurement, various sophisticated noise reduction techniques, and the utilization of cameras with high quantum efficiency for visible light [29-32]. However, when employing the ICSP-IR method in SWIR in vivo imaging, a resolution improvement of approximately 1.8 times was consistently achieved, nearly consistent with the highest reported resolution enhancement to date (Fig.4b3). Notably, variations in fluorescence beneath the test board resulted in differences in horizontal and vertical resolutions within smaller resolution intervals. Consequently, image resolution was standardized using the horizontal direction. The loss function curve exhibited rapid reduction and stability after 10,000 iterations (Fig.4b4). The imaging sequence of the binocular 3D system is detailed in Fig. 4(c), where the displacement stage pauses at four positions (P1 to P4). At each position, the 980 nm and 808 nm lasers alternately illuminate, and two cameras capture synchronized left and right views in 200ms exposure time. The total

imaging time for this sequence is 2.1 seconds. Using an RTX 3090 GPU and 64 GB of memory, the ICSP-IR achieves the performance of traditional methods at 2000 iterations, surpasses them at 3000 iterations, and reaches optimal results at 7000 iterations, with a total processing time of 4.55 minutes [Fig.4(d)].
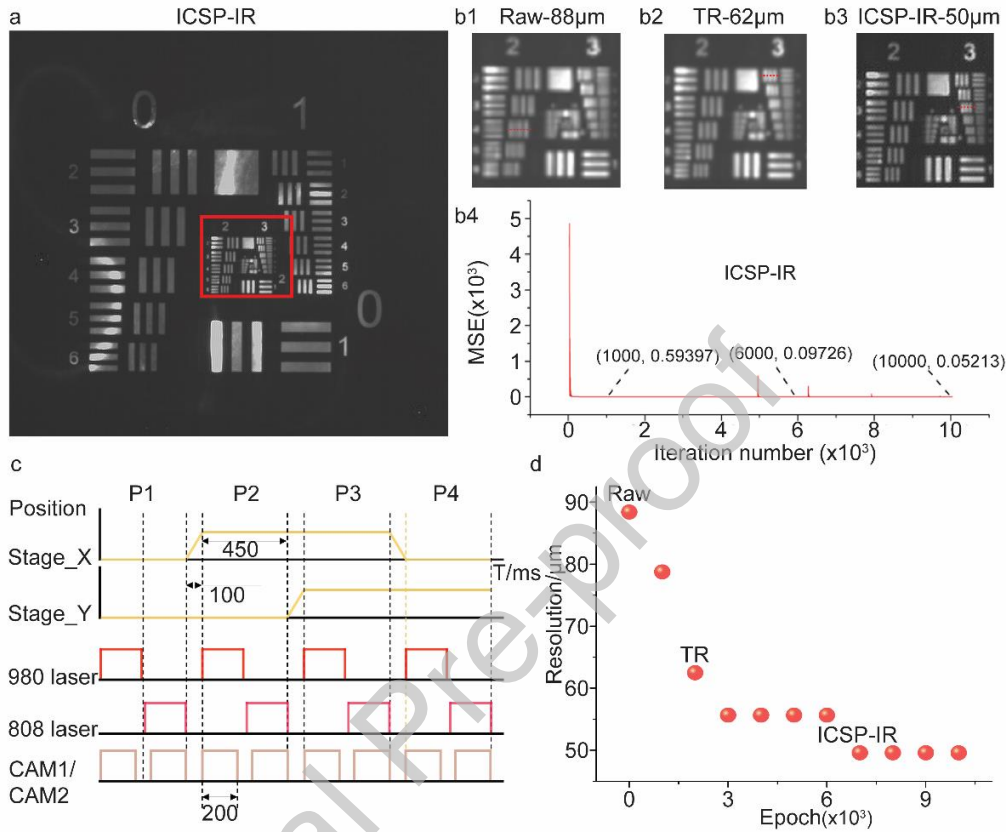
Fig. 4. Resolution enhancement. (a) Image output after ICSP-IR iteration. (b1) Raw image. (b2-3) The local enlargements of (a) were processed by TR and ICSP-IR, respectively. (b4) Loss function curve of ICSP-IR. Comparison between the raw image (b1) and the image (b3) processed by ICSP-IR using 4 half-pixel shifted images showing the final measured resolution enhanced from 88 μm to 50 μm. The number of iterations is displayed up to 1×10^4. (c) Data acquisition sequence. XY movement time between P1-P4 positions: 100 ms. Dwell time at each position: 450 ms. Camera exposure time: 200 ms. (d) ICSP-IR computation time. Stable optimal value reached at 7000 epochs. Time per epoch: 0.039s.

To assess the axial resolution of the binocular system after implementing the ICSP-IR algorithm, sub-diffraction-sized nanoparticles were employed as test specimens. Axial distances in three-dimensional space were measured by displacing a single point a specified distance HHH along the axial direction. (Fig.5). After imaging from both the left and right viewpoints (Fig.5b/d), the images obtained at these two positions were superimposed. The axial resolution of the binocular system was determined by the value of H at which the ICSP-IR algorithm (Fig.5a/c) could successfully identify and distinguish two points within a single view in the merge column. With an axial displacement (H) of 480 μm, the binocular system, which has a FOV of 42.5 x 53 mm and a viewing angle of 24°, demonstrated the ability to resolve

two discrete points separated by 126 μm. This empirical result confirms that, under the described conditions, the axial resolution of the system is established at 480 μm.
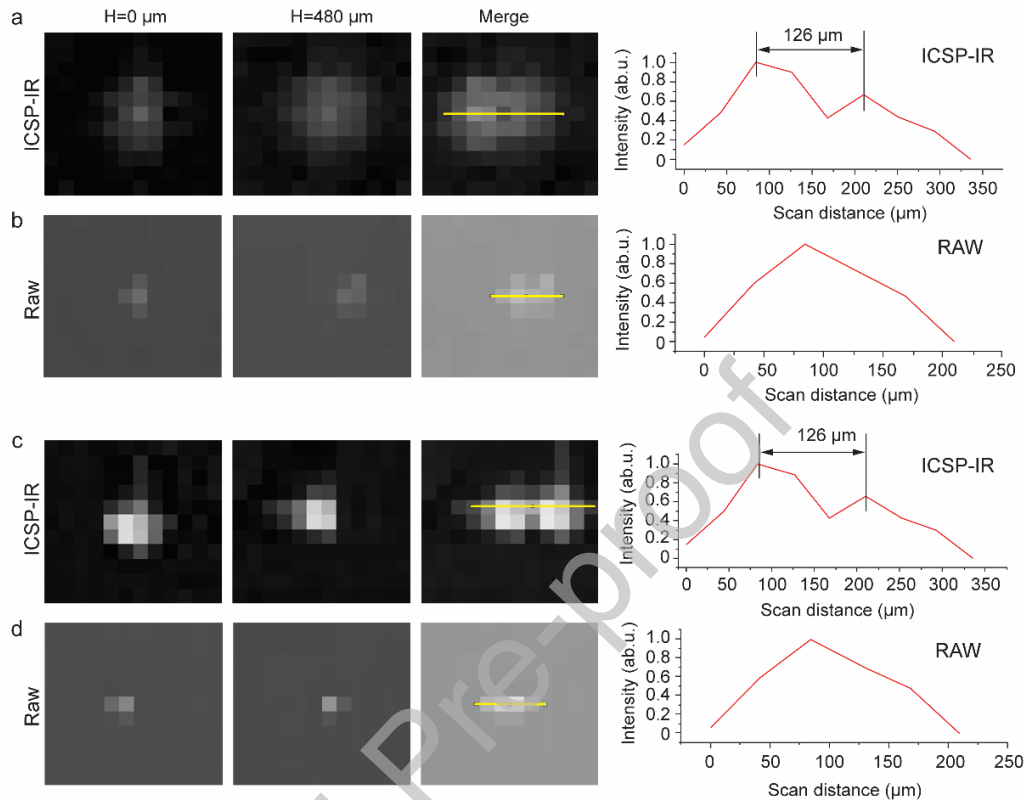
Fig. 5. Evaluation of the axial resolution of binocular systems. (a) and (c) represent the left and right views post-ICSP-IR processing. (b) and (d) correspond to the raw of left and right images. The last two columns depict the fused views at two different axial depths: H = 0 and H = 480, and alongside the positional profile curves for the line scan.

For practical applications, we applied the system for NIR-Ⅱ fluorescence imaging of the whole mouse leg (Fig.6). To assess imaging resolution, the microstructures of blood vessels, with and without the algorithm, were specifically examined. The grey levels of resolution were normalized, and full width at half maximum (FWHM) was utilized to compare the raw and ICSP-IR-based images. Given the limited presence of fine features in bones, resolution was assessed using the spacing distance between joints. The HR local bone and vascular images revealed finer details, such as joint space positions and vascular intersection locations (Fig.6a/b), showcasing smaller gaps. Vessel size alone provides a reliable measure of resolution. As demonstrated in Fig.6c, the resolution improvement reaches approximately 1.5~1.6 times. The resolution of the whole mouse leg is improved, which demonstrates the effectiveness of the algorithm.
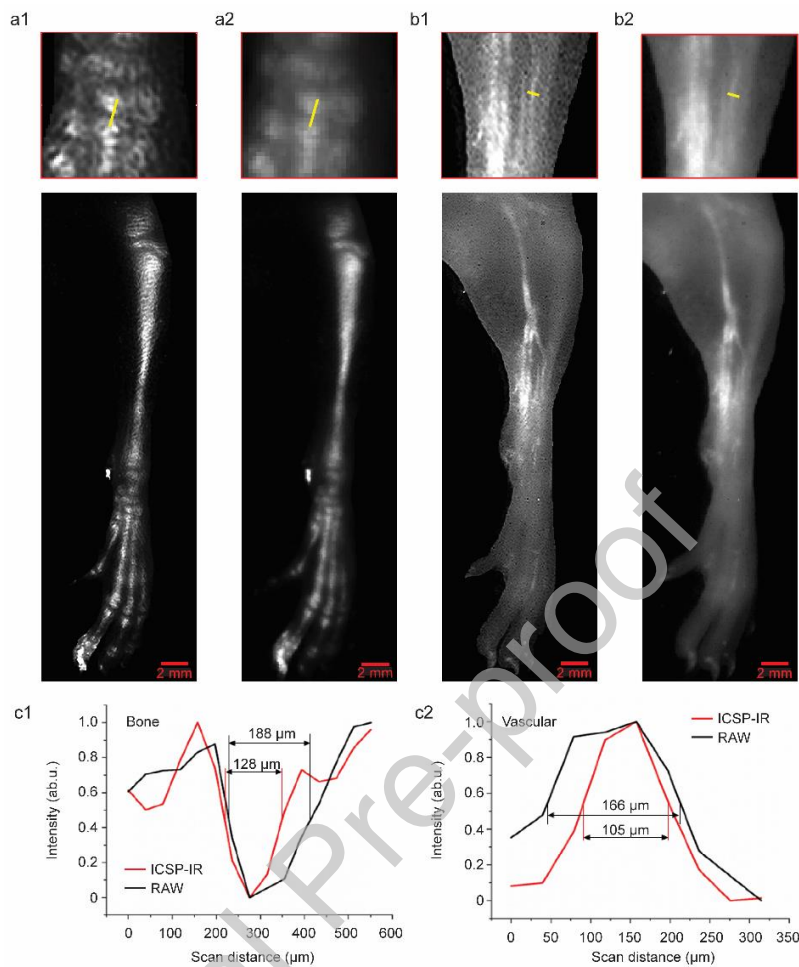
Fig. 6. The area of vessels and jo **Resolution enhancement evaluation by imaging the blood vessels of the whole mice leg.**int gaps in the red box was selected for comparison, and the position of the yellow line scan was taken to obtain the resolution evaluation profile. (a1) and (b1) HR image processed by ICSP-IR. (a2) and (b2) LR are raw images. (c1) The FWHM of the raw image and ICSP-IR image were 188 µm and 128 µm respectively under the bone image. (c2) In the vascular image, the corresponding FWHM values are 166 µm and 105 µm.

To comprehensively evaluate the effectiveness of the ICSP-IR method, we employed both a resolution target and actual biological samples. The resolution target quantified the theoretical limit of resolution enhancement under ideal conditions. However, due to the complex tissue structures and light scattering properties in in vivo samples, the actual imaging results may differ. Therefore, the in vivo imaging experiments demonstrate the application effectiveness of the method in real biological samples. This dual-validation approach ensures a thorough understanding of the ICSP-IR method's performance.

To expedite the overall imaging process, an automated control method was used to seamlessly switch between the two excitation light sources and precisely synchronize subpixel movement with imaging. Subsequently, ICSP-IR processed the subpixel images from both views, replacing the corresponding raw

images. Ultimately, we developed a binocular vision system for imaging a 42.5 mm x 53 mm area (FoV), demonstrating 3D volumetric imaging in 10 to 15 seconds. This resulted in a lateral resolution of 153 μm and an axial resolution of 480 μm, thanks to the integrated ICSP-IR sub-pixel-shifting resolution enhancement algorithm. To characterize imaging depth, careful screening revealed the deepest bone feature at 1011 μm below the skin's scattering layer. When comparing the 3D reconstructions obtained from the HR and LR (Fig. 7a/b), it becomes apparent that the representation of the tibia and foot joint in the skeletal image is notably superior in the former. In contrast, the Y-shaped vessel details in the 2D image are more faithfully reproduced in the latter. The 3D fusion (Fig. 7c1) presents a visual display of the relative positions of blood vessels and bones, with the flexibility to rotate around the Z-axis to provide varying perspectives. The two perspectives (Fig. 7c2/c3) provide a comprehensive representation of the 3D Visualization of 24 °, recapturing intricate details observed in the 2D image. In the zoomed-in portion of the HR_3D fusion (Fig. 7c4), the finest vessels are positioned above the tibia's central area.
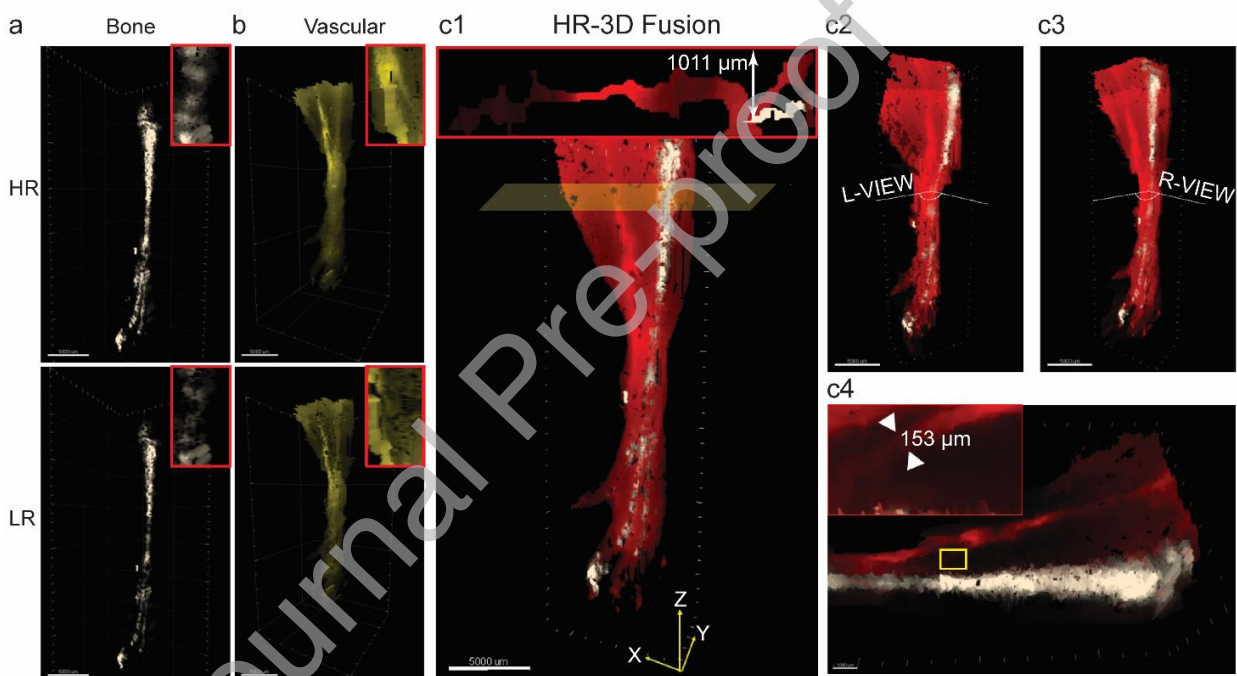
Fig. 7. 3D fusion imaging reconstruction of leg bone and vessels. (a-b) 3D bone image of HR and LR image. Bone and Vascular columns of (a) and (b) show 3D reconstructions using HR and LR atlases, respectively. Scale bar 5 mm (c1) HR_3D vascular and bone fusion. Scale bar 5 mm. (c2) and (c3) Left and right views of (c1). (c4) is an enlarged view of the upper part of (c1) and shows relative position of vessels and bones and the smallest distinguishable vessels features. Scale bar 1 mm.

4 Discussion and Conclusion

The maximum resolution enhancement of two folds through sub-pixel movement should be achievable, and using even smaller steps would theoretically result in unlimited enhancement. Nevertheless, the actual improvement achieved is limited by hardware and imaging system factors, resulting in a

discrepancy between the theoretical and practical outcomes. To evaluate the response of ICSP-IR to the movement step size, tests were conducted using both HR testing cards and simulated images (Fig. S4a). The enhancement achieved through half-pixel and 1/4-pixel movements was only 10 μm (Fig. S4b/c). The current framework of this study is more suitable for the half-pixel movement scheme, as the resolution improvement by 1/4-pixel movement has been limited, and imaging speed and subsequent data processing time can be significantly sacrificed.

We adopted the half-pixel movement strategy and used the ICSP-IR algorithm to obtain high-resolution 2D images. In contrast to the conventional camera resolution that suffers from the limitation of pixel size, the ICSP-IR enhancement algorithm is capable of efficient improvement of resolution by 1.6-fold in biological tissues. Fusion imaging of bones and blood vessels in vivo with 42.5 x 53 x 14 mm^3 was achieved with two types of NIR-Ⅱ contrast agents. The result provides an approach to independently study blood vessels and bone.

The enhancement achieved using NdNPs-PEG nanoparticles (at 1060 nm) was 1.5 times, slightly lower than that obtained using ErNPs-PEG nanoparticles (at 1530 nm). This discrepancy may arise from tissue scattering effects, the enhanced depth penetration capability of 1530 nm light, and variations in the quantum efficiency of the infrared camera across different spectral bands. Analyses of point spread function (PSF) images and assessments of camera quantum efficiency support this assertion (Fig. S3b/c), showing a larger PSF and lower quantum efficiency at 1060 nm than at 1530 nm.

Current NIR-II nanoparticles are limited by luminescence intensity, preventing efficient penetration through thick tissues, let alone fully penetrating the whole mouse body. Moreover, the translation of such two-view images into the whole 3D image remains difficult. Meanwhile, because the fluorescent image lacks texture features, it puts higher demands on the matching algorithm. The proposed ICSP-IR system faces several potential limitations and trade-offs. Theoretically, sub-pixel movements can offer unlimited resolution enhancement, but practical outcomes are constrained by hardware limitations such as the precision of displacement stages and sensor sensitivity. These limitations result in a gap between theoretical potential and practical performance. Additionally, while half-pixel movements have proven effective, smaller steps like 1/4-pixel movements introduce significant computational demands, leading to longer processing times and reduced imaging speed. The current NIR-II nanoparticles also limit luminescence intensity, affecting penetration through thick tissues.

Therefore, our future work will focus on high-resolution whole 3D imaging to ensure 360-range visualization. We believe that with the continuous advancement of in vivo fluorescent materials, the ability to image thick tissue is achievable. We anticipate that such 3D imaging technology using multiple views or structured light illumination based on NIR-Ⅱ contrast agents and resolution enhancement method will become a powerful imaging technique for deep-bone and vessels imaging. This may provide a new solution [40] to visualize tissue structures in a large volume for the study of bone-vessel associated diseases and tracking the pathway of biomolecules and vehicles.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Code, Data, and Materials Availability

The proposed control program for binocular data acquisition is written in C#. The reconstruction codes are written in MATLAB. The data have been analyzed using ImageJ and Imaris. The authors will provide additional details upon request. The code of ICSP-IR can refer to https://github.com/BigbrotherY/Code_ICSP-IR.git

Declaration of generative AI and AI-assisted technologies in the writing process
During the preparation of this work, the author(s) used Chatgpt to improve readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] H. Yu, M. Ji, Recent advances of organic near-infrared II fluorophores in optical properties and imaging functions, Mol. Imag. Biol., 23 (2021) 160-172.
[2] S. Diao, X. Zhang, L. Hu, Recent advances in near-infrared II fluorophores for biomedical imaging, Advanced Drug Delivery Reviews, 2023, 114697.
[3] Q. Wei, D. Xu, T. Li, et al., Recent Advances of NIR-II Emissive Semiconducting Polymer Dots for In Vivo Tumor Fluorescence Imaging and Theranostics, Biosensors, 12 (2022) 1126.
[4] N.-n. Zhang, C.-y. Lu, M.-j. Chen, et al., Recent advances in near-infrared II imaging technology for biological detection, Journal of nanobiotechnology, 19 (2021) 132.
[5] S.S. Liew, Z. Zeng, P. Cheng, et al., Renal-clearable molecular probe for near-infrared fluorescence imaging and urinalysis of SARS-CoV-2, J. Am. Chem. Soc., 143 (2021) 18827-18831.
[6] S. He, S. Chen, D. Li, et al., High affinity to skeleton rare earth doped nanoparticles for near-infrared II imaging, Nano Lett., 19 (2019) 2985-2992.
[7] Y. Li, X. Wen, Z. Deng, et al., In vivo high-resolution bioimaging of bone marrow and fracture diagnosis using lanthanide nanoprobes with 1525 nm emission, Nano Lett., 22 (2022) 2691-2701.
[8] C. Li, W. Li, H. Liu, et al., An activatable NIR-II nanoprobe for in vivo early real-time diagnosis of traumatic brain injury, Angewandte Chemie International Edition, 59 (2020) 247-252.
[9] M. Tan, F. Li, N. Cao, et al., Accurate in vivo nanothermometry through NIR-II lanthanide luminescence lifetime, Small, 16 (2020) 2004118.
[10] Y. Fan, P. Wang, Y. Lu, et al., Lifetime-engineered NIR-II nanoparticles unlock multiplexed in vivo imaging, Nature nanotechnology, 13 (2018) 941-946.
[11] X. Song, W. Zhu, X. Ge, et al., A new class of NIR-II gold nanocluster-based protein biolabels for in vivo tumor-targeted imaging, Angewandte Chemie International Edition, 60 (2021) 1306-1312.

[12] Z. Luo, D. Hu, D. Gao, et al., High-Specificity In Vivo Tumor Imaging Using Bioorthogonal NIR-IIb Nanoparticles, Advanced Materials, 33 (2021) 2102950.

[13] S. Yu, D. Tu, W. Lian, et al., Lanthanide-doped near-infrared II luminescent nanoprobes for bioapplications, Science China Materials, (2019).

[14] P. Wang, Y. Fan, L. Lu, et al., NIR-II nanoprobes in-vivo assembly to improve image-guided surgery for metastatic ovarian cancer, Nature communications, 9 (2018) 2898.

[15] S. Zhu, S. Herraiz, J. Yue, et al., 3D NIR-II molecular imaging distinguishes targeted organs with high-performance NIR-II bioconjugates, Advanced Materials, 30 (2018) 1705799.

[16] J. Qi, C. Sun, D. Li, et al., Aggregation-induced emission luminogen with near-infrared-II excitation and near-infrared-I emission for ultradeep intravital two-photon microscopy, ACS nano, 12 (2018) 7936-7945.

[17] F. Wang, H. Wan, Z. Ma, et al., Light-sheet microscopy in the near-infrared II window, Nat. Methods, 16 (2019) 545-552.

[18] S. Wu, Z. Yang, C. Ma, et al., Deep learning enhanced NIR-II volumetric imaging of whole mice vasculature, Opto-Electronic Advances, 6 (2022) 220105-220101-220105-220109.

[19] T. Chakraborty, M.K. Driscoll, E. Jeffery, et al., Light-sheet microscopy of cleared tissues with isotropic, subcellular resolution, Nat. Methods, 16 (2019) 1109-1113.

[20] J.F. Dekkers, M. Alieva, L.M. Wellens, et al., High-resolution 3D imaging of fixed and cleared organoids, Nature protocols, 14 (2019) 1756-1771.

[21] J. Coussement, A. Rouvié, E. Oubensaid, et al., New developments on InGaAs focal plane array, Infrared Technology and Applications XL, SPIE2014, pp. 39-47.

[22] R. Fraenkel, E. Berkowicz, L. Bikov, et al., Development of low-SWaP and low-noise InGaAs detectors, Infrared Technology and Applications XLIII, SPIE2017, pp. 11-18.

[23] S. Frasse-Sombet, T. Colin, C. Bonvalot, et al., The lowest cost and smallest footprint VGA SWIR detector with high performance, Infrared Technology and Applications XLV, SPIE2019, pp. 236-248.

[24] A. Rouvié, J. Coussement, O. Huet, et al., InGaAs focal plane array developments and perspectives, Infrared Technology and Applications XLI, SPIE2015, pp. 22-29.

[25] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, Cambridge university press2003.

[26] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on pattern analysis and machine intelligence, 22 (2000) 1330-1334.

[27] D.T. Larose, An introduction to data mining, Traduction et adaptation de Thierry Vallaud, (2005).

[28] M.P. Ekstrom, Digital image processing techniques, Academic Press2012.

[29] S. Peleg, D. Keren, L. Schweitzer, Improving image resolution using subpixel motion, Pattern Recog. Lett., 5 (1987) 223-226.

[30] H. Ur, D. Gross, Improved resolution from subpixel shifted pictures, CVGIP: Graphical models and image processing, 54 (1992) 181-186.

[31] H. Wei, T. Binnie, High-resolution image reconstruction from multiple low-resolution images, (1999).

[32] Y.P. Li, B. He, T.Y. Liu, Comparison and Simulation of Subpixel Imaging Modes for Linear CCD, Applied Mechanics and Materials, 236 (2012) 1032-1037.

[33] M. Bai, M. Li, A Presentation of Structures and Applications of Convolutional Neural Networks, Highlights in Science, Engineering and Technology, 61 (2023) 10291.

[34] M. Andries, F. Charpillet, O. Simonin, High resolution pressure sensing using sub-pixel shifts on low resolution load-sensing tiles, 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE2015, pp. 3890-3895.

[35] C. Dong, C.C. Loy, K. He, et al., Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence, 38 (2015) 295-307.

[36] X. Ai, C.J.H. Ho, J. Aw, et al., In vivo covalent cross-linking of photon-converted rare-earth nanostructures for tumour localization and theranostics, Nature communications, 7 (2016) 10432.

[37] Y. Zhong, Z. Ma, F. Wang, et al., In vivo molecular imaging for immunotherapy using ultra-bright near-infrared-IIb rare-earth nanoparticles, Nat. Biotechnol., 37 (2019) 1322-1331.

[38] F.P. Blommel, P.N. Dennis, D.J. Bradley, Effects of microscan operation on staring infrared sensor imagery, Infrared Technology XVII, SPIE1991, pp. 653-664.

[39] Y. Gu, Z. Guo, W. Yuan, et al., High-sensitivity imaging of time-domain near-infrared light transducer, Nature Photonics, 13 (2019) 525-531.

[40] C. Mi, X. Zhang, C. Yang, et al., Bone disease imaging through the near-infrared-II window, Nature Communications, 14 (2023) 6287.

Author Biography

Zhichao Yang is a PhD student at the Institute for Biomedical Materials & Devices (IBMD), School of Mathematical and Physical Sciences, University of Technology Sydney. His primary research interests are in the fields of bioimaging and microscopy.

Dayong Jin obtained his PhD from Macquarie University in 2007. In 2015, he joined UTS as a chair professor. He established the Institute for Biomedical Materials & Devices (IBMD), to transform advances in phonics and advanced materials into disruptive analytical, biomedical, and imaging technologies. He is an Australian Laureate Fellow, fellow of American Institute for Medical and Biological Engineering (AIMBE), and fellow of Australian Academy of Technology and Engineering (ATSE), with expertise covering biomedical engineering, photonics, nanotechnology, microscopy, and analytical chemistry, to enable rapid detection and

super resolution imaging of cells and molecules. He is the 2017 Physical Scientist of the Year among the Australian Prime Minister's Prize for Science.

Graphical Abstract
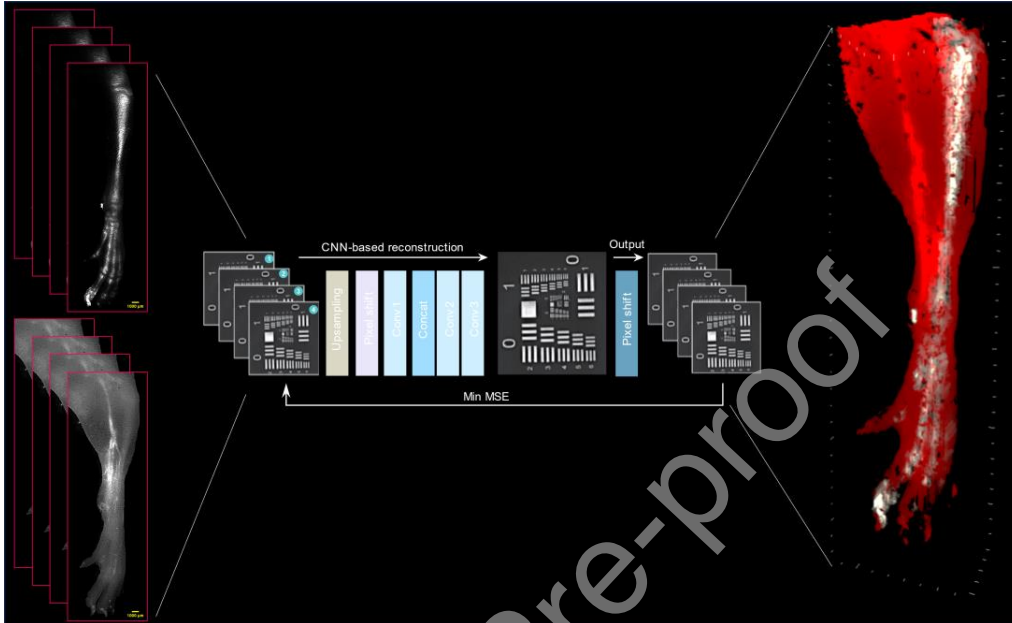
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: