國立臺灣大學電機資訊學院電子工程學研究所

碩士論文

Graduate Institute of Electronics Engineering
College of Electrical Engineering and Computer Science
National Taiwan University

Master's Thesis

TAG-SPARK: 深度學習提升高速雙光子體積成像系統應用於活體老鼠小腦運動神經之研究

TAG-SPARK: Empowering High-Speed Volumetric Imaging with Deep-Learning and Spatial Redundancy for the Cerebellar Neuronal Activities

謝尹慈

Yin-Tzu Hsieh

指導教授：林致廷 博士、朱士維 博士

Advisor: Chih-Ting Lin, Ph.D., Shi-Wei Chu, Ph.D.

中華民國 113 年 7 月

July, 2024

doi:10.6342/NTU202402518

誌謝

能完成這篇論文，首先我要特別感謝我的電子所指導教授 林致廷老師，願意讓我同時能以共同指導的身分與物理系 朱士維老師學習。這不僅讓我可以充分利用電資學院的各種課程與資源，同時也讓我可以選擇自己有興趣的研究——神經影像分析。更是謝謝致廷老師實驗室良好的團隊氣氛，在這兩年的期間，都受到了大家各種幫助，才可以讓我有更多的時間在研究上”Go ahead”。

與此同時，也感謝在中研院暑期實習時指導我的 陳壁彰老師，讓我充分的利用暑假的時間學習了很多基礎的影像處理知識與技術，才能夠一進入碩士研究就知道如何處理與分析。並且在後續兩年繼續指導我，也因為這樣的合作關係，讓我加強與學習了不同的操作方法。也感謝 李家銘博士，在過程中解答我很多光學知識、影像處理方法、問題解決的方向。因為壁彰老師實驗室，總是讓我覺得中研院就像一座後花園，可以在繁忙之中休息片刻。

而這一切，我想都是因為遇到了 朱士維老師，才發生的。即便當初因為疫情影響只能以線上的方式與老師面談，卻還是讓我感受到老師對於我想要以共同指導學生身分加入的支持，以及讓我在這兩年參加各種活動，例如前往美國參加 SPIE2024 進行報告；因為朱老師，我總是意外的達成很多原本都不敢想像的成就。還有像是去中研院暑期實習、李長榮基金會獎學金活動、公司暑期實習等等。都因為老師的支持與影響，讓我更善於走入人群，並且樂於尋找更多的機會，還有探索更多的目標與方向。

在研究上，除了朱老師的指導外，我也相當感謝實驗室的各位成員。無論是當初帶領我的展麟，讓我很快的熟悉儀器操作，與認識這個研究題目。或是在我還並未熟悉系統卻需要解決問題時，給予我許多幫助的 Kentaro 與 Ted，也是因為你們，總能與我分享天氣晴朗的時刻。以及在大大小小事情上，給予我建議的冠傑與采穎，替我這個初生之犢上了很多堂課。也是因為有各位的協助，才能讓我度過這兩年的時光，完成許多的任務。而這個龐大的研究題目中，能與不同專業領域的實驗室合作，也是讓我可以完成的原因。其中，促使老鼠實驗順利進行的，不外乎是有了 潘明楷老師與 李志昌博士的協助與教導，除了實驗更是讓我學習

doi:10.6342/NTU202402518

到許多生物方面的知識。而在數據分析方面，因為有 吳順吉老師的專業團隊，才能帶領我們一同討論，其中更是要感謝凱錞與文琪不斷的回應我各種疑問，協助我更清楚分析的技巧。

在這兩年期間，除了這個小腦的主題以外，也曾做過與生科系 陳示國老師合作的生理時鐘研究，特別感謝能夠藉由這個機會認識到柏廷學長。柏廷學長總是在實驗方面給我很多鼓勵與建議，可以認識同樣擁有底片嗜好，又是專業級樂手，讓我很開心。還有與義大利合作的 Adaptive Optics 題目，很高興因此可以認識 Antonio，以及學習到很多不同的技能。

我想，這兩年就好像玩了一款沉浸式互動遊戲，需要探索世界地圖，不斷解鎖任務與達成成就，控制不同技能與其他玩家合作，提升等級，持續旅行，認識全新角色。碩士肯定是一個龐大的章節，讓人穿過森林，越過沙漠，看過綠洲。

doi:10.6342/NTU202402518

中文摘要

揭開人體中複雜的神經網絡是理解大腦功能的重要一步。在神經活動的全面記錄中，同時取得大範圍神經群體，並擁有高時空解析度的體積成像技術至關重要，有助於理解神經訊號的傳遞與細胞之間的溝通。在神經科學領域中，雙光子高速鈣離子成像技術，以其亞微米空間解析度在捕捉神經活動中扮演重要角色。然而，由於取像速度與圖像質量之間的平衡，受限於有限的光子訊號流量，總導致低信噪比和對比度差的問題。

為了提升高速體積影像的品質，我們開發了一套深度學習降噪算法(TAG-lens-based SPAtial redundancy-driven noise Reduction Kernel, TAG-SPARK)，並將其與具有視頻速率的雙光子體積成像系統整合。利用可調式聲波梯度折射率透鏡（TAG lens）實現高速連續焦點變化，取得垂直（axial）空間冗餘資訊，進行自監督模型訓練的開發。結合此降噪深度學習演算法的影像處理方法，該技術實現了超過 300%的峰值信噪比（PSNR）和超過 700%的信噪比 （SNR）增強，同時保持神經活動的訊號變化特徵。

為了展示此技術的能力，我們以觀察活體小鼠的 Purkinje cells 個體與群體反應為例。Purkinje cell 在小鼠運動時的樹突神經訊號表現被視為其關鍵特徵。樹突細胞的鈣離子變化與 complex spike 之間存在顯著相關性。儘管相關的研究已經證實神經訊號由樹突到細胞本體的傳遞概念具有基礎性意義，但僅限於二維的平面影像研究，仍需要高速體積成像進行完整的驗證。

因此，我們的研究旨在探索神經訊號如何在獨立神經細胞中從樹突到細胞本體的傳遞，以及群體細胞之間的相互溝通。技術不僅以高品質影像捕捉神經活動，還有助於我們更深入地理解三維神經結構中神經訊號的傳遞途徑。

關鍵字: 多光子顯微鏡、可調式聲波梯度折射率透鏡 、高速體積成像、 深度學習降噪演算法、空間冗餘資訊、神經網絡、Purkinje cell

doi:10.6342/NTU202402518

ABSTRACT

Comprehensive recording of neural activities across large neuronal populations with high spatiotemporal resolution and volumetric imaging is crucial for understanding neural circuit signal transduction, information processing, and behavior generation. Two-photon high-speed fluorescence calcium imaging stands as a leading technique in neuroscience for capturing neural activities with sub-micrometer spatial resolution. However, challenges arise from the inherent trade-off between acquisition speed and image quality, resulting in a low signal-to-noise ratio (SNR) due to limited signal photon flux.

To enhance the quality of high-speed volumetric imaging, we developed a deep-learning denoising algorithm, TAG-SPARK (TAG-lens-based SPAtial redundancy-driven noise Reduction Kernel). We propose a home-built two-photon volumetric imaging system with video-rate capabilities, integrating image processing methods using a noise reduction deep learning algorithm. Leveraging the high-speed dense z-sampling at sub-micrometer-scale intervals of a TAG lens, we developed a self-supervised denoising algorithm that exploits the spatial redundancy of z-slices. This approach achieves >300% peak signal-to-noise ratio (PSNR) and >700% SNR enhancement while preserving the fast-spiking functional profiles of neuronal activities.

We demonstrated this technology's application by observing individual neurons and populations of Purkinje cells (PCs) in awake mice. PC activities, which induce dendritic calcium spikes, have been identified as a key feature of motor initiation. The correlation between dendritic and somatic complex spikes is significant. While the foundational importance of the concept of neuronal signal transmission from dendrites

doi:10.6342/NTU202402518

to soma is acknowledged, validating this requires high-speed volumetric calcium imaging.

Therefore, our study aims to explore how neural signals propagate from dendrites to cell bodies within individual neurons, as well as communication among neuronal populations. This tailored technique enables capturing neuronal activities with high SNR, advancing our understanding of neuronal signal transduction pathways within 3D neuronal architecture.

Keywords: Multi-photon microscopy, Tunable acoustic gradient index lens, High-speed volumetric imaging, Deep-learning noise reduction, Spatial redundancy, Neural networks, Purkinje cells

doi:10.6342/NTU202402518

CONTENTS

doi:10.6342/NTU202402518

doi:10.6342/NTU202402518

doi:10.6342/NTU202402518

LIST OF FIGURES

doi:10.6342/NTU202402518

doi:10.6342/NTU202402518

doi:10.6342/NTU202402518

LIST OF TABLES

doi:10.6342/NTU202402518

Chapter 1 Introduction

Regulating animals' behavioral functions, such as body movement coordination, maintenance of homeostasis, and control of the biological clock, the brain stands as one of the most crucial organs in animals. In 2005, *Science* published the top 125 questions worldwide, among them neuroscience and how the brain controls these functions garnered scientific attention. Despite continuous exploration by researchers, as evidenced by the updated 125 questions in 2021, the working principle of the brain and neural signaling remain among the unresolved issues (*1, 2*). This is primarily due to the absence of proper techniques that can simultaneously provide sub-cellular resolution and high-speed volumetric observation for in vivo brain studies. In this chapter, we will introduce the features of the brain along with the challenges associated with observing it in a living state. Subsequently, we introduce the observation techniques proposed in this thesis aimed at overcoming these obstacles. Additionally, we discuss how we integrate the tool with post-processing techniques to further enhance the breakthroughs achieved with this technique.

1.1 Introduction of brain, neuron, and neural signals

The human brain, an essential organ, oversees numerous functions, from basic survival to complex thought processes. Its complex architecture, encompassing a network of billions of neurons, facilitates a wide array of physiological processes and cognitive functions (*3, 4*). This centrality to human life has driven scientists to delve into its working principles.

A neuron consists of a nucleus-containing cell body (soma), signal-sending axon,

doi:10.6342/NTU202402518

and message-receiving dendrites. Neurons can vary in size and structure, with spherical cell soma ranging from 2 to 20 μm in diameter, axons that extend from a few micrometers to over a meter, and tree-like dendrites covering the region from sub-millimeter square to millimeters square.(*5*) The size variations in neurons enable the brain to send or receive neural signals throughout the body to perform different functions efficiently. For example, in Figure 1-1, sensory neurons typically have long dendrites to receive sensory signals from the external environment and transmit them through relatively short axons to the next neuron. In contrast, motor neurons have short dendrites while their axons are long enough to transmit signals over extended distances quickly (*6*).

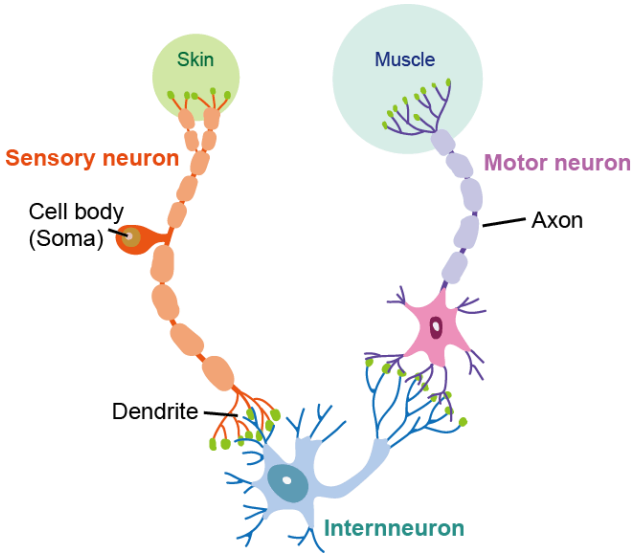


Figure 1-1 Different types of neurons.

(This illustration is a redraw based on the work of Prof. 羅竹芳, CH.8 大腦迷思-大腦與神經.)

Neural signals are electrical and chemical impulses (called action potentials) that traverse the brain's neural circuits, enabling the transfer of information among neurons. This dynamic interplay of signals forms the basis of neural communication, facilitating

doi:10.6342/NTU202402518

the integration of diverse stimuli and the generation of appropriate responses. To understand the complexities of the brain's communication system, it is necessary to recognize the pivotal role of neural signals in mediating sensory perception, motor control, learning, rest, and various higher-order cognitive functions. (*7*)

1.2 Common techniques for functional brain study and limitations

Upon stimulation, the brain sends neural signals, predominantly through action potentials, which are driven by changes in ion concentrations across the neuron's membrane. These action potentials are generated as ions move in and out of the neuron via voltage-gated sodium and potassium channels, which open in response to shifts in the neuron's electrical charge. As the electrical signal advances to the neuron's end, voltage-gated calcium channels permit the influx of calcium ions, facilitating the release of neurotransmitters and the propagation of the signal to downstream neurons **(**Figure 1-2). This precise coordination of ion flow and channel activity enables swift and efficient communication throughout the nervous system. Consequently, various methodologies are employed to record these action potentials and decipher their functional implications.

doi:10.6342/NTU202402518

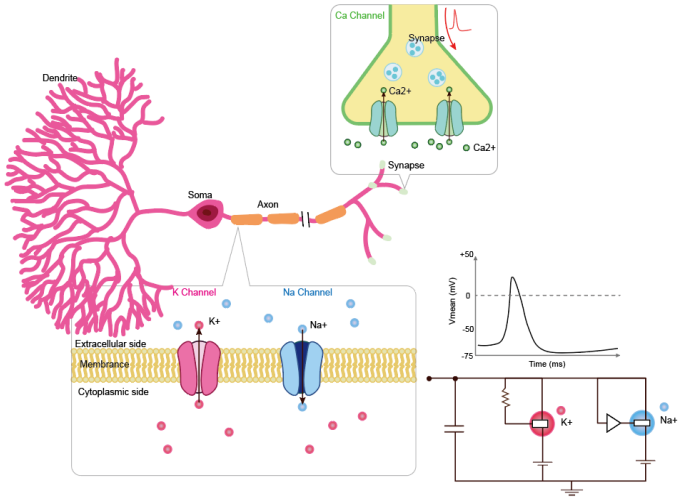


Figure 1-2, Neural ion channels and action potential.

The electrical potential changes induced by action potentials have allowed us to capture neural signals through electroencephalography (EEG) and electrophysiology (EP). EEG excels in capturing real-time brain electrical activity, at 1-30 Hz, enabling high temporal resolution. Although EEG is capable of recording brain-wide activity at large scales in centimeters, it is unable to distinguish signals coming from individual neurons. (*8*) On the other hand, EP allows for the direct and precise sampling of electrical signals from single neurons or even axons by inserting electrodes. Nevertheless, this invasive approach is limited to finite spacing and difficulty of identifying the origin of blindly targeted neuronal signals. In order to enhance the sampling density, EPs employ microarrays equipped with ten to hundreds of electrodes.(*9-12*) Recent advances, such as the Neuralink ASIC,(*13*) are capable of recording from thousands of electrodes. However, even with technologies like silicon probe technology that allow for the insertion of multiple probes to expand the brain's sampling range, they still suffer from similar invasiveness limitations as single-neuron electrode recordings.

For non-invasive techniques, medical imaging technologies such as magnetic resonance

doi:10.6342/NTU202402518

imaging (MRI) and positron emission tomography (PET) offer alternatives. Compared to EEG and EP, these imaging methods provide higher spatial resolution for observing 3D whole-brain images. Both MRI and PET provide functional imaging through indirect detection, such as measuring changes in hemodynamic responses and neurochemical molecular signals, to obtain information on brain neural activity. They lack the spatial and temporal resolution necessary to track individual neuronal activity and neither of them directly observe the flow of information among neurons to capture action potentials or ion concentrations.(*14*)

In summary, while MRI is adept at observing the complete signal flow of the entire brain with high SNR images, it struggles to capture signals at the level of neurons. EEG and EP, on the other hand, excel in capturing precise signals from individual neurons but cannot monitor the activity and spatial location of a large population of active neurons simultaneously. Thus, each traditional technology has inherent limitations that hinder the comprehensive study of individual neural activity across the brain.

1.3 Optical microscopy for brain study

Compared to medical imaging mentioned above, optical techniques offer sub-cellular resolution (in the micrometer range) and a wide range of temporal resolutions (1-30Hz) for individual neurons. Additionally, optical imaging covers large populations for multi-ROI (>1000), unlike EEG and EP with limited ROI density. This level of detail is crucial for conducting thorough analyses of neural networks, allowing us to understand the connectivity among neural signals.

The employment of optical imaging, coupled with the use of genetically encoded probes, such as voltage-sensitive dyes (VSD) and calcium indicators, targeted to specific

doi:10.6342/NTU202402518

cell types, allows for the precise capture of voltage or ion signals, offering a deeper understanding of how neurons interact.

Using VSD or genetically encoded voltage indicator (GEVI) enables the recording of action potential events directly within intact tissues. This technique is especially valuable for targeted imaging of population activity, due to their favorable voltage sensitivity, as well as sub-micrometer and sub-millisecond temporal resolution.(*15*) However, challenges remain, such as poor targeting of the plasma membrane,(*9*) leading to increased background noise, and low brightness, resulting in diminished photon detection.

Action potential firing frequently leads to alterations in the concentration of intracellular calcium. This provides a means to observe neural activity by monitoring fluctuations in intracellular calcium concentrations ([Ca2+]) with fluorescent calcium indicators. Typically, calcium imaging provides better brightness and SNR than VSD.

Optical calcium imaging encompasses a range of methods that hold promise for brain functional studies, particularly in the simultaneous capturing of hundreds of neurons with subcellular spatial resolution (*16-19*). However, to apply these techniques to *in vivo* brain imaging, two critical challenges need to be addressed: poor penetration depth and slow acquisition speed. In single-photon microscopy, the imaging depth is generally limited to tens of micrometers due to strong background scattering signals. The frame rate of conventional raster-scanning two-photon fluorescence microscopy is limited to tens of fps by galvanometric mirrors (*20, 21*).

To overcome the depth limitation, long wavelength excitation is often employed to minimize scattering and absorption in biological tissues, thereby enabling millimeter-level penetration into the living brain. However, the wavelength selection is also limited by fluorescence, where the excitation wavelength must be shorter than the

doi:10.6342/NTU202402518

emission. A breakthrough in 1990 was the invention of two-photon fluorescence laser scanning microscopy (*22*). Two-photon microscopy (2PM) has become a well-known and widely applied technique in the field of live brain imaging, thanks to its exceptional depth penetration around millimeters (*23-25*). The two-photon excitation process doubles the wavelength of excitation, thus giving better penetration. Furthermore, because of the higher efficiency of nonlinear multiphoton absorption at the focus of a laser beam, optical sectioning capability is significantly improved (*26*). This localized excitation ensures the acquisition of detailed three-dimensional structures of specimens. With its high spatial resolution and penetration depth, which allows for whole-brain recording in small animals such as flies. For relatively larger small animals like mice, when combined with other optical elements such as GRIN lenses (*27-30*), it is possible to acquire detailed neural signals in arbitrary ROIs across the entire brain. Therefore 2PM is the most suitable tool for imaging multiscale functional connectome. (*31-34*).

The development of 2PM has revolutionized the ability to measure the activity of deep neuronal populations. Regarding imaging speed, recent advancements at 2PM have notably enhanced lateral imaging speeds and achieved frame rates up to kHz, but the observations are restricted to a 2D lateral plane (*35, 36*). To thoroughly visualize the complex three-dimensional architecture of neurons and their dynamic signal transmissions, it is essential to extend this rapid control to include the z-axis through the development of high-speed volumetric scanning techniques.

1.4 High-speed volumetric imaging for brain study

Although typical scanning microscopy is able to achieve volumetric images, the scanning rate is not fast enough to capture the neuron transmission in milliseconds (*37*).

doi:10.6342/NTU202402518

To comprehensively explore and understand the intricate three-dimensional architecture of neuronal networks, several high-speed volumetric 2PM schemes have been developed (*38*, *39*), including multi-plane imaging (*40*-*43*), stereoscopic illumination with extended depth-of-field (*44*), and fast z-focus modulation mechanisms (*45*).

Multi-plane imaging techniques have emerged as a valuable tool in the arsenal of neuroscientists. The multiplexed scanning approach creates multiple beamlets that scan the sample in different depths simultaneously, enabling the direct acquisition of volumetric images. However, these strategies tend to be confined by their ability to image a relatively low number of layers (N < 30) and large axial interval (>15 μm), which restricts their utility in generating comprehensive volumetric data sets (*42*). On the other hand, stereoscopic illumination techniques broaden the depth of field through a Bessel beam, ensuring a robust focus when scanning the sample.(*46*) This not only increases information throughput but reduces the number of scans in the axial direction. However, this approach faces the challenge of losing precision in determining the exact z-position of structures within brain tissue, which is critical for accurate 3D reconstruction and analysis.

In response to these limitations, fast focus control techniques such as deformable mirrors(*47*), and tunable acoustic gradient (TAG) lenses (*48*-*54*) are able to acquire the full 3D information in detail. Both methods offer a degree of freedom in focusing modulation, yet TAG lenses perform even more promisingly by achieving the same task as deformable mirrors, but at a speed three orders faster (*37*). Using a TAG lens, it only requires five microseconds to sample hundreds of layers during volumetric imaging within a single transverse scan. By integrating TAG lens with 2PM, we not only achieve high vertical spatial resolution and penetration depth requirements but significantly improve temporal resolution and capture 3D spatial information (*55*). Furthermore, our

doi:10.6342/NTU202402518

prior research demonstrates that incorporating dual GRIN lens two-photon endoscopy with TAG lens enables *in vivo* imaging of neural circuits located 6.0 mm beneath the surface of the brain with high-contrast and 4Hz volume rate (*27*).

However, deep tissue observation and high-speed scanning sacrifice the contrast of images. This problem is exacerbated if a GRIN lens is used to reach into deeper tissues. We previously combined an adaptive optics (AO) module into a homebuilt 2PM that enhances image contrast by nearly two-fold in deep brain regions via correcting both systematic and sample aberrations while achieving 1 Hz volume rate (*56-59*). Nevertheless, this scanning rate is not fast enough to catch the neuronal signals, and if the volume rate is further increased, AO's capability diminishes.

In this study, we attempt to use post-image processing methods to enhance the image contrast. Compared to AO, image processing does not impose strict requirements on the original image quality. Additionally, there is greater flexibility for processing through various algorithms. Therefore, it is a practical approach toward high-speed, high-contrast volumetric imaging of deep brain tissues.

1.5 Image processing and deep-learning techniques for image contrast enhancement

Image contrast refers to the difference in brightness between different regions of an image, which is important for visualizing various cell structures. To enhance the quality of images, image processing plays a crucial role in allowing accurate analysis of data. Traditional image processing methods often rely on handcrafted algorithms and involve a series of predefined steps to process images, such as histogram equalization, image projection (average), smoothing filter, and deconvolution. Nevertheless, these

doi:10.6342/NTU202402518

approaches might prove ineffective when dealing with low signal-to-noise ratios (SNR), potentially leading to artifacts in functional images despite their effectiveness in enhancing image quality.

Deep learning, a subset of machine learning, employs multiple layers of neural networks to process large datasets. Inspired by the human brain, it has revolutionized image-processing tasks such as noise reduction, segmentation, recognition, and categorization. The adoption of deep learning in image processing entered the mainstream in the late 1990s with convolutional neural networks (CNNs), continuously evolved since then (*60-63*). Supervised, unsupervised, and self-supervised learning are three of the most important concepts in deep learning. A supervised learning model learns from labeled reference data provided by humans, enabling it to make accurate predictions. Conversely, unsupervised learning finds answers in the dataset without prior labels provided. Compared to supervised and unsupervised learning, self-supervised learning generates its own supervisory signal from the hidden structures within the data itself.

In the context of image-denoising algorithms, these methods typically utilize redundancy in the spatial or temporal domain and the similarity among input images for model training (*64-66*). Considering the live animal imaging, due to the rapid dynamic in neural signals, obtaining high-quality reference images (ground truth) is difficult, posing obstacles for conventional supervised learning. Since unsupervised learning relies on unlabeled data, it may impact the reliability of neural signals. To overcome these issues, self-supervised learning allows for the effective use of signal information and reduces artifacts from human factors. For example, DeepCAD-RT (*67*, *68*) is a well-known self-supervised learning model. It focuses on the noise reduction of calcium imaging. This model capitalizes on the temporal redundancy of consecutive image frames as the training dataset. By doing so, it significantly enhances the contrast and clarity of the

doi:10.6342/NTU202402518

images.

To demonstrate the effectiveness of this customized TAG-SPARK model, we showcase *in vivo* volumetric calcium observations of the mouse brain. In the following, we introduce Purkinje cells in the cerebellum due to their unique neural structural arrangement and high-speed signal spiking, which are ideal for this demonstration. Our goal is to incorporate noise reduction processing to uncover the uniqueness of neural signal performance.

1.6 Complex mouse brain study - Purkinje cells for example

By harnessing the power of deep learning on image processing, we offer a solution to the issues of low SNR images caused by high-speed volumetric scanning and deep tissue scattering. PC is a kind of neuron cell, which is suitable for this imaging system. Because PCs exhibit widely spread dendrites, forming a tree-like structure in 2D space. These complicated dendrites are arranged in parallel and extend vertically so that they require volumetric imaging to cover the whole individual neurons. We aim to utilize volumetric imaging to simultaneously record PC activity and its population recruitment. Additionally, we analyze the comprehensive signal transmission from dendrites to soma and discover significant variations among adjacent Purkinje cells.

PCs are located around 80-100 μm under the surface of the cerebellar cortex **(**Figure 1-3**)**, which causes background noise to interfere with the image contrast when doing in vivo optical imaging. In vivo PCs exhibit high-frequency action potentials, with simple spikes occurring at around 50 Hz, and accompanied by complex spikes by oscillations at a lower frequency range of 1-10Hz, which intermittently disrupt the regular

doi:10.6342/NTU202402518

spike activity. Therefore, high-speed volumetric scanning is important to capture neural signals in real-time.

PCs are intricately linked to motor control in the cerebellum. (*69-72*) Recent studies show that the cerebellum plays a role in a variety of brain functions, including learning, decision-making, social interactions, and memory. The working hypothesis is that the cerebellum acts as a deep learning and optimization engine within the brain. (*71*) The layer structure of PCs may be the reason why it can function as a convolution work, and the cerebellum is like a Graphical Processing Unit (GPU) (Schematic like Figure 2-6).

The theoretical expectation is that calcium influx in PC dendrites converges into the cell body, elevating the somatic calcium concentration necessary for the generation of complex spikes in PCs.(*73-77*) While the concept of dendritic-to-somatic calcium flow has been accepted as a foundation to justify the calcium imaging method for detecting PC complex spikes, However, the validation required high-speed volumetric calcium imaging, is yet to be achieved. Therefore, in this master thesis, we utilized the PC as our subject of application for the study of neurons with complex structures and functions. We believe this research provides a new option for understanding the computational mechanism of cerebellar neurons.

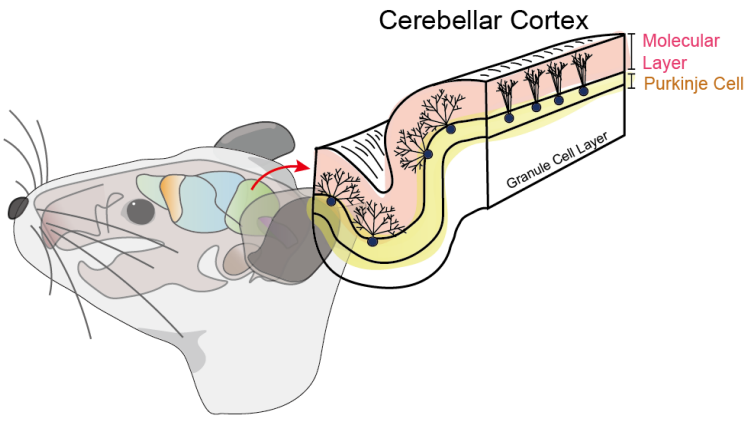


Figure 1-3, Structure of cerebellar cortex and Purkinje cells.

doi:10.6342/NTU202402518

(This illustration is a redraw based on the *Atlas Catalog* 2016*: Central Nervous System. Veterinary NeuroHistology*.[*])

1.7 Aim and structure of this thesis

This thesis aims not only to do a comprehensive observation of a large population of neuronal activities but enhance the image quality of high-speed volumetric images achieved from our home-built 2P-TAG system.

For this, we develop a self-supervised deep-learning algorithm that utilizes TAG lens to provide spatial redundancy of adjacent layers in the z-direction as training data to address the challenge of low image contrast. This algorithm is then integrated with a high-speed two-photon volumetric imaging system to capture clear neural signal transmission and retrieve *in vivo* neural network behavior within large cell populations in the mouse brain.

To examine the capability of our optical system and the noise-reduction algorithm, we observe the neuron signals transmitted through the PCs during various states of mouse movement. If our system works as intended, we expect to find the relationship between the neuron signals and the movement of the mouse. Additionally, we establish a comprehensive pipeline for image processing and analysis to meet the requirements of *in vivo* mouse brain studies. In **Chapter 2**, the physical principles of each essential component of this research are introduced. **Chapter 3** mentions the details of the system design and expected performance, along with experimental results for evidence, and sample preparation for experiments are discussed. In **Chapter 4**, we will discuss image processing and analysis methods, including how to apply these methods to complete a

[*] Website: http://vanat.cvm.umn.edu/neurHistAtls/cataPages/cataCNS.html.

doi:10.6342/NTU202402518

side project on the identification of individual synapses in the *Drosophila* antennal lobe by machine learning-based super-resolution image analysis. **Chapter 5** will present the results of image quality enhancement with noise reduction based on our developed deep-learning model. Furthermore, denoising unveils the intriguing discovery of dynamic neuron signal transmission in *in vivo* PCs. Finally, the conclusion summarizes our current progress and proposes future outlooks in **Chapter 6**.

doi:10.6342/NTU202402518

Chapter 2 Principle of the integration of two-photon TAG imaging system and deep-learning denoise method, TAG-SPARK

In this chapter, we introduce the general principles of three primary techniques. We begin with an overview of two-photon laser scanning microscopy, followed by an examination of TAG lens. Lastly, we discuss the concept of deep learning models and introduce our self-developed model, TAG-SPARK, which is used to enhance the image contrast.

2.1 Two-photon laser scanning microscopy

In contrast to classical single-photon wide-field fluorescence microscopy, 2PM presents distinctive advantages, including optical sectioning and penetration capabilities. The method is grounded in the principle of nonlinear excitation, where fluorophores are stimulated by the simultaneous absorption of two lower-energy photons and the excitation of a single photon with a longer wavelength (Figure 2-1). Two-photon absorption cross-section typically ranges from 1 to 300 GM ($1GM = 10^{-58} m^4 s^{-1} photon^{-1}$), which requires specific conditions and high-intensity laser sources due to the rarity of simultaneous photon-molecule interactions. To enhance two-photon absorption, pulse lasers with high peak energy at short pulse widths of ~100 fs are preferred. Such laser sources provide enough energy and are temporally concentrated, increasing the possibility of two-photon absorption occurring.

doi:10.6342/NTU202402518

Highly localized excitation of 2PM promotes intrinsic optical sectioning ability and allows high spatial resolution for the precise visualization of specific focal planes within biological specimens. This ensures clarity and reduces unwanted background signals, enabling detailed investigations at the cellular levels. Its longer wavelength enhances depth penetration capabilities, allowing for imaging deep into tissues without compromising resolution. Furthermore, since its focal plane is narrow, the risk of photodamage and phototoxicity in out-of-focus regions is minimized, thus prolonging sample viability, particularly during long-term imaging sessions.

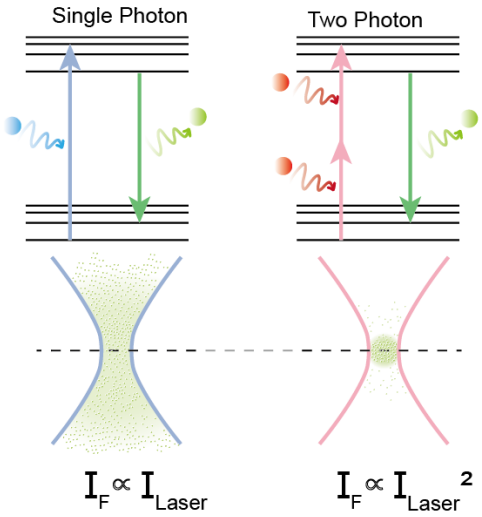


Figure 2-1, Diagram of single- and two-photon excitation and emission.

Through the absorption of a single photon or two lower-energy photons, fluorophores are excited from the ground state to the first excited state. In single-photon excitation, fluorescence is emitted along the focused light cone. In contrast, two-photon excitation exhibits a reduced out-of-focus signal due to its quadratic dependence on excitation intensity, thereby enabling optical sectioning.

(This illustration is a redraw based on the work of Helmchen, F. et al., *Nat Methods* 2005. (*24*))

In two-photon laser scanning microscopy, a pair of galvo mirrors are commonly utilized to achieve continuous raster scanning, allowing for sequential excitation of each position on the focal plane. Fluorescence signals emitted from each point are detected

doi:10.6342/NTU202402518

using a photomultiplier tube (PMT), and the signal is subsequently read by the computer. A 2D image is then reconstructed based on the intensity values assigned to the 2D matrix corresponding to the location of the laser focus (Figure 2-2A). By adjusting the focal plane in the axial direction within the sample cooperating with lateral scanning, 3D images composed of stacked layers of 2D images are achieved (Figure 2-2B).

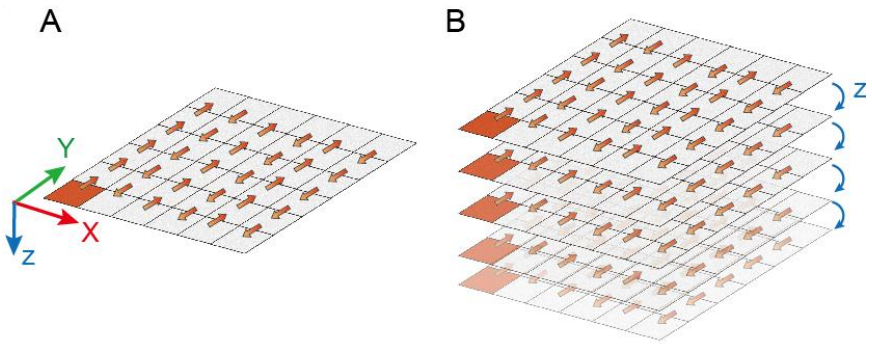


Figure 2-2, Laser raster scanning.

(A) Lateral point scanning and (B) volumetric scanning with translation stage in the z-direction.

Since the successful demonstration of the first 2PM in 1990 (*22, 78*), there has been remarkable progress in multiphoton microscopy (MPM) technology. The advancements include expansion of the field of view (FOV) (*79, 80*), implementation of multibeam optical designs (*81*), and integration of high-speed laser scanning techniques (*82*). Additionally, spatiotemporal multiplexing approaches have further enhanced the capabilities of 2PM. Miniaturized, head-mounted versions of 2PM were also developed and have reemerged (*83, 84*).

In summary, the principles of 2PM establish it as a powerful and versatile imaging method, offering unparalleled insights into the intricate structures and functions of biological specimens with minimal invasiveness and optimal precision.

doi:10.6342/NTU202402518

2.2 Tunable Acoustic Gradient Lens (TAG lens)

TAG lens is a fluid-filled (silicone oil) cylindrical cavity formed by a hollow piezoelectric tube with two flat glass windows on either side for optical access (Figure 2-3**)**. The piezoelectric transducer used to drive the TAG lens undergoes radial excitation with acoustic energy, inducing a periodic standing wave modulation in the fluid, thus altering the index of refraction and continuously varying its focal length. The specific focal lengths are selected with sub-microsecond temporal resolution.

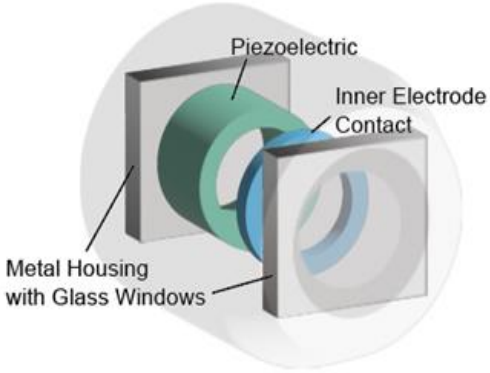


Figure 2-3, Expanded view of a circular TAG lens

(This illustration is a redraw based on the work of McLeod and Arnold, *Journal of Applied Physics* 2007. (*85*))

The mechanics of TAG lenses involve the dynamic interplay between the fluid-filled cylindrical cavity and the acoustic excitation generated by the piezoelectric transducer. The electrodes are placed on the inner and outer circumferences of the cylinder. The voltage signal is applied on the piezoelectric transducer, the vibration velocity of inner electrode contact is proportional to the driving voltage amplitude (*85*),

$$V = V_A sin(\omega t) \qquad \text{Eq. 2-1}$$

where ω is the driving frequency of the TAG lens.

The vibrations-induced standing wave patterns form distinct rings in the radial

doi:10.6342/NTU202402518

direction (*86*). The distribution of these rings follows the Bessel function and depends on the modulation of driving conditions. The rings are determined by the refractive index within the TAG lens, which takes the form of

$$n(\rho, t) = n_0 + n_A J_0\left(\frac{\omega\rho}{c_s}\right) sin(\omega t) \quad \text{Eq. 2-2}$$

assuming TAG lens with linearized fluid mechanics, where n_0 is the static refractive index of the filling fluid, n_A depends on the modulation parameters, ω is the driving frequency of TAG lens, c_s is the acoustic speed, and ρ is the radial coordinate in the lens plane.

At the lens plane, the transverse coordinates are given that $\rho = \sqrt{\xi^2 + \eta^2}$. The phase of the light passing through the lens is given by (*86*, *87*),

$$t_{lens}(\xi, \eta) = e^{ik_0(nL_0)} \quad \text{Eq. 2-3}$$

where $k_0 = 2\pi/\lambda$, the constant for free-space propagation, L_0 is the thickness of a TAG lens. (Here, the thin lens approximation is valid for the system, i.e. a light ray exits the lens at the same transverse location where it entered. Because the deflection of a ray traveling through the lens (<50 μm) is much smaller than the acoustic wavelength within the lens.)

Assuming the beam wave completes the entire TAG lens to arrive L_0, then moves forward or backward by z (Figure 2-4). For a wavefront within the TAG lens, where the phase shift on this wavefront is a constant C, the shape of such wavefront is calculated by the following equation,

doi:10.6342/NTU202402518

$$\because\ t_{lens} = e^{ik_0(n(L_0+z(\rho)))} = const.$$

$$n(\rho)(L_0 + z(\rho)) = C \ \leftrightarrow\ z(\rho) = \frac{C}{n(\rho)} - L_0 \qquad \text{Eq. 2-4}$$

where z<0 is the distance from the exit window to the wavefront.

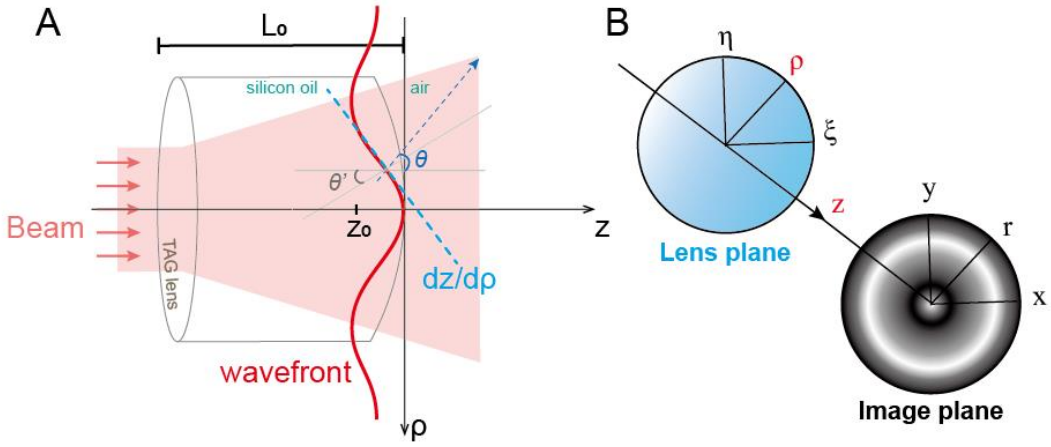


Figure 2-4, Schematic of the beam passing through the TAG lens.

(A) Schematic of beam and TAG lens. (B) Coordination of lens plane and image planes, which refers to the paper by McLeod and Arnold, *Applied Optics* 2008.(*86*)

With Eq. 2-4, we can calculate the angle of deflection θ' at the exit of TAG lens (before exiting), by assuming |z| << L_0,

$$tan(\theta'(\rho)) = \frac{\partial z}{\partial \rho} = \frac{-C\ \partial n}{n^2\ \partial \rho}$$

$$= -\frac{n(L_0 + z)\ \partial n}{n^2\ \partial \rho}$$

$$= -\frac{L_0\ \partial n}{n\ \partial \rho} \qquad \text{Eq. 2-5}$$

To get the angle of a light ray after exiting the lens, Snell's law is applied to the

doi:10.6342/NTU202402518

fluid–air interface,

$$sin(\theta(\rho)) = n(\rho)sin(\theta'(\rho)) \quad \text{Eq. 2-6}$$

where θ is the angle of light after leaving the lens. Assume $n(\rho)/n_{air} = n(\rho)$ and apply Eq. 2-5 and Eq. 2-6 with small angle approximation, we yield,

$$\because\ \theta' = -\frac{L_0}{n}\frac{\partial n}{\partial \rho}, \text{ and } \theta = n\theta'$$

$$\therefore\ \theta(\rho) = -L_0\frac{\partial n}{\partial \rho} \quad \text{Eq. 2-7}$$

Figure 2-5A shows an approximation of a linear slope (red and green dash line in Figure 2-5A) about the inflection point of the central peak. Based on these relations, by modulating the voltage signal periodically, the refractive index will change from maximum to minimum periodically. The angle of light after leaving the lens also changes periodically, with respect to the gradient of the refractive index.

With Eq. 2-2, we know the major rings occur near the extrema of $J_0\left(\frac{\omega\rho}{c_s}\right)$, as illustrated in the gray area of Figure 2-5A. Increasing the driving frequency compresses the radial of the major rings (*85*), and the first major ring, ρ^* is relative to the effective numerical aperture (NA) of the TAG lens (*55*), which is approximately given by,

$$\rho^* = \frac{3.832c_s}{\omega} \quad \text{Eq. 2-8}$$

where 3.832 is the radial coordinate of the first minimum of $J_0(\rho)$.

doi:10.6342/NTU202402518

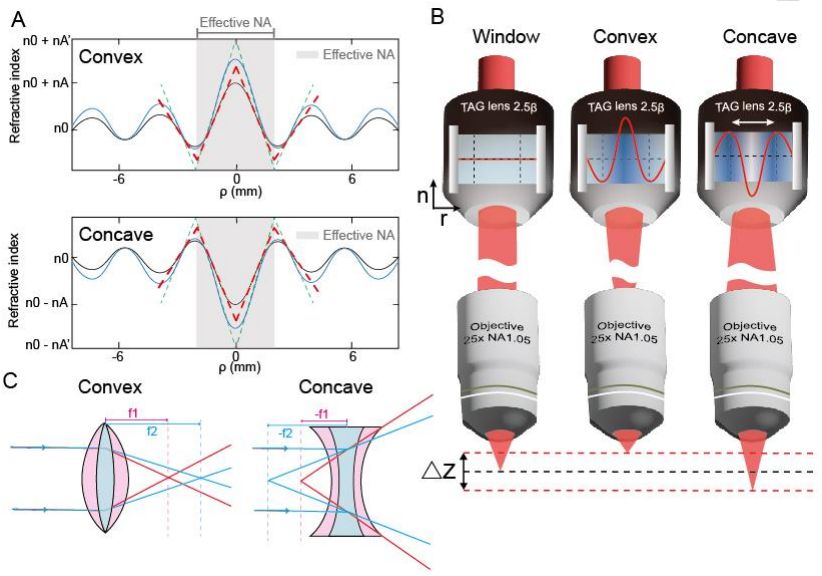


Figure 2-5, Schematic of TAG lens operation in three representative states.

(A) (From top to bottom) The refractive index profile in the radial direction shows the convex and concave lens effects, respectively. The black and blue lines represent wavefronts with different optical powers, while the red and green dashed lines indicate the gradient of the profile. The gray area represents the major ring of the wavefront, with its width corresponding to the effective NA. (B) The convergent and divergent states of the beam corresponding to a convex and concave lens effect. (Red and blue lines represent different optical powers of the lens). (C) (From left to right) TAG lens combined with an objective for axial scanning, where TAG lens acts as a window, convex lens, and concave lens. The focus shifts △z.

(The illustration of (A) is a redraw based on the work of McLeod and Arnold, *Applied Optics* 2008.(*86*))

With Eq. 2-1 and Eq. 2-2, during a single oscillation period, three representative states occur in the TAG lens, which corresponds to zero, positive, and negative $sin\,(\omega t)$.

It results in convex and concave equivalent effects (*88*),

$$sin\,(\omega t) = 0, n = n_0\,, unchanged$$

$$sin\,(\omega t) > 0, n = n_{max}, converged$$

$$sin\,(\omega t) < 0, n = n_{min}, diverged$$

doi:10.6342/NTU202402518

at $sin\,(\omega t) = 0$, the refractive index remains uniform, preserving the incident beam's wavefront. When $sin\,(\omega t) > 0$, resulting in a local maximum, the TAG lens acts as a convex lens, converging the beam. Conversely, when $sin\,(\omega t) < 0$, it experiences a local minimum with low curvature, making the TAG lens equivalent to a concave lens, diverging the beam (Figure 2-5C).

Based on the theoretical analysis above, we know that the greater the applied voltage amplitude, the greater the degree of angular deformation. In practical implementation, the calculation of the depth of field often relies on the optical power of the TAG lens (*47*). Additionally, the relationship between the optical power of the TAG lens and the applied voltage varies among different TAG lens modulation frequencies. We use a TAG lens from "TAG Lens LLC," and the voltage-focal length relationship is documented in this reference. [spec[1]].

The TAG lens should be placed onto the conjugate plane of the objective lens, which varies the axial position of the focus. Different refractive power results in different depth of field captured, i.e. the largest axial moving range of the focus will be changed. The depth of field is calculated by (*37*, *89*),

$$DOF = n \times f_{obj}^2 \times \delta_{total} \times \left(\frac{f_2}{f_1}\right)^2 \qquad \text{Eq. 2-9}$$

$$\delta_{total} = \frac{1}{f_{TAG}}$$

where n is the refractive index of the sample, f_{obj} is the focal length of the objective length, δ_{total} is the optical power of the TAG lens which is equal to the reciprocal of the

doi:10.6342/NTU202402518

focal length of the TAG lens, f_1 and f_2 is the focal length of the relay lens between TAG lens and objective lens, and they are related to the reciprocal of the magnification factor of the beam. The DOF will be increased by increasing the focal length of the objective lens or decreasing the magnification factor caused by the relay lens.

It is noteworthy that a higher driving frequency extends the DOF by increasing optical power (*88*), the related equation is written as,

$$\delta(t) = \frac{L_0 n \omega^2 sin(\omega t)}{2v^2} \qquad \text{Eq. 2-10}$$

where ω is the driving frequency, v is the acoustic speed in the medium. The DOF of a TAG lens demonstrates a quadratic dependency on the driving frequency. However, the driving frequency in a TAG lens is inversely proportional to the size of effective NA and results in a smaller beam size. As a result, increasing the driving frequency to enhance optical power diminishes the effective NA and lateral resolution. Therefore, a trade-off exists between DOF extension and spatial resolution, determined by the choice of driving frequency within the resonant capabilities of the TAG lens.

TAG lenses offer numerous advantages over traditional technologies. First, its response time is limited only by acoustic resonance frequency within the lens material, enabling rapid adjustments in focal length. Second, TAG lenses have higher damage thresholds compared to delicate liquid crystals, enhancing their durability and longevity in practical applications. Additionally, its analog nature ensures it is free of pixelation, resulting in smooth and continuous focal length transitions, which is our most advantageous characteristic. In the next section, we will describe how this advantage will be applied to address our problems (*48-50*, *90*).

doi:10.6342/NTU202402518

2.3 Introduction of deep-learning model

In 2P-TAG lens high-speed volumetric microscopy systems, we encounter issues with poor image contrast. The issue of insufficient SNR mainly results from the reduction in voxel dwelling time of the scanner to capture high-frequency biological signals, which limits the number of photons and results in shot noise that further affects the accuracy of data analysis. To enhance SNR, traditional image processing methods are employed. For example, using Gaussian filters to retain the main features of the image (detailed in **Chapter 4.1.2**) to eliminate shot noise. However, these methods may lead to further blurring and loss of information regarding the original intensity of the neuron signal.

In recent years, CNNs have provided a new solution for improving image quality. Through iterative learning and updating parameters of multiple convolutional layers, neural networks are able to capture complex features and details in images, preserving signals and reducing noise to enhance SNR.

In this research, we were inspired by DeepCAD-RT proposed by Xinyang et al (*67, 68*) for real-time image denoising. Our modified version, TAG-SPARK, uses self-supervised deep-learning, following the Noise2Noise algorithm (*91*) and U-Net architecture (*92*), performs spatial denoising based on TAG lens dense z-sampling. In the following sections, I will first introduce the concept of deep-learning models, then self-supervised learning, and finally U-Net architecture.

2.3.1 The concept of creating deep-learning models

The root concept of deep learning is creating an appropriate function into which we input data and expect it to produce the desired results. For example, a dog recognition

doi:10.6342/NTU202402518

function that distinguishes different kinds of dogs.

A **model** is created by combining multiple functions, to achieve a larger set of objectives. Using an animal recognition system as an example, we combine different functions to recognize cats, dogs, and other animals, ultimately creating a model capable of identifying a variety of animal species. In deep learning, the ultimate goal is to find a suitable model that extracts complex features from simple input data and performs complicated tasks. This process is what we call **training**.

Deep learning models typically consist of layers of network structures called neural networks (NN). In deep learning neural structures, we also have an analogous system: the input data, through weight calculation and signal integration, transforms into an output that continues to the next layer **(**Figure 2-6). We often use GPUs to accelerate the training process of this big calculation.

Interestingly, the structure of these NN layers is similar to the structure of PCs, where synapses, dendrites, soma, and axons are in charge of input, weight calculation, signal integration, and output. This neural signal transmission allows neurons to communicate with each other. Moreover, the biological structure of the parallel arrangement of PCs in the cerebellum resembles the architecture of GPU used in deep learning. This is precisely why in **Chapter 1**, we described the PCs as the GPU in the cerebellum (**Chapter 1.6**).

doi:10.6342/NTU202402518

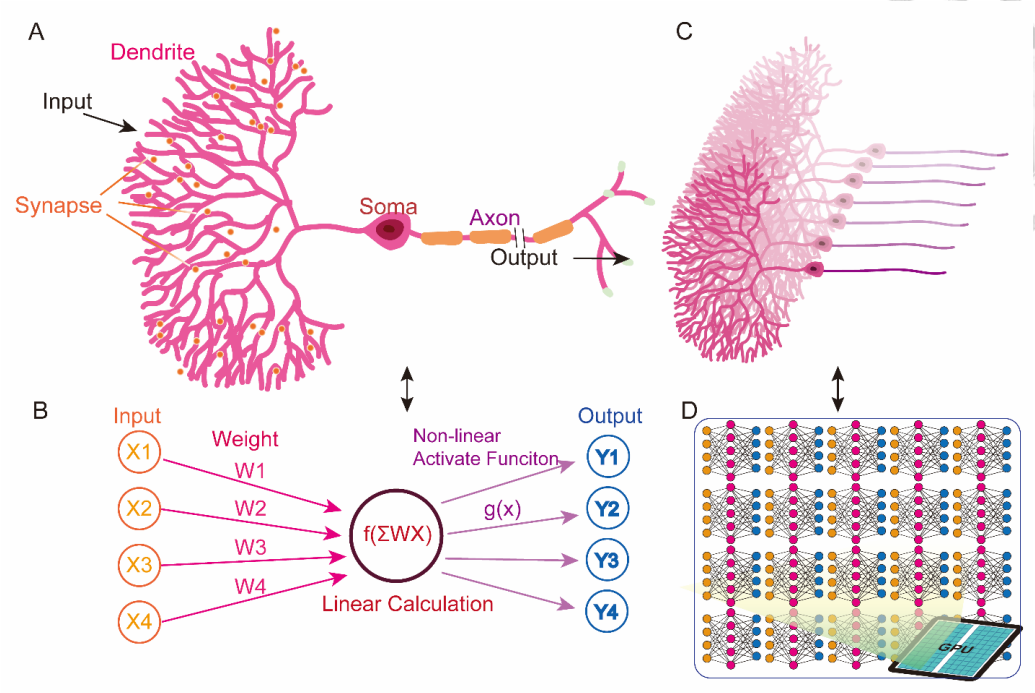


Figure 2-6, Diagram of neuron network in biology and deep-learning.

(A) A PC consists of a tree-like dendrite, circular soma, and a long axon to transmit the signal. (B) A deep-learning neuron network is composed of input dataset, calculation, and output data. (C) The parallel arrangement of PCs creates a complex neuron network to work in the cerebellum. (D) GPU is used to perform the parallel calculation which deals with big data at the same time.

After training the model, we need a standard to evaluate the effectiveness of the training. **Loss function** is a crucial component that quantifies the difference between the predicted outputs of an algorithm and the target values. The loss value obtained from the model indicates its predictive accuracy. Algorithms like backpropagation utilize the loss function's gradient concerning the model's parameters to minimize the loss, thereby enhancing the model's performance on the dataset.

In machine learning, one of the most commonly used loss functions for regression is the Mean Square Error (MSE), which calculates the average of the squared differences between predicted and target values.

doi:10.6342/NTU202402518

$$MSE = \frac{1}{n}\Sigma(y_i - \hat{y}_i)^2$$

where y_i is the input value, $\hat{y}_i$ is the calculated output, and n means the number of data.

Due to the square, the outliers heavily affect loss values. Therefore, Mean Absolute Error (MAE) provides a more moderate measure of the difference between predicted and target values.

$$MAE = \frac{1}{n}\Sigma|y_i - \hat{y}_i|$$

In this study, we implemented the average of MSE and MAE as our loss function to try to enhance the effectiveness of the model training.

As the loss value converges to a minimum, the training of the model is completed. However, sometimes the model may overlearn the training data, i.e. misjudging its complexity and progressing in the wrong direction, which is **overfitting**. To prevent overfitting, we limit the epochs of the training or employ the regularization methods to make the model resistant to noise.

2.3.2 Supervised learning and self-supervised learning

Supervised learning is a machine learning approach that has a clearly defined ground truth. The resulting algorithm minimizes the loss function between ground truth and predicted outcome through gradient descent.

In a common supervised learning setup for image processing, we would have k pairs of clear images (x_i) alongside noisy images $(\hat{x}_i)$. The neural network's learning objective would be to,

doi:10.6342/NTU202402518

$$argmin_\theta \sum_i^k L(f_\theta(\hat{x}_i), x_i)$$

During training, multiple input-target pairs affected by different noises are provided. Therefore, in each epoch, the network model f_θ encountered different sets of noisy input-target pairs to estimate the optimal parameters θ, which will be fine-tuned to achieve convergence. It is done under the loss function L, which measures the dissimilarity between output and clean target (x_i).

Supervised deep learning applications span image recognition, natural language processing, and more, which drive innovation in artificial intelligence. However, in dynamic calcium imaging, it is challenging to obtain the images with high image contrast due to high-speed scanning in deep tissue.

Unlike traditional supervised deep-learning, self-supervised deep-learning discovers hidden patterns in data without ground truth. The basic assumption of self-supervised learning is that the data itself contains the necessary answers. The model's task is to focus on the invariant features amid the difference, using them to discern the answers.

Noise2Noise is a well-known example. The concept of Noise2Noise was introduced by Jaakko et al. in 2018. It directly uses observed images containing random noise, allowing the model to learn and remove noises. It is advantageous in situations where obtaining high-quality target images is challenging. The neural network's learning objective would be,

$$argmin_\theta \sum_i^k L(f_\theta(\hat{x}_i), \hat{y}_i)$$

doi:10.6342/NTU202402518

where, $\hat{x}_i$ and $\hat{y}_i$ are two independent noisy images obtained from the same scene but affected by independent instances of noise.

Noise2Noise efficacy hinges on strict prerequisites: the input and target images must exhibit substantial similarity. The noise should be uncorrelated with an average of zero. This ensures that the model only handles noise and does not interfere with the signal. Independence between input and target noise prevents erroneous parameter learning, while noise unrelated to input features maintains robustness. These conditions guarantee that Noise2Noise effectively learns and removes noise from images (*93, 94*).

2.3.3 Introduction of U-net

U-Net, a CNN architecture introduced by Ronneberger et al. in 2015 (*92*, *95*), has emerged as a pivotal framework in the field of computer vision. It was initially invented and first used for biomedical image segmentation. The distinctive U-shaped design of U-Net encompasses the encoder, decoder, and the shortcut connections between them. The encoder, or analysis path, extracts features and reduces dimensions through pooling, thereby preserving the foundational features of the image. (Figure 2-7A) The decoder, or synthesis path, involves up-sampling and uses additional convolutional layers to reconstruct spatial details lost during the encoding process. (Figure 2-7B) Noteworthy features include a bottleneck layer that retains high-level semantic information and shortcut connections traverse the entire network, linking corresponding layers of the encoder and decoder, facilitating the retention of detailed information and effectively addressing the issue of gradient vanishing.

doi:10.6342/NTU202402518

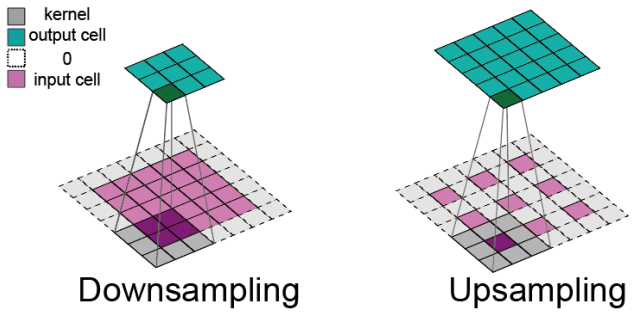


Figure 2-7, Down-sampling and up-sampling.

The goal of CNN is to transform the input and try to find the feature patterns from the previous layer by down-sampling and up-sampling. A down-sampling convolutional neural attempts to compress the input, while an up-sampling tries to expand the input, conversely. For example, (A) Down-sampling is when a 3x3 kernel slides over a 5x5 input to give a 3x3 output. (B) Up-sampling is a kind of transposed convolution, which uses strides 3x3 to increase the output dimension to 5x5.

The U-Net architecture utilized in TAG-SPARK consists of specific steps. In Figure 2-8, downward arrows denote down-sampling achieved through max pooling, with the resulting features highlighted in orange. Conversely, upward arrows indicate up-sampling, followed by interpolation with features from shortcut connections, where previous step images are directly incorporated, and the resulting features are marked in yellow. Throughout this process, convolution with Rectified Linear Unit (ReLU) is applied to generate feature maps, with the values displayed on the colored box indicating the number of feature maps. The final output amalgamates these actions, effectively distinguishing the core characteristics of the input image from unwanted noise. In simpler terms, it conducts down-sampling to identify features, followed by up-sampling to reintegrate features into output layers.

doi:10.6342/NTU202402518

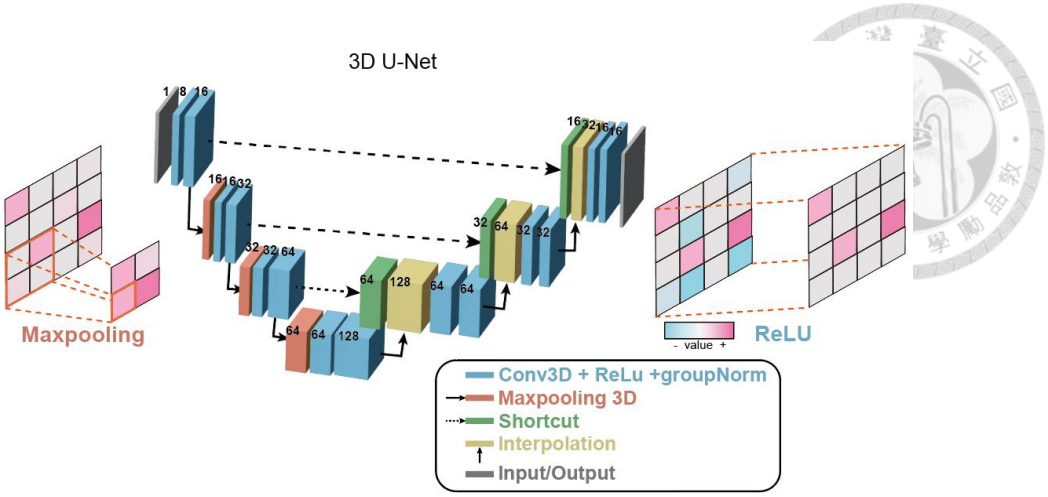


Figure 2-8, 3D U-Net model architecture.

The operation of max pooling and ReLU for convolution.

This architecture has proven to be particularly effective in medical image segmentation tasks, demonstrating its versatility and widespread adoption in diverse applications due to its ability to capture both global context and local details in images.

2.3.4 TAG-lens-based SPAtial redundancy-driven noise Reduction Kernel (TAG-SPARK)

In accordance with the Noise2Noise concept (see **Chapter 2.3.2** for details), our proposed denoising process, TAG-SPARK, was designed to exploit the spatial redundancy inherent in TAG-based image slices for effective noise reduction. TAG-SPARK focuses on utilizing spatial redundancy facilitated by TAG lens volumetric imaging. This leverages dense z-sampling capabilities inherent to TAG lens technology.

To avoid the discontinuity of structure, we utilize oversampling in the axial direction, i.e. spatial redundancy. The raw data of high-speed scanning images achieves 265 layers in z, which spans across extended DOF. Therefore, the gap between each frame is much less than the point spread function in the axial direction (see **Chapter 3.3.1**

doi:10.6342/NTU202402518

for details of system resolution). This capability effectively utilizes adjacent z-layer images that demonstrate similar and continuous structures, with independent noise characteristics that enhance TAG-SPARK.

To utilize the higher-dimensional information (e.g., correlations across frames) inherent in the 3D data stacks for noise reduction, we replaced the original 2D U-Net with a 3D U-Net (Figure 2-8). As Figure 2-9 illustrates, instead of feeding frame pairs individually, we stacked two separate frame stacks (red and yellow frames) as the input and target stacks to the model. Therefore, each set of adjacent z-direction image slices was treated as independent samples that represent the same scene. These pairs served as input-target pairs in the training of a 3D U-Net model.

Due to the 3D nature of these frame stacks, we utilize a 3D U-Net for their processing, which allows us to exploit the information across different frames with, for example, the 3x3x3 convolution kernels. Compared to the original Noise2Noise 2D U-Net model which takes 34 layers, the 3D U-Net model we used has fewer layers (25). Additionally, the sequence and the number of different types of layers vary between the two models. Furthermore, the loss function we used is the average of the L1 (MSE) and L2 (MAE) norms to retain the advantages of both. (see **Chapter 2.3.1** for details) The L1 norm enhances resilience to outliers, while the L2 norm promotes convergence during training. In contrast, Noise2Noise employs the L2 norm solely.

doi:10.6342/NTU202402518

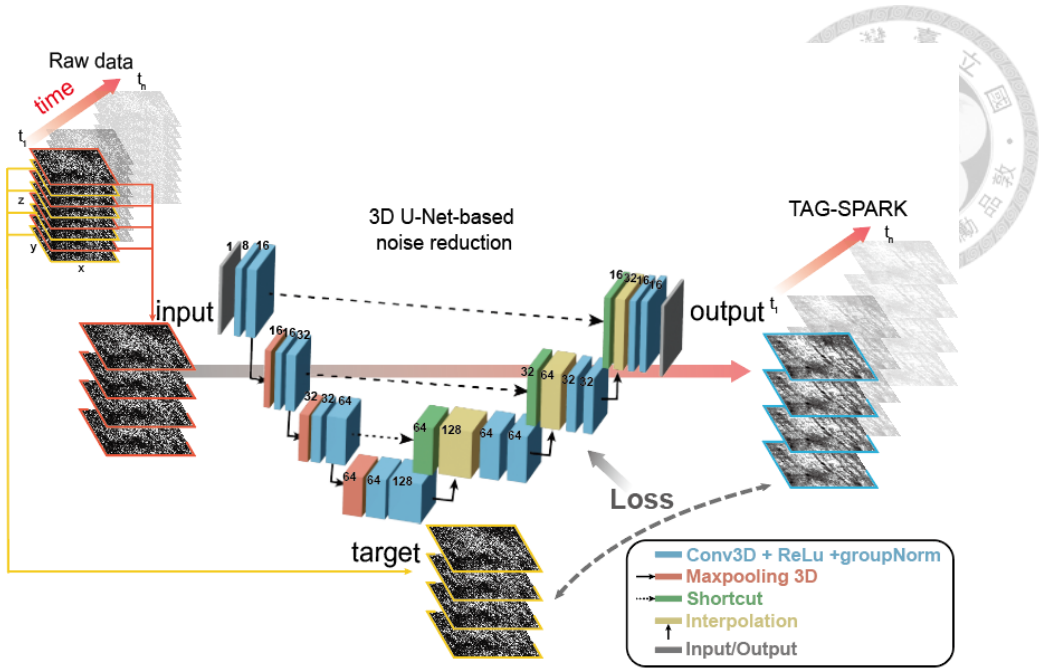


Figure 2-9, Schematic of 3D U-Net model in TAG-SPARK denoising algorithms.

TAG-SPARK utilized a 3D U-net model. Multiple raw volumetric images at various time points are used for model training, where even layers (indicated by a yellow frame) and odd layers (indicated by a red frame) of each sub-stack are extracted to create two 3D tiles: one for input and the other for target volumes, which are used for training the network. The loss function is defined by comparing the output sub-stack (highlighted by a blue frame) with the target.

In **Chapter 4.3**, we will discuss how to set the parameters for training the dataset acquired from the 2P-TAG system. Although TAG-SPARK relies on optimizing the structure using spatially adjacent layers, it also significantly improves image quality (see **Chapter 5** for details).

doi:10.6342/NTU202402518

Chapter 3 Optical system design and sample preparation

This chapter will cover elements related to experiments, including optical setup, system performance, calculation and simulation of theoretical and practical values, as well as relevant methods for sample preparation. We highly recommend thoroughly understanding this chapter's theory before beginning experiments, as it will assist in experiment designs and anticipated results.

3.1 Optical setup

The home-built high-speed two-photon volumetric imaging system (Figure 3-1**,** Table 3-1) is driven by a tunable Ti: Sapphire oscillator (Chameleon Ultra II, Coherent). The excitation wavelength was tuned to 940 nm for two-photon excited fluorescence of green fluorescent protein GCaMP6f. With a variable beam expander (#87-564, Edmund), the laser beam size was adjusted to match the effective aperture of the tunable acoustic gradient lens (TAG lens 2.5β, TAG Optics), which is 4 mm. The laser beam was guided through a lens pair (AC254-060-B-ML & LA1422-B, Thorlabs), and then was relayed to a pair of galvanometric mirrors (6215H, Cambridge Technology) for 2D raster scanning. Through a telecentric scan lens and tube lens pair (SL50-2P2 & TL200-2P2, Thorlabs), the scanning pattern was reflected by a dichroic beam splitter (FF749-SDi01-25x36×3.0, Semrock) and directed onto the back aperture of a 25x objective (XLPLN25XWMP2, Olympus). The emitted two-photon fluorescence was

epi-collected by the same objective, passing through the dichroic beam splitter, and then

doi:10.6342/NTU202402518

focused by an imaging lens (LA4306-A, Thorlabs) onto a photomultiplier tube (PMT) module (H14119-40, Hamamatsu). Two bandpass filters (FF01-520/15-25, Semrock) were placed in front of the PMT to filter out the reflected excitation laser and undesired background signals.

Instrument	Manufacturer	Model number
Ti: Sapphire Laser	Coherent	Chameleon Ultra II
Half-wave plate	Thorlabs	AHWP05M-980
Polarizing beam splitter	Thorlabs	CM1-PBS252
Variable beam expander	Edmund	#87-564, Broadband NIR 1X-3X Research-Grade Variable Beam Expander
TAG lens	TAG Optics	TAG lens 2.5β
4f system, lens 1	Thorlabs	AC254-60-B
4f system, lens 2	Thorlabs	LA1422-B
Galvanometric mirrors	Cambridge Technology	6215HM40B
Scan lens	Thorlabs	SL50-2P2
Tube lens	Thorlabs	TL200-2P2
Dichroic	Semrock	FF749-SDi01-25×36×3.0
Objective lens	Olympus	XLPLN25XWMP2
Image lens	Thorlabs	LA4306-A
Bandpass filter	Semrock	FF01-520/15-25

doi:10.6342/NTU202402518

PMT	Hamamatsu	H14119-40
Mouse tilting stage	NARISHIGE	MAG-2
X/Y stage	OptoSigma	TSD-601S
Z stage	Thorlabs	MGZ30, self-modified
Clinical Microscope	Olympus	BX51
HBO Mercury Microscope Lamp	Olympus	U-LH100HG
Microscope HBO Light Controller	Olympus	BH2-RFL-T2

Table 3-1, List of each instrument with manufacturer and model number.

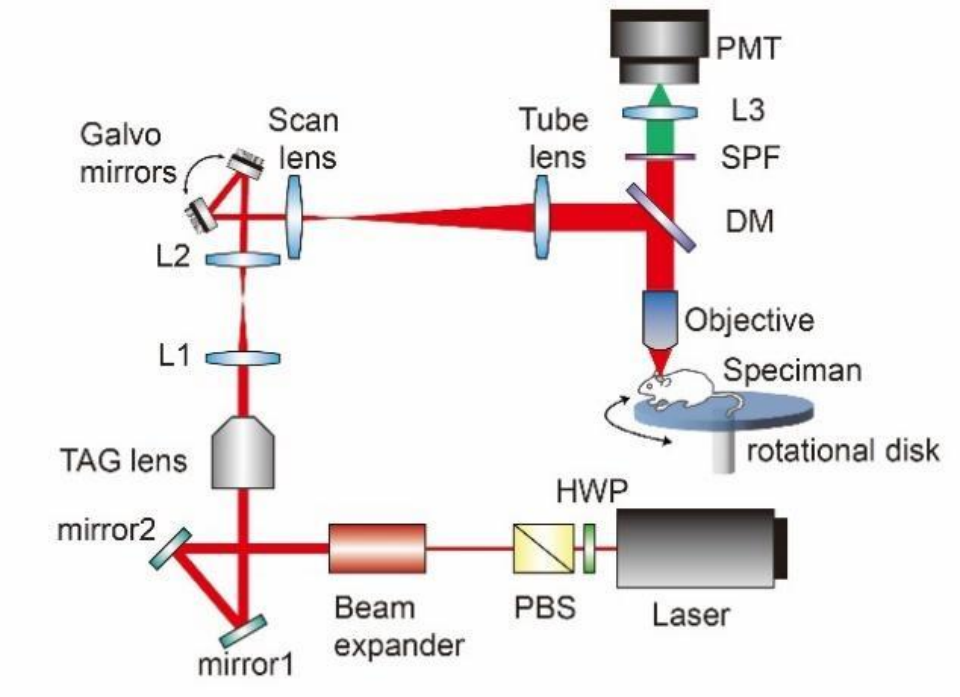


Figure 3-1, Schematic of home-built high-speed two-photon volumetric imaging system.

HWP: half-wave plate; PBS: polarization beam splitter; TAG: tunable acoustic gradient; L: lens; DM: dichroic mirror; SPF: short-pass filter; PMT: photomultiplier tube.

doi:10.6342/NTU202402518

3.2 Sample preparation

In this section, we will introduce the preparation of samples used in the study, which are relevant to the selection of components in the next section and research.

3.2.1 Standard sample preparation - fluorescent microspheres

To characterize the performance of our system, fluorescent microspheres are used to be standard test samples. Measurement of both spatial resolution and DOF under the objective lens is performed with 200 nm diameter yellow-green fluorescent microspheres (F8811, Thermo Fisher Scientific). On the other hand, 10 μm diameter yellow-green fluorescent microspheres (F8836, Thermo Fisher Scientific), with comparable size to mouse neurons, are chosen to characterize volumetric images.

2D sample

200 nm diameter fluorescent microspheres are diluted by deionized (DD) water with a 100-dilution factor. Drip it onto the slide and cover it with a coverslip, then wait for it to dry. Nail polish is applied around the edges of the coverslip to seal it. (Figure 3-2)

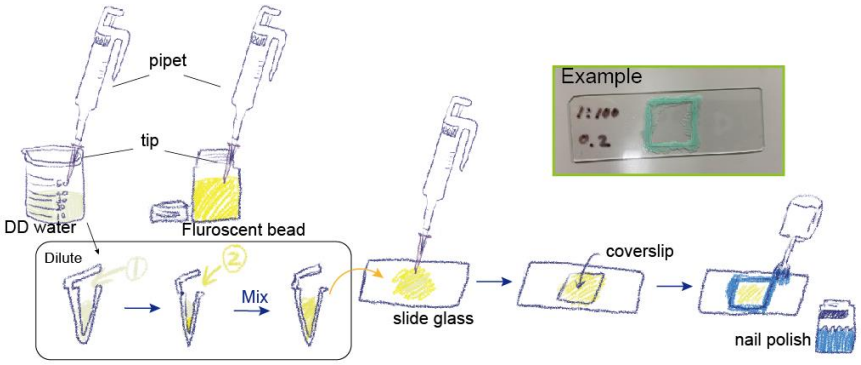


Figure 3-2, Experimental procedure diagram for 2D FLB sample preparation.

3D Sample

To characterize the extended DOF for characterizing volumetric images, a stock of

doi:10.6342/NTU202402518

10 μm diameter fluorescent microspheres is mixed into 0.7% agarose hydrogel with a ratio of 1:3 to mimic mouse neurons. The agarose solution (LE Agarose, refractive index = 1.3329) is heated and stirred until reaching 80℃ and achieving transparency, then cooled to ~ 45℃ to become a jelly-like sample. The diluted fluorescent microspheres are mixed into the agarose hydrogel and placed in the refrigerator for solidification (Figure 3-3).

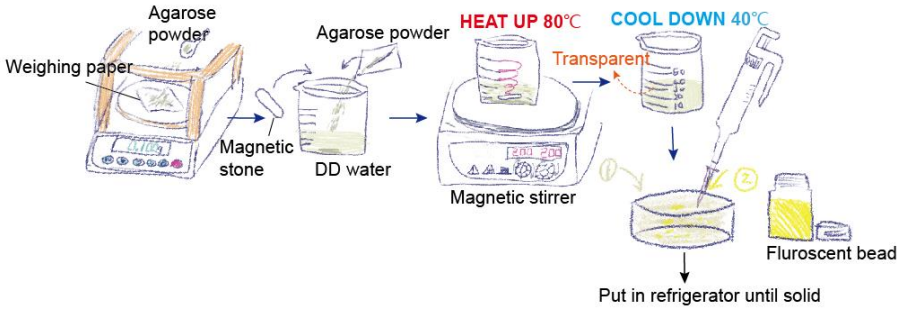


Figure 3-3, Experimental procedure diagram for 3D FLB sample preparation.

3.2.2 Mouse preparation

All surgical and experimental procedures are conducted according to the guidelines for the care and use of laboratory animals approved by the Institutional Animal Care and Use Committees of National Taiwan University (approval number NTU110-EL-00023). The adult male Pcp2-GCaMP6f mice, produced by crossing Pcp2-cre (Jackson Lab. #: 004146) with tTA2-GCamP6 (Jackson Lab. #: 030328) will be employed for the experiment. The mice are housed in standard conditions with a 12-hour light/12-hour dark cycle and provided ad libitum food and water. A surgical procedure is conducted to implant a cranial window on the cerebellar surface. Initially, a 4×4 mm piece of the skull is removed from the midline, 6 mm posterior from bregma. Subsequently, a 5×5 mm cover glass is placed on the cerebellar surface and affixed with glue to the skull. After the glue has healed, dental cement (Super-Bond, Japan) is applied to the skull and the cranial

doi:10.6342/NTU202402518

window's edge. Finally, a head plate (Neurotar, US) is secured atop the cranial window using dental cement. The mouse undergoes a one-week recovery period before engaging in *in vivo* optical imaging experiments (Figure 3-4).

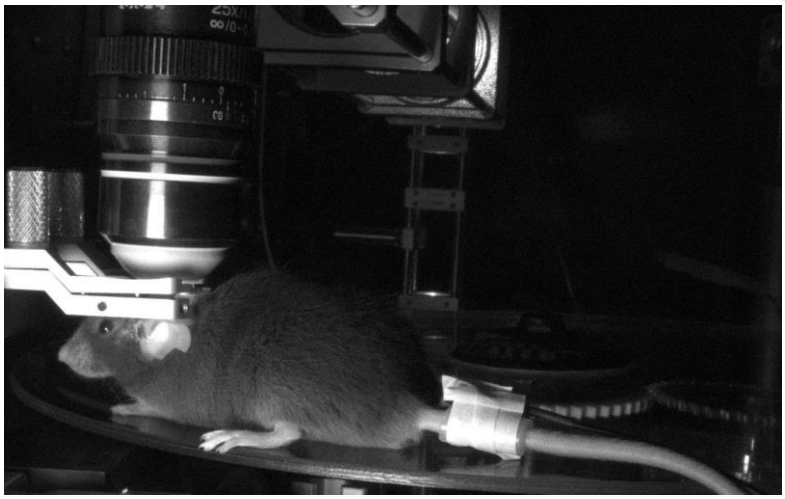


Figure 3-4, Recorded photo of mouse experiment under IR camera[2].

3.3 Theoretical and experimental performance

In this section, the performance of the system is evaluated and discussed from multiple aspects, which can be broadly categorized into optical parameters and acquisition rate. Besides theoretical value, we employed relatively simple samples such as fluorescent beads and *in vitro* mouse brain slices to assess the system's experimental performance

3.3.1 Spatial resolution

2PM employs nonlinear excitation to induce highly localized fluorescence within the sample, usually confined to a thin plane, facilitating high-resolution imaging. The lateral and axial resolution of the imaging system is determined by the illumination point spread function (PSF), IPSF(x,y,z). Unlike confocal microscopy, 2PM requires distinct

doi:10.6342/NTU202402518

formulations for the definition of the focal volume. (*26*)

The theoretical spatial resolution in 2PM can be calculated by

$$FWHM_{xy} = 2\sqrt{ln2} \times \frac{0.32}{\sqrt{2}} \frac{\lambda}{NA},\ NA \leq 0.7$$

$$FWHM_{xy} = 2\sqrt{ln2} \times \frac{0.325}{\sqrt{2}} \frac{\lambda}{NA^{0.91}},\ NA > 0.7$$

$$FWHM_z = 2\sqrt{ln2} \times \frac{1}{\sqrt{2}} \times \frac{0.532\lambda}{n - \sqrt{n^2 - NA^2}}$$

where $FWHM_{xy}$ and $FWHM_z$ are the full width at half maximum of the PSF in the lateral and axial direction, λ is the excitation wavelength, NA is the numerical aperture of the final imaging lens, and n is the refractive index inside the sample. By substituting in the parameters of our system, which are 940 nm for λ, 1.33 for n of the refractive index of the sample (water), and 1.05 for NA of the objective lens (Olympus, XLPLN25XWMP2),

$$FWHM_{xy} = 2\sqrt{ln2} \times \frac{0.325}{\sqrt{2}} \times \frac{940 \times 10^{-9}}{1.05^{0.91}} = 344\ nm$$

$$FWHM_z = 2\sqrt{ln2} \times \frac{1}{\sqrt{2}} \times \frac{0.532 \times 940 \times 10^{-9}}{1.33 - \sqrt{1.33^2 - 1.05^2}} = 1146.3\ nm$$

However, in practical implementation, as the beam size didn't fill the whole aperture (*96*), the NA of the objective part will be influenced by the effective aperture size of the laser beam. **(**Figure 3-5) For this, we measure the laser beam size with a CMOS camera at the back aperture of the objective lens, the FWHM of the Gaussian profile shows 6.3 mm, and the corresponding to the effective NA will become,

doi:10.6342/NTU202402518

$$FWHM = D = 6.3\ mm$$

$$NA = n \times \sin\theta\ , where\ n\ is\ refractive\ index$$

$$2 \times tan\theta = \frac{D}{f}, where\ f\ is\ the\ focal\ length\ of\ objective\ lens$$

$$NA_{eff} = n \times sin(tan^{-1}(D/2f)) = 1.33 \times sin(tan^{-1}(\frac{6.3}{2\times 7.2})) = 0.53$$

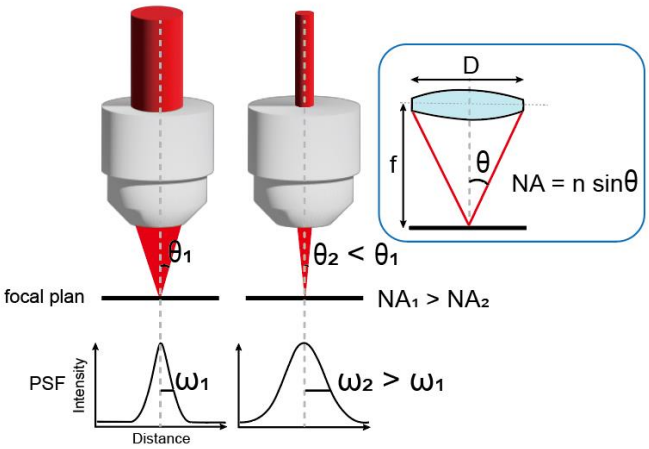


Figure 3-5, Effect of laser beam size with NA.

When the objective lens is filled to meet the full aperture requirement, the effective NA of the objective lens increases, resulting in higher image resolution.

According to the calculated effective NA of the objective part is 0.53, we get:

$$FWHM_{xy} = 2\sqrt{ln2} \times \frac{0.32}{\sqrt{2}} \frac{940 \times 10^{-9}}{0.53} = 668.2\ nm$$

$$FWHM_z = 2\sqrt{ln2} \times \frac{1}{\sqrt{2}} \times \frac{0.532 \times 940 \times 10^{-9}}{1.33 - \sqrt{1.33^2 - 0.53^2}} = 5.3\ \mu m$$

To evaluate the performance of the two-photon system, we choose fluorescent microspheres as our sample. Due to the convolution of the PSF with the object, the size

doi:10.6342/NTU202402518

of fluorescent microspheres should be smaller than three times the theoretical PSF, then the observed spot size will be roughly the same as the theoretical PSF (Figure 3-6, see **Chapter 4.1.2** for details).

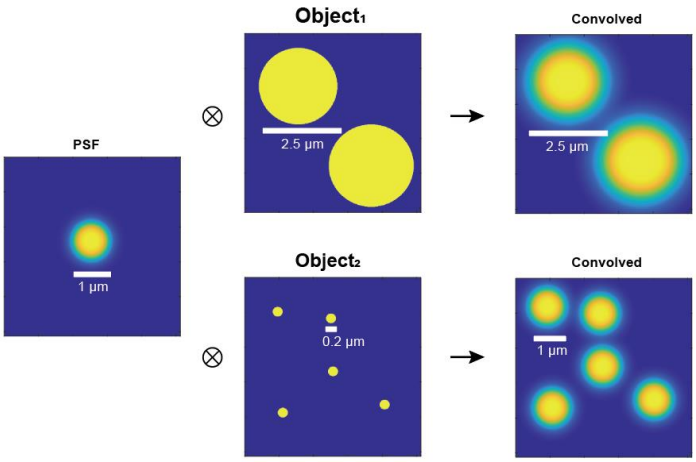


Figure 3-6, The image acquisition results from the convolution of objects with the PSF.

Therefore, 200 nm FLB was selected for analysis, and the results are shown in Figure 3-7. We average a total of 85 measurements for lateral and axial profiles, and the resulting resolution is 673.9 ± 0.52 nm and 4.96 ± 0.28 μm respectively. The error compared to the theoretical value is 1.12% and 6.6% for lateral and axial measurement. This result indicates that both the lateral and axial resolution in our system is sufficient to resolve single neurons (~10 μm) in the mice brain.

$$FWHM_{xy} = 673.9 \pm 0.52\, nm$$

$$FWHM_{z} = 4.96 \pm 0.28\, \mu m$$

doi:10.6342/NTU202402518

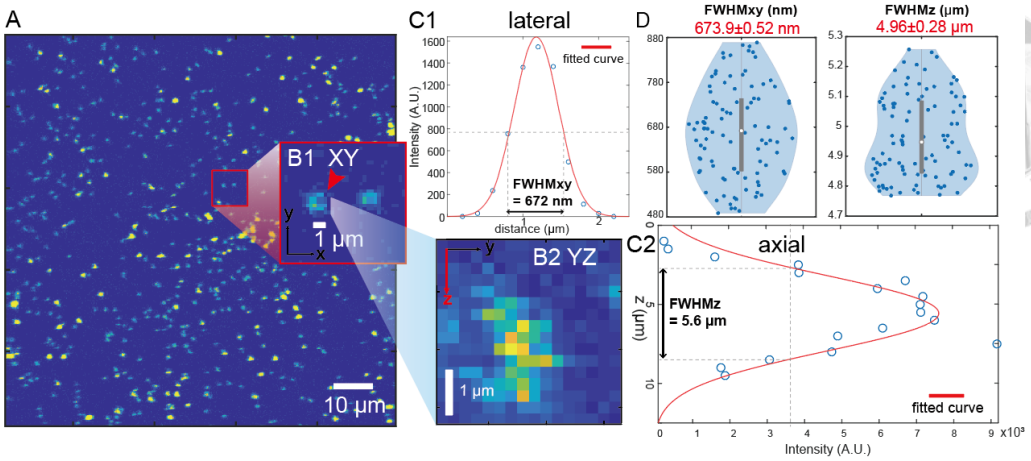


Figure 3-7, Optical resolution of images of standard 200 nm beads.

(A) A two-dimensional (2D) image of 200-nm beads. (B1) The XY image shows the red square region zoomed in from (A), with a single bead indicated by a red arrowhead. (B2) The YZ image of the same bead. (C1) The signal intensity profile shows that the lateral size of the bead is 672 nm. (C2) The signal profile demonstrates the axial size is 5.6 μm. (D) Average of 85 beads for FWHM values of lateral and axial profiles, indicating that the resolution is 673.9 ± 0.52 nm and 4.96 ± 0.28 μm, respectively.

3.3.2 Field of view (FOV)

The FOV of a microscope determines the size of the imaged area, which depends on the magnification and the field number (F.N.) of the objective lens. F.N. is referred to as the diaphragm size of an objective lens in mm unit which defines the maximum range of observable area. (Figure 3-8A) If the observed sample area is larger than the objective FOV, considerable vignetting would appear at the corners of the image. To calculate the maximum practical FOV, FN should be divided by the objective magnification,

$$\text{Field of View} = \frac{\textit{Field Number}}{\textit{Objective Magnification}}$$

$$F.N. = 18\ mm$$

doi:10.6342/NTU202402518

$$FOV = \frac{18}{25} = 720\ \mu m$$

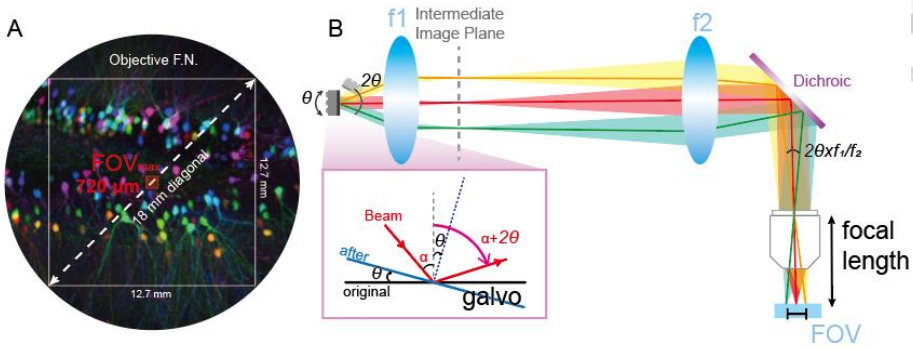


Figure 3-8, Theoretical FOV of the system.

(A) FN of the objective lens, practical FOV. (B) FOV is determined by Galvo angle within the system.

In practical implementation, the FOV of images also depends on the angle of Galvo scanners (Figure 3-8B), scan lens, tube lens, and the focal length of the objective (*97*). The initial deflection angle range of the beam, denoted as α, is reduced by the lens system before the objective. This reduction in angle effectively controls the beam's spread across the FOV. The size of FOV is proportional to the ratio of the focal length of the scan lens (f_{SL}) to that of the tube lens (f_{TL}), expressed as,

$$FOV = 2 \times tan(\theta_{Galvo} \times \frac{f_{SL}}{f_{TL}}) \times f_{Obj}$$

For our system, scanning range setting for Galvo mirrors at total range θ_{Galvo}= 6.5 degrees ($\pm 3.25\ degree$) (Figure 3-9), considering the effect of the lens system and objective (Olympus XLPLN25XWMP2), the corresponding FOV is,

$$FOV = 2 \times tan\left(6.5 \times \frac{50}{200}\right) \times 7.2 = 408\ \mu m$$

doi:10.6342/NTU202402518

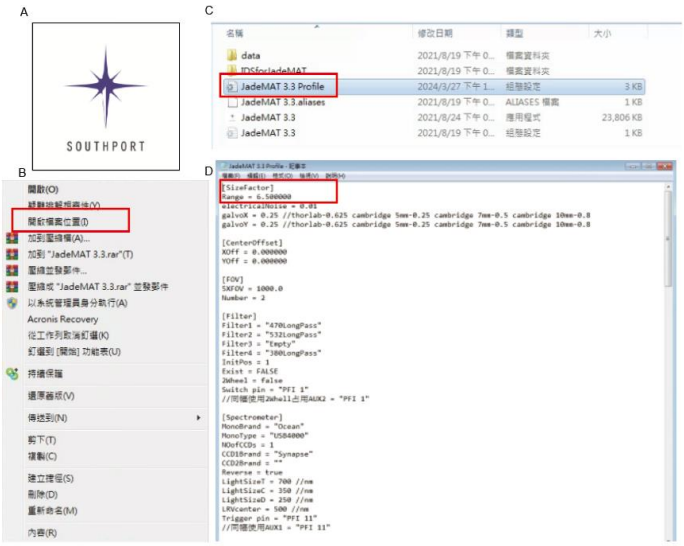


Figure 3-9, Control panel of "JadeMAT3.1", low-speed scanning software.

In the practical implementation (Figure 3-10), we used a micrometric ruler to measure the FOV through the reflection of visible light. The obtained result was 419.8 μm with an error of 2.8%.

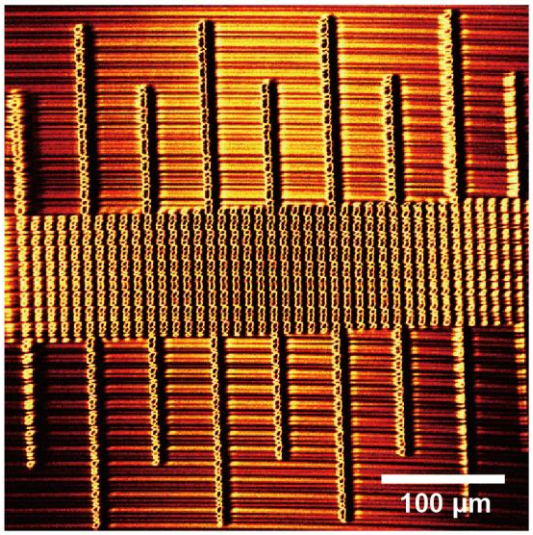


Figure 3-10 FOV measurement with a micrometric ruler.

Glass slide with a micrometric ruler. The total is 1 mm divided into 100 units, that is 10 μm per division.

Scale bar: 100 μm

doi:10.6342/NTU202402518

While the FOV size depends on the ratio $\frac{f_{SL}}{f_{TL}}$, the beam diameter at the back aperture of the objective depends on $\frac{f_{TL}}{f_{SL}}$. This results in a trade-off between high NA (resolution) or a large FOV.

3.3.3 Extension depth of field (DOF)

In imaging systems, the DOF determines the distance between the closest and furthest point capable of generating an image. For laser scanning microscopy, the focus is typically adjusted by moving the sample stage. Although DOF isn't limited, imaging speed is limited by the speed of the stage. In our system, we increase imaging speed with the use of a TAG lens (see **Chapter 2.2** for details), which poses additional limitations on the DOF of the system.

Expected DOF

As mentioned in **Chapter 2.2**, the DOF of TAG lens can be calculated by,

$$DOF = n \times f_{obj}^2 \times \delta_{total} \times \left(\frac{f_2}{f_1}\right)^2 \qquad \text{Eq. 2-9}$$

$$= n \times f_{obj}^2 \times \delta_{total} \times \frac{1}{M^2}$$

Additionally, we can use the ABCD matrix to simulate the theoretical DOF (Simulation code in GitHub[3]). In optics, the complex beam parameter, often represented by the symbol q, is a complex number that defines the characteristics of a Gaussian beam at a specific position z along the beam's axis. The determination of the length of the extended DOF is commonly employed in ray transfer matrix analysis, a method used to

doi:10.6342/NTU202402518

compute the properties of the beam at various points as it undergoes propagation through an optical system.

For the Gaussian beam, the complex beam parameter q is defined as,

$$\frac{1}{q(z)} = \frac{1}{R(z)} - \frac{i\lambda}{\pi n \omega(z)^2} \quad \text{Eq. 3-1}$$

where $R(z)$ is the curvature radius of the wavefront, λ is the wavelength, n is the index of refraction, $\omega(z)$ is the beam radius (defined at $1/e^2$ intensity). Given the initial beam parameter, q_0, one can use the ray transfer matrix of an optical system, $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, to find the resulting beam parameter, q_1, after the beam has traversed the system,

$$q_1 = \frac{Aq_0 + B}{Cq_0 + D}$$

$$\rightarrow \frac{1}{q_1} = \frac{C + \frac{D}{q_0}}{A + \frac{B}{q_0}} \quad \text{Eq. 3-2}$$

$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, where A and D are the magnification and inverse magnification, respectively. B and C are related to the focal length of the lens or the distance of free space propagation. For free space propagation over a distance d, the matrix is, $\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$. For a thin lens with focal length f, the matrix is, $\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}$. For a TAG lens with optical power δ, the matrix is, $\begin{bmatrix} 1 & 0 \\ -\delta & 1 \end{bmatrix}$. For light passing through media with different refractive

doi:10.6342/NTU202402518

indices, from n_1 to n_2, the matrix is, $\begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}$.

To calculate the propagation of a light ray through multiple optical elements then enter the sample, we multiply the matrices for each element in the order of their occurrence. As illustrated, the total ray transfer matrix M_{total} ($M=\begin{bmatrix} A & B \\ C & D \end{bmatrix}$) can be described as,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_{obj}} & 1 \end{bmatrix} \times \begin{bmatrix} 1 & d_1 \\ 0 & 1 \end{bmatrix} \times \ldots \times \begin{bmatrix} 1 & 0 \\ -\delta & 1 \end{bmatrix} \quad \text{Eq. 3-3}$$

where $f_{objective} = \frac{f_{tube}}{M}$, f_{tube} is the tube lens which designed to pair with the objective lens, M is the magnification of the objective lens.

$$M_{total} = M_z M_n M_{obj} M_{d1} M_{TL} M_{d2} M_{SL} M_{d3} M_{L1} M_{d4} M_{L2} M_{d5} M_{TAG}$$

where $M_n$is refractive index transfer matrix, M_{d_n} is free space transfer matrices, and $M_{obj}, M_{TL}, M_{SL}, M_{L1}, M_{L2}, M_{TAG}$ are respectively the transfer matrices of the objective lens, tube lens, scan lens, relay lens 1, relay lens 2 and TAG lens, each approximated in the thin lens transfer matrix form. The matrix form of each element is presented below, with each parameter inside listed in Table 3-2.

$$M_n = \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix}, n_1 = 1\ (air), n_2 = 1.33\ (water)$$

$$M_{f_i} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f_i} & 1 \end{bmatrix},\ f_i = f_{obj}, f_{TL}, f_{SL}, f_{L1}, f_{L2}$$

doi:10.6342/NTU202402518

$$M_{d_n} = \begin{bmatrix} 1 & d_n \\ 0 & 1 \end{bmatrix}, n = 1, \ldots 5, z$$

$$M_{TAG} = \begin{bmatrix} 1 & 0 \\ -\delta & 1 \end{bmatrix}$$

With the assumption that the incident beam enters from the back aperture of the TAG lens, the focus condition of this system is that the beam parameter (q) must be purely imaginary (i.e. $\frac{1}{R} = 0$). As the optical power of the TAG lens (δ) changes, the corresponding focal position z also changes. Thus, by inserting the possible optical powers that the TAG lens has, we obtain a range of focal position z which represent the DOF of the system. By substituting all the parameters into the equations above (Eq. 2-9), and the simulation with ABCD matrix (combining Eq. 3-2 & Eq. 3-3) (Figure 3-11), both results show the expected DOF of our system, with optical power at 10 D, is 97 μm (Assume the refractive index of the sample is 1.33).

$$DOF = 1.33 \times 7.2^2 \times 10 \times \left(\frac{60}{40}\frac{50}{200}\right)^2 = 97\ \mu m$$

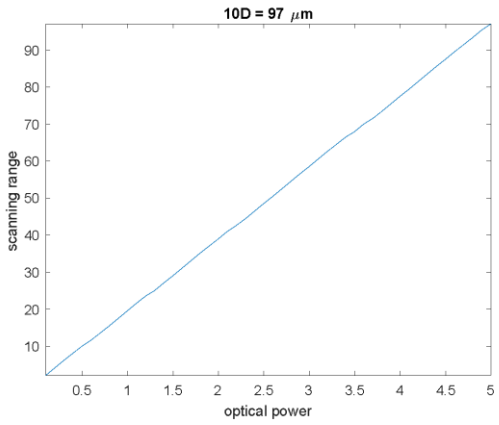


Figure 3-11, Simulation of different optical power and corresponding scanning range.

doi:10.6342/NTU202402518

The focal intensity calculates the FWHM of z about its focal intensity. As shown in Figure 3-12, the axial intensity is shaped like a dumbbell due to the resonant sinusoidal trend of the TAG lens. The focal point stays longer at the endpoints of the focus extension, and the intensity at those positions is greater than that of the center. Additionally, Figure 3-12A considers the effect of the system's axial resolution, which influences the DOF result from 97 μm to 100.9 μm. Let's further consider the impact of the convolution of the beam with the object, as shown in Figure 3-12B. It simulates the result of the DOF for the system after convolution with a 10 μm FLB (representing the size of the cell body), and the DOF becomes 103.1 μm. In the following comparison, we assume the theoretical DOF is 103.1 μm.

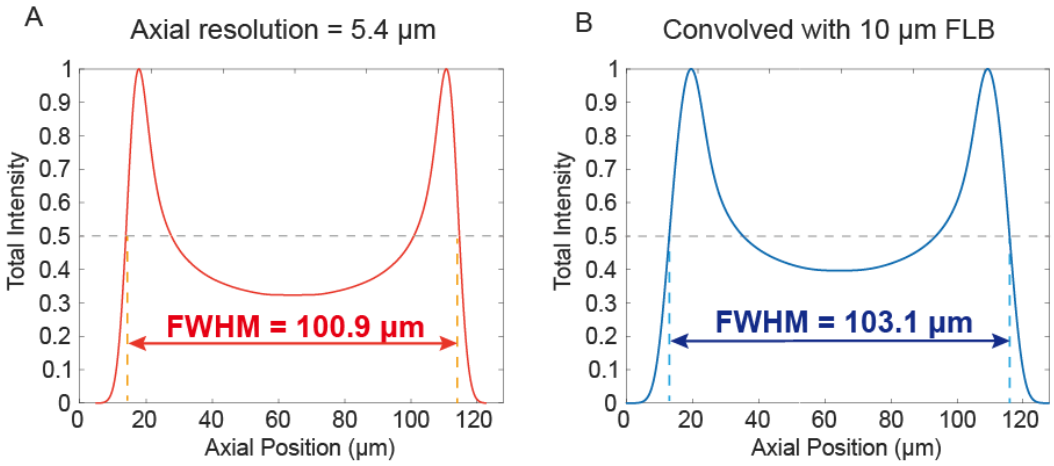


Figure 3-12, Simulation of the intensity distribution of Gaussian beam passes through different axial positions.

(A) FWHM of the Gaussian beam matches the FWHM of the axial resolution. (see Chapter 3.3.1 for details) (B) FWHM of the Gaussian beam matches the result after convolution with the 10 μm FLB.

Parameter	Description	Value
λ	Excitation wavelength	940 nm
ω	Incident beam radius	6.3 mm

doi:10.6342/NTU202402518

R	Radius of curvature of incident wave front	∞
$d5$	Distance between TAG and relay lens 2	60 mm
$d4$	Distance between relay lens 2 and relay lens 1	100 mm
$d3$	Distance between relay lens 1 and scan lens	90 mm
$d2$	Distance between scan lens and tube lens	250 mm
$d1$	Distance between tube lens and objective	200 mm
δ	Optical power of TAG lens	$-5 \sim +5\ m^{-1}$ (D)
f_{L2}	Focal length of relay lens 2	60 mm
f_{L1}	Focal length of relay lens 1	40 mm
f_{SL}	Focal length of scan lens	50 mm
f_{TL}	Focal length of tube lens	200 mm
f_{obj}	Focal length of objective	7.2 mm
f_{tube}	Focal length of tube lens paired to objective	180 mm

Table 3-2, Parameters for DOF calculation.

Practical DOF

To verify the practical DOF of the system, we conducted three different experiments. Different samples and methods are used to correctly access the multiple errors that may occur when calculating the DOF.

The first experiment is scanning 10 μm FLB with low-speed scanning and calculating the DOF via the extension of the image in z-direction. The second experiment

doi:10.6342/NTU202402518

compares low-speed scanning and high-speed scanning results of the same FLB sample and measures the extension of the TAG lens with correlation map analysis. The third experiment is the same correlation map analysis with a mouse brain slice sample.

We set the resonance frequency of the TAG lens at 188k Hz and its voltage amplitude at 40% which corresponds to the optical power of $\pm 5\ D$. The scan frame rate is 1 Hz for the first experiment, and the volume rate is 11 Hz for the second and third experiment. We use Olympus XLPLN25XWMP2 immersion objective lens which corresponds to a theoretical FOV of $400 \times 400 \times 97\ \mu m^3$.

10 μm FLB with low-speed scanning

We experimented with 10 μm FLB to measure the intensity profile along the axial direction. In the experiment, we turn on the TAG lens collect a stack image with low-speed scanning, and change the focus plane with the translation stage. Then, we calculate the DOF with the intensity profile and average a total of 10 measurements for the axial profile, the resulting DOF is $136 \pm 6.23\ \mu m$. (Figure 3-13) The error compared to the theoretical value (103.1 μm) is about 32% for the axial measurement.

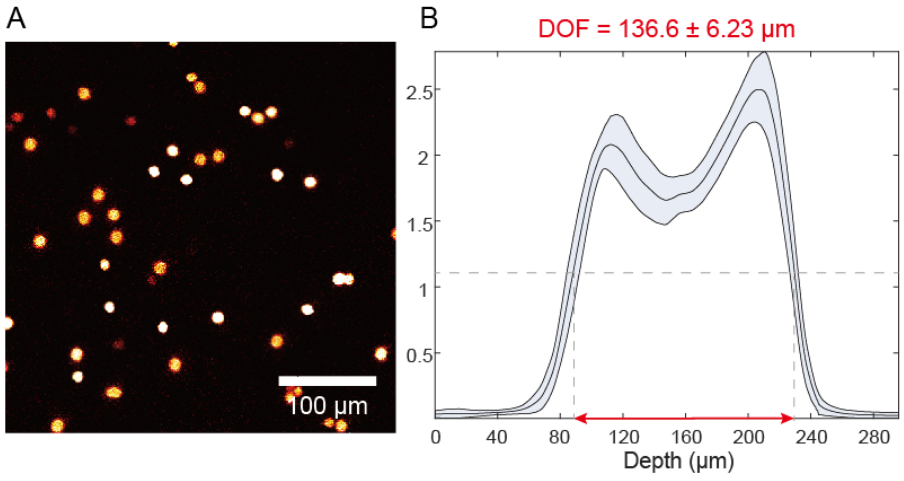


Figure 3-13, Experimental extended DOF.

doi:10.6342/NTU202402518

(A) 3D 10 μm FLB sample with z-projection of a volume with a depth of about 137 μm. Scale bar: 100 μm (B) Collect the z-projection images with the translation stage. The intensity profile along the z direction.

Since the TAG lens operates on a sinusoidal track mechanism, we observe that it exhibits a profile resembling that of a dumbbell, like the simulated appearance (Figure 3-12).

Another interesting finding is that the dumbbell-like PSF is not uniform. The TAG lens creates convex and concave effects, which change the spot size at the objective lens aperture and consequently affect the intensity of the focal points. When it acts as a concave lens (red beam light in Figure 3-14), the light spot size entering the objective is divergent, making the light spot at the objective lens aperture larger, leading to an increase in the effective NA, and thus, the light is focused with higher intensity in deep position (red side, z2). Conversely, with a convex lens, the convergent beam results in lower intensity (blue side, z1).

In our experiment, we obtain complete three-dimensional images by moving the stage of the sample in the z-direction to cover the whole structure. Therefore, when we move the FLB from z1 to z2, we achieve a complete image. As a result, the energy is relatively focused, leading to a slight increase in image intensity.

doi:10.6342/NTU202402518

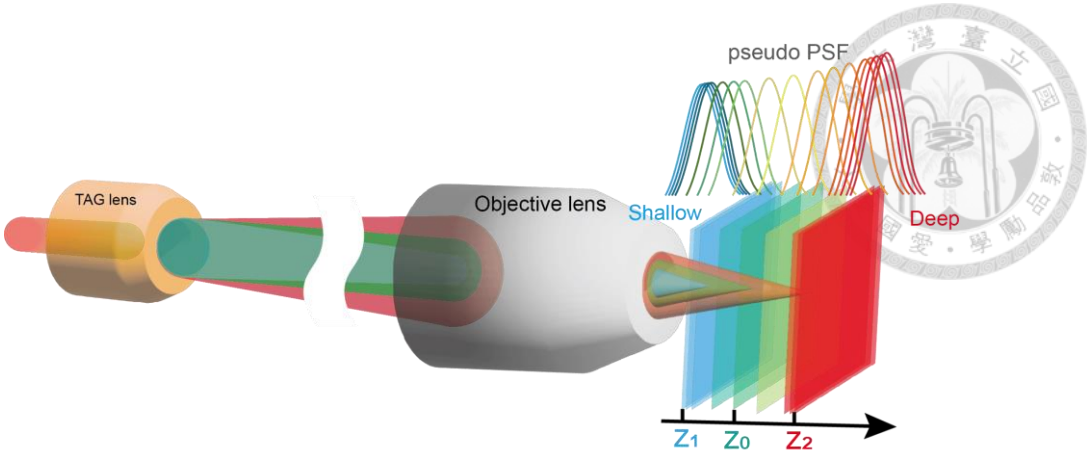


Figure 3-14, Diagram of the intensity of PSF in different axial positions.

TAG lens generates convex and concave effects, which make the beam converge (blue beam) and diverge (red beam). While the focus shifts at different depths in the sample, corresponding to the position z1 to z2 (from blue to red plane), the spot size at the back aperture of the objective will change which consequently affects the NA of the objective, i.e. the intensity of the profile. The deeper it goes, the higher the intensity.

Correlation map analysis

In this study, we will collect low-speed scanning images as structural images, and high-speed scanning images as functional images. To ensure the extended DOF, we will use correlation analysis (see **Chapter 4.2.2** for detailed correlation analysis and code in GitHub[4]) to find the most corresponding region.

We first obtain a volume stack with an accurate z range by scanning with stage. Then, we use the TAG lens to obtain another volume stack in a similar region. Assuming the stage scanned stack has an accurate range, we can obtain the experimental DOF by comparing these two volume stacks with correlation map analysis. This analysis method is precise for volumetric image reconstruction. Before confirming the corresponding physical DOF, we will first reconstruct the high-speed volumetric images with a total of 265 voxels in z-direction. Then, we perform correlation analysis to compare the corresponding range of slow-scan images. We use the contour map to illustrate the

doi:10.6342/NTU202402518

correlation coefficient values, yellow indicates higher values, while blue indicates values close to 0. (The color bar shows the corresponding coefficient values.) Throughout this process, to enhance the precision of the correlation calculation, we utilized a 10-second temporal projection to enhance the contrast of fast-scan images.

The following shows the results of two kinds of the samples, fluorescent bead, and mouse brain slice, under the measurement and analysis.

10 μm FLB

Figure 3-15A reveals that the first and last layers of the fast scan images correspond most closely to specific layers of the slow scan images, respectively. Figure 3-15B demonstrates the center layer of the fast scan image (purple) overlaps significantly with the most closely corresponding layer from the slow scan (green), and the correlation coefficient value is up to 0.83. Figure 3-15C demonstrates corresponding layers between slow scan images and the 10-second time-projected fast scan images exhibit a sinusoidal trend (white dash line), consistent with the principle of TAG lens following a temporally sinusoidal trend as described in **Chapter 2.2.**

Therefore, it determined that the first and last layers of the fast scan images within the physical DOF span 116 μm. Compared with the theoretical value of 103.1 μm, it exhibits an error of approximately 12.5%.

doi:10.6342/NTU202402518

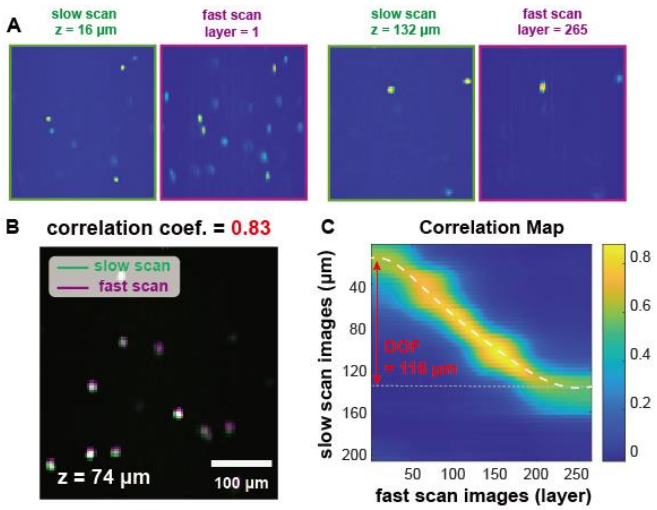


Figure 3-15, Correlation analysis for DOF with 10 μm FLB.

(A) Slow scan image and corresponding 10s projection of fast scan image of FLB. (B) Two images overlap to check the correlation coefficient value. (C) Correlation contour map of the fast scan image and slow scan image.

However, to further confirm the effective DOF of the TAG lens, we propose repeating the same experiment with mouse brain slices. Because of the structure of FLB, even out of the focal plane, the circular fluorescent beads still form a blurred circle shape, which may lead to inaccuracies in comparing high- and low-speed scan images. For example (Figure 3-15C), in the fast scan, layers 140-170 correspond to the depth of 100-120 in the slow scan, both showing significantly high correlation coefficient values.

Cleared mouse brain slice

In order to simulate the *in vivo* mouse experiment, we used cleared mouse brain slices to do the experiment and the correlation analysis to confirm the extended DOF. In Figure 3-16, we repeat the analysis step and the result presents the sine wave trend as well. From the result, we know the extended range of the mouse brain slice is 116 μm, which compared with the theoretical value (103.1 μm) has an error of around 12.5%. The findings align with those of the 10 μm FLB examination, and Figure 3-16C vividly

doi:10.6342/NTU202402518

depicts a sinusoidal trend in the correlation map.

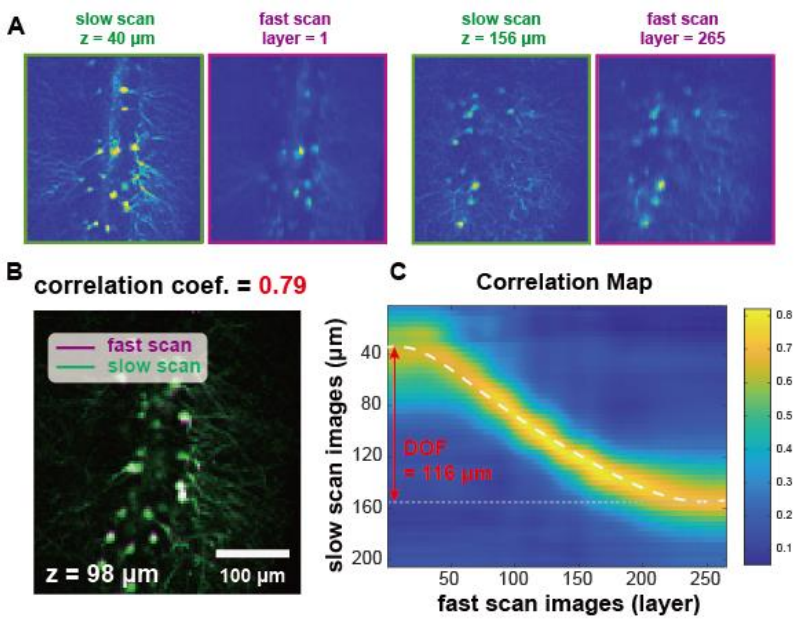


Figure 3-16, Correlation analysis for DOF with *in vitro* mouse brain slice.

(A) Slow scan image and corresponding 10s projection of fast scan image of *in vitro* mouse brain slice. (B) Two images overlap to check the correlation coefficient value. (C) Correlation contour map of the fast scan image and slow scan image.

Based on the analysis of brain slice experiments, a clear sinusoidal trend in the correlation map between high- and low-speed scans, is consistent with the operation method of the TAG lens. The structural differences in brain slices across different layers make the comparison between the two scanning modes easier, thereby reducing the probability of repeated matches. Therefore, before simulating live experiments, brain slices should be used to estimate the DOF.

Discussion

The interesting thing is that the measured DOF varies in different driving amplitudes as **Chapter 2.2** mentioned the mechanism of TAG lens. Figure 3-17 shows evidence of a

doi:10.6342/NTU202402518

linear dependence on the driving amplitude (or optical power, Table 3-3) and DOF.

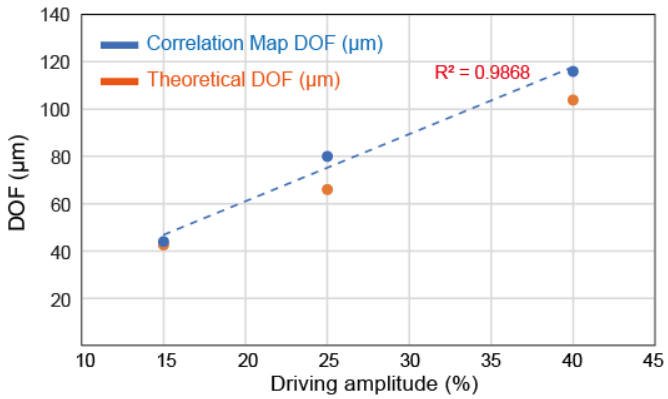


Figure 3-17, DOF with respect to different driving amplitude.

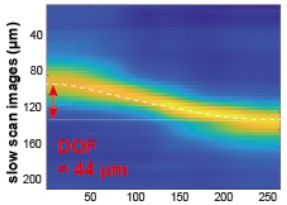
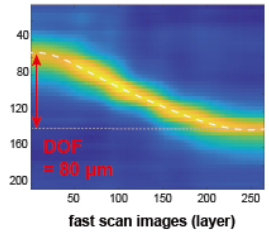
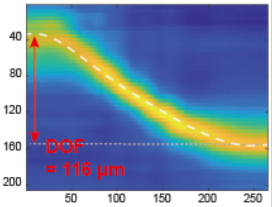
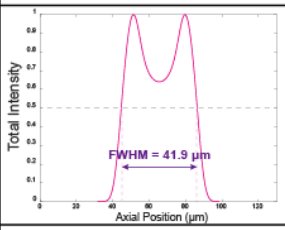
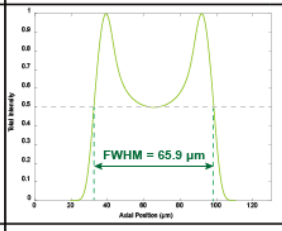
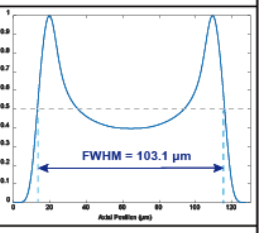
DOF (μm)			
Amplitude seting (%)	15	25	40
Optical power (D)	3.75	6.25	10
Correlation map DOF (μm)			
	44	80	116
Theoretical DOF (μm)			
	41.9	65.9	103.1
Error (%)	5	21.4	12.5

Table 3-3, DOF with respect to different optical power.

While the three analytical methods mentioned above are generally reliable, hidden

doi:10.6342/NTU202402518

issues should be taken into consideration. For example, the DOF analysis from the slow scan on FLB may encounter issues due to the convolution of the PSF and the FLB sample, which could potentially affect the DOF calculations. For the correlation map analysis, although the FLB and mouse brain slice examination come with similar results, as mentioned earlier, there is concern about identifying corresponding layers because the structure of the FLB does not change much within layers. Therefore, the DOF measured using fast scan images from cleared mouse brain slices serves as the optimal template for our future experiments on *in vivo* mouse brain imaging.

However, the measured DOF is approximately 20% larger than the theoretical value. We suspect that this discrepancy may result from the convolution effect of the beam, as illustrated in Figure 3-12, which simulates the Gaussian beam passing through the DOF.

Optical power	Theoretical value (convolved)	3D 10um FLB				Mouse brain	
		Low-speed scan / axial intensity		High-speed scan / correlation analysis			
		Experiment	Error	Experiment	Error	Experiment	Error
10 D	103.1 μm	136 μm	32%	116 μm	12.5%	116 μm	12.5%

Table 3-4, Different DOFs result from different experiments.

In summary, with correlation analysis, the extended DOF of the TAG lens corresponds to low-speed scanning images. The low-speed scanning image plays a crucial role in providing structural information for comparison with corresponding regions in the high-speed scan images. Figure 3-18 shows the 3D structure of the mouse brain slice. The slow-scanning process was achieved with a size of 400x400×300 μm^3

doi:10.6342/NTU202402518

and a voxel dwelling time of 5 μs (Figure 3-18A). Since it was sequentially acquired, several minutes were needed to obtain the whole volume. The same volume was obtained by our high-speed volumetric scanning process (Figure 3-18B), which utilized a TAG lens for rapid z-modulation at a resonant frequency of 188 kHz. The resulting volume dimensions were 400 x 400 x 116 μm^3 with a voxel dwelling time of 0.02 μs, and to enhance the contrast we performed a 10-second time average of images. Compared with these two scanning modes, it possesses the same neuronal structure, verifying our system's capability of acquiring and reconstructing correct 3D volumetric images.

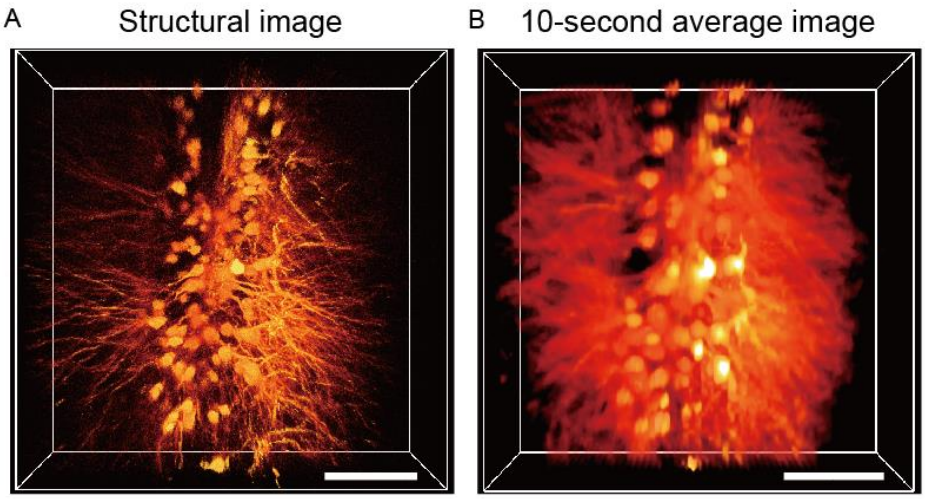


Figure 3-18, Reconstructed volumetric images of mouse brain sample.

The volumetric images were captured by: (A) Low-speed scanning image as the structural image (voxel dwelling time of 5 μs). (B) High-speed volumetric image with a 10-second average. (voxel dwelling time of 0.02 μs). Scale bar:100μm.

However, when we applied this technique to *in vivo* mouse experiments, the tradeoff of the TAG lens should be considered. For the TAG-lens-based high-speed volumetric system, the axial scanning speed and DOF are primarily limited by the TAG lens parameters, such as resonance frequency and optical power. As **Chapter 2.3** mentioned,

doi:10.6342/NTU202402518

due to the principle of the TAG lens, the higher driving frequency offers longer DOF extension but reduces the effective NA, resulting in lower spatial contrast.(*54*) In the experiment, we should carefully consider this trade-off to meet critical requirements, including the need for adequate temporal resolution to capture neuronal spikes, spatial contrast for resolving sub-cellular structures, and sufficient DOF to encompass the three-dimensional neuronal architecture.

3.3.4 Imaging volume rate

To complete an axial scan under each voxel, the TAG lens oscillation frequency was set to twice the resonant frequency at $f_{TAG} = 188k\ Hz$. The number of voxels in the z-direction is limited by the acquisition rate of DAQ cards. In our system, the acquisition rate of DAQ card is set to 100 MHz, which is the maximum speed of DAQ card,

$$voxel_z = \frac{DAQ\ Acquisition\ Rate}{2 \times Resonant\ Frequency} = \frac{100 \times 10^6}{2 \times 188 \times 10^3} = \frac{531}{2} = 265\ pixels$$

Therefore, the number of voxels in the z direction must be less than 265 pixels. Consequently, we will achieve axial scanning rate and voxel dwelling time,

$$axial\ scanning\ rate\ = \frac{1}{188 \times 10^3} = 5\ \mu s$$

$$voxel\ dwelling\ time = \frac{1}{voxel_z \times Resonant\ Frequency}$$

$$= \frac{1}{265 \times 188 \times 10^3} = 0.02\ \mu s$$

On the other hand, the number of voxels in the y direction is limited by galvo response time [Cambridge Technology Inc., “62XXH Series Galvanometer Scanners”]

doi:10.6342/NTU202402518

and TAG lens resonant frequency. Since the TAG lens is continuously scanning, there must be enough pixels in the y direction so that the TAG lens completes a scanning cycle within a single pixel. In terms of specifications of galvo mirrors, the response time for scanning angles from full angle to minimum angle is 850 microseconds for 20 degrees and 270 microseconds for 0.1 degrees. With a linear approach, the response frequency corresponding to 6.5 degrees is 2190 Hz. The lower limit of the number of voxels is,

$$y_{px} \geq \frac{188\times10^3}{2190} \cong 86 \text{ pixels}$$

There is no limitation on the number of voxels in the x direction. We can calculate the imaging volume rate by selecting an appropriate pixel number that satisfies the limitations listed above. For example, an image with $128 \times 128 \times 265$ voxels in three dimensions will have a volume rate of

$$volume\ rate = \frac{188 \times 10^3}{128 \times 128} = 11.5\ Hz$$

doi:10.6342/NTU202402518

Chapter 4 Image processing and analysis

This chapter is separated into two main parts: image processing and image analysis. For image processing, a diverse range of methodologies is devoted to various tasks such as denoising, enhancement, segmentation, feature extraction, and classification. For image analysis, we will discuss methods such as clustering, quantifying image quality, and statistical analysis. These methodologies are crucial in research concerning the variation in group responses and the enhancement of image quality.

4.1 Image processing

In this section, some useful image processing methods will be introduced, including how to reconstruct high-speed volumetric images from 2D files, basic refining techniques to resolve the blur and motion artifact problems due to the system, and automatic segmentation to speed up the labeling and quantification analysis.

4.1.1 High-speed volumetric image reconstruction

In our high-speed scanning system, the output file type is TDMS. To recover the 3D images, converting the image type is necessary. After converting TDMS files into multiple yz-sectioned TIFF images, the reconstruction of the 3D images will be processed with MATLAB (Reconstruction code in GitHub[5]). Depending on the data acquisition and galvo scanning setting, not only does the sorting sequence of images need to be arranged, but different types of scanning artifacts also need to be accounted for. The scanning artifacts include calibration of lateral (XY plane) images due to galvo scanning, configuring the interval of physical z dimension, and time correction between pixels.

doi:10.6342/NTU202402518

Sorting images to 3D volume

In this setup, we capture yz images at different x positions with TAG lens, then recombine them into xyz volume. During the scanning process, the number of pixels in the y direction is 128, and 265 in the z direction, which corresponds to a resonant frequency of 188kHz.

Suppose our reconstruction volume size is 128x128x265 voxels, the scanning direction is depicted in Figure 4-1. In actuality, our system collects one yz image per period of the galvo mirror, meaning that each image is a combination of two mirror images, mirroring each other and differing in the x direction by 1 pixel. When reconstructing the volume, we divide these images into two sets, flip the images, and insert them in the correct position. We repeat this process until all 2D TIFF images are reconstructed, resulting in the complete xyzt 4D data.

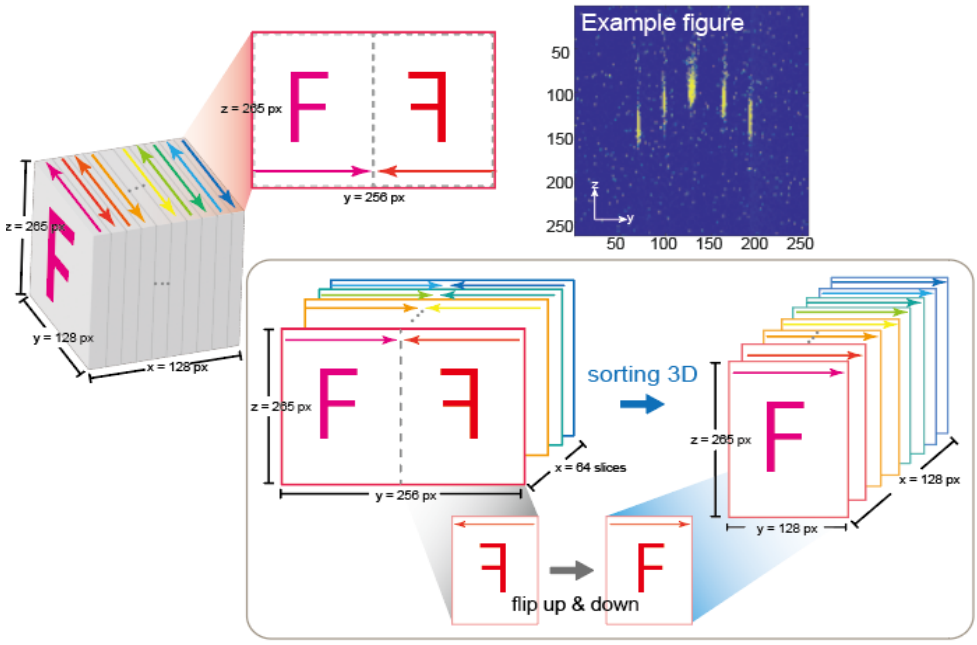


Figure 4-1, Reconstruction of high-speed volumetric images from 2D images.

Reconstruct one frame

doi:10.6342/NTU202402518

After reconstruction, we encounter a discontinuity issue between columns on each xy plane (Figure 4-2), which arises because of the galvo scanner's acceleration. In the lateral dimensions of the image, the galvo sequentially scans the sample at each sampling point to acquire fluorescence data. As the galvo approaches the boundary of the region of interest, it needs to decelerate to switch to the next column. Therefore, at the edges of the fast axis, the galvo scanner undergoes acceleration and deceleration processes (Figure 4-2). This results in non-uniform scanning intervals, leading to nonlinear stretching distortion of the actual image along the fast axis edges, i.e. the boundary of the y-axis. Additionally, if there is a horizontal displacement between the y-axis, it causes aliasing artifacts resembling jagged edges in the image.

Therefore, we need to calibrate for distortions on the xy-plane to reconstruct images unaffected by edge deformations. To address this, we can adjust parameters in MATLAB to fine-tune the degree of distortion in odd and even columns, as well as the starting position of the alignment, enabling effective correction. Once these coefficients are obtained, they are applied to the interpolation process of the original images to eliminate the aliasing artifacts generated by the system. We correlate the processed images with low-speed scanning images, to ensure the artifact is removed.

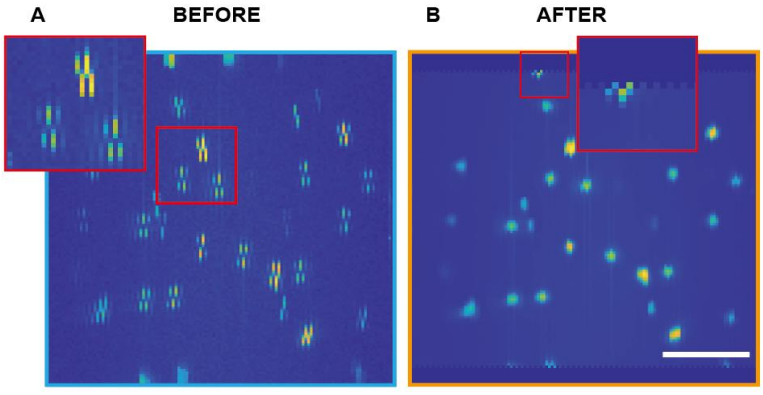


Figure 4-2, y-axis calibration of high-speed volumetric image reconstruction.

(A) Before calibration, there are aliasing artifacts with adjacent columns in the image. (B) After removing

doi:10.6342/NTU202402518

the stretching distortion, there are jagged edges in the boundary of the image. Scale bar: 100 μm

Configuring the interval of z

Due to the rapid focusing capabilities of the TAG lens, the spacing between sampling points along the z-direction varies in the sinusoidal wave track. As a result, in the acquired high-speed volumetric images, the physical spacing between each z-layer image is unequal (Figure 4-3). This results in the top and bottom parts being more densely packed, while the middle region is sparser. When reconstituting physical spacing, the top and bottom would have a greater number of images available for interpolation to do the z-direction resampling. Therefore, the top and bottom of the images have higher SNR than the middle region.

The interval of z-direction sampling also affects the SNR of the reconstructed image. The larger the sampling spacing, the more images are included in the interpolation process, and the consequential SNR becomes higher. With this in mind, it is important when defining the DOF to follow the method mentioned in **Chapter 3.3.3** to not overextend, so there are ample images to provide high SNR images after image reconfiguration.

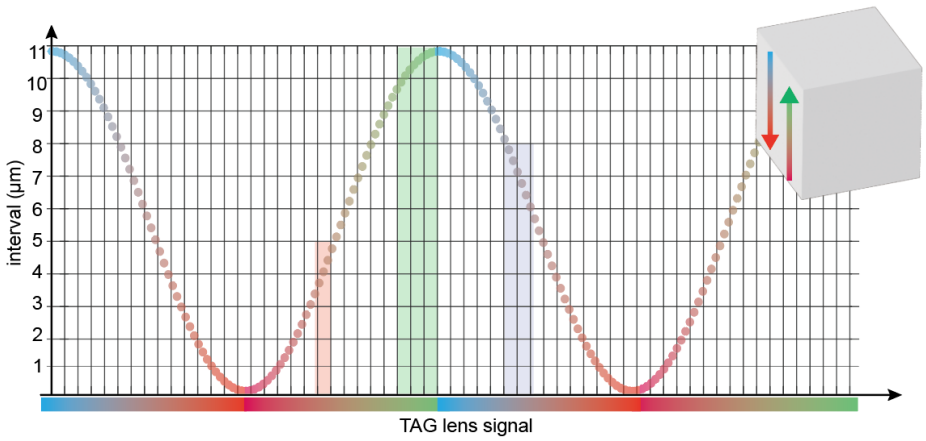


Figure 4-3, Schematic of the sampling strategy.

doi:10.6342/NTU202402518

The trajectory of the focus transmission from top to bottom and back to top (blue to red and red to green).

Following the results of **Chapter 3.2.3**, we reconstruct the volume images with interpolation based on the physical unit. Here we set the z voxel size 2 μm to enhance the image SNR and repeat the correlation map analysis. In Figure 4-4, the trend of the correlation contour map becomes a linear shape after interpolation. The resulting DOF is 116 μm, which is the same as the previous result (Figure 3-16).

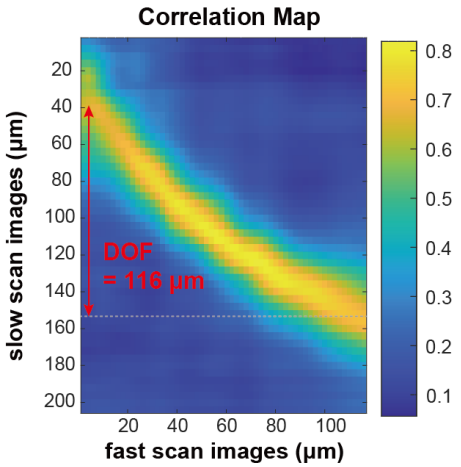


Figure 4-4, Correlation analysis for DOF with micrometers unit.

Time correction

When processing functional volumetric images, we assume that all voxels are accurately acquired at the beginning of the acquisition time for the respective volumetric image. However, due to the finite scanning speed, each voxel in a volumetric image is sampled at slightly different time points (Figure 4-5). To account for these delays between voxels, we use linear interpolation to assign a time t_i for each voxel in the volumetric image,

doi:10.6342/NTU202402518

$$y_j(\tau_i) = \frac{(\tau_i - t_{(i-1)j})x_j(t_{ij}) + (t_{ij} - \tau_i)x_j(t_{(i-1)j})}{t_{ij} - t_{(i-1)j}}, i = 1,2, \ldots, N-1$$

where i is the sequence of volume, there are total N volumes in the image. j is the sequence of xy sampling voxels, $j = 1,2, \ldots, 128 \times 128$. $x_j(t_{ij})$ means the original intensity of j voxel at time series t_{ij}, and $y_j(\tau_i)$ is corrected intensity.

Our goal is to correct the voxels from acquired time points (t_{ij}) into the reference time points (τ).

$$t_{ij} = i \times T + j \times t_{voxel\ dwelling\ time}$$

$$\tau(i) = \tau_i = i \times T, \qquad i = 1,2, \ldots, N-1,$$

where T represents the acquisition time of each functional volume, $t_{voxel\ dwelling\ time}$ represents the voxel dwelling time. For example, the resonance frequency at 188 kHz, would take $0.02\ \mu s$ per voxel.

After utilizing linear interpolation to achieve time correction, each voxel in the volume is back to its reference time point. Nevertheless, if the sampling rate is much faster than cellular reactions, such as the experiments for SCN observation, the evaluation of computer performance and biological information may proceed without this calibration function.

doi:10.6342/NTU202402518

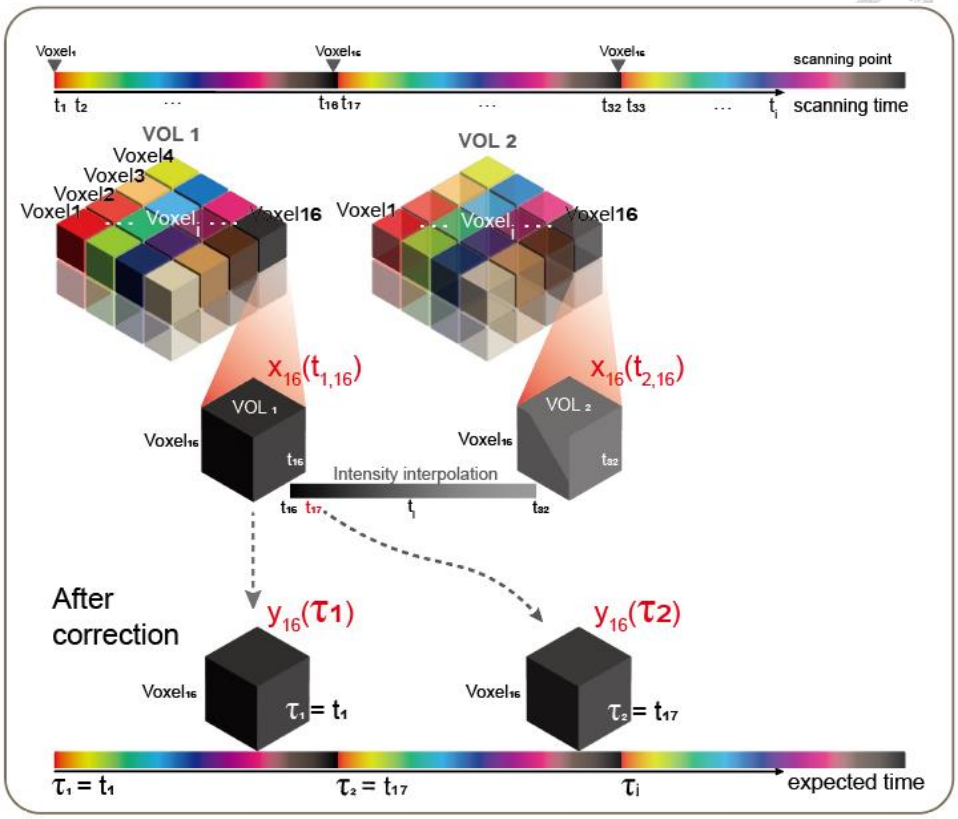


Figure 4-5, Voxel-dependent time correction for high-speed scanning images.

4.1.2 Image processing: Brightness adjustment, deconvolution, and motion correction

Brightness adjustment- Gaussian filter

When there's a significant disparity in intensity distribution across the image, it becomes challenging to observe details in both bright and dark regions simultaneously. Hence, we employ a normalization technique on the image to redistribute intensity values. In image processing, normalization refers to adjusting the range of pixel intensity values, which is also known as contrast stretching or histogram. Here, we'll introduce one useful method of brightness adjustment, which involves applying a Gaussian filter to the image. We directly utilize the Gaussian blur filter available in ImageJ and apply it to the original images. Then, divide the original image with the filtered image to achieve the results.

doi:10.6342/NTU202402518

(Figure 4-6)

$$X_{norm} = \frac{X}{Gaussian\ filter \otimes X}$$

Because Gaussian filters remove high-frequency components, thereby reducing the overall contrast of the image. The result, X_{norm} often yields a smoother brightness distribution, leading to more uniform contrast across the image. (Figure 4-6) For the Gaussian filter, σ represents the standard deviation of the Gaussian distribution, which determines the blur level of the filter. A larger σ value results in a more blurred effect, while a smaller one produces a sharper effect. (Figure 4-6C)

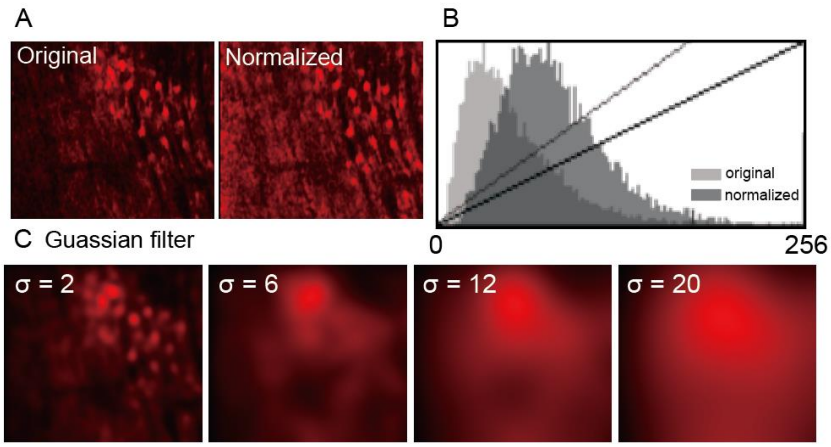


Figure 4-6, The effect of Gaussian filter with different sigma values.

(A) Original and normalized image after applying a Gaussian filter. (B) Histograms of images before (light gray) and after (dark gray) normalization. (C) Different effects were observed with varying sigma (σ).

Deconvolution

In microscopy, light passing through the sample's optical path causes spreading or diffraction, resulting in blur due to the microscope's finite point spread function (PSF) and optical aberrations. PSF represents the system's response to light, affecting how a

doi:10.6342/NTU202402518

point in the specimen is imaged. (Figure 4-7) Convolution impacts image quality, especially resolution, critical for studying small biological structures.

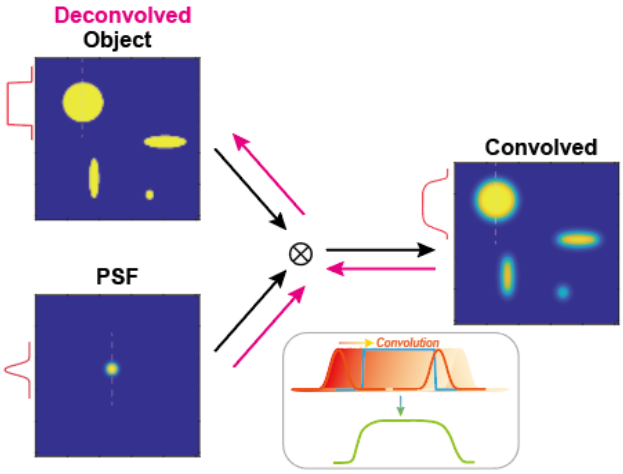


Figure 4-7, Diagram of objective convolved and deconvolved with PSF.

The black arrow indicates the convolution process, while the pink arrow indicates the deconvolution process. The red line represents the intensity profile at the side.

Deconvolution algorithms reverse image distortions, aiming to estimate the original sample appearance by removing blur and aberrations. This is crucial for high-resolution and high-contrast microscopy. To achieve deconvolution, it is necessary to figure out the point spread function of the optical system. We can estimate the theoretical PSF based on a mathematical model or use algorithms to derive the optimal PSF solution. However, the actual PSF of the system may be different due to system aberration, shot noise, and other practical errors. Therefore, it is generally advised to experimentally measure the PSF, which involves capturing images of fluorescent beads that are one-third smaller than the theoretical value. (see **Chapter 3.2.1** for details). Crucially, the PSF lacks perfect uniformity since it fluctuates based on imaging location, depth, and even over time. Conventional deconvolution methods typically assume a consistent PSF across the entire image and fail to obtain satisfactory results. Ongoing

doi:10.6342/NTU202402518

research is dedicated to developing methods for more precise estimation and accommodation of PSF variations (*98-100*).

Deconvolution algorithms in microscopy are categorized into inverse filters and iterative methods. (Table 4-1) Inverse filters rapidly produce clear images by reversing blurring using the PSF, but they are noise-sensitive and prone to artifacts. Iterative algorithms, like the Richardson-Lucy algorithm, repeatedly estimate the original image by comparing it with the blurred image, updating the estimate through iterations. They outperform inverse filters in deblurring but are computationally intensive.

Inverse Filters	**Iterative Algorithms**
Wiener Deconvolution	Richard-Lucy
Naïve Inverse Filtering	Jansson-Van Cittert
Linear Least Squares	Landweber
Tikhonov Filtering	Nonlinear Tikhonov-Miller

Table 4-1, Two types of deconvolution algorithms.

Richardson-Lucy algorithm

The Richardson-Lucy algorithm, a type of iterative deconvolution, refines the initial estimate of the true image iteratively by estimating the blur kernel and minimizing the difference between the observed and estimated images through convolution and adjustment. This process continues until convergence is achieved or for a set number of iterations.

ImageJ provides the Richard-Lucy deconvolution with the “Deconvolutionlab2” plugin.(*101*) First, a simulated PSF is generated by the “PSF generator”. (Detail in GitHub[6]) This simulated 3D PSF is then configured within the deconvolution. After the deconvolution process, the optimized result is obtained. In Figure 4-8C, the profile

doi:10.6342/NTU202402518

analysis (blue and orange curves present before and after deconvolved respectively) reveals that the neuronal cells have become clearer, achieving a higher contrast.

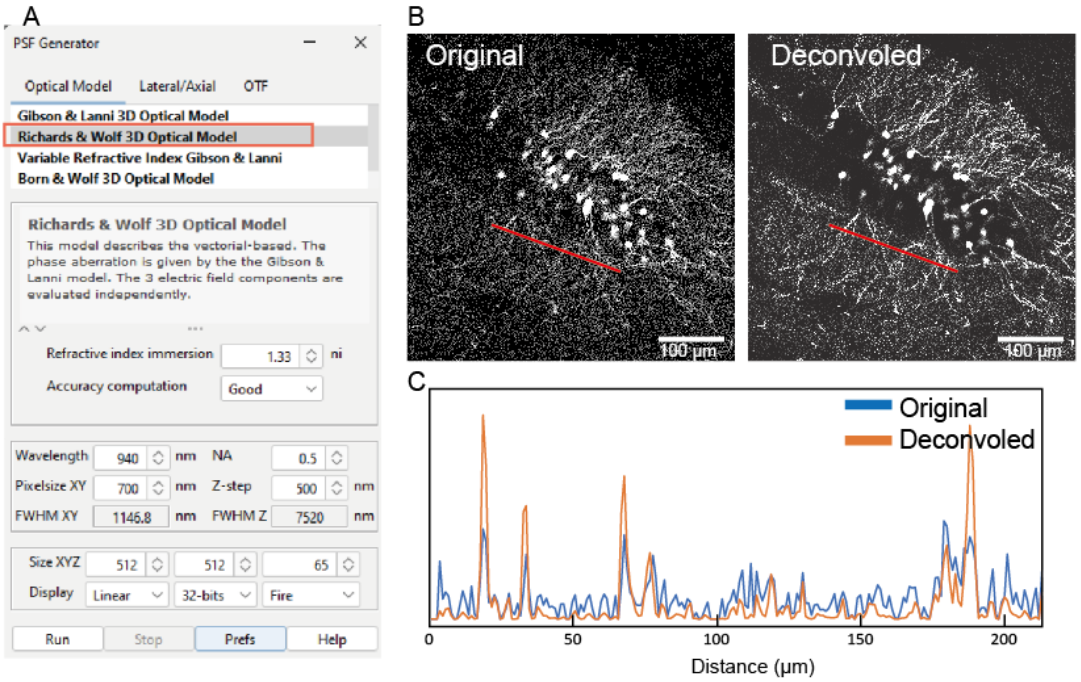


Figure 4-8, Deconvolution with ImageJ.

(A) The panel setting of "PSF Generator" in the ImageJ plugin. (B) The image before and after deconvolution. (C) Intensity profile from the red line plot in (B).

Motion correction

During *in vivo* measurement, tissue movement would introduce unwanted fluctuations in recorded fluorescence signals, resulting in misinterpretation of calcium traces. To address lateral image position variations occurring at each time point due to tissue movement, lateral motion correction was applied to the denoised functional images. Assuming that head movement did not induce shear or scale transformation, a rigid body transformation using an affine transformation was employed. The transformation is expressed as follows:

$$u = T_{motion} w$$

doi:10.6342/NTU202402518

where $w = \begin{bmatrix} w_x & w_y \\ w_z & 1 \end{bmatrix}^T$ and $u = \begin{bmatrix} u_x & u_y \\ u_z & 1 \end{bmatrix}^T$ are the space coordinates of the functional images and the reference images, respectively. T_{motion} is the transformation matrix

$$T_{motion} = T_{motion}(t_x, t_y) = \begin{bmatrix} 1 & 0 & t_x & t_y \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where t_x and t_y specify the displacement along the x- and the y-axis. The parameters in T_{motion} were determined through an iterative optimization method:

$$\hat{T}_{motion} = arg \min_{T_{motion}} C(T_{motion}, V_{functional}, V_{reference})$$

where C denotes the similarity metric to be optimized during motion correction, which is the Mattes mutual information metric. $V_{reference}$ and $V_{functional}$ are the first functional and any one of the remaining functional scans, respectively.

4.1.3 Automated image segmentation and labeling

Image segmentation, a fundamental task in computer vision, involves the partitioning of an image into distinct regions, each representing a specific ROI. This process aids in simplifying the complex visual data of an image and is crucial for object recognition, scene understanding, and image analysis. Semantic segmentation is a pixel-wise classification of a given image. Instance segmentation takes this a step further by not only categorizing pixels but also distinguishing between individual instances of

doi:10.6342/NTU202402518

objects. Various techniques, including color-based segmentation, edge detection, clustering algorithms, and deep learning approaches like the U-net model (see **Chapter 2.3.2** for details), contribute to achieving accurate and efficient segmentation. This capability finds applications across diverse fields, from medical imaging to autonomous vehicles, where extracting meaningful information from images is essential for analysis. Here, two useful software for automatic segmentation will be introduced, one is for circular shape objects, Cellpose (*102-104*), and the other one is for irregular shape objects, ilastik (*105*).

Cellpose[7]

Cellpose is a versatile machine learning-based algorithm designed for segmenting cellular structures, such as cells, nuclei, and other similar objects, in biological images. One of its key innovations lies in its approach to segmenting objects by treating them as sources of diffusion, rather than directly fitting their outlines. This unique strategy allows Cellpose to effectively process images and predict gradient vector fields, facilitating the extraction of object shapes with remarkable precision.

Cellpose is based on U-net architecture but deviates from the conventional U-net design in several aspects. These modifications include (1) additional convolutional layers, novel activation functions, and (2) the integration of residual blocks to address challenges like class imbalance and vanishing gradients. It captures the global context and fine details through (3) carefully crafted skip connections, ensuring accurate segmentation of cellular structures in diverse imaging conditions. Figure 4-9 demonstrates the operation of Cellpose with FLB image. The UI system offers various display modes (Figure 4-9A) for users to confirm whether the trained masks meet expectations. Additionally, it provides over 10 pre-trained models stored in the model zoo (Figure 4-9C) for direct

doi:10.6342/NTU202402518

application. However, in specific cases where users need to retrain models, the system also offers a convenient drawing tool (Figure 4-9B) for segmenting ROIs to train customized models (Figure 4-9D). Integrated with Python to automate big data processing, including segmentation, labeling, spatial distribution analysis, and quantification, without limitations on 2D images.

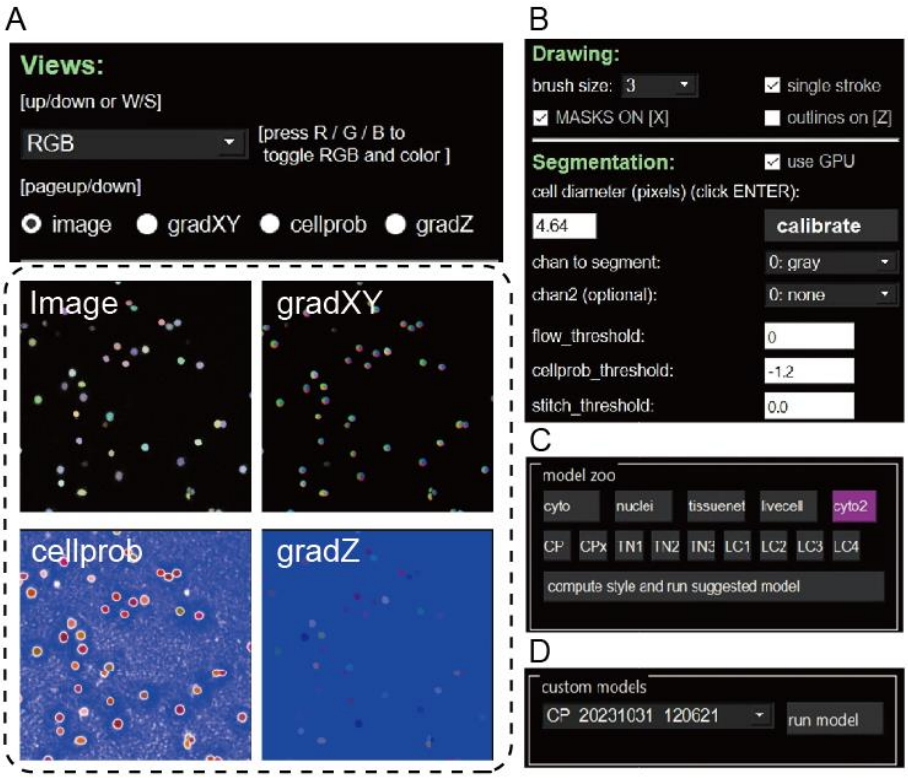


Figure 4-9, The software panel of Cellpose.

Ilastik[8]

Another powerful segmentation tool is Ilastik, which features predefined workflows for image segmentation, object classification, counting, and tracking. Users customize workflows by providing sparse training annotations. It handles data in up to five dimensions (3D, time, channels), with on-demand computational operations for interactive predictions on large datasets. Different from Cellpose mentioned above,

doi:10.6342/NTU202402518

Ilastik is suitable for irregular shapes of bio-image analysis, because it accomplishes semantic segmentation rather than instance segmentation.

Ilastik's pixel classification workflow categorizes pixels by assigning user-defined labels (Figure 4-10, red arrows) into semantic classes (e.g., foreground versus background) rather than individual objects. To identify individual objects, connected components analysis is applied to the pixel classification results. Classification of pixels utilizes image filter outputs as features, with Random Forest as the classifier. Random Forest over other nonlinear classifiers due to its minimal parameterization and robustness to parameter choices (Figure 4-10A). Additionally, its generalization performance makes it particularly suitable for training by non-experts. In Figure 4-10B, the line-shape dendrite image as an example, the result shows that the structure and background are separated with little labeling.

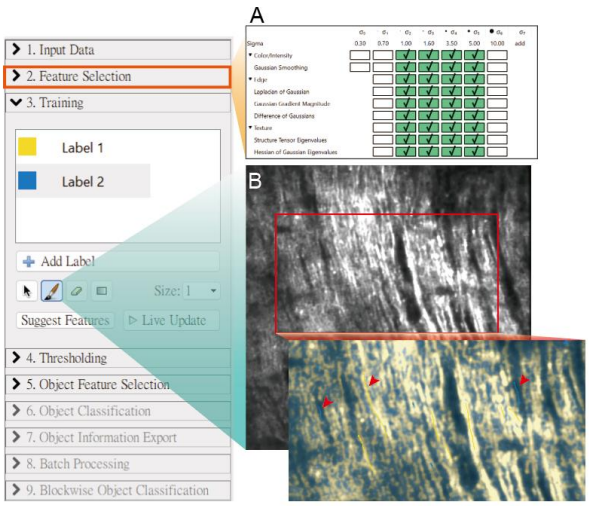


Figure 4-10, The software panel of Ilastik.

doi:10.6342/NTU202402518

4.1.4 Appendix: Identification of individual synapses in *Drosophila* antennal lobe by machine learning-based super-resolution image analysis

In this section, we will introduce how to utilize the powerful tool, Cellpose, for the quantitative analysis of big data and what information was achieved from the analysis. This research collaboration was conducted with Dr. Bi-Chang Chen from Academia Sinica.

Understanding how the nervous system releases neurotransmitters and how organisms respond to different stimuli has emerged as a critical inquiry. The development of super-resolution optical microscopy techniques enables the visualization of detailed synaptic distributions within large-scale neural structures at nanometer-level spatial resolution. However, the massive number of synapses within biological organisms poses a challenging task when combining biological structures with big data for quantitative analysis. Quantitative analysis serves as a fundamental tool in deciphering the intricate workings and functions of the nervous system. This research focuses on revealing the number of neural circuits responsible for *Drosophila*'s antennal lobe using Expansion Light-Sheet Microscopy (ExLSM) (*106-108*). This offers detailed insights into the olfactory neural system of *Drosophila*, with a specific emphasis on examining the quantity and distribution of synapses within the antennal lobe (AL).

Drosophila melanogaster serves as a prominent model organism in neuroscience research, especially the olfactory system. The AL is a pivotal center for olfactory information processing in *Drosophila*, which is pivotal in enabling insects to seek food, evade predators, and locate mating partners. Among numerous neurons, which typically

doi:10.6342/NTU202402518

create tightly packed synaptic connections within the glomeruli. Each glomerulus functions as an independent unit in the processing of olfactory information and encodes specific odors. From the previous research (*109-114*), it was discovered the AL comprises 56 glomeruli, which receive and share signals with other receptors. Therefore, how to form the connection and to link any behavioral activity, has so far been assumed by complete quantitative mapping of the neuron network.

To investigate whether ExM is amenable to the whole AL, we applied the tissue with X10 expansion microscopy protocol. The representative of the segmented image of the expanded AL of the fruit fly is shown. Figure 4-11 shows the whole AL of the fruit fly stained with antibodies against the postsynaptic PSD 95 homolog, disc large (DLG), and the post-synaptic active zone protein, Brp to identify the active zone at neuronal synapses. With ExM, enhanced image resolution through physical expansion is 10 times larger, which allowed us to magnify synapses originally ~200 nm in size to nearly 2 μm.

Consequently, we achieved super-resolution with a voxel size of 0.325 μm. Then, we used Cellpose to perform a quantitative analysis of synapses (Figure 4-11B) and Amira for manual segmentation of glomeruli. After post-processing and calculation, in our preliminary results, we classified a total of 56 glomeruli of one of the antennal lobes and compared left- and right-side distributional differences. (Figure 4-11C) This approach ensures the complete protocol to analyze patterns. Glomerular volume may influence synapse density. Therefore, we assess the volume of each glomerulus on both sides, providing a crucial context for interpreting synapse count data, and further, we calculate the synapse density for each glomerulus. (Figure 4-11D) To delve deeper, we embark on a comprehensive investigation, aiming to definitively map and quantify these

doi:10.6342/NTU202402518

differences across the entire AL.

After creating this complete analysis pipeline, we perform the experiments with different parameters set of fruit flies, such as sexual difference, and food sources, and to observe the neuron distribution. In addition, to ensure statistical robustness, we not only raise the number of fruit fly examinations but also employ appropriate statistical tests to compare synapse counts and densities between different flies. This approach solidifies the scientific foundation of the findings.

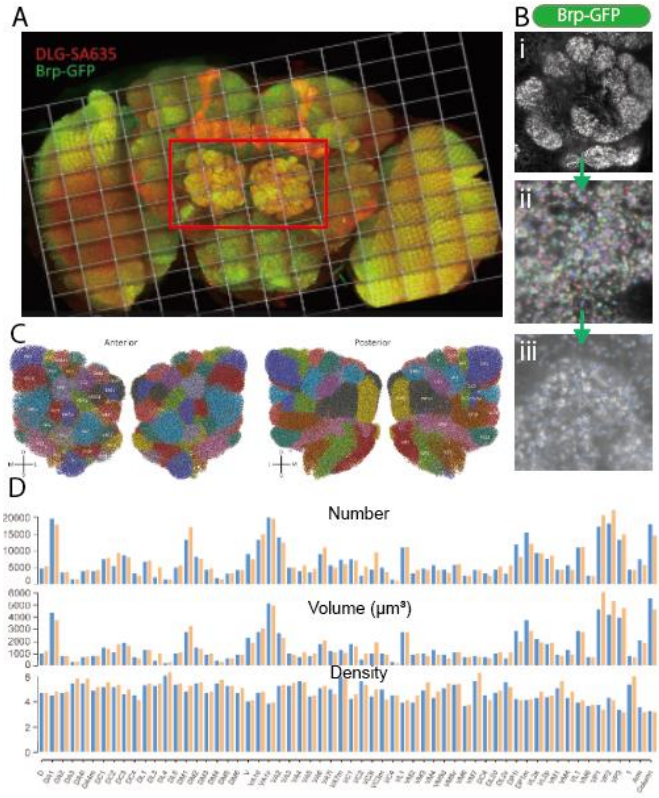


Figure 4-11, Quantitative analysis of synapse in the *Drosophila* antennal lobes.

(A)Whole brain of Drosophila, which was stained with DL and Brp to identify the structure and the active zone at neuronal synapses. (B)Utilize Cellpose to do automatic segmentation of individual post-synapses. (C)56 glomeruli in the AL. (D)Quantitative analysis of the number of synapses, the volume of a single glomerulus, and the density of synapses.

doi:10.6342/NTU202402518

Compared to traditional microscopy techniques which prevent images from resolving details at the level of individual synapses in *Drosophila* due to the diffraction limit. 3D reconstruction of serial EM sections offers the best solution for measuring active synapses in the brain. However, this technique is time-consuming, labor-intensive, and requires specialized expertise (*115*). ExM provides an elegant solution to this challenge. This method physically expands the specimen within a swellable polymer hydrogel, effectively increasing its resolution and allowing for clear visualization of individual synapses using standard microscopes. We will use our new ExM strategies to evaluate the feasibility of using expansion microscopy in the antennal lobe of fruit flies. Integrating with machine-learning-based software, Cellpose, and our analysis pipeline, which allowed for quantification in mere hours. This method overcomes constraints posed by regional samples and potential errors in manual calculations. Moreover, its low time cost enables us to conduct statistical experiments and analyses.

4.2 Image analysis

Images contain rich structural information beyond visual observation. Effective image analysis extracts this information for various applications in fields like medicine, biology, astronomy, and engineering, aiding tasks such as object detection, and quality control. Here, we'll introduce analysis methods including clustering, quantification, and statistics, each with unique advantages and applications tailored to different aspects.

4.2.1 Clustering analysis: PCA, k-means

Cluster analysis uncovers patterns in data, aiding segmentation, feature detection, and data understanding. In this section, two useful clustering methods will be introduced,

doi:10.6342/NTU202402518

each based on different criteria.

Principal Component Analysis (PCA) (Analysis code in GitHub[9])

PCA is a dimensionality reduction technique widely used in statistics, machine learning, and data analysis. PCA transforms the original data into a new coordinate system, finding a new set of orthogonal axes (principal components) along which the data varies the most. (Figure 4-12, red axes) To capture the most important features in the data while discarding less important information. The first principal component accounts for the largest variance in the data, the second principal component (which is orthogonal to the first) captures the second largest variance, and so on.

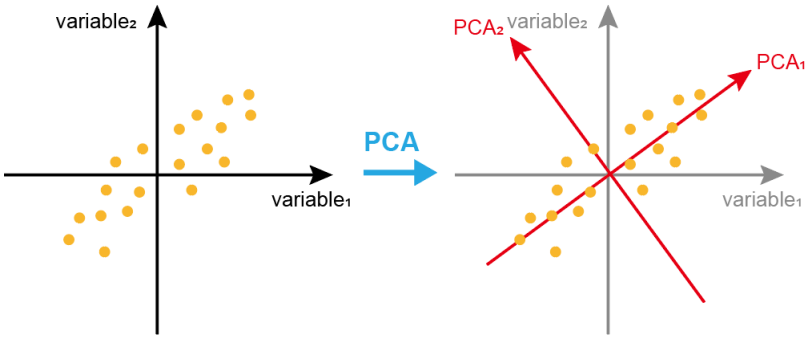


Figure 4-12, Schematic of PCA analysis.

To ensure that all features contribute equally to PCA, the required data is standardized to have a mean of zero and a standard deviation of one across each feature dimension. First, we compute the covariance matrix to obtain information about the relationships between different features and their variances.

$$Cov(X) = \frac{1}{n-1}\left(X - \underline{X}\right)^T\left(X - \underline{X}\right),$$

where X is a $n \times m$ matrix of standardized data, $\underline{X}$ is $m \times 1$ column of average X, n

doi:10.6342/NTU202402518

is the number of observations, m is the number of features.

Second, decompose the covariance matrix into its eigenvectors, $Q_{m\times m}$, and eigenvalues, A. They represent principal components and the amount of variance respectively.

$$Cov(X) = QAQ^T$$

Then, based on their corresponding eigenvalues in descending order to sort the eigenvectors. Finally, PCA projects the original data onto the selected principal components' subspace.

$$Y = XQ_k,$$

where k means the first k principal components, $Y_{n\times k}$ is projected data.

This transformation reduces the data's dimensionality while preserving the maximum amount of variance possible. It is widely used for various purposes, such as noise reduction, feature extraction, and data preprocessing (*41*).

K-means cluster analysis

K-means cluster analysis (also known as c-means clustering) is another popular clustering algorithm used in machine learning and data analysis. *"Birds of a feather flock together"*, is the key concept of k-means clustering. The goal of k-means is to partition a dataset into K clusters, where each data point belongs to the cluster with the nearest mean. (Figure 4-13) It is an unsupervised learning algorithm since it doesn't require labeled data for training.

doi:10.6342/NTU202402518

The algorithm works as follows:

1. Initialization: Choose K initial cluster centroids randomly from the feature space in the dataset. These centroids represent the initial cluster centers.

$$\mu_c^0 \in R^d, c = 1,2, \dots, K$$

2. Assignment: Each data point calculates the Euclidean distance to all K centroids. Assign each data point to the cluster whose centroid is the closest and creates K clusters. Here, the distance can be replaced with other distance formulas, but typically Euclidean distance is the primary choice.

$$S_c^t = \left\{x_i : \left||x_i - \mu_c^t|\right| \leq \left||x_i - \mu_{c*}^t|\right|, \forall i = 1, \dots, n\right\}$$

3. Update Centroids: Recalculate the centroids of each cluster by taking the mean of all data points assigned to that cluster.

$$\mu_c^{t+1} = \frac{sum(S_c^t)}{n_c} = \sum_{i=1}^{n_c} x_i|_{x_i \in S_c^t}$$

4. Repeat Steps 2 and 3: Repeat the assignment and centroid update steps until convergence, i.e., until the centroids no longer change significantly, or a predefined number of iterations is reached.

$$S_c^{t+1} = S_c^t, \forall c = 1, \dots K$$

doi:10.6342/NTU202402518

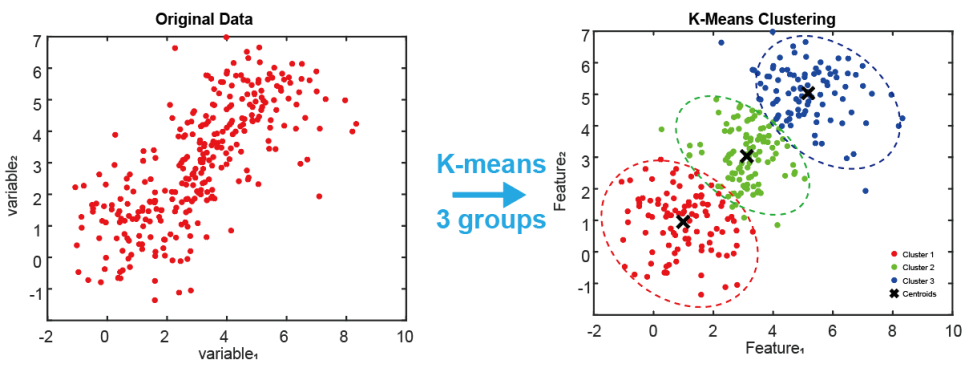


Figure 4-13, Schematic of k-means clustering.

4.2.2 Quantify image enhancement: SNR, PSNR, PCC

After applying image processing, a suitable quantification approach for the improvement of image quality and information is important. Quantitative analysis of image enhancement involves comparing specific image characteristics, such as pixel intensity distributions or feature detection results. Metrics like SNR, PSNR, Pearson correlation coefficient (PCC), and root mean square error (RMSE) are commonly used to quantify the effectiveness of image enhancement algorithms. In this section, we will introduce SNR and PSNR for analyzing intensity quantification and PCC for structural feature quantification. Both methods help to objectively evaluate the performance of image enhancement and its impact on image quality.

SNR and PSNR are two metrics commonly used to assess the quality of images. SNR focuses on the ratio of signal to noise, while PSNR focuses on the amount of distortion in the image. They are often expressed in decibels (dB), which facilitates easy comparison across different imaging systems. Although both metrics evaluate the relationship between the signal and noise, they differ in their calculations and applications.

doi:10.6342/NTU202402518

Signal to Noise Ratio (SNR)

SNR (*116*) measures the ratio of the measured signal strength to the overall measured noise intensity at each pixel. It provides a general indication of how much the signal stands out from the noise, reflecting the clarity of the signal relative to the background noise level.

$$SNR = 10\log_{10}\left(\frac{I - bg}{\sqrt{I - bg + \sigma^2}}\right)$$

where I is the average intensity of the signal in the image, bg represents the average intensity of the region without signal, i.e. background region, and σ is the standard deviation of background intensity. dB scale uses a logarithmic representation.

Peak Signal to Noise Ratio (PSNR)[10]

On the other hand, PSNR is specifically used to evaluate how much the quality of the signal is affected by distortion. It calculates the ratio between the maximum intensity of the image and the power of the noise affecting the image. Therefore, it needs a noise-free image as ground truth (or reference) to calculate the MSE, which quantifies the average squared difference between two images. Smaller MSE suggests less disparity between the images, indicating lower noise. Higher PSNR values correspond to lower levels of noise or distortion in the image, indicating higher image quality or fidelity.

$$PSNR = 10log\left(\frac{MAX^2}{MSE}\right) = 20log\left(\frac{MAX}{\sqrt{MSE}}\right),$$

where MAX represents the maximum intensity of the image.

doi:10.6342/NTU202402518

Pearson Correlation Coefficient (PCC)

The PCC is used to assess the linear relationship between two variables. It is denoted by the symbol r and ranges from -1 to 1, which provides valuable insights into the degree of association between two sets of data.

The formula to calculate the PCC (r) between two variables of X and Y with n data points,

$$r = \frac{\Sigma_{i=1}^{n}(X_i - \underline{X})(Y_i - \underline{Y})}{\sqrt{\Sigma_{i=1}^{n}(X_i - \underline{X})^2}\sqrt{\Sigma_{i=1}^{n}(Y_i - \underline{Y})^2}}$$

, $\underline{X}$ *and* $\underline{Y}$ are the average of variables X *and* Y respectively.

As r = 1 indicates a perfect positive linear relationship, meaning that as one variable increases, the other variable increases proportionally. As r = -1 indicates a perfect negative linear relationship, meaning that as one variable increases, the other variable decreases proportionally. As r = 0 indicates no linear relationship between the variables. (Figure 4-14)

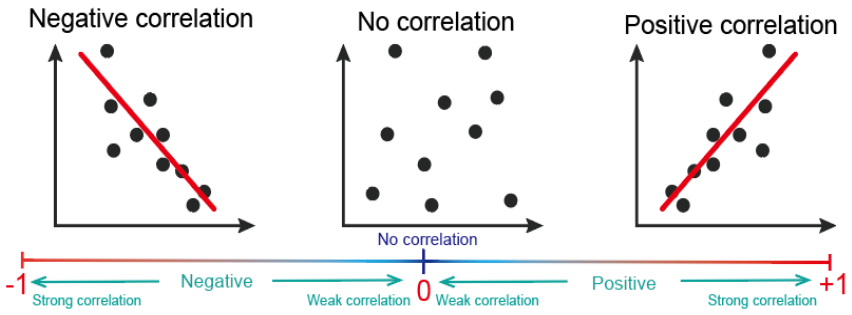


Figure 4-14, Schematic of Pearson correlation coefficient.

Applying to images, PCC quantifies how similar the intensity values are in two images, i.e. the features of two images. When a value close to 1 indicates a strong

doi:10.6342/NTU202402518

similarity between the two images, a value close to -1 indicates a strong negative correlation. A value near 0 suggests no linear relationship between the pixel intensities of the two images.

We may utilize the PCC for various image comparison tasks, for example, in image registration, PCC is used as a similarity metric to measure the alignment accuracy between two images. Higher values indicate better alignment between the structure of images. As mentioned in **Chapter 3.3.3**, analyzing the correlation contour maps of low- and high-speed scanning images is convenient for determining the DOF of the TAG lens.

4.2.3 Statistics analysis: Z-test, T-test and ANOVA analysis

Statistical significance implies that observed data are not random but attributable to a specific cause, lending credibility to experimental results. It underscores the critical role of statistical analysis in identifying meaningful patterns, aiding researchers in drawing informed interpretations and dependable conclusions. In this section, we will introduce three essential statistical analysis methods: Z-test, T-test, and ANOVA analysis[11].

Z-test

A Z-score quantifies the number of standard deviations a data point is from the mean, indicating its position above or below the mean in a normal distribution. Essentially, it represents how far a raw score is from the mean in terms of standard deviations. Z-score is the number representing the result from the Z-test, which is situated on a normal distribution curve calculated with the formula,

$$z = \frac{x - \mu}{\sigma}$$

doi:10.6342/NTU202402518

$\underline{x}$ = observed mean of the sample

μ = mean

σ = standard deviation

If a z-score is 0, it indicates that the data point's score is identical to the mean score. A z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

T-test

The t-test is a statistical method commonly used to compare the means of two groups and determine if there is a significant difference between them. It is particularly useful when dealing with small sample sizes (less than 30), unknown population standard deviation, and normally distributed data. The test examines if the observed differences between groups result from actual variations or could be mere random chance.

T-tests come in two main variants: independent samples t-test, and paired samples t-test, each serving a crucial role in different contexts. Here, we'll focus on the independent samples t-test for comparing two independent groups.

Independent samples t-test

The independent samples t-test is used to compare the means of two independent groups of continuous variables to determine if there is a significant difference. In the context of the independent samples t-test, the null hypothesis plays a crucial role by providing a benchmark against comparing the two groups. A null hypothesis (H_0) is a

doi:10.6342/NTU202402518

claim that the researcher hopes to “reject”. To determine how significant the difference in returns is, the t-score is always represented by the letter p (probability).

The p-value serves as an alternative to rejection points to provide the smallest level of significance at which the null hypothesis would be rejected. The lower the p-value, the stronger the evidence that the null hypothesis is false, and we use asterisks (*) to denote significance levels: For example, P values less than 0.05 indicate statistical significance (*p < 0.05; **p < 0.01; ***p < 0.005; and ****p < 0.001).

T-test can be applied to one sample, the formula is:

$$t = \frac{\underline{x} - \mu}{\sigma/\sqrt{n}}$$

$\underline{x} =$ observed mean of the sample

$\mu =$ assumed mean

$\sigma =$ standard deviation

$n =$ sample size

Additionally, it can be applied to two samples, the formula is:

$$t = \frac{\underline{x_1} - \underline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\underline{x_{1,2}} =$ observed mean of the sample 1 and sample 2

$\sigma_{1,2} =$ standard deviation of the sample 1 and sample 2

$n_{1,2} =$ size of sample 1 and sample 2

doi:10.6342/NTU202402518

Comparing the test statistic by transforming it into a p-value. Confirming the rejection of the null hypothesis. This indicates that statistically, there is a significant difference in the means of these two independent samples.

After the t-score, the overall degrees of freedom is the sum of the individual degrees of freedom (f) for the two samples:

$$f = n_1 + n_2 - 2$$

Independent samples t-test is separated into one-tailed test and two-tailed test. **One-tailed test** shows that the sample mean would be higher or lower than the population mean, but not both. (Figure 4-15A) In other words, the critical area of a distribution is one side, which is run to determine whether a claim is true or false.

When using a one-tailed test, the analyst tests for the possibility of the relationship in one direction of interest and completely disregards the possibility of a relationship in another direction. For example, in the image quality examination, a one-tailed test appropriately shows whether the PSNR (or PCC) of denoised and original images is increased or not. For the null hypothesis, we assume the mean value of PSNR (or PCC) of denoised images is less than the original images.

On the other hand, **a two-tailed test** is performed to determine whether the population parameter of the sample is greater than or less than a specific range of values. (Figure 4-15B) Therefore, there are two rejection regions, one in either direction, left and right, towards each curve tail. It is used to compare two entirely different and mutually independent groups.

doi:10.6342/NTU202402518

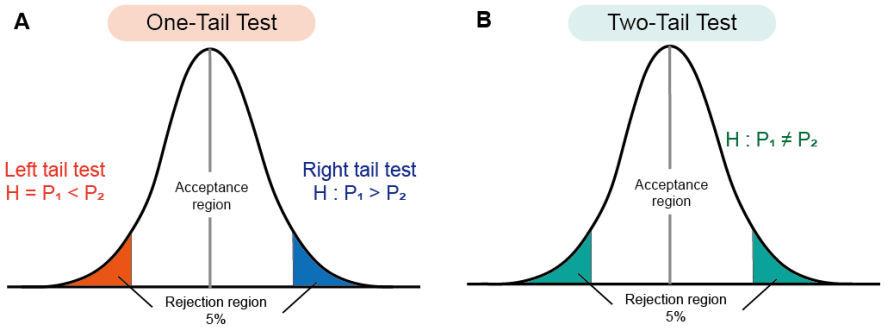


Figure 4-15, Schematic of t-test.

(A) One-tail test. The orange area is for the left tail test, blue area is for the right tail test. (B) Two-tail test. The green area shows the two rejection regions on both sides.

ANOVA analysis

Analysis of variances (ANOVA), also called the Fisher analysis of variance, is a statistical technique used to test for differences between independent variables on a dependent variable. It is the extension of the t- and z-tests.

The basic assumptions of ANOVA are that the samples within each group are independently and identically distributed from a normal distribution and that the variances of the groups are equal. In other words, ANOVA can only determine if there are differences between groups. If you need to compare the means of more than two groups, ANOVA is a better choice than a t-test for statistical analysis.

ANOVA analysis typically involves the following steps:

ANOVA decomposes the variance into different components, including between-group variance and within-group variance. Between-Groups Sum of Squares (BSS) indicates the difference between the expected values of all observations and those within each group. BSS = 0, which suggests no difference in average values between groups:

$$BSS = \Sigma_i n_i \left(\underline{Y_i} - \underline{Y_{total}} \right)^2,$$

doi:10.6342/NTU202402518

where $\underline{Y_i}$ represents the i-th average value

Within-Groups Sum of Squares (WSS) measures the variation within each group. WSS = 0 implies no difference between observations within each group:

$$WSS = \Sigma_i \Sigma_j \left(Y_{ij} - \underline{Y}_{total}\right)^2$$

Comparing BSS and WSS helps assess differences in expected values between groups. However, adjustments for degrees of freedom are crucial, calculated using mean square values to compare within-group and between-group variance.

$$BMSS = \frac{\Sigma_i n_i \left(\underline{Y_i} - \underline{Y}_{total}\right)^2}{k - 1}$$

$$WMSS = \frac{\Sigma_i \Sigma_j \left(Y_{ij} - \underline{Y}_{total}\right)^2}{N - k}$$

For the F-value of ANOVA:

$$F = \frac{BMSS}{WMSS}$$

Then, we can conduct hypothesis testing based on the F-value and degrees of freedom to determine if the between-group differences are significant. ANOVA results typically include an F-value and its corresponding p-value. If the p-value is below the significance threshold (usually 0.05), we reject the null hypothesis, indicating a significant difference between groups. Conversely, if the p-value exceeds the threshold, we fail to reject the null hypothesis, suggesting no significant difference between groups.

4.3 Parameter setting for deep-learning denoise model:

doi:10.6342/NTU202402518

TAG-SPARK

In **Chapter 2.3.4,** we mentioned the structure and concept of TAG-SPARK[12]. Here we will discuss how to set the parameters for training and testing this denoise model. To set the parameters in the script for training, we need to check the path of the folder containing training datasets, the number of iterations (or epochs) for training, and the GPU used for computation. For example, following the experiment of mouse brain slice (Figure 4-4), there are 1145 volumetric images generated by high-speed scanning within 90 seconds at a volume rate of 11.5 Hz (practical volume rate may achieve, # of volumetric images/ experiment time = 12.7 Hz). The FOV of images is $400 \times 400 \times 116\ \mu m^3$. After reconstructing to micrometer interval, each volumetric image had dimensions of $128 \times 128 \times 66$ voxels, with the voxel size of $3.125 \times 3.125 \times 2\ \mu m^3$. Due to the small digital resolution, each functional volumetric image was used to train the model without additional segmentation, and all data in the datasets folder was used for direct training. In our specific scenario, we typically limited the total number of training epochs to prevent overfitting. To confirm whether the training results had converged, the recoding of the loss function across epochs was saved in the same folder as the models. Therefore, the parameters set for training the model are as follows:

n_epochs = 10

GPU = '0'

show_loss_record = True

The model was trained using PyTorch on a GPU computing platform equipped

doi:10.6342/NTU202402518

with a GeForce RTX 2080 Ti and 128 GB of DRAM. CUDA 11.1 was employed to harness the GPU's computational capabilities during training.

After successful training, we proceed to the testing step. During this step, we input the path to the folder containing the test datasets and the trained model folder. The TAG-SPARK model demonstrates specificity, as its performance on the test dataset heavily depends on its similarity to the training data, influenced by the characteristics of the training dataset. Therefore, the test dataset must be the same as the training dataset.

Figure 4-16A shows that after TAG-SPARK processing, the SNR of high-speed scanning image at a single depth increases by 4.4 dB, i.e. the effective noise reduction is about 275%. In addition, compared with the 30-times photon average image, the violin plot in Figure 4-16B illustrates the PCC value enhanced from 0.64 to 0.85. The increased positive PCC value evidences that TAG-SPARK has effectively maintained the neuronal structures and features.

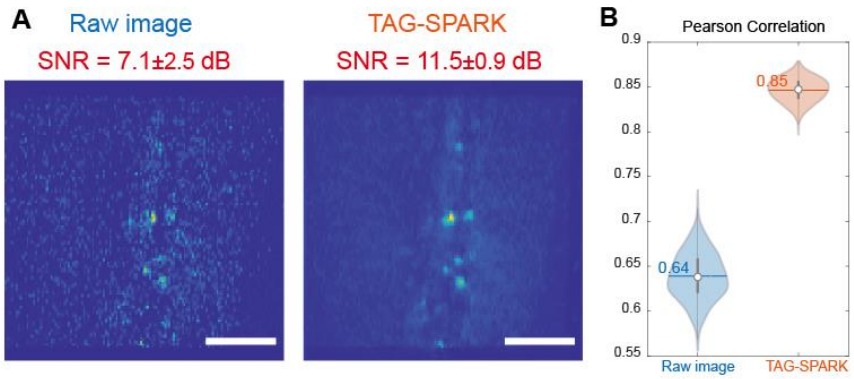


Figure 4-16, Image quality enhancement before and after TAG-SPARK processing.

(A) Average SNR of volumetric image before and after TAG-SPARK processing, increase from 7.1±2.5 dB to 11.5±0.9 dB. (B) Pearson correlation violin plot of the raw and TAG-SPARK denoised images to the 30x photons average.

Because of the trade-off between TAG lens optical power and image contrast, the

doi:10.6342/NTU202402518

quality of the image affects whether our TAG-SPARK model can be effectively trained. Therefore, before conducting *in vivo* experiments, we need to plan carefully to ensure that the premise of not causing damage to mice. We should test the image quality under different optical powers and examine whether TAG-SPARK can enhance image quality under that SNR of images.

doi:10.6342/NTU202402518

Chapter 5 Results and analysis

According to **Chapter 3**, we have established knowledge of optical system setup. To understand the evidence of the system's performance (see **Chapter 3.3** for details). In this chapter, we focus on applying our techniques to investigate the activity of PCs in the cerebellar cortex, serving as an experimental model. Our goal is to observe how neural signals are vertically transmitted under different response conditions and examine the impact on the population of cells. We recorded the behavior of the mouse during rest and movement states. When the neurons are sensory-evoked, we expect a change in the fluorescence intensity. Through these experiments[13], we seek a comprehensive insight into the understanding of the dynamic nature of neural networks and the impact of technological improvements on studying brain function and neurotransmission.

5.1 High-speed volumetric image related to *in vivo* mouse behavior recording

In this section, we will elaborate on the significance of the brain region and neural cells observed in the experiment. Additionally, we will present the structural and functional images achieved through the utilization of low- and high-speed scanning techniques respectively. Preliminary analyses were conducted on the obtained images, revealing valuable insights into the distribution of neural signals within the observed in vivo mouse brain.

To understand the intricate dynamics of living brains, we constructed a high-speed volumetric 2PM system, which combines low-speed and high-speed axial scanning using a linear translation stage and TAG lens. Low-speed scanning provides detailed structural mapping with a high SNR and optimizes localization as analysis. High-speed scanning

doi:10.6342/NTU202402518

captures not only architecture information but also rapid neuronal dynamics in real time.

Structural image of PC in the mice

To study the dynamic neuronal signals of PCs in mice, we fixed the mouse's head onto a rotating disk, allowing for free movement during the experiment. In the low-speed scanning mode, the linear translation stage (MGZ30, Thorlabs) was employed to achieve axial translation with a spacing of 1 μm. Laser power after the objective lens is set around 100 mW. Each frame was acquired with a pixel dwelling time of 5 μs, three times averaging (frame rate takes 3 s), and a volume size of 400 x 400 x 152 (μm^3), featuring 512 x 512 pixels per 2D image and 152 frames in the stack (Figure 5-1).

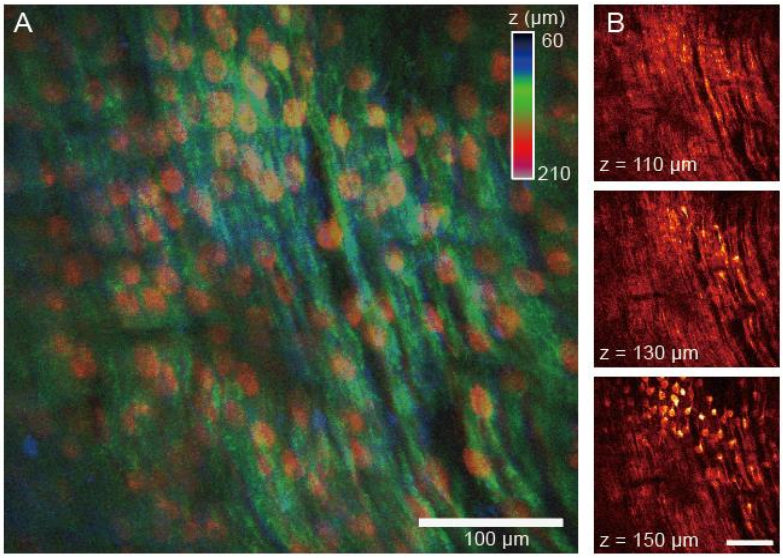


Figure 5-1, Low-speed Structural image of PCs in *in vivo* mouse.

(A) Volumetric image with z-projection. (B) 2D images in different depths. From top to bottom: dendritic layer to somatic layer. (C) Schematic of PCs from the dendritic layer to the somatic layer. In the xy sagittal planes, the dendrite branches exhibit parallel line structures, while the cell bodies display circular images.

Mouse behavior experiment

To ensure the accuracy of neural signals and the monitoring of mouse movement

doi:10.6342/NTU202402518

during high-speed scanning experiments, we implemented a rigorous approach. To precisely capture the three-axis velocity of the mouse tail movement over time, a tail accelerometer[14] was incorporated (Figure 5-2). Furthermore, a high-speed infrared camera was employed to record the entire experimental process[2], enabling the observation of the mouse's movement states and tracking behavior with DeepLabCut[15] (*117*, *118*).

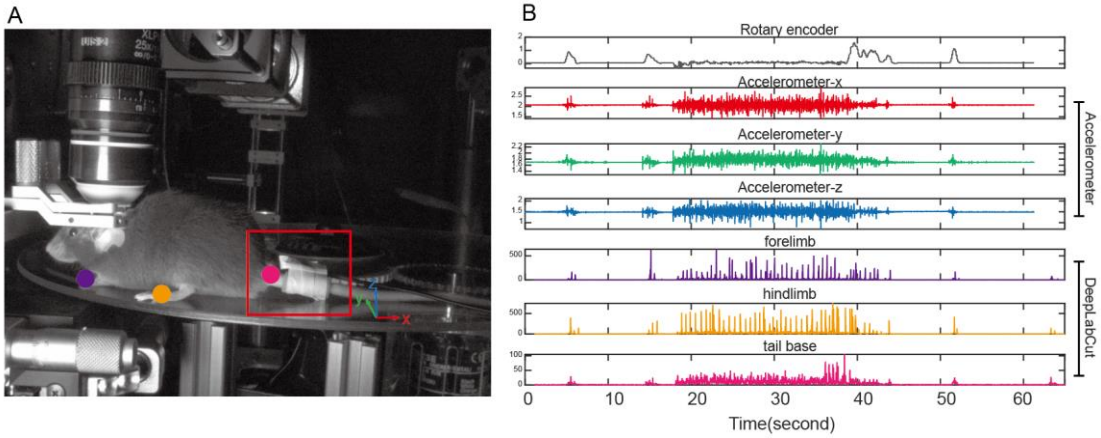


Figure 5-2, Photograph of the mouse head-fixation and rotating disc system.

(A) The red square highlights the miniaturized three-axis accelerometer installed at the base of the mouse's tail. Purple, orange, and pink points were detected from DeepLabCut software. We performed the tracking analysis for the foot, hint position, and tail position as tracking points to record the mouse's locomotion. (B) Signal analysis revealed that during the middle 20 seconds, the mice maintained a movement state. From top to bottom, the signal is recorded from the rotation disk, accelerometer, and tracking analysis.

For behavioral experiments, distinct states were designed for the mouse, encompassing periods of complete rest and active movement. During a 65-second experiment in a state of complete rest, without external intervention, mice may exhibit spontaneous movements. Subsequently, for the movement state, external recording was initiated by activating the LED light 5 seconds after the start of recording. After an initial 20-second period of inactivity, the rotation disk was then shaken for 20 seconds to induce

doi:10.6342/NTU202402518

movement in the mouse. Following this, the disk remained stationary for another 20 seconds before the LED light was activated again 5 seconds before the end of the recording, serving as an external intervention marker (Figure 5-3).

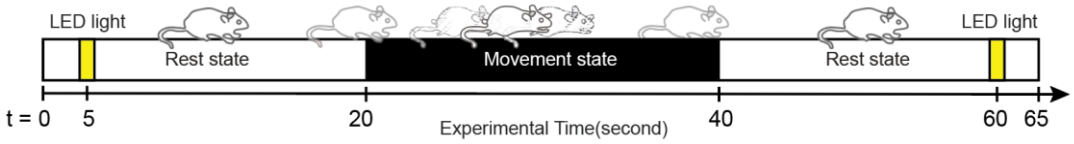


Figure 5-3, Experimental design for PCs.

The yellow bar presents the moment when the LED light. The black region presents the movement state of the mouse. During the middle 20 seconds, we will shake the disk to force the mice to run.

In high-speed scanning mode, we utilized a TAG lens to rapidly modulate z-positions at a resonant frequency of 188 kHz, with a voxel dwelling time of 0.02 µs. The volume rate varied based on differences in lateral pixel counts, resulting in a diverse dataset capturing neural structures and behavioral responses across different scanning speeds. In our experimental conditions, we set the TAG lens driving amplitude at 15%, corresponding to an optical power of approximately 3.75 D. This adjustment extended the DOF to encompass the neuronal architecture of PCs within a physical dimension of 62 µm. With this driving amplitude, both dendritic and somatic architecture are captured, and z resolution achieves sub-micrometer resulting in notably homogeneous representation between adjacent layers.

Figure 5-4A shows the structural images, at a depth of 110 µm, parallel line structures revealed dendrite branches in distinct planes. As the depth reaches 130 µm, dendrite branches converge into circular soma, consistent with the typical spacing between dendrite and soma in PCs. At the bottom layer (150 µm), the soma exhibited flask-like shapes. Figure 5-4B left side shows the functional image at a single time point.

doi:10.6342/NTU202402518

However, the SNR of images is too low to directly compare with structural images. A 3-second temporal projection was applied to the images to enhance the image quality (Figure 5-4B right side). The resulting high-speed scanning images undergo visual co-registration and low-speed scanning images. Accordingly established in **Chapter 4.1.2** that both low- and high-speed scanning corresponds to specific FOV, and the structure of neurons is clear. Additionally, Figure 5-4B demonstrated high-speed imaging examples at three distinct depths with adjacent pairs (109 μm & 110 μm; 129 μm & 130 μm; 149 μm & 150 μm), presenting the similarity in each pair. These image properties establish a spatial redundancy for the application of the TAG-SPARK model.

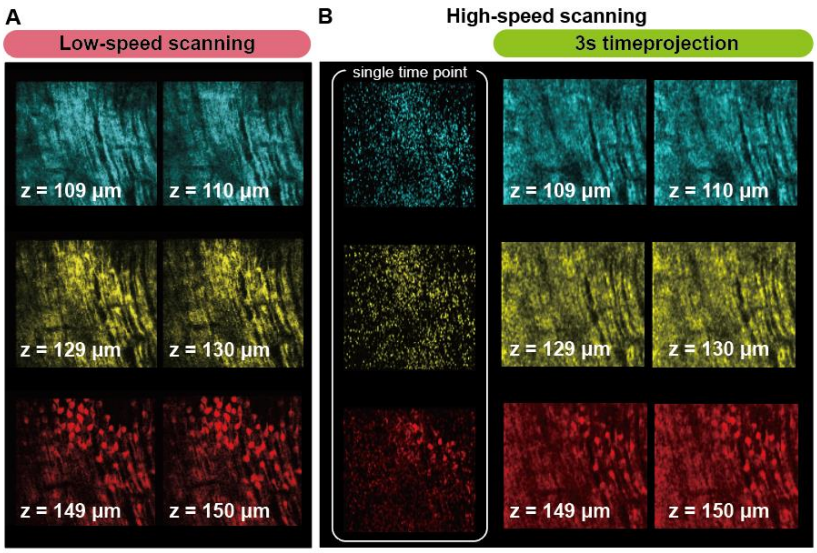


Figure 5-4, Adjacent layers of image of *in vivo* PCs at different depths.

(A) Low-speed scanning structural images. (B) Left: High-speed scanning of images at a single time point. Right: 3-second time projection images.

Preliminary dynamic neuronal signal analysis

To identify individual neuron behaviors, we combined information from structural images, which were treated as ROI masks, and functional images, providing time-lapse data for dynamic analysis (Figure 5-5A). For normalized signal, we set the

doi:10.6342/NTU202402518

average of intensity during the rest state as baseline,

$$\frac{\Delta F}{F} = \frac{F - \bar{F}}{\bar{F}}$$

where F is the mean of intensity at a single time point, $\bar{F}$ is the baseline of intensity.

Our preliminary analysis focused on the distinct cells located at the deep somatic layer (z=130, 140, 150 um), revealing diverse responses across different cells. Additionally, the response is related to mouse behavior, as indicated by the orange area in Figure 5-5B.

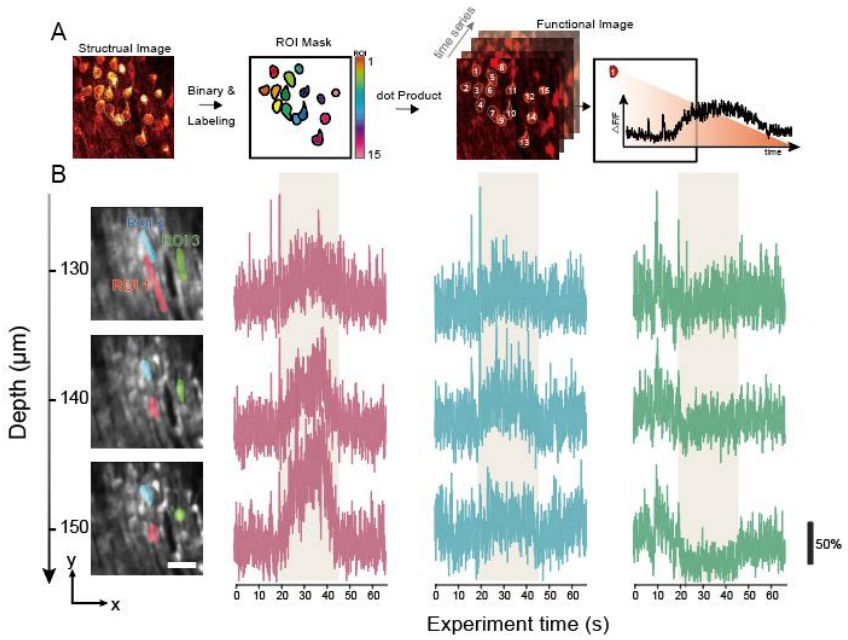


Figure 5-5, Preliminary analysis of PCs functional image.

(A) The process of ROI mask segmentation and applied to functional images for analysis. (B) (Left) Images of the mouse cortex of three individual PCs in different observation depths from 130 μm to 150 μm. The locations of three individual PCs are marked as ROI 1 (Red), ROI 2 (Blue), and ROI 3 (Green). Scale bar: 50 μm. (Right) Corresponding calcium traces at various depths with raw data analysis.

doi:10.6342/NTU202402518

5.2 Evaluation of functional images: comparing before and after

In this section, we will compare the original and denoised images through quantitative analysis of image quality using the methods mentioned in **Chapter 4.2**. To evaluate the improvement in image quality after applying TAG-SPARK, a comparison is made between the images before and after denoising and a high-quality reference image.

5.2.1 Pre-image processing for analysis

Figure 5-6 shows the pipeline of pre-image processing for functional analysis after denoising images. After TAG-SPARK processing, the enhancement of high-speed scanning image quality allows the distinguishable structures. Therefore, applied it with motion correction at different time points to ensure that the structures are aligned at the same position. (see **Chapter 4.2.1** for details) Then, low-speed scanning images as structural images underwent a series of steps for co-registration. Initially, a 3-second time projection was applied to the high-speed scanning images. Subsequently, a rough manual alignment was performed through visual inspection to ensure both the high-speed and low-speed images were within the same field of view. For further refinement, the “Register Images” feature in the Amira software was utilized. A “Rigid” transformation with disabled rotation was selected. These procedures ensure consistent alignment between the high-speed scanning and low-speed scanning images, enhancing overall reliability in data analysis and interpretation. At last, we segmented the individual cells as ROI masks and applied them to the functional imaging for time-lapse analysis.

doi:10.6342/NTU202402518

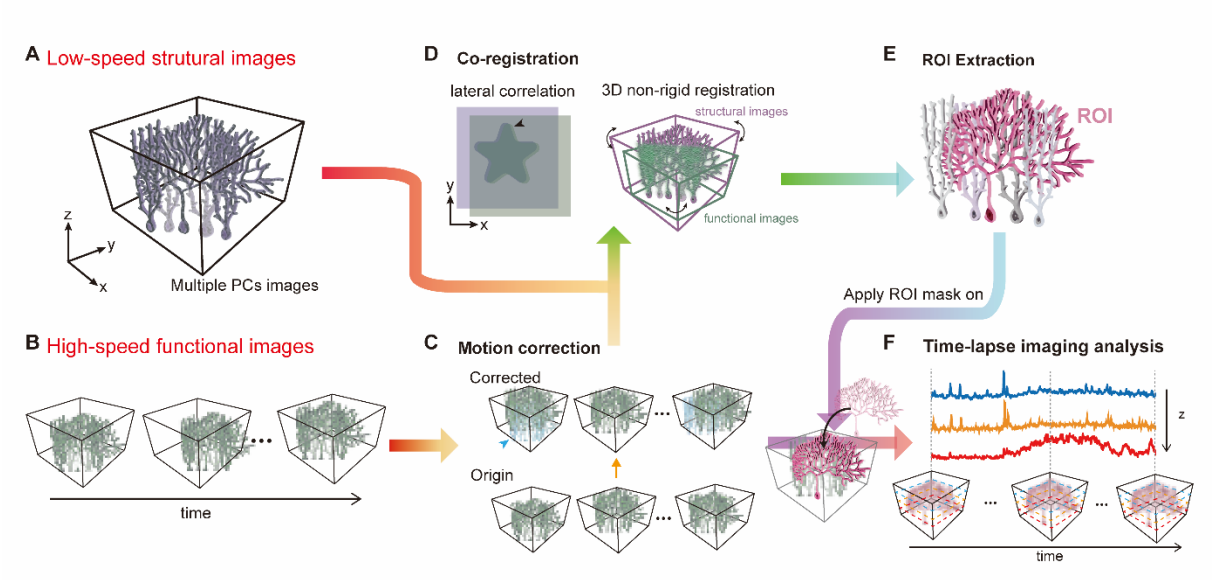


Figure 5-6, Schematic of cell segmentation and co-registration of calcium imaging via low-speed structural and high-speed functional scanning modalities.

(A) Low-speed scan images serve as the structural image. (B) High-speed functional images after TAG-SPARK processing. (C) Motion correction on functional images at different time points to ensure that the structures are aligned at the same position. (D) Co-registration of low-speed structural and high-speed functional images using lateral correlation analysis and 3D non-rigid registration to identify the best correspondence. (E) Segmentation of individual neurons using the structural images as ROI masks. (F) Applying the ROI masks to time-lapsed functional images and analyzing signal changes at different time points to capture calcium dynamic responses on a 3D basis.

5.2.2 Evaluation of denoising method: TAG-SPARK for structural similarity and SNR optimization

In Figure 5-7, we analyzed differences between raw data and the images after TAG-SPARK processing. The results (the blue and orange lines present raw and denoised data respectively) show that TAG-SPARK effectively reduced noise intensity while preserving the peak values and trends of the signals.

doi:10.6342/NTU202402518

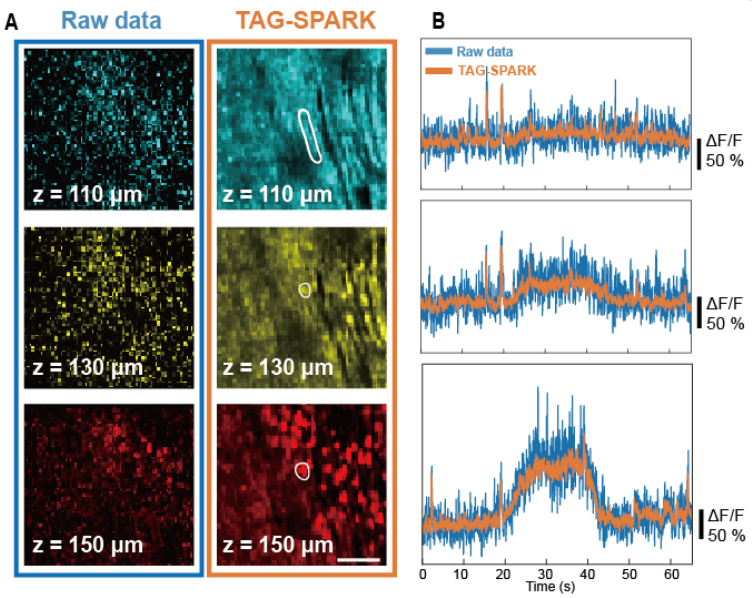


Figure 5-7, TAG-SPARK denoising performance on calcium imaging of PCs.

(A) *In vivo* calcium imaging of PCs at different depths of high-speed and denoised high-speed scanning mode, employing the TAG-SPARK model, respectively. Scale bars: 100 μm. (B) From top to bottom: Calcium imaging of PCs in the dendritic and somatic layers, respectively. Corresponding calcium traces of a selected dendrite and soma within a 65-second observation time window.

The results of the effectiveness of the denoising methods, TAG-SPARK are quantitatively assessed to measure the extent of improvement in image quality. We utilize the PCC to analyze structural similarity and the PSNR to optimize the PSNR of temporal signals. To validate if the quality of the denoised images is significantly greater than the original images, a one-tailed paired t-test is employed (see **Chapter 4.1.2** for details). Before initiating the quantitative analysis, it is imperative to establish a reference point for comparing the original and denoised images. Relying solely on static structural images for comparison is not advisable, given the dynamic information in high-speed scanning images.

To account for the variability in neural signals, the assessment involves both PCC and PSNR for denoised images concerning the moving average of N photon images. The

doi:10.6342/NTU202402518

denoised image exhibits comparable quality to the 60-times averaged image, or the 60x photons image, as indicated by significantly elevated PCC and PSNR values. (Figure 5-8)

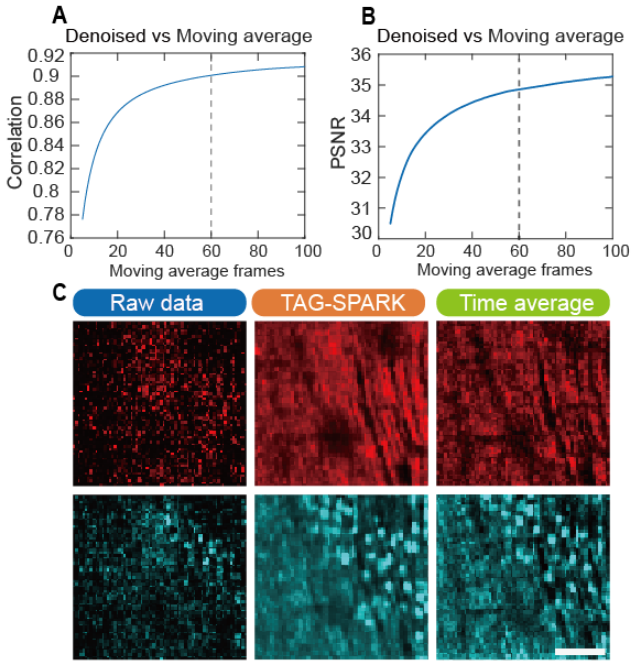


Figure 5-8, Denoise and moving average comparison.

(A)Pearson's correlation coefficient (PCC) and (B) peak signal-to-noise ratio (PSNR) of denoised images to the moving average of N times photons image. The 60-times averaged image, i.e. 60x photons image, has approached reasonably high PCC and PSNR versus the denoised image (gray dash line). That is, the quality of the denoised image is similar to that of the 60x average. (C) From left to right show the images with raw data, TAG-SPARK processed, and 60x photon average image. The top and bottom demonstrate the different depths, with dendritic and somatic layers respectively.

In Figure 5-9 A and B, we compared the volumetric images of raw data, after TAG-SPARK processed, and images overlaid with a moving average of 60x photons. The violin plots illustrate that denoised images displayed a significantly higher PCC distribution, with a median of 0.81, in contrast to the raw data with a PCC of 0.48. The increased positive PCC value suggests that the denoising process has effectively maintained the integrity of neuronal structures and features. In addition, to present the

doi:10.6342/NTU202402518

improvement of image quality and noise reduction, we calculate the distribution of the PSNR before and after denoising. PSNR value increases by around 5.6 dB, i.e. the effective noise reduction is about 360%. Both PCC and PSNR demonstrated statistically significant improvement, as evidenced by the results of the one-tailed paired t-test (p-value < 0.001). These observations emphasize the effectiveness of our method in suppressing fluctuating noise across spatial and temporal domains, while accurately preserving the rapid neural activities evident in the calcium traces.

To further confirm the effectiveness of noise reduction at various depths, we analyze the quality of 2D images at every 8 μm, as depicted in Figure 5-9 C and D. The distribution of PCC shows that at the edges of volumetric images (90 μm and 150 μm), higher correlation coefficients are due to TAG lenses with various focusing points along the z-direction, varying in a sinusoidal wave track. On the other hand, the distribution of PSNR increases with depth. This is because PSNR depends on the maximum intensity of the images, and the dendritic layer typically has lower intensity compared to the somatic layer. Figure 5-9B shows a violin plot of 3D PSNR like the result of the depth-dependent PSNR at 150 μm, incorporating both dendritic and somatic regions where the intensity of the soma influences the peak signal value for calculation.

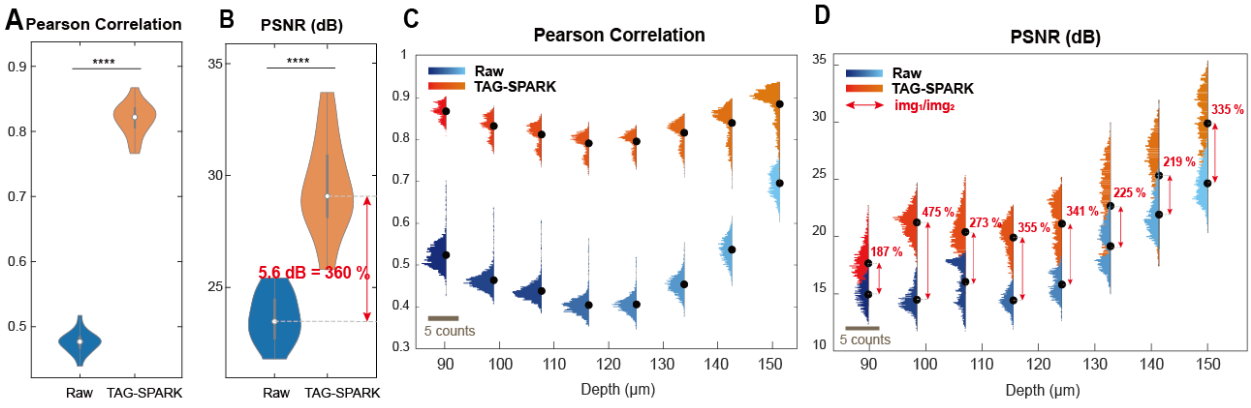


Figure 5-9, TAG-SPARK denoising performance on calcium imaging of PCs.

Calcium imaging of PCs in the dendritic and somatic layers, with pursued blue and red color respectively.

doi:10.6342/NTU202402518

(A) Pearson correlation violin plot of the raw and TAG-SPARK denoised images to the 60x photons average. (B) Distribution of PSNR before and after TAG-SPARK processing. (C) Pearson correlation and the corresponding histogram distribution at different depths before and after TAG-SPARK processing to the 60x photons average. The PCCs of raw data range between 0.4-0.7, while those of TAG-SPARK are 0.8-0.9. (D) Distribution of PSNR and the corresponding histogram distribution at different depths before and after TAG-SPARK processing. The PSNRs of raw data range between 14-24 dB, while those of TAG-SPARK are 17-30 dB. The maximal enhancement is 475%.

As introduced in **Chapter 4.2.2,** SNR and PSNR play distinct roles in evaluating image quality. SNR compares the signal and noise within the image itself, while PSNR compares the maximum intensity and MSE between the ground truth (or noise-free image) and the processed image. To further confirm the effectiveness of noise reduction at various depths, we analyze the depth-dependent SNR variation at every 8 μm. Figure 5-10 shows a significant improvement in SNR, before TAG-SPARK processing, the SNR of the layered original image ranged between 1-5 dB (average 2.3 dB, blue color histograms). After denoising, the SNR values improved to 9-12 dB (average 10.4 dB, orange color histograms). The red double arrows and numbers indicate the ratio of SNR improvement of TAG-SPARK, ranging from 7.2 dB (527%) to 8.9 dB (771%). On average, the overall SNR enhancement is 8.1 dB, corresponding to about 645% improvement in image quality.

doi:10.6342/NTU202402518

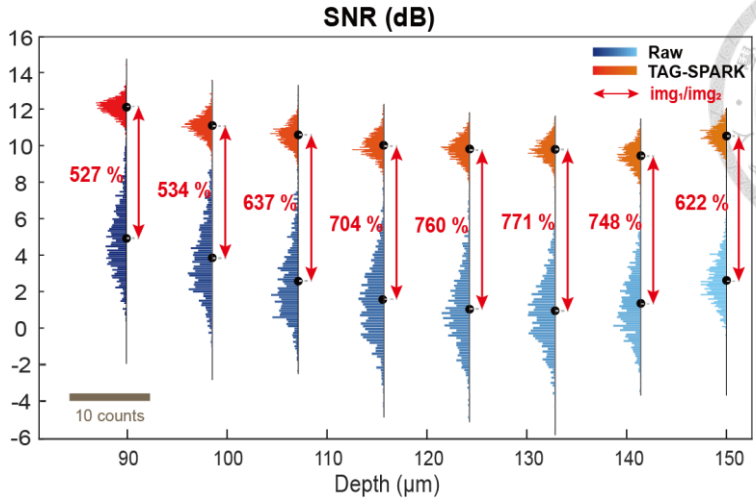


Figure 5-10, Comparison of raw and TAG-SPARK images with depth-dependent SNR analysis.

Distribution of SNR across 2D layers at various depths before and after TAG-SPARK processing. SNR enhancement of TAG-SPARK among different depths of 2D layers, the best approaching 800%.

5.3 Individual neural signal transmission and population neural behavior

In this section, we apply TAG-SPARK to functional imaging and uncover intriguing results. We not only analyze four-dimension (xyz-t) neural signal variations within individual neurons but also examine changes in large cell populations, revealing insights into both individual neural dynamics and collective behavior.

5.3.1 Signals transmission through individual neurons

The theoretical expectation is that calcium influx in PC dendrites converges into the cell body, elevating the somatic calcium concentration necessary for the generation of complex spikes in PCs. To observe the variations in neural signals over time and depth, manual segmentation was performed, and three individual neural cells were selected.

doi:10.6342/NTU202402518

(Figure 5-11) We utilized the “Volume Edit” function in the Amira software to track individual cells in 3D imaging.

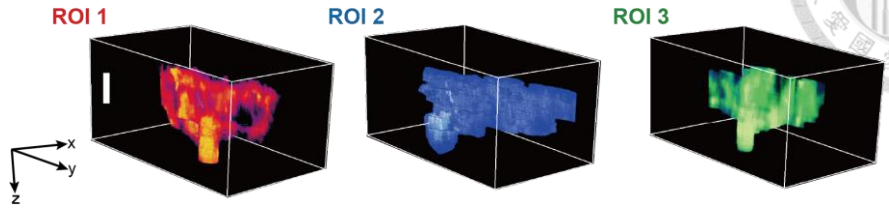


Figure 5-11, Three-dimensional image reconstruction of three individual neurons.

In this study, we extracted the x-axis acceleration variations from the three-axis accelerometer (Figure 5-12A) to compare with the neural dynamic signals captured by functional images. In Figure 5-12B, we observed that as the mice initiated movement, there were corresponding alterations in neural signals. Furthermore, these changes differed with depth. In superficial dendritic layers, the response manifested as peaks within one second, while in deeper somatic layers, peaks lasted over ten seconds.

Notably, the three individual neural cells exhibited distinct expression patterns, particularly evident during the mouse movement state, where positive, flat, and negative responses were observed in the 10-second last peaks, i.e. slow signal response. These states were also identifiable in the analysis of raw images as Figure 5-5B, yet the intensity variations corresponding to short-duration spontaneous movements became obvious after denoising, i.e. fast signal response. Additionally, the contour maps (Figure 5-12C) were employed to visually represent continuous changes, highlighting the differences in intensity with red and blue colors, i.e. positive and negative values, in the responses of the three neural cells in deeper layers (4D dynamic neural signal[16]).

doi:10.6342/NTU202402518

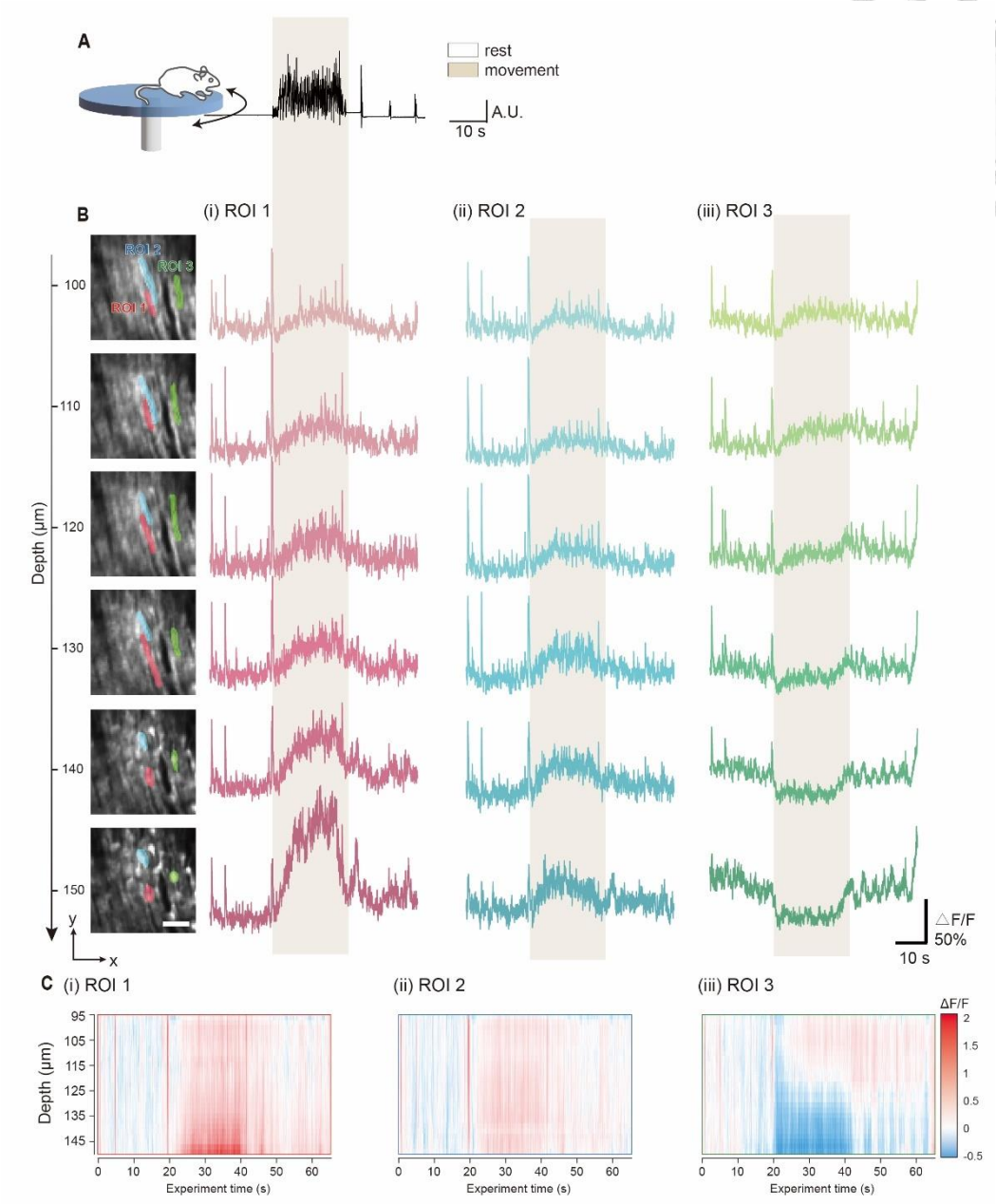


Figure 5-12, Dendrite-to-soma signal tracing along three individual neurons.

(A) Schematic of a mouse placed on the rotating disc and the intensity profile of the leg movements. (B) (Left) Images of the mouse cortex of three individual PCs in different observation depths from 100 μm to 150 μm. The locations of three individual PCs are marked as ROI 1 (Red), ROI 2 (Blue), and ROI 3 (Green). Scale bar: 50 μm. (Right) Corresponding calcium traces at various depths. (C) Corresponding depth-dependent calcium response of three individual neurons.

In a summary of the image analysis conducted, we confirm that TAG-SPARK,

doi:10.6342/NTU202402518

even after denoising, effectively preserves the inherent trends and intensity variations in the signals. We successfully preserve the peaks in neural signals of individual neurons concerning both depth and temporal information. This elucidation allows us to establish the relationship between neural signals and behavior.

5.3.2 Relation between large cell populations and depth of brain

To further explore dendritic-to-somatic calcium evolution within PC populations, as a proof-of-concept experiment, the 27 cells extracted in the dendritic and somatic layers (Figure 5-13), and all time-lapse signals depicted in Figure 5-13C-F. We perform the clustering analysis on both dendrite and soma responses based on the threshold value $\Delta F/F = 0.2$ of the signals during the phase of the movement (the period colored with gray shadow). The non-intuitive convolution result is classified into three types, including green: $\Delta F/F > 0.2$, purple: $-0.2 < \Delta F/F < 0.2$, and blue: $\Delta F/F < 0.2$. For a complete visualization of the overall calcium dynamics, Figure 5-13G and H depict the distribution of calcium traces for all 27 cells. These traces are presented as colored shadows in the dendritic and somatic layers, respectively, while the average calcium traces are represented by solid curves. All dendritic responses were averaged together due to their consistent behaviors. Thus, in Figure 5-13A and B, corresponding ROI markers are labeled with colors corresponding to their respective functional groups. It's noteworthy that even neighboring cells within the somatic layer exhibited varying behavioral patterns.

In line with previous research (*70*, *119-121*), PCs show similar dendritic calcium spikes during prolonged movement, as indicated by the gray area (Figure 5-13C). However, there was a divergence in calcium dynamics within the PCs at the somatic layer (Figure 5-13D-F). While some adhered to the conventional signal convolution theory,

doi:10.6342/NTU202402518

differing only in signal strength, others exhibited negative responses, as highlighted by the red arrowhead (Figure 5-13J**)**. This suggests a decrease in calcium concentration, contradicting the common belief of additive convolution from dendrites to soma.

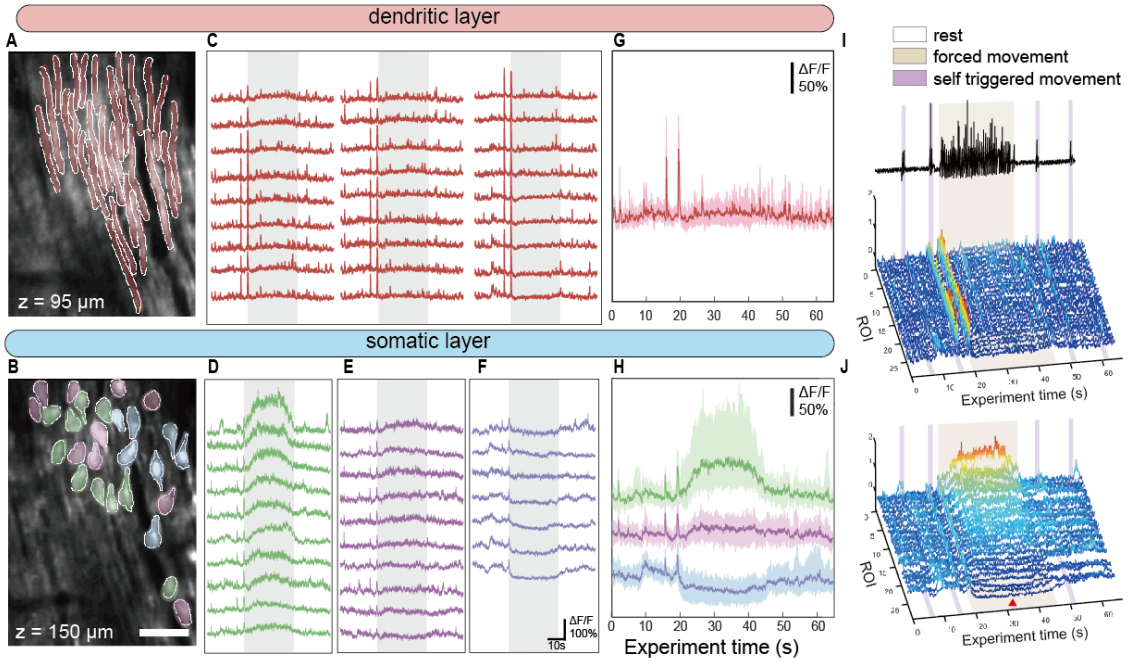


Figure 5-13, Calcium dynamics of PC populations at different depths.

Low-speed scanning images in dendritic and somatic layers, respectively. Corresponding calcium traces of 27 cells, marked in (A) and (B). (C)-(F) 27 cells with different time-lapse analyses were marked in the red/green, blue, and purple shadows in the dendritic/somatic layers. (G-H) Collective calcium response in the dendritic/somatic layers. Three distinct behaviors are distinguished based on the threshold value of $\Delta F/F = 0.2$ during the phase of the movement stimulation phase, represented by three different colors: green ($\Delta F/F < 0.2$), purple ($-0.2 < \Delta F/F < 0.2$), and blue ($\Delta F/F > 0.2$). In (D), ROI markers are labeled with colors corresponding to their respective groups. (I-J) Corresponding calcium traces of 27 cells, marked in (C-F).

For the functional disparity analysis, we performed pairwise correlation comparisons of calcium traces among the 27 cells in each layer with 5-μm depth intervals from the dendritic to somatic layers. In Figure 5-14 the map shows the averaged correlation coefficients decline from 0.7 to 0.1, and the histogram distribution illustrates

doi:10.6342/NTU202402518

the overall differences in each layer, particularly revealing more distinct variations in the deeper layers. At depths ranging from 95 to 115 μm, the correlation coefficient distributions were consistently positive, indicating similar calcium dynamics in the dendritic layers. However, beyond 120 μm depth, the distribution of the correlation coefficient ranged from positive to negative regions, suggesting the existence of at least three kinds of behaviors in the PC populations in the somatic layers, i.e., positive, zero, and negative values of the correlation coefficients. The volumetric imaging results provided direct evidence that a significant portion of somatic PC responses do not follow their dendritic calcium activities.

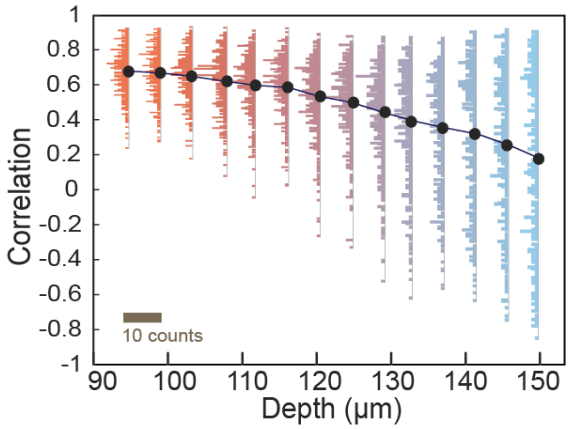


Figure 5-14, Depth-dependent correlation analysis.

Correlation calculation at different depths over a 65-second observation time window and the corresponding histogram distribution.

While imaging calcium activities of PC dendrites has been a standard approach to representing somatic complex spike activities of PCs, this discovery leads to a strong biological impact, and the dendritic-to-somatic interactions require revisiting. This phenomenon could only be accurately discerned through the simultaneous detection of dendritic calcium spikes and somatic calcium fluctuations facilitated by our volumetric imaging technique along with the contrast-enhanced TAG-SPARK algorithm.

doi:10.6342/NTU202402518

In summary, this analysis confirms the critical role of volumetric imaging in analyzing neural signals. The responses observed in a single layer of cells do not necessarily reflect similar reactions in cells located in different layers. Moreover, as the PC exemplifies, the differences in neural signal changes become more disparity with increasing depth.

doi:10.6342/NTU202402518

Chapter 6 Discussion

In this study, we integrate a TAG lens 2P high-speed volumetric system with a deep-learning denoising algorithm, TAG-SPARK, to demonstrate calcium functional imaging across a large population of PCs in awake mice. Our high-speed volumetric system acquires hundreds (265) of axial layers per volume, facilitating effective training of the TAG-SPARK model in the spatial domain. We present the applicability to neuroscience by examining calcium dynamics of PCs in the mouse cerebellar cortex, exploring signal transmission from dendrites to soma at various depths, and revealing variations among individual neurons. This ensures the preservation of temporal information in calcium dynamics and achieves around 400% (5.6 dB) PSNR and over 700% (10.4 dB) SNR enhancement.

Photobleaching

For in vivo mouse experiments, we control the laser power after an objective lens of around 100 mW to protect the mice from damage and photobleaching. The utilization of TAG-SPARK for image quality improvement eliminates the need for high-power illumination, addressing concerns about photobleaching (Figure 6-1).

doi:10.6342/NTU202402518

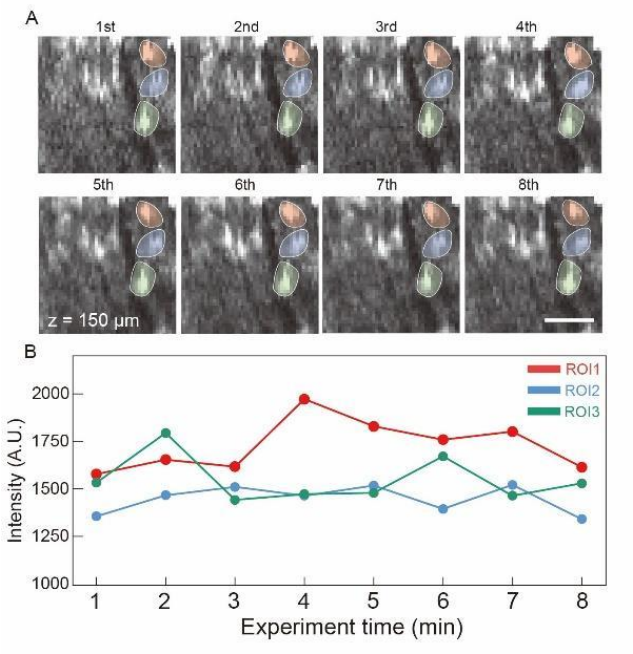


Figure 6-1, Photobleaching examination within 8 minutes.

(A) Raw calcium images of one optical section in the eight continuous high-speed volumetric imaging trials. (B) Average fluorescence intensity of the three ROIs across the eight trials, featuring that no significant photobleaching is observed.

Increasing the laser power further improves contrast but may lead to sample bleaching or even damage. In Figure 6-2A and B, we applied different laser powers to *in vitro* mouse brain slices. It clearly shows that with a two-fold increase in laser power, the SNR of the image increases by approximately 3.1 dB, corresponding to a 204% enhancement in image quality. However, we scanned the same FOV of the sample over 1 minute, capturing an image every 2 seconds, and analyzed the intensity-time dependence. The results (Figure 6-2C) indicate that high laser intensity causes a gradual decrease in overall image intensity within one minute, whereas low-intensity laser settings maintain relatively stable image intensity.

doi:10.6342/NTU202402518

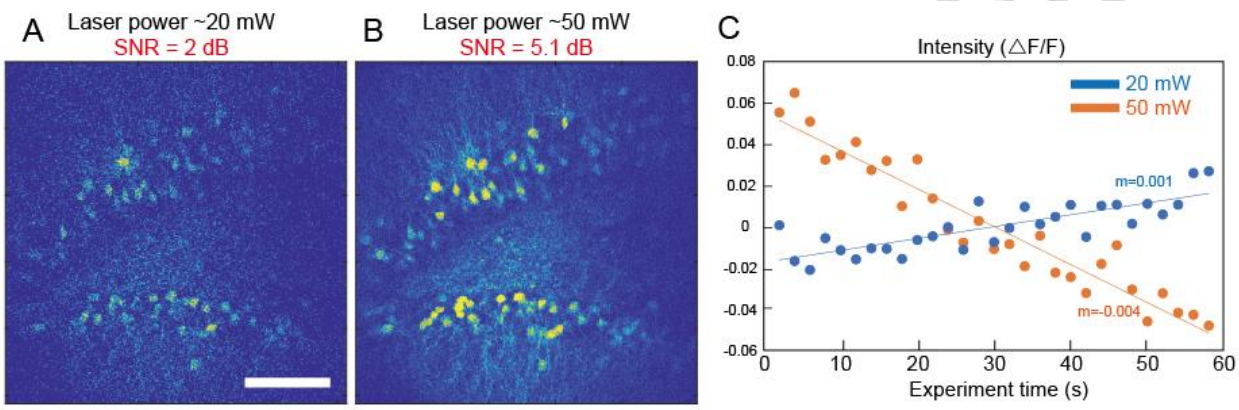


Figure 6-2, Photobleaching results from different laser power.

Low-speed scanning image of *in vitro* mouse brain slice (A) with 20mW laser power. SNR of the image is 2 dB. (B) with 50mW laser power. SNR of the image increases to about 5 dB. (C) Time-dependent intensity difference observed over 1 minute.

We discussed two pivotal technologies in our study: the TAG lens, which facilitates high-speed volumetric scanning, and TAG-SPARK, a key player in enhancing image quality.

TAG lens: DOF and digital resolution

Starting with the TAG lens, which has been widely utilized in volumetric 2PM systems. The advancement of TAG lenses necessitates consideration of two important parameters: DOF and digital resolution, which correspond to the strength of its optical power (diopters) (see **Chapter 2.2** for details), and data acquisition rate (see **Chapter 4.1.1** for details). TAG lens has been widely utilized in volumetric 2PM systems. TAG lens has been widely utilized in volumetric 2PM systems. One early milestone paper was in 2015,(*122*) reporting 130 μm extension and 40 voxels in the z-direction. Over the past decade, advancements in the application of 2PM with TAG lenses have substantially improved DOF extension and axial digital resolution. In 2018, Piazza et al.

doi:10.6342/NTU202402518

reported a DOF of 120 μm and axial digital resolution of 80 voxels,(*48*) and Har-Gil et al., achieved an impressive volume rate of 73.4 Hz, with each volume capturing a DOF of 330 μm using 150 voxels.(*123*) In 2019, our group obtained functional imaging of fruit flies with a depth of 100 μm and 70 axial voxels.(*54*) In 2021, the number became 200 μm DOF extension and 73 axial voxels, when we incorporated a GRIN lens endoscopy for volumetric functional inspection at 6 mm below the skull.(*27*) In 2022, we integrated plug-and-play AO to address aberrations during high-speed volumetric scanning.(*56*) However, these previous reports did not demonstrate sub-micrometer axial layer gaps, which is necessary to implement the TAG-SPARK idea. In 2023, Hsu et al. introduced a resonant galvo together with the TAG lens, pushing the numbers toward an axial resolution of 256 voxels, with a DOF of only 30 μm (*124*). Although the spatial redundancy should be adequate, their double-resonant design resulted in Lissajous sampling, which might be problematic when combined with the Noise2Noise model. In this work, we demonstrated more than 200 axial layers with 60 μm in DOF extension, tailor-made to realize the idea of deep-learning denoise of TAG-SPARK.

Except for the TAG lens, there are many other approaches for volumetric 2PM, but in general, the axial layer density is much less than the TAG lens approaches. For example, reverberation 2PM achieves video-rate (30Hz) observation across a DOF spanning 500 μm, but with a gap as large as 90 μm between each plane.(*43*) Light bead microscopy (LBM), creates axially distinct foci at different depths simultaneously, enabling the direct acquisition of large volumetric images on a millimeter scale,(*15, 41, 42*) but the layers are separated by at least 10 μm. Another recent multi-plane imaging technique is dual-objective 2PM, which offers a clear view of 380 μm thick brain slices with 50 axial layers; nevertheless, it is restricted to *in vitro* experiments due to the geometrical design.(*125*) We envision that our TAG-SPARK algorithm applies to other

doi:10.6342/NTU202402518

volumetric imaging systems as long as the density in the axial direction is adequate to fulfill the spatial redundancy requirement.

Trade-off in the application of TAG lens

For TAG lens application, its axial scanning speed, i.e. the resonance frequency, is primarily constrained by readily available parameters as well. Functioning as a resonant component, the TAG lens operates at multiple harmonic frequencies, where a higher driving frequency extends the DOF but compromises the effective NA, leading to a lower spatial resolution (*55*). This trade-off is carefully navigated in our experiments to meet essential criteria such as (1) adequate temporal resolution for capturing dendritic spikes, (2) sufficient spatial resolution for identifying sub-cellular structures, (3) enough DOF to cover the 3D architecture of neurons, and (4) a dense z-sampling dataset for training the TAG-SPARK model.

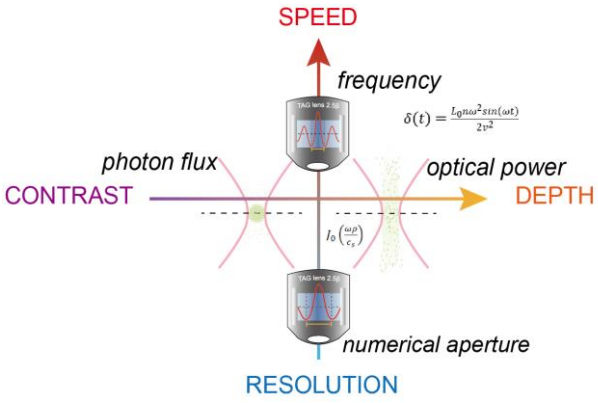


Figure 6-3, Four parameters of TAG lens.

When approaching 'Depth,' one moves away from 'Contrast' because photons are spread over a larger depth of field. Similarly, when near 'Speed,' 'Resolution' becomes distant due to effective NA decreased. Increasing 'Speed' also increases 'Depth' since optical power correlates with resonance frequency. However, augmenting 'Speed' diminishes 'Contrast' because high-speed scanning reduces photon flux at each voxel.

doi:10.6342/NTU202402518

While it's feasible to further extend the DOF by amplifying the TAG lens's optical power with driving voltage, this adjustment decreases the excitation photons per voxel, diminishing the SNR and potentially hindering model training. If larger volumes are required, beyond the capabilities of the deep-learning denoising algorithm, incorporating hardware solutions like adaptive optics can be advantageous. Adaptive optics addresses optical elements and deep-tissue aberrations, potentially improving the image contrast at 2PM without compromising the efficiency of photon usage (*56*, *57*).

TAG-SPARK versus temporal redundancy training model

We integrate our TAG-lens-based high-speed volumetric system with a deep-learning denoising algorithm, TAG-SPARK, a spatial redundancy-based algorithm, effectively tackles the challenge of artifacts in denoised output from neighboring time points. In contrast to temporal redundancy-based algorithms such as DeepCAD-RT[17] (Figure 6-4).

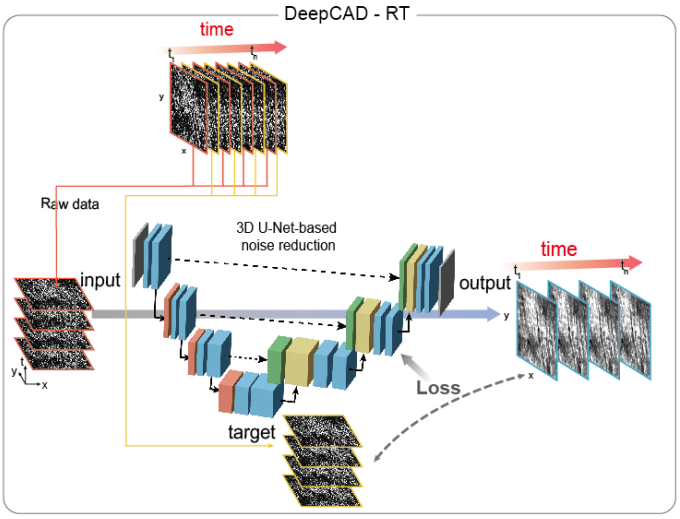


Figure 6-4, General principle of DeepCAD-RT.

The self-supervised training strategy involves using input-target pairs based on temporal redundancy. Consecutive frames in the temporal domain of the original stack are divided into two sub-stacks (red and yellow frames). These sub-stacks serve as the input and target volumes for training the deep neural network.

doi:10.6342/NTU202402518

During subsequent acquisitions, a 3D (x–y–t) window traverses the entire stack to generate the output stack. The loss function is defined by comparing the output sub-stack (blue frame) with the target and trains the model until it converges.

Here, we present a comparative analysis of the same set of cerebellar images to evaluate the denoising capabilities of both methods for structural and functional analysis. In Figure 6-5A, both TAG-SPARK and DeepCAD-RT demonstrate their effectiveness in enhancing image quality, with structural details becoming significantly clearer.

Further, for the functional dynamic analysis, we divide our exploration into two aspects: the signal (ROI1) and the noise (ROI2). The ROI1 depicts cellular responses captured at a single time point. The ROI2 depicts that noise interference enhances the background intensity at a single time point. From Figure 6-5 B and C, it can be observed that the denoised images by TAG-SPARK faithfully capture the cellular response activity and minimize noise interference at other time points. However, in the case of DeepCAD-RT, which dilutes a sharp signal peak, and spreads shot noise contamination to other frames. This is due to the utilization of x, y, and t inputs during the training of the DeepCAD-RT model, enabling the model parameters to learn data from different time points simultaneously. As a result, the output images mix information from neighboring time points, especially when tracing neuronal dynamics under high-speed, low-contrast acquisition, which introduces artifacts.

doi:10.6342/NTU202402518

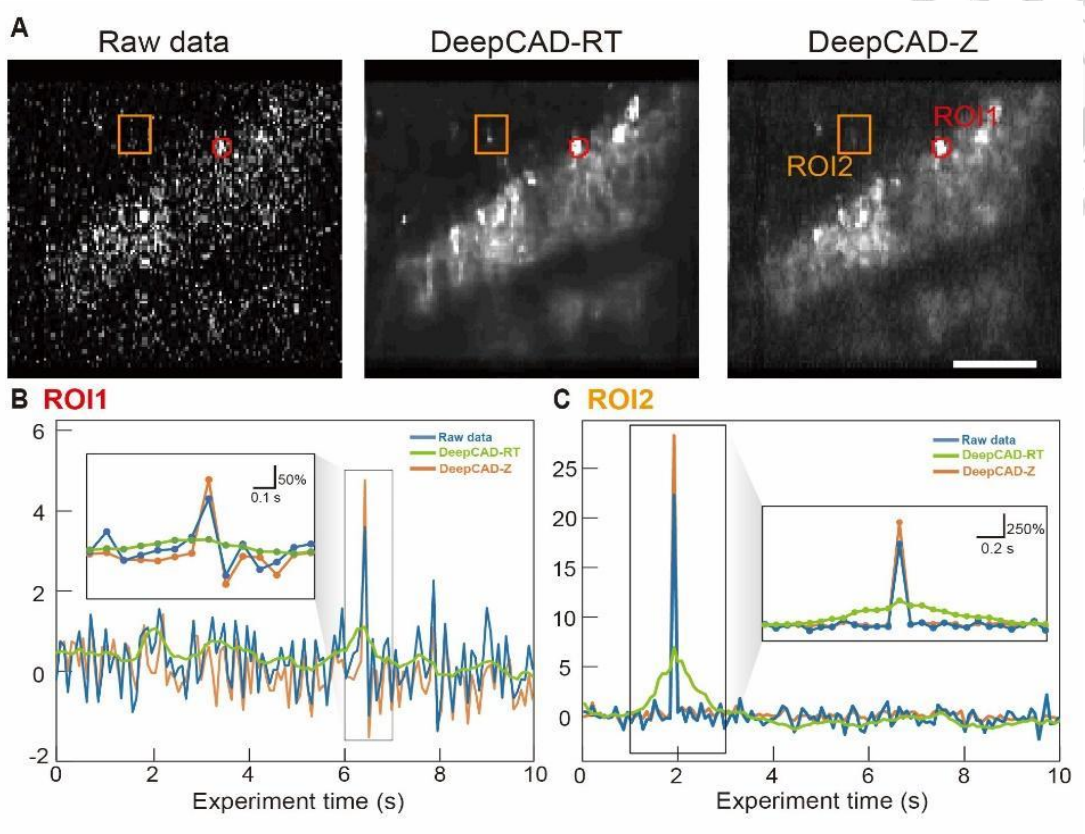


Figure 6-5, Functional analysis with different denoised algorithms.

(A) *In vivo* calcium imaging of PCs: (Left) high-speed scanning image (raw data); (Middle) high-speed scanning image denoised with DeepCAD-RT algorithm, based on temporal redundancy self-supervised learning; (Right) high-speed scanning image denoised with TAG-SPARK. Both DeepCAD-RT and TAG-SPARK satisfactorily improve the SNR. (B) A calcium spike in ROI1 encircles a bright neuron. Although the DeepCAD-RT (green curve) seems to remove noise more efficiently, it smears the spike. On the other hand, TAG-SPARK preserves the spike dynamics nicely. (C) Temporal response of ROI2, which is a region with pure noise. There is a particularly strong noise peak at ~ 2 second recording time, possibly due to cosmic ray irradiation, but DeepCAD-RT spread the noise distribution into other time points.

Therefore, TAG-SPARK emerges as the preferred solution when capturing transient responses, such as rapid spiking in PC dendrites, which is crucial due to its ability to preserve the fidelity of fast-changing events. On the other hand, for relatively slow dynamics, such as the convolutional result in PC soma, DeepCAD-RT offers higher noise

doi:10.6342/NTU202402518

reduction. These two approaches coexist, with neither fully substituting the other. Thus, the mixture of temporal and spatial redundancies, as well as the corresponding deep-learning denoising tools for different compartments of neurons, may generate optimal results with the best SNR for high-speed 3D mapping. For example, in situations where images show spatial continuity without rapid temporal changes. In such cases, a new model combining the strengths of both approaches could potentially achieve better SNR. One straightforward approach is to first apply spatial denoising to the images, followed by temporal denoising for further SNR improvement. Another approach involves developing a new data stacking scheme to create x-y-z-t image stacks as input and target stacks for a network model, such as a 4D U-Net, to learn noise reduction. Therefore, further refinement of TAG-SPARK is possible by integrating advanced network architectures, such as an attention mechanism.

TAG-SPARK versus spatial redundancy supervised training model

In addition to the comparison with the temporal training model DeepCAD-RT, we compare TAG-SPRAK with other "supervised" deep-learning models that utilize spatial information for training, i.e. content-aware image restoration (CARE[18]).(*126*) Since CARE needs ground truth structural images as the training data source to improve the SNR, not a Noise2Noise model, we expect that CARE is superior in denoise performance. In Figure 6-6, we compare the SNR performance (see **Chapter 4.2.2** for SNR analysis) as well as the functional responses of raw data, TAG-SPARK, DeepCAD-RT, and CARE, respectively. CARE offers the best SNR. Nevertheless, because the training focuses on structural information retrieval, CARE completely loses functional information, both fast and slow responses. Therefore, CARE is not suitable for reducing noise in dynamic neural signal imaging. Regarding DeepCAD-RT, if the

doi:10.6342/NTU202402518

functional response is not fast, it gives SNR improvement comparable to CARE, while keeping the functional observation. However, if fast spiking behavior is the target of interest, such as the red area and the inset in Figure 6-6, DeepCAD-RT might create artifacts, and TAG-SPARK gives the best balance between SNR and functional preservation.

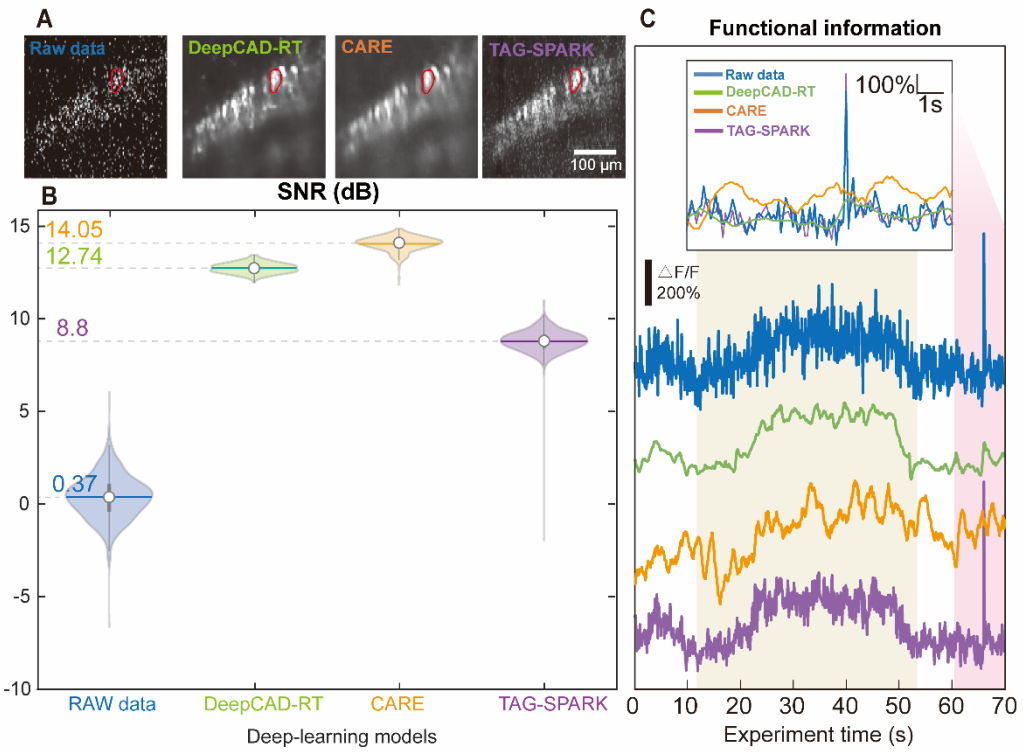


Figure 6-6, Evaluation of different denoise deep-learning models on 2D calcium imaging of PCs.

(A) *In vivo* calcium imaging of PCs: (Left to right) high-speed scanning image of raw data; denoised with DeepCAD-RT algorithm, based on temporal redundancy self-supervised learning; denoised with CARE, based on spatial redundancy supervised learning; denoised with TAG-SPARK. Scale bar: 100 μm. (B) Corresponding SNR analysis shows that DeepCAD-RT and CARE provide comparable enhancement that is better than TAG-SPARK. (C) A calcium spike in ROI encircles a bright neuron in (A). Although both DeepCAD-RT (green curve) and CARE (orange curve) seems to remove noise more efficiently, DeepCAD-RT smears the spike and CARE completely loses the functional information. On the other hand, TAG-SPARK preserves the spike dynamics, especially the spike at ~ 66 seconds of recording time.

doi:10.6342/NTU202402518

In summary, our volumetric system enhanced by the TAG-SPARK algorithm offers significant value beyond technical advancements in biomedical imaging. Besides the example of PCs, most principal neurons and interneurons form complex microcircuits in three dimensions. Two-dimensional imaging techniques struggle to fully capture these microcircuit interactions, limiting our understanding of crucial biological functions, such as the day-night cycle calculated by neurons in the suprachiasmatic nucleus (*127, 128*), or the reward-processing by the interaction between cholinergic interneurons and the medium spiny neurons in the striatum (*129*). High-speed volumetric microscopy integrated with deep-learning noise reduction algorithms provides a deeper insight into the interactions between large neuronal populations and the intricate dynamics of subcellular calcium at the millimeter scale and millisecond level.

Data analysis strategies

Currently, our analysis method involves selecting multiple ROIs and examining the average intensity within each region as it changes with depth and time. However, we determine whether the denoised signal represents noise or a functional signal based solely on behavioral records. This determination assumes that behavioral performance correlates with sensory evocation. To enhance our analysis, we could incorporate additional domain knowledge. For instance, in observing neural responses, a single neuron typically spans multiple pixels. If adjacent pixels display the same neural signal changes, we consider it a true signal. Conversely, if pixel-dependent analysis of adjacent pixels reveals differing response peaks, we should classify it as noise. Therefore, we may achieve effective signal analysis.

doi:10.6342/NTU202402518

REFERENCE

1. D. Kennedy, C. Norman, What Don't We Know? *Science* **309**, 75-75 (2005).
2. 125 questions: Exploration and discovery, In honor of Shanghai Jiao Tong University's 125th Anniversary. *Science*, (2021).
3. M. C. A. Paul Alivisatos1, George M. Church3, Ralph J. Greenspan4, Michael L. Roukes5, and Rafael Yuste6,, The Brain Activity Map Project and the Challenge of Functional Connectomics. *Neuron* **74**, 970–974. (2012).
4. M. E. Raichle, A brief history of human brain mapping. *Trends Neurosci* **32**, 118-126 (2009).
5. G. Stuart, N. Spruston, M. Häusser, *Dendrites* (Oxford University Press, 2016).
6. M. F. Bear, B. W. Connors, M. A. Paradiso, *Neuroscience : exploring the brain* (Wolters Kluwer, Philadelphia, ed. 4th , international, 2016).
7. M. E. Raichle, Two views of brain function. *Trends Cogn Sci* **14**, 180-190 (2010).
8. A. Biasiucci, B. Franceschiello, M. M. Murray, Electroencephalography. *Curr Biol* **29**, R80-R85 (2019).
9. M. Scanziani, M. Hausser, Electrophysiology in the age of light. *Nature* **461**, 930-939 (2009).
10. J. J. Jun, N. A. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. A. Anastassiou, A. Andrei, C. Aydin, M. Barbic, T. J. Blanche, V. Bonin, J. Couto, B. Dutta, S. L. Gratiy, D. A. Gutnisky, M. Hausser, B. Karsh, P. Ledochowitsch, C. M. Lopez, C. Mitelut, S. Musa, M. Okun, M. Pachitariu, J. Putzeys, P. D. Rich, C. Rossant, W. L. Sun, K. Svoboda, M. Carandini, K. D. Harris, C. Koch, J. O'Keefe, T. D. Harris, Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232-236 (2017).
11. C. Forro, D. Caron, G. N. Angotzi, V. Gallo, L. Berdondini, F. Santoro, G. Palazzolo, G. Panuccio, Electrophysiology Read-Out Tools for Brain-on-Chip Biotechnology. *Micromachines (Basel)* **12**, (2021).
12. M. Dipalo, H. Amin, L. Lovato, F. Moia, V. Caprettini, G. C. Messina, F. Tantussi, L. Berdondini, F. De Angelis, Intracellular and Extracellular Recording of Spontaneous Action Potentials in Mammalian Neurons and Cardiac Cells with 3D Plasmonic Nanoelectrodes. *Nano Lett* **17**, 3932-3939 (2017).
13. E. Musk, An Integrated Brain-Machine Interface Platform With Thousands of Channels. *J Med Internet Res* **21**, e16194 (2019).
14. E. Berg, S. R. Cherry, Innovations in Instrumentation for Positron Emission Tomography. *Semin Nucl Med* **48**, 311-331 (2018).
15. T. D. Weber, M. V. Moya, K. Kılıç, J. Mertz, M. N. Economo, High-speed multiplane confocal microscopy for voltage imaging in densely labeled neuronal populations. *Nature Neuroscience* **26**, 1642-1650 (2023).
16. Q. Chen, J. Cichon, W. Wang, L. Qiu, S. J. Lee, N. R. Campbell, N. Destefino, M. J. Goard, Z. Fu, R. Yasuda, L. L. Looger, B. R. Arenkiel, W. B. Gan, G. Feng, Imaging neural activity using Thy1-GCaMP transgenic mice. *Neuron* **76**, 297-308 (2012).
17. S. Weisenburger, A. Vaziri, A Guide to Emerging Technologies for Large-Scale and Whole-Brain Optical Imaging of Neuronal Activity. *Annu Rev Neurosci* **41**, 431-452 (2018).
18. J. T. Russell, Imaging calcium signals in vivo: a powerful tool in physiology and

doi:10.6342/NTU202402518

pharmacology. *Br J Pharmacol* **163**, 1605-1625 (2011).
19. C. Grienberger, A. Giovannucci, W. Zeiger, C. Portera-Cailliau, Two-photon calcium imaging of neuronal activity. *Nat Rev Methods Primers* **2**, (2022).
20. N. Ji, J. Freeman, S. L. Smith, Technologies for imaging neural activity in large volumes. *Nat Neurosci* **19**, 1154-1164 (2016).
21. W. Yang, R. Yuste, In vivo imaging of neural activity. *Nature Methods* **14**, 349-359 (2017).
22. W. Denk, J. H. Strickler, W. W. Webb, Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73-76 (1990).
23. P. Theer, W. Denk, On the fundamental imaging-depth limit in two-photon microscopy. *J Opt Soc Am A Opt Image Sci Vis* **23**, 3139-3149 (2006).
24. F. Helmchen, W. Denk, Deep tissue two-photon microscopy. *Nat Methods* **2**, 932-940 (2005).
25. T. Wang, C. Xu, Three-photon neuronal imaging in deep mouse brain. *Optica* **7**, 947-960 (2020).
26. W. R. Zipfel, R. M. Williams, W. W. Webb, Nonlinear magic: multiphoton microscopy in the biosciences. *Nat Biotechnol* **21**, 1369-1377 (2003).
27. Y. F. Chien, J. Y. Lin, P. T. Yeh, K. J. Hsu, Y. H. Tsai, S. K. Chen, S. W. Chu, Dual GRIN lens two-photon endoscopy for high-speed volumetric and deep brain imaging. *Biomed Opt Express* **12**, 162-172 (2021).
28. C. He, J. Chang, Q. Hu, J. Wang, J. Antonello, H. He, S. Liu, J. Lin, B. Dai, D. S. Elson, P. Xi, H. Ma, M. J. Booth, Complex vectorial optics through gradient index lens cascades. *Nat Commun* **10**, 4264 (2019).
29. H. Uhlirova, M. Stiburek, T. Pikalek, A. Gomes, S. Turtaev, P. Kolbabkova, T. Cizmar, "There's plenty of room at the bottom": deep brain imaging with holographic endo-microscopy. *Neurophotonics* **11**, S11504 (2024).
30. J. C. Jung, A. D. Mehta, E. Aksay, R. Stepnoski, M. J. Schnitzer, In vivo mammalian brain imaging using one- and two-photon fluorescence microendoscopy. *J Neurophysiol* **92**, 3121-3133 (2004).
31. R. S. Fischer, Y. Wu, P. Kanchanawong, H. Shroff, C. M. Waterman, Microscopy in 3D: a biologist's toolbox. *Trends Cell Biol* **21**, 682-691 (2011).
32. L. Bai, Z. Zhang, L. Ye, L. Cong, Y. Zhao, T. Zhang, Z. Shi, K. Wang, Volumetric Imaging of Neural Activity by Light Field Microscopy. *Neurosci Bull* **38**, 1559-1568 (2022).
33. W. Yang, R. Yuste, In vivo imaging of neural activity. *Nat Methods* **14**, 349-359 (2017).
34. G. Sancataldo, L. Silvestri, A. L. Allegra Mascaro, L. Sacconi, F. S. Pavone, Advanced fluorescence microscopy for in vivo imaging of neuronal activity. *Optica* **6**, 758-765 (2019).
35. J. Wu, Y. Liang, S. Chen, C. L. Hsu, M. Chavarha, S. W. Evans, D. Shi, M. Z. Lin, K. K. Tsia, N. Ji, Kilohertz two-photon fluorescence microscopy imaging of neural activity in vivo. *Nat Methods* **17**, 287-290 (2020).
36. R. Li, S. Wang, J. Lyu, K. Chen, X. Sun, J. Huang, P. Sun, S. Liang, M. Li, M. Yang, H. Liu, S. Zeng, X. Chen, L. Li, H. Jia, Z. Zhou, Ten-kilohertz two-photon microscopy imaging of single-cell dendritic activity and hemodynamics in vivo. *Neurophotonics* **10**, 025006 (2023).
37. S. Kang, M. Duocastella, C. B. Arnold, Variable optical elements for fast focus control. *Nature Photonics* **14**, 533-542 (2020).
38. J. Wu, N. Ji, K. K. Tsia, Speed scaling in multiphoton fluorescence microscopy.

doi:10.6342/NTU202402518

Nature Photonics **15**, 800-812 (2021).
39. R. Lu, W. Sun, Y. Liang, A. Kerlin, J. Bierfeld, J. D. Seelig, D. E. Wilson, B. Scholl, B. Mohar, M. Tanimoto, M. Koyama, D. Fitzpatrick, M. B. Orger, N. Ji, Video-rate volumetric functional imaging of the brain at synaptic resolution. *Nat Neurosci* **20**, 620-628 (2017).
40. Z. Qin, Z. She, C. Chen, W. Wu, J. K. Y. Lau, N. Y. Ip, J. Y. Qu, Deep tissue multi-photon imaging using adaptive optics with direct focus sensing and shaping. *Nat Biotechnol* **40**, 1663-1671 (2022).
41. J. Manley, S. Lu, K. Barber, J. Demas, H. Kim, D. Meyer, F. M. Traub, A. Vaziri, Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron* **112**, 1694-1709 e1695 (2024).
42. J. Demas, J. Manley, F. Tejera, K. Barber, H. Kim, F. M. Traub, B. Chen, A. Vaziri, High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beads microscopy. *Nat Methods* **18**, 1103-1111 (2021).
43. D. R. Beaulieu, I. G. Davison, K. Kilic, T. G. Bifano, J. Mertz, Simultaneous multiplane imaging with reverberation two-photon microscopy. *Nat Methods* **17**, 283-286 (2020).
44. A. Song, A. S. Charles, S. A. Koay, J. L. Gauthier, S. Y. Thiberge, J. W. Pillow, D. W. Tank, Volumetric two-photon imaging of neurons using stereoscopy (vTwINS). *Nat Methods* **14**, 420-426 (2017).
45. N. Olivier, A. Mermillod-Blondin, C. B. Arnold, E. Beaurepaire, Two-photon microscopy with simultaneous standard and extended depth of field using a tunable acoustic gradient-index lens. *Opt Lett* **34**, 1684-1686 (2009).
46. G. Thériault, M. Cottet, A. Castonguay, N. McCarthy, Y. De Koninck, Extended two-photon microscopy in live samples with Bessel beams: steadier focus, faster volume scans, and simpler stereoscopic imaging. *Front Cell Neurosci* **8**, 139 (2014).
47. W. J. Shain, N. A. Vickers, B. B. Goldberg, T. Bifano, J. Mertz, Extended depth-of-field microscopy with a high-speed deformable mirror. *Opt Lett* **42**, 995-998 (2017).
48. S. Piazza, P. Bianchini, C. Sheppard, A. Diaspro, M. Duocastella, Enhanced volumetric imaging in 2-photon microscopy via acoustic lens beam shaping. *J Biophotonics* **11**, 11:e201700050. (2018).
49. M. Duocastella, G. Sancataldo, P. Saggau, P. Ramoino, P. Bianchini, A. Diaspro, Fast Inertia-Free Volumetric Light-Sheet Microscope. *ACS Photonics* **4**, 1797-1804 (2017).
50. J. T. Lingjie Kong, Justin P. Little, Yang Yu, Tim Lämmermann, Charles P. Lin Ronald N. Germain, and Meng Cui, Continuous volumetric imaging via an optical phase-locked ultrasound lens. *Nat Methods*, (2015).
51. M. D. B. S. C. B. Arnold, Simultaneous imaging of multiple focal planes for three-dimensional microscopy using ultra-high-speed adaptive optics. (2012).
52. M. Duocastella, S. Surdo, A. Zunino, A. Diaspro, P. Saggau, Acousto-optic systems for advanced microscopy. *Journal of Physics: Photonics* **3**, (2020).
53. C. W. Hsu, C. Y. Lin, Y. Y. Hu, S. J. Chen, Dual-resonant scanning multiphoton microscope with ultrasound lens and resonant mirror for rapid volumetric imaging. *Sci Rep* **13**, 161 (2023).
54. K. J. Hsu, Y. Y. Lin, Y. Y. Lin, K. Su, K. L. Feng, S. C. Wu, Y. C. Lin, A. S. Chiang, S. W. Chu, Millisecond two-photon optical ribbon imaging for

doi:10.6342/NTU202402518

small-animal functional connectome study. *Opt Lett* **44**, 3190-3193 (2019).
55. K. J. Hsu, K. Y. Li, Y. Y. Lin, A. S. Chiang, S. W. Chu, Optimizing depth-of-field extension in optical sectioning microscopy techniques using a fast focus-tunable lens. *Opt Express* **25**, 16783-16794 (2017).
56. C.-L. Chung, T. Furieri, J.-Y. Lin, T.-C. Chang, J.-C. Lee, Y.-F. Chen, M.-K. Pan, S. Bonora, S.-W. Chu, Plug-and-play adaptive optics for two photon high-speed volumetric imaging. *Journal of Physics: Photonics* **4**, 024003 (2022).
57. T. Furieri, A. Bassi, S. Bonora, Large field of view aberrations correction with deformable lenses and multi conjugate adaptive optics. *J Biophotonics* **16**, e202300104 (2023).
58. N. Ji, Adaptive optical fluorescence microscopy. *Nature Methods* **14**, 374-380 (2017).
59. C. Rodriguez, A. Chen, J. A. Rivera, M. A. Mohr, Y. Liang, R. G. Natan, W. Sun, D. E. Milkie, T. G. Bifano, X. Chen, N. Ji, An adaptive optics module for deep tissue multiphoton imaging in vivo. *Nat Methods* **18**, 1259-1264 (2021).
60. C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, C.-W. Lin, Deep learning on image denoising: An overview. *Neural Networks* **131**, 251-275 (2020).
61. R. F. Laine, G. Jacquemet, A. Krull, Imaging in focus: An introduction to denoising bioimages in the era of deep learning. *Int J Biochem Cell Biol* **140**, 106077 (2021).
62. C. Belthangady, L. A. Royer, Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat Methods* **16**, 1215-1225 (2019).
63. K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* **26**, 3142-3155 (2017).
64. C. Qiao, D. Li, Y. Liu, S. Zhang, K. Liu, C. Liu, Y. Guo, T. Jiang, C. Fang, N. Li, Y. Zeng, K. He, X. Zhu, J. Lippincott-Schwartz, Q. Dai, D. Li, Rationalized deep learning super-resolution microscopy for sustained live imaging of rapid subcellular processes. *Nat Biotechnol* **41**, 367-377 (2023).
65. J. Platisa, X. Ye, A. M. Ahrens, C. Liu, I. A. Chen, I. G. Davison, L. Tian, V. A. Pieribone, J. L. Chen, High-speed low-light in vivo two-photon voltage imaging of large neuronal populations. *Nat Methods* **20**, 1095-1103 (2023).
66. J. Lecoq, M. Oliver, J. H. Siegle, N. Orlova, P. Ledochowitsch, C. Koch, Removing independent noise in systems neuroscience data using DeepInterpolation. *Nat Methods* **18**, 1401-1408 (2021).
67. X. Li, G. Zhang, J. Wu, Y. Zhang, Z. Zhao, X. Lin, H. Qiao, H. Xie, H. Wang, L. Fang, Q. Dai, Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. *Nat Methods* **18**, 1395-1400 (2021).
68. X. Li, Y. Li, Y. Zhou, J. Wu, Z. Zhao, J. Fan, F. Deng, Z. Wu, G. Xiao, J. He, Y. Zhang, G. Zhang, X. Hu, X. Chen, Y. Zhang, H. Qiao, H. Xie, Y. Li, H. Wang, L. Fang, Q. Dai, Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit. *Nat Biotechnol* **41**, 282-292 (2023).
69. J. Eccles, R. Llinas, K. Sasaki, Excitation of Cerebellar Purkinje Cells by the Climbing Fibres. *Nature* **203**, 245-246 (1964).
70. M. A. Gaffield, S. B. Amat, H. Bito, J. M. Christie, Chronic imaging of movement-related Purkinje cell calcium activity in awake behaving mice. *J*

doi:10.6342/NTU202402518

Neurophysiol **115**, 413-422 (2016).
71. D. Gatti, L. Rinaldi, L. Ferreri, T. Vecchi, The Human Cerebellum as a Hub of the Predictive Brain. *Brain Sci* **11**, (2021).
72. M. Mansour, Role of Adrenergic Neur gic Neurons in Mot ons in Motor Contr or Control: Examination of ol: Examination of Cerebellar Purkinje Neurons in Mice Following Selective Adrenergic Cell Ablation in Vivo. *Honors Undergraduate Theses* **53**, (2016).
73. K.-S. Han, C. H. Chen, M. M. Khan, C. Guo, W. G. Regehr, Climbing fiber synapses rapidly and transiently inhibit neighboring Purkinje cells via ephaptic coupling. *Nature Neuroscience* **23**, 1399-1409 (2020).
74. M. A. Gaffield, S. B. Amat, H. Bito, J. M. Christie, Chronic imaging of movement-related Purkinje cell calcium activity in awake behaving mice. *J. Neurophysiol.* **115**, 413-422 (2015).
75. C. J. Roome, B. Kuhn, Simultaneous dendritic voltage and calcium imaging and somatic recording from Purkinje neurons in awake mice. *Nat. Commun.* **9**, 3388 (2018).
76. M.-K. Pan, Y.-S. Li, S.-B. Wong, C.-L. Ni, Y.-M. Wang, W.-C. Liu, L.-Y. Lu, J.-C. Lee, E. P. Cortes, J.-P. G. Vonsattel, Q. Sun, E. D. Louis, P. L. Faust, S.-H. Kuo, Cerebellar oscillations driven by synaptic pruning deficits of cerebellar climbing fibers contribute to tremor pathophysiology. *Science Translational Medicine* **12**, eaay1769 (2020).
77. S. H. Kuo, C. Y. Lin, J. Wang, P. A. Sims, M. K. Pan, J. Y. Liou, D. Lee, W. J. Tate, G. C. Kelly, E. D. Louis, P. L. Faust, Climbing fiber-Purkinje cell synaptic pathology in tremor and cerebellar degenerative diseases. *Acta Neuropathol* **133**, 121-138 (2017).
78. R. Mostany, A. Miquelajauregui, M. Shtrahman, C. Portera-Cailliau, Two-photon excitation microscopy and its applications in neuroscience. *Methods Mol Biol* **1251**, 25-42 (2015).
79. N. J. Sofroniew, D. Flickinger, J. King, K. Svoboda, A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* **5**, (2016).
80. J. N. Stirman, I. T. Smith, M. W. Kudenov, S. L. Smith, Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat Biotechnol* **34**, 857-862 (2016).
81. C. Maibohm, O. F. Silvestre, J. Borme, M. Sinou, K. Heggarty, J. B. Nieder, Multi-beam two-photon polymerization for fast large area 3D periodic structure fabrication for bioapplications. *Scientific Reports* **10**, 8740 (2020).
82. J. L. Fan, J. A. Rivera, W. Sun, J. Peterson, H. Haeberle, S. Rubin, N. Ji, High-speed volumetric two-photon fluorescence imaging of neurovascular dynamics. *Nature Communications* **11**, 6020 (2020).
83. W. Zong, R. Wu, M. Li, Y. Hu, Y. Li, J. Li, H. Rong, H. Wu, Y. Xu, Y. Lu, H. Jia, M. Fan, Z. Zhou, Y. Zhang, A. Wang, L. Chen, H. Cheng, Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice. *Nat Methods* **14**, 713-719 (2017).
84. F. Helmchen, M. S. Fee, D. W. Tank, W. Denk, A Miniature Head-Mounted Two-Photon Microscope: High-Resolution Brain Imaging in Freely Moving Animals. *Neuron* **31**, 903-912 (2001).
85. E. M. a. C. B. Arnold, Mechanics and refractive power optimization of tunable acoustic gradient lens. *Journal of Applied Physics* **102**, 033104 (2007).

doi:10.6342/NTU202402518

86. E. M. a. C. B. Arnold, Optical analysis of time-averaged multiscale Bessel beams generated by a tunable acoustic gradient index of refraction lens. (2008).
87. J. W. Goodman, M. E. Cox, Introduction to Fourier Optics. *Physics Today* **22**, 97-101 (1969).
88. E. M. Alexandre Mermillod-Blondin, and Craig B. Arnold, High-speed varifocal imaging with a tunable acoustic gradient index of refraction lens. *Opt. Lett.* **33**, (2008).
89. G. Sancataldo, L. Scipioni, T. Ravasenga, L. Lanzanò, A. Diaspro, A. Barberis, M. Duocastella, Three-dimensional multiple-particle tracking with nanometric precision over tunable axial ranges. *Optica* **4**, 367-373 (2017).
90. A. M.-B. Nicolas Olivier, Craig B. Arnold, and Emmanuel Beaurepaire, Two-photon microscopy with simultaneous standard and extended depth of field using a tunable acoustic gradient-index lens. *Opt. Lett.* **34**, (2009).
91. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila, paper presented at the Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2018.
92. Ö. Ç. a. A. A. a. S. S. L. a. T. B. a. O. Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. (2016).
93. A. F. Calvarons, Improved Noise2Noise Denoising with Limited Data. (2021).
94. W. Jung, H. S. Lee, M. Seo, Y. Nam, Y. Choi, N. Y. Shin, K. J. Ahn, B. S. Kim, J. Jang, MR-self Noise2Noise: self-supervised deep learning-based image quality improvement of submillimeter resolution 3D MR images. *Eur Radiol* **33**, 2686-2698 (2023).
95. O. Ronneberger, P. Fischer, T. Brox, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi, Eds. (Springer International Publishing, Cham, 2015), pp. 234-241.
96. F. Helmchen, W. Denk, Deep tissue two-photon microscopy. *Nature Methods* **2**, 932-940 (2005).
97. J. R. Bumstead, J. J. Park, I. A. Rosen, A. W. Kraft, P. W. Wright, M. D. Reisman, D. C. Côté, J. P. Culver, Designing a large field-of-view two-photon microscope using optical invariant analysis. *Neurophotonics* **5**, 025001 (2018).
98. H. Jang, Y. Li, A. A. Fung, P. Bagheri, K. Hoang, D. Skowronska-Krawczyk, X. Chen, J. Y. Wu, B. Bintu, L. Shi, Super-resolution SRS microscopy with A-PoD. *Nat Methods* **20**, 448-458 (2023).
99. Y. Li, Y. Su, M. Guo, X. Han, J. Liu, H. D. Vishwasrao, X. Li, R. Christensen, T. Sengupta, M. W. Moyle, I. Rey-Suarez, J. Chen, A. Upadhyaya, T. B. Usdin, D. A. Colon-Ramos, H. Liu, Y. Wu, H. Shroff, Incorporating the image formation process into deep learning improves network performance. *Nat Methods* **19**, 1427-1437 (2022).
100. K. Agarwal, R. Machan, Multiple signal classification algorithm for super-resolution fluorescence microscopy. *Nat Commun* **7**, 13752 (2016).
101. D. Sage, L. Donati, F. Soulez, D. Fortun, G. Schmit, A. Seitz, R. Guiet, C. Vonesch, M. Unser, DeconvolutionLab2: An open-source software for deconvolution microscopy. *Methods* **115**, 28-41 (2017).
102. M. Pachitariu, C. Stringer, Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634-1641 (2022).
103. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist

doi:10.6342/NTU202402518

algorithm for cellular segmentation. *Nat Methods* **18**, 100-106 (2021).
104. B. Yang, C. J. Richards, T. B. Gandek, I. de Boer, I. Aguirre-Zuazo, E. Niemeijer, C. Åberg, Following nanoparticle uptake by cells using high-throughput microscopy and the deep-learning based cell identification algorithm Cellpose. *Frontiers in Nanotechnology* **5**, (2023).
105. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**, 1226-1232 (2019).
106. C. M. Lee, X. Tian, C. Tsao, P. Chen, T. N. Huang, Y. P. Hsueh, B. C. Chen, Macro Photography with Lightsheet Illumination Enables Whole Expanded Brain Imaging with Single-cell Resolution. *Discoveries (Craiova)* **9**, e133 (2021).
107. Y. Chen, X. Li, D. Zhang, C. Wang, R. Feng, X. Li, Y. Wen, H. Xu, X. S. Zhang, X. Yang, Y. Chen, Y. Feng, B. Zhou, B. C. Chen, K. Lei, S. Cai, J. M. Jia, L. Gao, A Versatile Tiling Light Sheet Microscope for Imaging of Cleared Tissues. *Cell Rep* **33**, 108349 (2020).
108. Y. C. Tsai, W. C. Tang, C. S. L. Low, Y. T. Liu, J. S. Wu, P. Y. Lee, L. Q. Chen, Y. L. Lin, P. Kanchanawong, L. Gao, B. C. Chen, Rapid high resolution 3D imaging of expanded biological specimens with lattice light sheet microscopy. *Methods* **174**, 11-19 (2020).
109. T. Lin, C. Li, J. Liu, B. H. Smith, H. Lei, X. Zeng, Glomerular Organization in the Antennal Lobe of the Oriental Fruit Fly Bactrocera dorsalis. *Front Neuroanat* **12**, 71 (2018).
110. C. M. Root, K. I. Ko, A. Jafari, J. W. Wang, Presynaptic facilitation by neuropeptide signaling mediates odor-driven food search. *Cell* **145**, 133-144 (2011).
111. A. F. Silbering, R. Rytz, Y. Grosjean, L. Abuin, P. Ramdya, G. S. Jefferis, R. Benton, Complementary function and integrated wiring of the evolutionarily distinct Drosophila olfactory subsystems. *J Neurosci* **31**, 13357-13375 (2011).
112. P. Stockinger, D. Kvitsiani, S. Rotkopf, L. Tirian, B. J. Dickson, Neural circuitry that governs Drosophila male courtship behavior. *Cell* **121**, 795-807 (2005).
113. L. B. Vosshall, R. F. Stocker, Molecular architecture of smell and taste in Drosophila. *Annu Rev Neurosci* **30**, 505-533 (2007).
114. D. Task, C. C. Lin, A. Vulpe, A. Afify, S. Ballou, M. Brbic, P. Schlegel, J. Raji, G. Jefferis, H. Li, K. Menuz, C. J. Potter, Chemoreceptor co-expression in Drosophila melanogaster olfactory neurons. *Elife* **11**, (2022).
115. J. A. Horne, C. Langille, S. McLin, M. Wiederman, Z. Lu, C. S. Xu, S. M. Plaza, L. K. Scheffer, H. F. Hess, I. A. Meinertzhagen, A resource for the Drosophila antennal lobe provided by the connectome of glomerulus VA1v. *Elife* **7**, (2018).
116. Y. Zhao, M. Zhang, W. Zhang, Y. Zhou, L. Chen, Q. Liu, P. Wang, R. Chen, X. Duan, F. Chen, H. Deng, Y. Wei, P. Fei, Y. H. Zhang, Isotropic super-resolution light-sheet microscopy of dynamic intracellular structures at subsecond timescales. *Nat Methods* **19**, 359-369 (2022).
117. T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, M. W. Mathis, Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols* **14**, 2152-2176 (2019).

doi:10.6342/NTU202402518

118. J. Lauer, M. Zhou, S. Ye, W. Menegas, S. Schneider, T. Nath, M. M. Rahman, V. Di Santo, D. Soberanes, G. Feng, V. N. Murthy, G. Lauder, C. Dulac, M. W. Mathis, A. Mathis, Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods* **19**, 496-504 (2022).
119. M. H. Hastings, E. S. Maywood, M. Brancaccio, Generation of circadian rhythms in the suprachiasmatic nucleus. *Nat Rev Neurosci* **19**, 453-469 (2018).
120. R. Cachope, Y. Mateo, B. N. Mathur, J. Irving, H. L. Wang, M. Morales, D. M. Lovinger, J. F. Cheer, Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. *Cell Rep* **2**, 33-41 (2012).
121. Y. Zang, E. De Schutter, The Cellular Electrophysiological Properties Underlying Multiplexed Coding in Purkinje Cells. *J Neurosci* **41**, 1850-1863 (2021).
122. L. Kong, J. Tang, J. P. Little, Y. Yu, T. Lämmermann, C. P. Lin, R. N. Germain, M. Cui, Continuous volumetric imaging via an optical phase-locked ultrasound lens. *Nat Methods* **12**, 759-762 (2015).
123. H. Har-Gil, L. Golgher, S. Israel, D. Kain, O. Cheshnovsky, M. Parnas, P. Blinder, PySight: plug and play photon counting for fast continuous volumetric intravital microscopy. *Optica* **5**, 1104-1112 (2018).
124. C.-W. Hsu, C.-Y. Lin, Y. Y. Hu, S.-J. Chen, Dual-resonant scanning multiphoton microscope with ultrasound lens and resonant mirror for rapid volumetric imaging. *Scientific Reports* **13**, 161 (2023).
125. M. Zhai, J. Yu, Y. Hu, H. Yu, B. Xie, Y. Yu, D. Li, A. Wang, H. Cheng, Dual-objective two-photon microscope for volumetric imaging of dense scattering biological samples by bidirectional excitation and collection. *Photonics Research* **12**, 1351-1361 (2024).
126. M. Weigert, U. Schmidt, T. Boothe, A. Muller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, E. W. Myers, Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* **15**, 1090-1097 (2018).
127. C. C. Lee, F. Liang, I. C. Lee, T. H. Lu, Y. Y. Shan, C. F. Jeng, Y. F. Zou, H. T. Yu, S. K. Chen, External light-dark cycle shapes gut microbiota through intrinsically photosensitive retinal ganglion cells. *EMBO reports* **23**, e52316 (2022).
128. H.-Y. Hsiao, C.-M. Liu, H.-w. Lin, T.-C. Liu, C.-L. Lu, Y.-S. Huang, C. Chen, K. N. Tu, Unidirectional growth of microbumps on (111)-oriented and nanotwinned copper. *Science* **336**, 1007-1010 (2012).
129. G. Laverne, J. Pesce, A. Reynders, E. Combrisson, E. Gascon, C. Melon, L. Kerkerian-Le Goff, N. Maurice, C. Beurrier, Cholinergic interneuron inhibition potentiates corticostriatal transmission in direct medium spiny neurons and rescues motor learning in parkinsonism. *Cell Rep* **40**, 111034 (2022).

doi:10.6342/NTU202402518

FOOTNOTE†

[1] TAG lens manuals (TAG lens LLC): https://reurl.cc/0v7bW9

(Mitutoyo TAG lens: https://reurl.cc/9v2Akx)

[2] Behavior recording-IR camera: https://reurl.cc/6vrWQM

[3] ABCD matrix for DOF simulation:

https://github.com/AllyHsieh/SWChulab/tree/main/SimulationDOF

[4] Correlation analysis: https://github.com/AllyHsieh/SWChulab/tree/main/Correlation

[5] Reconstruction of high-speed volumetric images:

https://github.com/AllyHsieh/SWChulab/tree/main/reconstruction_fastscan_v9

[6] ImageJ-Deconvolutionlab2:

https://github.com/AllyHsieh/SWChulab/blob/main/Deconvolution.md

[7] Cellpose: https://github.com/MouseLand/cellpose

[8] Ilastik: https://www.ilastik.org/

[9] PCA: https://github.com/AllyHsieh/SWChulab/tree/main/PCA

[10] PSNR and PCC:

https://github.com/AllyHsieh/SWChulab/tree/main/Moving%20average%20analysis%20for%20PSNR

[11] T-Test and ANOVA analysis: https://www.graphpad.com/quickcalcs/contMenu/

[12] TAG-SAPRK: https://github.com/AllyHsieh/SWChulab/tree/main/DeepCAD-Z

[13] 29-4 mouse experiment data: https://reurl.cc/lQG8W6

[14] Behavior recording- Accelerometer (OpenEphys): https://reurl.cc/DjGaj6

[15] DeepLabCut: https://github.com/DeepLabCut/DeepLabCut

---

† Please log in to the lab account to ensure access to the data save in Google cloud.

doi:10.6342/NTU202402518

[16] 4D high-speed scanning image movie: https://reurl.cc/LWRY54

[17] DeepCAD-RT: https://github.com/cabooster/DeepCAD-RT

(Drive: https://reurl.cc/z1xGj6)

[18] CARE (Drive): https://reurl.cc/yLzbbO

doi:10.6342/NTU202402518