MICROSCOPY

Microscopy, 2020, 79–91
doi: 10.1093/jmicro/dfaa007
Advance Access Publication Date: 26 March 2020

Invited Special Issue

Generative and discriminative model-based approaches to microscopic image restoration and segmentation

Shin Ishii[1,2,3,*], Sehyung Lee[1,3], Hidetoshi Urakubo[1], Hideaki Kume[3,4] and Haruo Kasai[3,4]

[1]Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, [2]ATR Neural Information Analysis Laboratories, Kyoto 619-0288, Japan, [3]International Research Center for Neurointelligence, The University of Tokyo, Tokyo 113-0033, Japan, and [4]Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

*To whom correspondence should be addressed. E-mail: ishii@i.kyoto-u.ac.jp

Received 1 October 2019; Revised 2 February 2020; Editorial Decision 10 February 2020

Abstract

Image processing is one of the most important applications of recent machine learning (ML) technologies. Convolutional neural networks (CNNs), a popular deep learning-based ML architecture, have been developed for image processing applications. However, the application of ML to microscopic images is limited as microscopic images are often 3D/4D, that is, the image sizes can be very large, and the images may suffer from serious noise generated due to optics. In this review, three types of feature reconstruction applications to microscopic images are discussed, which fully utilize the recent advancements in ML technologies. First, multi-frame super-resolution is introduced, based on the formulation of statistical generative model-based techniques such as Bayesian inference. Second, data-driven image restoration is introduced, based on supervised discriminative model-based ML technique. In this application, CNNs are demonstrated to exhibit preferable restoration performance. Third, image segmentation based on data-driven CNNs is introduced. Image segmentation has become immensely popular in object segmentation based on electron microscopy (EM); therefore, we focus on EM image processing.

Key words: image processing, image super-resolution, Bayesian estimation, maximum likelihood estimation, deep learning, image segmentation

Introduction

How do we recover image features of our interests from the degraded images acquired by optical/electron microscopy? Because the microscopic image acquisition essentially includes a degradation process, their recovery is not straightforward, and hence, it becomes a difficult inverse problem. The optics of fluorescent microscopy suffers from the diffraction of photons, which is represented by point spread function (PSF). Although one may think that deconvolution with the PSF would be able to recover the original image, it is not the case indeed. Even if the PSF of the microscope is available, the detailed process, from which precise points in the target tissue a set of photons measured at a single detector (like a photon multiplier) have been produced, includes uncertainty, that is, the photon diffraction process is generally irreversible. Furthermore, the detectors may suffer from shot noises, which are also irreversible because of the difficulty in identifying every detail of individual shot noises. Such irreversibility makes the process to obtain microscopic images, which is called forward optics, a data degeneration process. Thus, reverse optics

(opposite of forward optics) should be an ill-posed inverse problem. In other words, numerous possible original images can be obtained from a single observed image.

In this review, we discuss three kinds of microscopic image processing techniques which fully utilize the recent advancements in machine learning (ML) technologies. Their targets are to recover high-resolution (HR) images based on blurred and low-resolution (LR) images, which is termed image super-resolution or image restoration, and to recover image features such as cellular attribution, which is a type of image segmentation. In the last two or three decades, a number of ML methods with constrained optimization have been proposed, and very recently, they have partly been replaced with data-driven deep learning methods. In terms of statistical ML, these two are different in their approaches; the former is a generative model-based approach and the latter is a discriminative model-based approach [1], but they share the same objective, which is to recover original image features by solving ill-posed inverse problems stemming from the irreversible forward optics.

The first topic discussed in this review is the integration of a set of (that is, multiple) degraded images into a single HR image. This is software-based, multi-frame image super-resolution [2], which is different from hardware-based super-resolution [3,4]. Software-based image super-resolution could be economical and is applicable in various situations irrespective of the specifications of the measurement hardware, but it requires prior knowledge of the measurement environments. In this case, we presume that the set of multiple images comprises more information than a single image. The second topic is data-driven image-restoration of a LR image, even from a single shot, by fully employing a database of pairs of low- and HR images [5]. Recent advancements in convolutional neural networks (CNNs) [6,7] have enabled the development of non-linear filters with the full usage of huge databases. We here demonstrate, with quantitative evaluation, how a variant of CNN, 3D U-Net, can restore blurred and noise-contaminated images. Even if the acquired images have sufficiently rich information, moreover, recovering the attribution of each pixel to either of multiple cells or sub-cellular structures is not an easy task, which is a typical instance of image segmentation and the third topic of this review. The most prominent example can be seen in image processing of electron microscopy (EM). Although EMs can clearly capture membranes, they do exhibit not only neuronal/glial membranes but also those of sub-cellular structures such as mitochondria. Such complicated situations make the image segmentation from EM a non-trivial task. We would like to discuss that 2D/3D CNNs and their variants are now state-of-the-art in such an EM-based image segmentation problem.

The remainder of this review is organized as follows. First, the conventional ML technologies with the generative model-based approach for, in particular, multi-frame super-resolution is discussed in section 'Generative-model based approaches to multi-frame super-resolution'. This section is also important for showing mathematical formulations to solve difficult inverse problems that underly in many feature extraction problems from microscopic images. In section 'Deep learning-based image super-resolution and restoration', image restoration from single frames, which is a similar problem to the super-resolution from multiple frames, is discussed. Here, data-driven approaches like those with deep learning are in recent trends. We show quantitative comparisons between model-based and deep learning-based approaches. In section 'EM image segmentation', EM-based image segmentation is discussed, with a particular interest in the usage of data-driven 2D/3D CNNs. Since there have been many comparison studies including open challenges in this topic, we put our focus on the qualitative descriptions of the current state-of-the-art methods.

Generative model-based approaches to multi-frame super-resolution

The objective of this section is to discuss generative model-based statistical technologies to deal with microscopic images, in a particular application to multi-frame super-resolution. It is also important to show mathematical formulations to solve difficult inverse problems that underly in many feature extraction problems from microscopic images. We first discuss the difficulty underlying the inverse-optic problems. After presenting the conventional maximum likelihood formulation, we discuss a constrained optimization approach, i.e. maximum a posteriori (MAP) formation, and more advanced integration approach, i.e. Bayesian formulation (Table 1). We show some demonstrations when applied to 2D natural images and 4D microscopic images.

Preliminaries

Let x be a HR image and y_t be the tth LR image downsampled from x. We assume that there are multiple observations, i.e. $t = 1,...,T$, where T is the number of observations. As we attempt to restore a single HR image x based on multiple LR images $\{y_1,...,y_t,...,y_T\}$, this problem is known as multi-frame image super-resolution. We assume that each LR image has been observed after different combinations of image rotation, image shift, image blurring due to the photon diffraction, and image downsampling; these processes can be collectively represented as an imaging parameter θ_t. As the photon energies measured by each photon detector are additive, the entire process becomes linear for the HR image x, if the parameters θ_t are known in advance, that is

$$y_t = W(\theta_t)\, x + n_t \quad (t = 1, \ldots, T). \tag{1}$$

Note that W is a linear matrix whose row and column dimensionalities are the numbers of pixels of the low- and HR images, y and x, respectively, and is a product of matrices each representing image rotation, image shift, image blurring or image downsampling. Usually, however, it is non-linear with respect to parameter θ. We also assume that the shot noise n_t obeys a spatio-temporally independent (i.e. white) Gaussian noise

$$n_t \sim N(0, \beta^{-1} I), \tag{2}$$

where $\beta > 0$ denotes a spatially uniform precision, and I is the identity matrix with the same dimensionality as that of the measured image.

According to the formulation of Bayesian super-resolution [8], the inverse problem of Eqs. (1) and (2) can be described as

$$p\left(x \middle| \{y_1, \ldots, y_T\}\right) = \frac{p(x) \prod_{t=1}^{T} \int p\left(y_t | x, \theta_t\right) p\left(\theta_t\right) d\theta_t}{p\left(\{y_1, \ldots, y_T\}\right)}, \tag{3}$$

where $p(x)$ is the prior for the HR image, which represents the naturality of the target objects in the image plane, and $p(\theta_t)$ is the prior for the observation parameters. In the subsequent subsections, we often either do not define the prior $p(\theta_t)$ explicitly or assume it to be uniform. Furthermore, we resolve the integration with respect to

the parameter θ using the expectation-maximization algorithm [9]. The probability $P(y_t|x,\theta_t)$ is known as the likelihood and is defined by the forward optics, Eqs. (1) and (2). Typical natural images comprise edges and smooth textures surrounded by multiple edges; this characteristic may be represented as a stochastic process called line process [10]. If the target objects observed by microscopy have such a characteristic, we use the prior $p(x)$ to represent the line process. In contrast, if the target has only a few blight points, like in the dark-field microscopy, we may employ sparseness prior to obtain a HR image with few white pixels. Conventionally, popular prior $p(x)$ has been defined as a band-diagonal Gaussian, which represents the local smoothness of the given image. If the prior and the likelihood are given by smooth Gaussian and the spatially uniform Gaussian, respectively, the deconvolution process with these models can be considered as the well-established normalized linear filtering [2,11]. If the prior is presented as a line process, the posterior distribution of x becomes a mixture of Gaussians, with a large, finite number of Gaussian components. As a result, the inverse problem becomes intractable; thus, an approximation for obtaining the posterior, Eq. (3), is required [8]. Note that if we use the simplest Gaussian prior, the naive application of Eq. (3) cannot break the diffraction limit. To realize subdiffraction limit, we need appropriate priors [8], other kinds of information like that in the temporal domain [12] and/or uncertainty resolution based on Bayesian integration.

Maximum likelihood and MAP estimation

According to the simplest idea to solve the super-resolution problem, Eqs. (1) and (2), we obtain the combination of x and the set of parameters, $\{\theta_1,...,\theta_T\}$, that maximizes the likelihood

$$[X,\{\theta_t\}] = \underset{x,\{\theta_t\}}{\operatorname{argmax}} \sum_{t=1}^{T} \left\| y_t - W(\theta_t)\, x \right\|^2, \tag{4}$$

which is known as the maximum likelihood estimation (MLE) and equivalent to the standard least square estimation. As the dependence of the objective function on the parameter set is non-linear, an iterative algorithm, such as the expectation-maximization algorithm [13], is required to solve this problem. Moreover, it is noteworthy that if the blurring function (i.e. PSF) is just Gaussian, which may be different from the realistic microscopic diffraction process, the reduced problem becomes much simpler and has been used in many single-molecular morphology estimation problems. This scenario is known as super-resolution based on registration, because the estimation of the parameter set $\{\theta_t\}$ corresponds to the registration (estimation of rotation and shift mostly, because the blur is just Gaussian) over multiple LR images.

However, this type of MLE may not work well because the non-linear combination of the rotation and shift makes the indeterminacy of the HR image x severe. This serious indeterminacy would make the estimation algorithm vulnerable. One possible modification to address this issue is to introduce regularization.

$$[X,\{\theta_t\}] = \underset{x,\{\theta_t\}}{\operatorname{argmax}} \sum_{t=1}^{T} \left\| y_t - W(\theta_t)\, x \right\|^2 + \lambda \left\| \theta_t - \theta_O \right\|^2 + \eta \|Lx\|^2, \tag{5}$$

where θ_0 is a rough estimation of the observation parameter, and L represents a high-pass filter, similar to the Laplacian filter. The optimization of Eq. (5) corresponds to the MAP estimation, which is an advanced version of the previous MLE. The second and third terms in Eq. (5) present the preference to estimate the rather consistent shift/rotation between different images and that to estimate HR image comprising a smaller number of edges and several smooth textures. These terms can also be seen as our constraints on the estimation and are called priors. Here, one of the simplest priors for the HR image, Laplacian, is presented. A more powerful one is to prefer sparseness in spatial gradients of the HR image x, which is called total variation (TV) regularization [14]. In the context of data-driven image restoration (see section 'Deep learning-based image super-resolution and restoration'), we will use this TV regularization method as a representative of generative model-based methods like what we discussed in this section. Although the MAP estimation resembles the original Bayesian formation of the used multi-frame super-resolution, Eq. (3), it does not involve the integration as in the case with Bayesian formation.

Bayesian super-resolution

Because of the regularization stemming from the prior, the MAP estimation often produces stable solutions more than the MLE. However, it could be still difficult to obtain a good restored image using the MAP estimation because of the complicated and non-linear relationship between the registration parameters and the HR image to be estimated. Such non-linearity often produces enumerable combinations of sub-optimal registration parameters, thereby making the estimation of the HR image difficult, even though the priors could ease the difficulty to some extent. A possible solution to further resolve this indeterminacy is to introduce integration to the HR image. Because the integration copes with the uncertain estimation of the HR image, the indeterminacy when estimating the registration parameters is expected to be reduced.

The Bayesian super-resolution dates back to the seminal work by Tipping and Bishop [15]. They attempted to apply the expectation-maximization algorithm to increase the marginal likelihood, often known as evidence, Eq. (3), in terms of x and the parameter set $\{\theta_1,...,\theta_T\}$, simultaneously. After estimation, the posterior distribution of the HR image is given by

$$p\left(x|\left\{y_1,\ldots,y_T\right\},\left\{\theta_1,\ldots,\theta_T\right\}\right) = \frac{p(x)p\left(\{t_t\}\,|\,\{\theta_t\},x\right)}{\int p(x)p\left(\left\{y_t\right\}|\,\{\theta_t\},x\right)dx} = N\left(x|\mu,\Sigma\right), \tag{6}$$

where

$$\Sigma = \left(\alpha L^T L + \beta \sum_{t=1}^{T} W(\theta_t)^T W(\theta_t)\right)^{-1} \text{ and } \mu = \beta\Sigma\left(\sum_{t=1}^{T} W(\theta_t)^T y_t\right). \tag{7}$$

This estimation is equivalent to the following optimization

$$\left\{\hat{\theta}_t\right\} = ¥\operatorname{argmin}_{\{\theta_t\}} \beta \sum_{t=1}^{T} \left\| y_t - W(\theta_t)\, x \right\|^2 + \alpha \|L\mu\|^2 - \log|\,\Sigma\,|\,. \tag{8}$$

Although this objective function resembles a typical constrained square loss, it includes the estimation of the covariance matrix Σ of the posterior distribution of x. This non-linear effect stemming from the uncertain estimation of x is helpful to resolve the complex

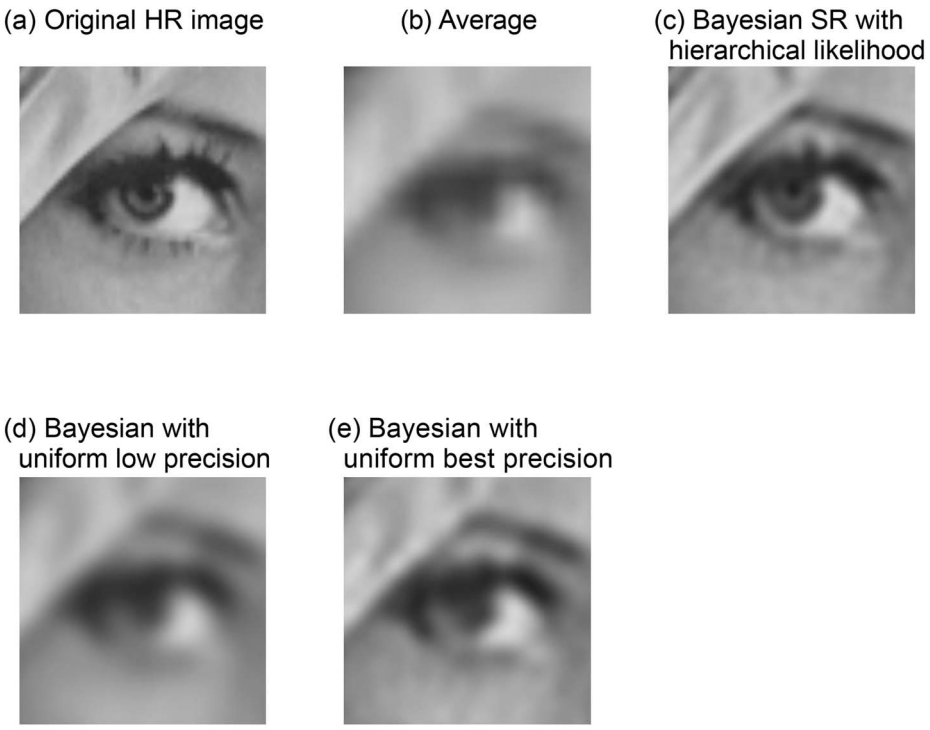

Fig. 1. Multi-frame image super-resolution applied to the eye part of a natural 'LENA' image. (a) The original high-resolution (HR) image; 15 ($T = 15$) low-resolution (LR) images with different registration parameters and different occlusion patterns are obtained from the HR image. As this is a simulation study, we applied a Gaussian PSF, whose precision (inverse variance) was small and large on occluded and non-occluded LR pixels, respectively, and one-sixteenth down-sampling to obtain a single LR image. Shift and rotation were also applied individually to each LR image. (b) An average image over the 15 partially occluded LR images, PNSR = 20.85 dB. (c) Bayesian super-resolved image with a hierarchical likelihood model (i.e. a non-uniform PSF) to deal with the partial occlusion, 31.38 dB (highest). (d) Bayesian with a non-hierarchical likelihood model (i.e. a uniform PSF) whose noise precision was lowly estimated, 25.57 dB. (e) Bayesian with a uniform PSF whose noise precision was best tuned, 28.06 dB [17]. The figure reproduction was permitted by the Journal of Systems Science and Complexity (Springer Nature Switzerland AG).

relationship between the registration parameters $\{\theta_t\}$ and the HR image x, which would stabilize the estimation of the registration parameter.

In contrast, Pickup *et al*. [16] presented the integration over the registration parameters. Their objective function was defined as

$$p\left(x|\left\{y_t\right\}\right) = \frac{p(x)}{p\left(\left\{y_t\right\}\right)} \int p\left(\left\{\theta_t\right\}\right) p\left(\left\{y_t\right\}|\left\{\theta_t\right\}, x\right) d\left\{\theta_t\right\}, \qquad 9)$$

which is the marginalized version of the posterior distribution (Eq. (3)) with respect to the registration parameters $\{\theta_t\}$. To perform the marginalization with respect to the registration parameters, one needs a prior for them, and Pickup *et al*. used the Huber distribution that enhances the edge preservation of the HR image. This is a suitable approach because the uncertain estimation of the registration parameters could critically affect the estimation of the HR image x. They used the Laplace approximation to estimate the parameter posterior and simplified the integration of the marginal likelihood Eq. (9) with respect to the parameter. Although the Laplace approximation cannot address the multi-modality of the complicated (i.e. probably multi-modal) distribution of the registration parameters, it is a practical solution for performing the Bayesian super-resolution within a reasonable computation time.

Figure 1 demonstrates the multi-frame super-resolution of two-dimensional natural images, LENA, where $T = 15$. Here, a hierarchical likelihood model was used to cope with the structural occlusion (represented as a highly noisy region) independently applied to each image [17]. Figure 2 shows another demonstration when applied to the 4D (XYZT) two-photon microscopic image of neurites from an *in vivo* (anesthetized) mouse. In this *in vivo* case, we regarded a series of images taken in the timelapse manner as a set of multiple images,

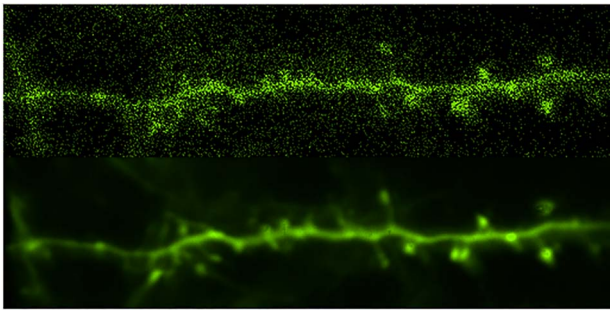

Fig. 2. Multi-frame super-resolution applied to the fluorescent 4D (two photon microscopy, XYZT) image sequences of a neuronal fiber from an *in vivo* (anesthetized) mouse. Upper: one frame image (integrated over the Z-direction). There are blurs and shot noises. Original image sequence was taken with dual colors (green: morphologies, red: calcium activities), with the image size of 1024 (pixels) × 256 (pixels) × 21 (slices) × 760 (frames), and the voxel size of 72.2 × 72.2 × 1008.5 (nm^3). We used only the green channel for this demonstration. Red color was removed. Lower: super-resolved image. Neural fiber structure is colored green, showing the registered image over time, but with enhancement in pixel number (that is super-resolution). Because the image movements have been calibrated, some blurs can be observed due to the imperfect registration over time. Regardless, the image has higher resolution and reduced noise than the original one frame image shown in the upper panel.

that is the registration parameters are correlated between adjacent image frames. Considering such a time-series factor, we applied a linear dynamical system-based modeling to the parameter estimation, whereas we avoided the integration with respect to the parameters or the HR image, due to the complexity of the integration. However, a very HR image has been obtained by the used simplified Bayesian (indeed, MAP) estimation.

Deep learning-based image super-resolution and restoration

This section introduces the ML technologies of image restoration from singleframe microscopic images, this is a mathematical similar problem to the image super-resolution from multiple frames that was discussed in section 'Generative-model based approaches to multi-frame super-resolution'. For this problem, the constrained optimization approach like the MAP estimation with TV regularization was the conventional standard, but data-driven deep learning-based approach has been developed as a new standard. Reflecting this trend, we show quantitative comparisons between generative model-based like TV regularization and discriminative model-based (i.e. deep learning-based) approaches in this section.

Image restoration has generally relied on supervised ML, that is on a database comprising pairs of original images and their corresponding annotated (labeled) images. Each annotated image can be a clean image denoised/deblurred from the original image. Several works address the effectiveness of image restoration in the field of fluorescent microscopy [18–20]. Also, there are recent deep learning-based denoising methods [21,22] that also share their source codes including pre-trained models to the public. In bio-medical image segmentation, U-net [23], an advanced CNN-based deep learning with an encoder–decoder type of architecture successfully yields fine segmentation results while preserving image details. This is achieved by the skip connections in the encoder–decoder architecture that enable decoding from features that were encoded with different spatial scales. Here, we demonstrate the different results obtained using three methods: deconvolution with TV regularization, as a typical method in the generative model-based approach, plain encoder–decoder style CNN and U-net. The latter two are based on discriminative models, although the objective here is the same among the three methods. We employed peak-signal-to-noise ratio (PSNR) and a structural similarity index measure (SSIM) [24] to evaluate the performance of denoising. When calculating SSIM, we obtained the local SSIM for image patches with $5 \times 5 \times 5$ pixels and then averaged the obtained values over all the patches.

To evaluate the denoising performance of the three types of methods, we prepared the ground-truth images and their respective noisy images through simulation. We obtained five two-photon images in 3D (three images from transparent brain imaging and two images from *in vivo* imaging), using scanning microscopy, with a single channel, in which the pixels were represented in 16 bits; each image size was approximately $512 \times 512 \times 400$. Based on the obtained real images, we produced ground-truth images $\hat{I}$ by manually removing possible noises from the real images. Subsequently, we prepared simulated two-photon images based on the abovementioned manually annotated ground-truth images. To simulate the forward optics in two-photon microscopy, we assumed Gaussian PSF (blur) G and pixel-wise Gaussian shot noises; the ground-truth images were blurred by convolving Gaussian PSF with three different standard deviations (SDs): 0.4 (condition B1, in Table 2), 0.8 (B2) and 1.2 (B3). Pixel-wise white noise $\mathcal{N}$ was subsequently added by a normal distribution of mean zero and three different values of SD: 500 (condition N1, in Table 2), 1000 (N2) and 1500 (N3). According to $I_{\text{test}} = \hat{I} * G_{\text{PSF}} + \mathcal{N}$ where $*$ refers to 3D convolution, we prepared 45 3D test images from five different 3D images disturbed by 9 (= 3 × 3) different noise settings. During restoration, the performance of the algorithm was estimated by verifying the similarity of the denoised image and its corresponding ground-truth image in terms of the SSIM and PSNR metrics. Notably, these test images were not used for training the denoising models and were only used for testing.

Denoising performance comparison

We compared the three methods: deconvolution with TV regularization, and two different CNN models, baseline CNN and U-net. In the first method, the following objective function was minimized:

$$I^* = \text{argmin}_I \left\| I * G - I_{\text{test}} \right\| + \lambda_{\text{TV}} \left\| \Delta I \right\|, \tag{10}$$

where I_{test} is the input noisy image and $*$ denotes convolution with a 3D Gaussian filter G and a fixed SD of 0.9. The remaining parameters were set heuristically as follows. The smoothness, ΔI, was calculated by averaging the difference between the center and surrounding pixels. The weight of the TV regularization was set to $\lambda_{\text{TV}} = 0.15$. We obtained the deconvolved image I^* by minimizing the abovementioned equation using a gradient descent method. Note that this method is a type of MAP estimation introduced in section 'Maximum likelihood and MAP estimation'.

The used U-net (implementation of original U-net can be found at https://github.com/zhixuhao/unet) architecture comprised seven encoding and decoding blocks where each block was comprised of three convolutional layers (Fig. 3, lower). ReLU was added after each block except for the last one. U-net used batch normalization. Similarly, the baseline CNN comprised seven encoding and decoding blocks and therefore had an architecture similar to that of the U-net (Fig. 3, upper); it was based on the encoder–decoder architecture but without skip connections. The loss function L for training CNN-based models was as follows:

$$\mathcal{L} = \left\| \hat{I} - \phi(I) \right\|_1. \tag{11}$$

This loss function is the Manhattan distance (L1 distance) between the trained model output $\varphi(I)$ and the desired output $\hat{I}$. For training, the ADAM optimizer was employed with a mini-batch having three images, each of $96 \times 96 \times 96$ pixels. All networks were trained with a learning rate of 0.001 with a decaying factor 0.9. We performed the optimization for 25 epochs where each epoch was composed of 1300 mini-batches.

Table 2 summarizes the performance of restoration found by averaging the results of the test images; the upper and lower tables detail the SSIM and PSNR values, respectively. The input images are fairly noisy, and this can be seen in the overall SSIM and PSNR values of 0.0642 and 17.51 dB, respectively. The averaged SSIM and PSNR values of the denoised images by Deconv with TV, Baseline CNN and U-net were 0.8567/27.21, 0.9339/27.07 and 0.9514/29.11 dB, respectively. Therefore, data-driven deep learning-based methods, i.e. the Baseline CNN and U-net, generally demonstrated superior denoising performance than that of the classical deconvolution-based (or generative model-based) method (i.e. Deconv with TV). The deconvolution-based method presents the results that are comparable to those of the deep learning-based methods, only if the input image

Table 1. Glance at difference in methodologies between generative model-based approaches to multi-frame image super-resolution

Inference method	Objective function	Likelihood	Prior	Integration	Reference
Maximum likelihood	Equation (4)	Linear transform, Gaussian noise	None	No	
Maximum a posteriori	Equation (5)	Linear transform, Gaussian noise	Laplacian/Gaussian for HR image	No	[2,11]
		Linear transform, Gaussian noise	Total variation for HR image	No	[14]
Bayesian	Equation (8)	Linear transform, Gaussian noise	Laplacian for HR image	Yes, over HR image	[15]
Bayesian	Equation (9)	Linear transform, Gaussian noise	Huber for parameter	Yes, over parameter	[16]

Most of them were originally developed for 2D image processing. Since the maximum likelihood is just a seminal technique, we do not provide literature.

Table 2. The denoising performances of the three methods change when different noise was artificially applied to the input images

Noise level	N1-B1	N1-B2	N1-B3	N2-B1	N2-B2	N2-B3	N3-B1	N3-B2	N3-B3	Overall
Input (SSIM)	0.1035	0.0946	0.0835	0.0664	0.0583	0.0496	0.0475	0.0406	0.0339	0.0642
Deconv with TV	0.9573	0.9413	0.9285	0.9599	0.9457	0.9326	0.6916	0.6823	0.6713	0.8567
Baseline CNN	0.9368	0.9365	0.9334	0.9377	0.9354	0.9298	0.9373	0.9327	0.9251	0.9339
U-net	0.9825	0.9705	0.9557	0.9670	0.9504	0.9333	0.9506	0.9337	0.9187	0.9514
Noise level	N1-B1	N1-B2	N1-B3	N2-B1	N2-B2	N2-B3	N3-B1	N3-B2	N3-B3	Overall
input (PSNR)	22.89	22.47	21.99	16.89	16.75	16.60	13.41	13.33	13.26	17.51
Deconv with TV	29.09	27.69	26.61	28.99	27.79	26.77	26.58	25.98	25.37	27.21
Baseline CNN	27.33	27.33	27.11	27.41	27.22	26.79	27.28	26.89	26.28	27.07
U-net	34.58	31.64	29.42	30.53	28.58	26.99	28.10	26.68	25.51	29.11

The test images were gradually degraded as the level of white noise (N) and Gaussian blur (B) increased from 1 to 3 (Please refer to the main text for more details). The table presents the average performances in terms of SSIM (upper) and PSNR (lower).

is not exceedingly noisy, e.g. when the SSIM value was 0.9573 for the N1-B1 condition, but its performance rapidly decreased with the increase in the noise level. The benefit of using the CNN-based methods became evident with the increase in the level of noise. Table 2 presents the results that confirm the effectiveness of U-net. On the contrary, the baseline CNN demonstrated more stable performance but lost several image details.

Figure 4 presents the results of denoising when noisy data were used in condition N2-B2. The deconvolution-based method (Deconv with TV) introduced vagueness to the edge pixels due to its regularization parameter that prefers the smaller value between adjacent pixels. Although the baseline CNN presented images that were closer to the ground-truth, the resulting image was still blurry, due to which several small structures disappeared. U-net demonstrated a better performance in terms of the restoration of structures in a more detailed manner. Therefore, U-net often exhibits superior robustness to noise as compared with the other models, including the ability to preserve detailed structures.

Performance improvement in neural tracking

Here, we evaluate the denoising process for its usefulness in pre-processing for post-analyses of neural images. As a typical example of such post-analyses, we chose structure reconstruction problems of neurons and neural networks that were both evaluated by the performance of neural tracking. We applied a popular neural tracking algorithm [25], which is actually the same algorithm used in the preparation of ground-truth images to noisy two-photon microscopy images and the pre-processed images obtained by the denoising methods. Here, the performance was examined based on the similarity between the tracking result and the ground-truth, which was obtained by visual inspection refereeing to the software-based tracking result. As for the similarity measure between two topographic maps, we used the performance classification that classified if each pixel belonged to foreground or background; here, the ground-truth topography was the target. More importantly, tracking topography obtained from noisy or denoised images was first projected onto the ground-truth image, and then, the region growing algorithm [26] (the same method used in the production of ground-truth image) was run to fill the topographic gaps. Note that two tracking results obtained from the noisy and denoised images were used as seed points in this segmentation, and all factors other than the sets of seed points were common, which enabled the sole comparison between noisy and denoised images. The overall flow of the experiment is graphically presented in Figure 5.

Tracking performance was evaluated by the foreground/background segmentation accuracy in terms of the precision:

$$\frac{\#\text{correct foreground pixels}}{\#\text{estimated foreground pixels}}$$

and the recall:

$$\frac{\#\text{correct foreground pixels}}{\#\text{GT foreground pixels}}.$$

We introduced this test procedure to N2-B2 and N3-B3 image datasets, which were so noisy that many topographic structures were not easily identified by the tracking algorithm. Table 3 presents the precision and recall. As shown in Table 3, the recall rate was greatly improved by the image denoising process, especially in

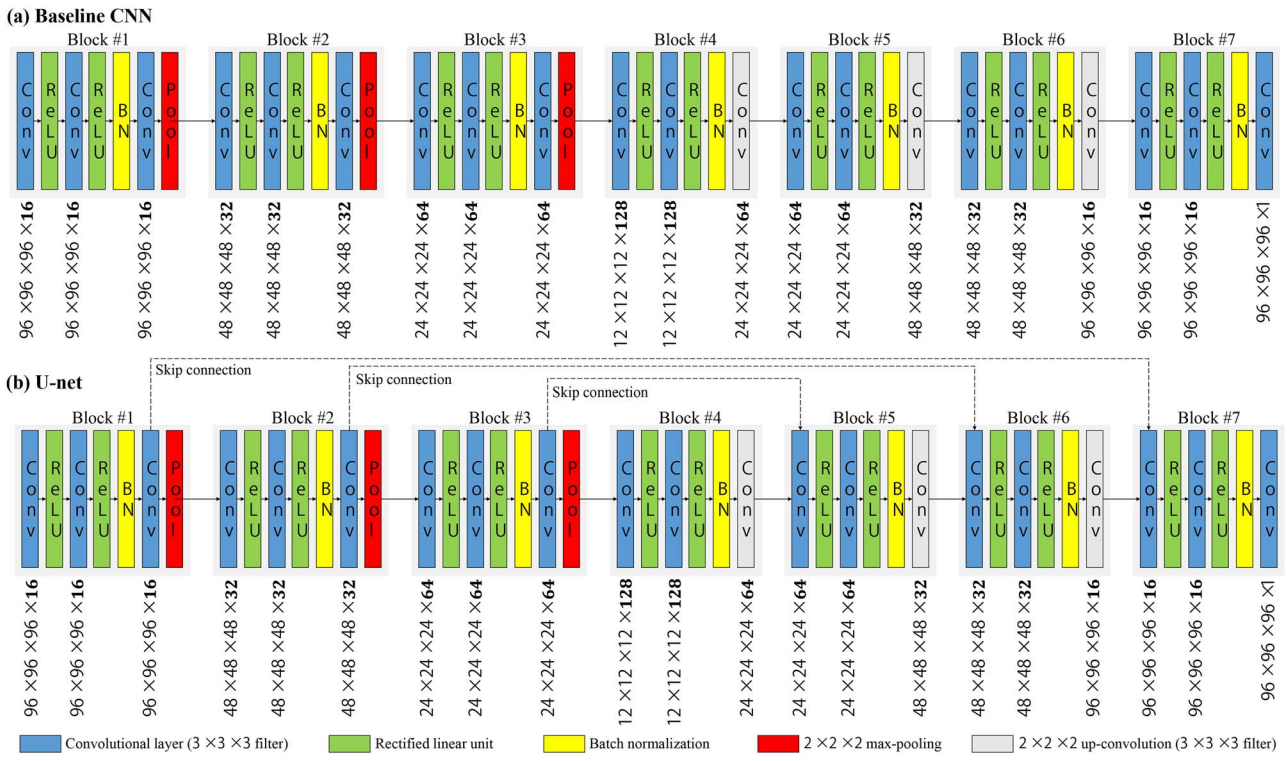

Fig. 3. (a) Baseline CNN and (b) U-net used in our image denoising test. Unlike to baseline CNN, Unet has additional skip connections between encoders and decoders to bypass deeper convolutional layers. Except for these connections, the remaining parts were identical. The numbers located at below of convolutional layers mean the resolution of features where bold numbers indicate the number of filters used in each layer.

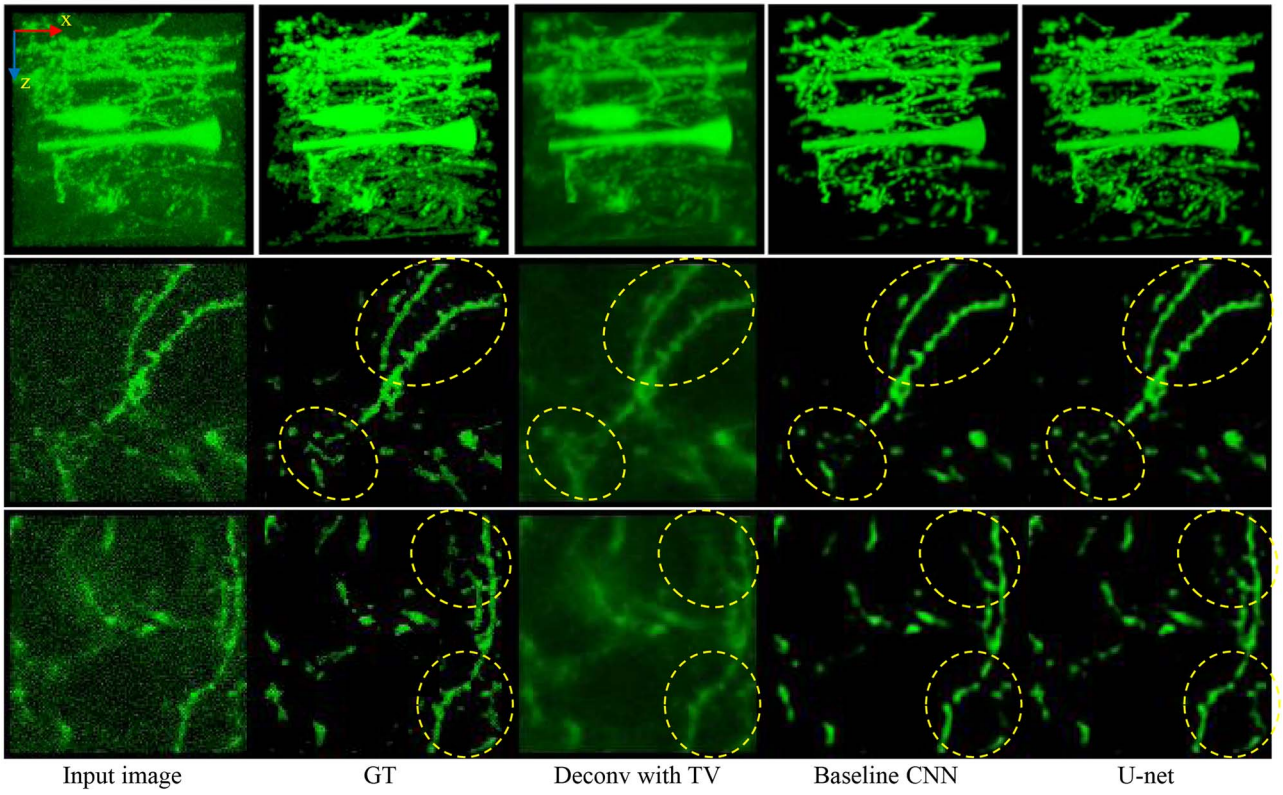

Fig. 4. The first three rows illustrate the input 3D images and two sliced XY plane images, respectively. Moving from the leftmost to the rightmost column, the input images, ground-truth (GT) images, denoised images by TV-regularized deconvolution, baseline CNN and U-net are presented.

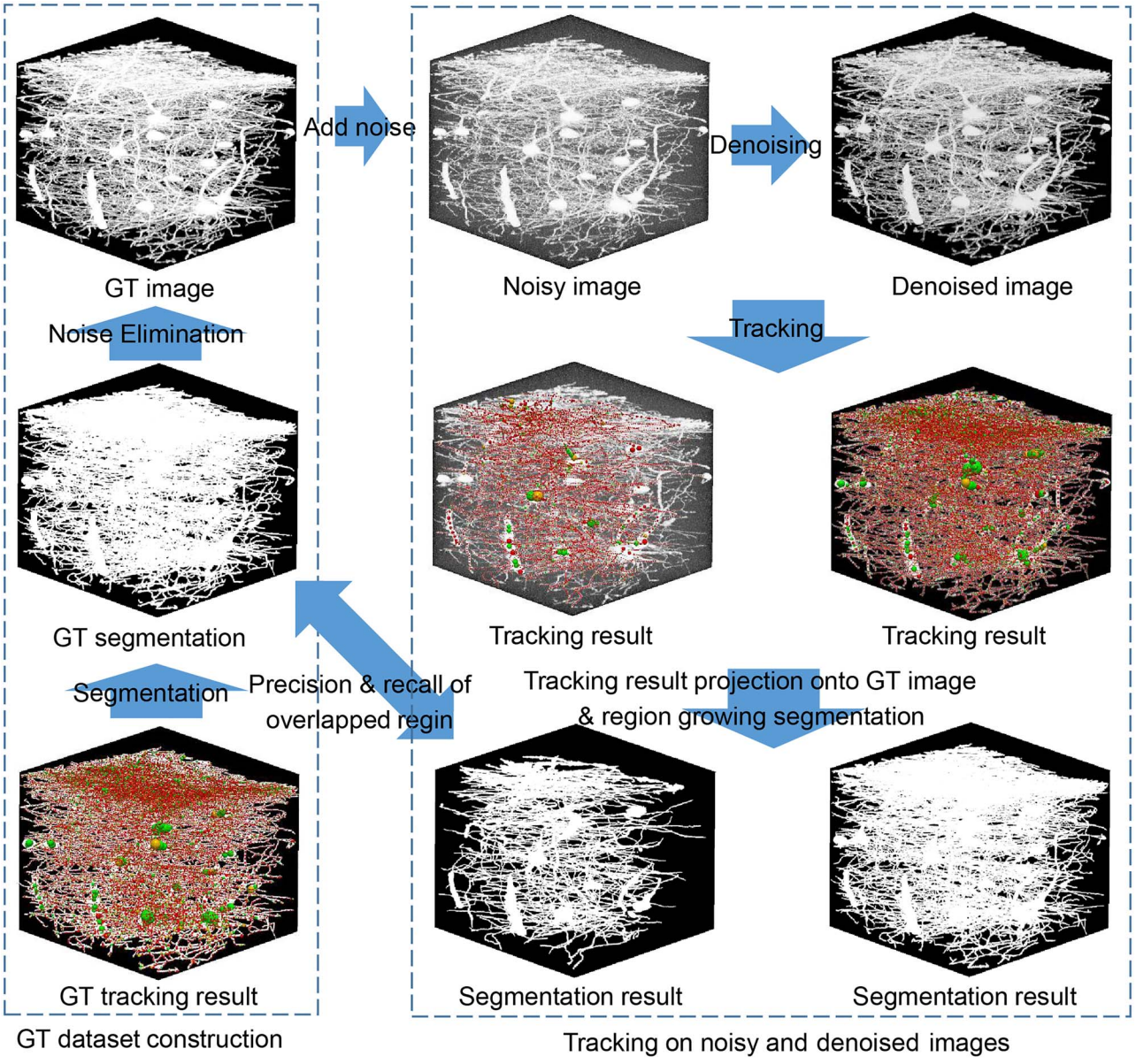

Fig. 5. The flowchart of neural tracking. Green, yellow and red points in the tracking results mean branch, terminal and bridge nodes, respectively. White and black regions in the results denote the foreground and background, which are classified by the seed growing segmentation.

Table 3. The precision and recall rate of tracking results

	N2-B2				N3-B3			
	Noisy	Deconv	CNN	U-net	Noisy	Deconv	CNN	U-net
Precision	0.9832	0.9844	0.9836	0.9860	0.9757	0.9847	0.9825	0.9838
Recall	0.4862	0.4927	0.4945	0.6008	0.1198	0.3349	0.4484	0.5088

N3-B3. When we applied Neutube tracking algorithm to each image, we found that it was normally stuck and failed to track at some branches and bridges that were seemingly ambiguous because of severe noises. It was also observed that, whereas many neurites on the noisy images were separated into small fractions by the tracking algorithm (tracking software + region growing), these neurites were appropriately concatenated by the same tracking algorithm after applying the denoising process. We consider that better denoising process improved the neural tacking performance by removing such ambiguity. As presented in Tables 2 and 3, the tracking performance in terms of recall and precision was improved if a noisy image was replaced by a denoised image by U-net. This is evidence that the denoising method is useful for pre-processing in biological applications such as neural tracking.

EM image segmentation

In this section, we summarize the architecture and performance of supervised ML technologies applied to the segmentation problem from volumetric EM images. CNNs became popular for this problem since a 2D CNN outperformed other methods in an EM segmentation challenge [27]. Two-dimensional CNNs were soon extended into 3D CNNs because of the target dimension, and currently, two state-of-the-art 3D CNNs are widely used for large-scale segmentation [28,29]. We discuss such technologies with a particular focus on 2D/3D CNNs and their variants. We also introduce our software package to easily conduct CNN-based segmentation [30].

Optical/fluorescent microscopy can provide extensive information on the localization and molecular activities of neurons. However, this tool has two fundamental limitations. First, there is an upper bound in the spatial resolution (∼200 nm) of microscopes, which is determined by the diffraction limit [31]. A majority of the fluorescent microscopes available these days have this limitation. However, 'optical' super-resolution technologies can be used to further decrease this upper bound to ∼60 nm in certain cases [32]. Second, only sparsely labeled targets can be observed in dense brain tissues, because the contrasts disappear if all the densely distributed targets are fluorescently labeled. One typical class of targets that are dense in nature is the neuronal fibers of the brain. EM is the best modality for observing such densely distributed neuronal fibers. With EM, all the membrane structures of a target tissue can be observed with a spatial resolution of < 1 nm (Fig. 6A) [35]. Based on the observed neuronal boundaries, researchers can extract neuronal fibers by utilizing the analyzed informatics and image processing technologies. Such reconstruction of neuronal objects is now known as EM connectomics, which is, however, challenging one faced during image segmentation.

In the field of EM connectomics, 3D reconstruction of all neuronal objects from a volumetric EM image has attracted attention [36], because such reconstruction can provide information of the entire neural circuit where all the brain functions are embedded. The deep learning-based ML techniques such as CNNs play an indispensable role in this type of 3D reconstruction, especially in the process of neuronal boundary detection, which is an important step in image segmentation, from EM images (Fig. 6A and B). This is not a trivial task that can be solved by conventional filters (e.g. edge detection filter), because EM images contain numerous boundary-like structures such as mitochondria and endoplasmic reticulums (ER), which must be excluded based on contextual information (Fig. 6A and B). Obtained probability maps of neuronal boundaries are further processed by an object-detection method, typically a combination of a 3D watershed and fragmental-segment agglomeration, leading to the segmentation of neuronal objects (Fig. 6C) from other background objects.

In early days, data-driven CNNs were considered as one of the options for neuronal boundary detection in EM images, because other methods showed similar accuracies in this type of segmentation (e.g. random forest classifier) [37,38]. CNNs have become accepted as a standard since the ISBI2012 EM image segmentation challenge [27,39], because a 2D CNN proposed by the team IDSIA achieved a state-of-the-art performance in this competition [40]. Another 2D CNN, called U-net, showed further improvement in this performance, thereby becoming a standard of biomedical image segmentation [23]. U-net has a U-shaped encoder–decoder network organization comprising a contracting path to capture context and a symmetric expanding path (Fig. 6D). It also has the same-scale skip connections to provide fine spatial information during upsampling. This unique architecture enables the simultaneously capturing of the global shapes of large objects as well as detailed edge information. Additionally, several other 2D CNNs have been proposed for neuronal boundary detection, such as FusionNet [41], fully convolutional networks (FCNs) with skip connections [42], and M^{2}FCN [43].

All the CNNs introduced thus far are 2D in nature. The 2D CNNs have two advantages: computationally inexpensive training process and the requirement of only 2D ground-truth. However, a 2D EM image has many ambiguous structures that can only be identified by referring to adjacent Z-slice images. Thus, to further improve the accuracy of image segmentation, 3D CNNs are applied naturally to the volumetric EM images. We here introduce two representative 3D CNNs for image segmentation used for neuronal EM images.

The first one is a variant of the 3D U-net that was developed by a research group of Princeton University [28]. The network architecture is similar to the original U-net (Fig. 6D) [23]; however, each fully connected module also has a residual skip connection. Because their target EM images have lower Z resolution, the convolution in Z-direction is sparser than those in X- and Y-directions. The residual 3D U-net combined with a two-step object identification method (watershed plus the fragmented-segment agglomeration based on long-range affinity prediction) achieved the state-of-the-art performance in the ISBI 2013 challenge on the 3D segmentation of neurites in EM images (SNEMI3D, 11/10/2019; Table 4) [33,48], and its accuracy was even beyond human performance. Recently, another 3D U-net having 'isotropic' 3D connections with an object identification method also achieves good performance when applied to isotropic volumetric EM images (FIB-25, 6 nm per XYZ-pixel) [49,50].

Another famous 3D CNN architecture is the flood-filling network (FFN), which was developed by a Google research team [29]. FFNs exhibited the best performance in the segmentation of isotropic volumetric EM images (FIB-25) [50] and also showed the second-best in the case of anisotropic volumetric images (SNEMI3D, Table 4) [33]. FFNs do not produce a probability map of neuronal boundaries but directly infer the shapes of the neuronal objects. FFNs have two types of inputs: a patch of a target volumetric image (typically $33 \times 33 \times 33$ voxels) and a predicted shape at the precedent step, and therefore, FFNs predict the shape of the centered object of the target patch. The targeted neuronal objects are spatially and iteratively tracked to determine their overall shapes. The FFN itself has a cascaded architecture of CNN modules, each of which is a 3D CNN with a skip connection. FFNs have been adopted for the fly brain project in Janelia Research Campus [51], aiming to obtain neuronal circuitry from the volumetric EM image of a complete adult Drosophila brain [52].

Notably, 3D image segmentation does rely not only on deep learning such as CNNs for neuronal boundary detection but also on other object identification methods. For example, predicted neuronal boundaries are processed using a seeded watershed algorithm, and the subsequently produced over-segmented objects are connected by means of an agglomeration method, such as the graph-based active learning of agglomeration [53] and the globally optimal objectives (MULTICUT) [54]. Recently, a multi-object tracking technique [45] and agglomeration introducing biological constraints [47] also achieve high performance in neuronal EM image segmentation (Table 4) [33].

The primary advantage of such 3D CNNs is the level of accuracy in their predictions. The current leaders' board of the SNEMI3D is occupied by the combinations of novel 3D CNNs and watershed/ag-

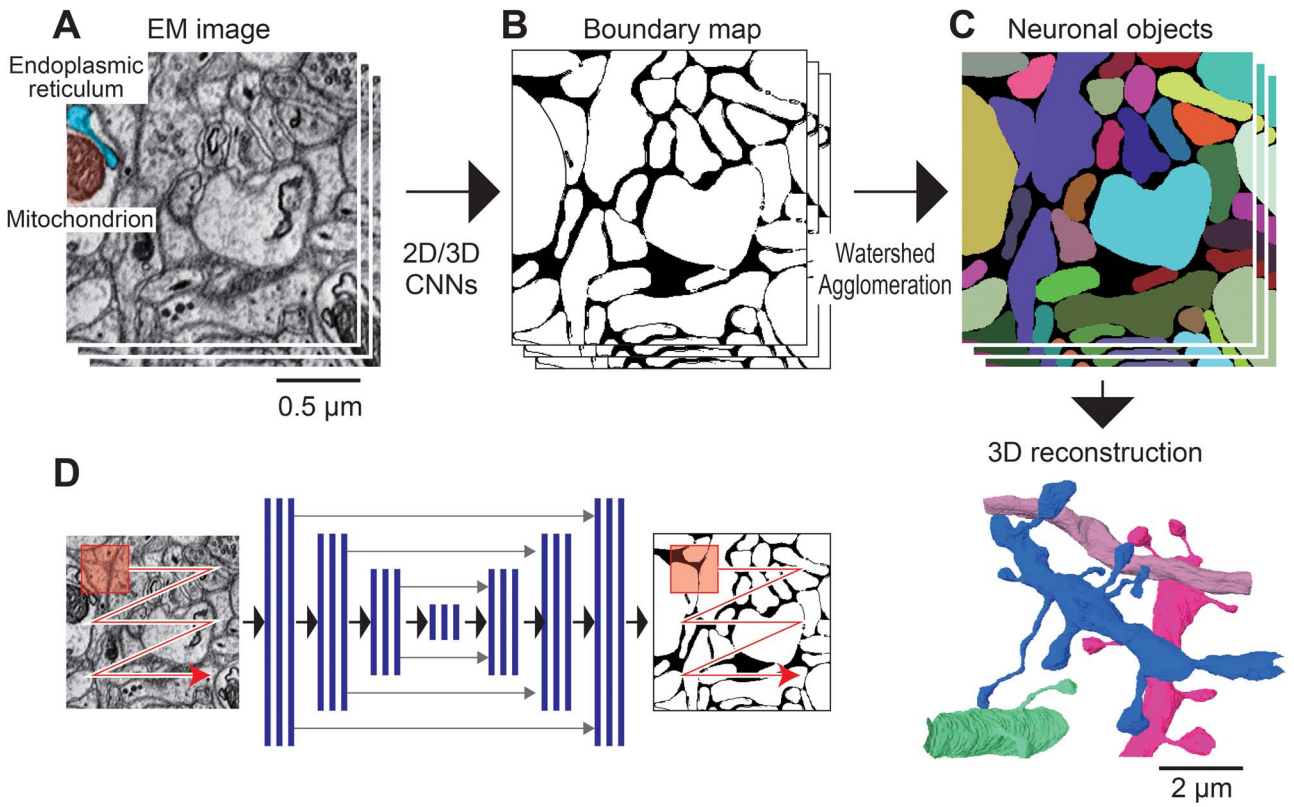

Fig. 6. Three-dimensional reconstruction of neuronal objects from a volumetric EM image. (A) A volumetric EM image from the SNEMI3D [33,34]. In addition to neuronal boundaries, it contains several objects such as mitochondria (red) and ERs (blue). (B) Neuronal boundary map. Two-D/3D CNN should only detect neuronal boundaries while ignoring the other objects. (C) Derived neuronal objects (top) and 3D reconstruction (bottom). (D) Architecture of U-net that predicts a patch of neuronal boundary map (left, red rectangular area) from an EM image patch (left, red rectangular area). Multi-step contracting (down-sampling) path is followed by the expanding (upsampling) path (black arrows). It also has the same-scale skip connections (gray arrows). Each of the three blue lines denotes a fully connected module.

Table 4. SNEMI 3D leaders showing performance beyond human segmnentation (11/10/2019) [33]

Rank	CNN/Agglomeration	Rand error	Group	Reference
1	Residual 3D U-net/long-range affinity prediction	0.0249	PNI	[28]
2	FFNs/FFN-based agglomeration	0.0291	GAIP	[29]
3	Rhoana pipeline	0.0350	S&T	[44]
4	LSTM-U-net/3C	0.0410	CCG	[45]
5	Unopen	0.0461	LZL-USTC	
6	Residual 3D U-net/entropy policy	0.0468	CS17	[46]
7	3D U-net/biologically-constrained graphs	0.0584	VCG	[47]
8	Human values	0.0600		

The RAND score is a measure of segmentation accuracy, and the smaller value denotes the higher accuracy in the segmentation of neuronal objects [27].

glomeration algorithms (Table 4) [33]. However, 3D CNNs have two disadvantages: large computational costs and the requirement of 3D ground-truth. Even when a latest GPU computing card is used, the aforementioned models take more than 1 week for training both 3D residual U-net and FFN. Additionally, 1 week is required for drawing 3D ground-truth for FFNs, and the requirement of 3D ground-truth may be a bottleneck if test trials are being performed by general experimentalists. Even worse, the ground-truth needs to be re-drawn if the image acquisition condition is altered. To address this, researchers are now developing domain adaptation technologies, i.e. the improvements of the applicability of a trained CNN to the EM images that were obtained under an unknown condition [55–57]. Such a technology will drastically decrease the efforts to prepare the ground truths for specific EM images obtained by end users.

ML technologies have therefore enormously contributed to EM connectomics. The spatial scale of the 3D reconstruction from EM images grows rapidly over 1 mm^3 [58], and research to widen the scale is also ongoing [34,52,58,59]. Notably, CNNs are a key technology for neuronal boundary detection, but the methods for EM connectomics are not limited to this and includes: the alignments of 2D EM image stack to generate a volumetric EM image, imperfect EM image handing, fragmental segment agglomeration, synapse detection and neuronal circuit reconstruction [38,60], and each of the steps involves challenges in the fields of informatics and image processing. It is also important to develop software environment to manage these

processes. The amount of EM imaging data being generated should not be underestimated as the rate at which these data are increasing is beyond the scale of peta-byte [61], and advanced research teams have developed their in-house software pipelines to execute various software on their huge EM imaging data [44,58]. On the contrary, software environments available for the general experimentalists are inferior in terms of their effectiveness. Because the best CNN for 3D neuronal image segmentation depends on image acquisition conditions, end-user should be able to test multiple CNNs for their own EM images. EM segmentation has been targeted by many standalone software packages, such as Reconstruct [62], Ilastik [63], Knossos [64], Microscopy Image Browser [65] and VAST lite [66]. They are useful for manual segmentation, but they currently do not support CNN-based segmentation. Very recently, plug-ins for the widely used ImageJ software were developed to handle CNN-based segmentation [67,68]. The use of these plug-ins is advantageous, but users still need to launch a Linux server to train target CNNs. We also developed standalone software to test CNN-based segmentation, called UNI-EM [30]. This software is designed for researchers who have limited programming skills, so users can easily follow the procedure of CNN-based segmentation, such as ground-truth generation, training, inference, proof reading and visualization, without the knowledge of computer languages or CNN frameworks. This software has developed based on Python and a CNN framework Tensorflow, and thus, developers can easily incorporate new CNN models into UNI-EM. The developed source code with an online manual is available at the public repository GitHub (https://github.com/urakubo/UNI-EM). Such efforts for connecting users and developers will further activate the field of CNN-based EM connectomics as well as the field of image feature extraction.

Concluding remarks

Image processing methodologies for microscopy have shown great advancements in the last two or three decades, thank to the developments in various ML technologies. Generative model-based approaches, typically seen in the multi-frame super-resolution discussed in section 'Generative-model based approaches to multi-frame super-resolution', was developed in parallel with the improvements of statistical ML techniques such as (approximated) Bayesian inference. These approaches are suitable especially when the knowledge of optics and/or target characteristics is available. When such knowledge is not available, one possible idea is to rely on the data themselves. With the accumulations of data, especially annotated data, supervised learning-based approaches have become realistic. Deep learning methods enable the non-linear filters to be trained based on the set of given input and output images. These approaches require someone to prepare the transformed/annotated images. Although such preparation is not easy in realistic applications, the deep learning-based methods like U-net can identify the non-linear filters even without the knowledge of optics and/or target characteristics. As a typical application of such an approach, we discussed image restoration from optical microscopic images in section 'Deep learning-based image super-resolution and restoration'. Because deep learning-based methods have generality in terms of optics and/or targets, they can be applied to other image transformation problems. As a typical problem, we discussed EM image segmentation in section 'EM image segmentation'. Not surprisingly, deep learning-based methods, CNNs and their variants, have become the state of the art even in this problem.

Toward the next generation, the combination of different approaches seems important. Low-level features can be extracted, in terms of non-linear filters, from unannotated images. After estimating such low-level features, image transformation can be trained based on relatively small amount of data. In the field of ML, this kind of technique is called semi-supervised learning. Otherwise, one deep learning-based image processing tool might be transferred by calibrating it with a relatively small amount of calibration data, which is called transfer learning. Such combination of different approaches will be applicable to more difficult situations like when the amount of annotated images is completely lacking or very much lacking. Another idea is to combine generative model-based and discriminative model-based approaches. Such a combination is especially fruitful when there is some knowledge of the optics/target, but it is not very much reliable. One example can be seen in the image segmentation from MRI images [69].

Funding

This work was supported by CREST (JPMJCR1652 to all the authors) from Japan Science and Technology Agency, Japan, Grants-in-Aid for Scientific Research (17H06310 to SI) from Japan Society for Promotion of Science (JSPS), Japan, Exploratory Challenge on Post-K computer (to SI and UK) from Ministry of Education, Culture, Sports, Science and Technology, Japan, and World Premier International Research Center Initiative (WPI) from JSPS, Japan (to SI, SL, H.Kume and H.Kasai).

Conflict of interest

There are no conflicts of interest.

References

1. Bishop C M (2006) *Pattern Recognition and Machine Learning* (Springer, New York, USA).
2. Elad M, and Feuer A (1997) Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* 6 (12): 1646–1658.
3. Hell S W, and Wichmann J (1994) Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Optics Lett.* 19 (11): 780–782.
4. Neice A (2010) Methods and limitations of subwavelength imaging. *Adv. Imag. Elect. Phys.* 163: 117–140.
5. Freeman W T, Jones T R, and Pasztor E C (2002) Example-based superresolution. *IEEE Comput. Graph.* 2: 56–65.
6. Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36: 193–202.
7. LeCun Y (1986) Learning processes in an asymmetric threshold network. In: *Disordered Systems and Biological Organization*, pp. 233–240 (Springer, Berlin-Heidelberg, Germany).
8. Kanemura A, Maeda S, and Ishii S (2009) Superresolution with compound Markov random fields via the variational EM algorithm. *Neural Netw.* 22 (7): 1025–1034.
9. Dempster A P, Laird N M, and Rubin D B (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B* 39 (1): 1–22.
10. Geman S, and Geman D (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intel.* 6: 721–741.
11. Hardie R C, Barnard K J, and Armstrong E E (1997) Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.* 6 (12): 1621–1633.

12. Dertinger T, Colyer R, Iyer G, Weiss S, and Enderlein J (2009) Fast, backgroundfree, 3D super-resolution optical fluctuation imaging (SOFI). *Proc. Natl. Acad. Sci. USA*. 106 (52): 22287–22292.
13. Fish D, Brinicombe A, Pike E, and Walker J (1995) Blind deconvolution by means of the Richardson–Lucy algorithm. *J. Opt. Soc. Am. A* 12 (1): 58–65.
14. Farsiu S, Robinson M D, Elad M, and Milanfar P (2004) Fast and robust multiframe super resolution. *IEEE Trans. Image Process*. 13 (10): 1327–1344.
15. Tipping M E, and Bishop C M (2003) Bayesian image superresolution. In: *Adv. Neural Inf. Process. Syst*: Vol. 15 pp. 1279–1286.
16. Pickup L C, Capel D P, Roberts S J, and Zisserman A (2007) Overcoming registration uncertainty in image super-resolution: Maximize or marginalize? *EURASIP J. Adv. Sig. Process*. 2007 (2): 23565.
17. Kanemura A, Maeda S, Fukuda W, and Ishii S (2010) Bayesian image superresolution and hidden variable modeling. *J. Syst. Sci. Complex*. 23(1): 116–136.
18. Coupé P, Munz M, Manjón J V, Ruthazer E S, and Collins D L (2012) A CANDLE for a deeper in vivo insight. *Med. Image Anal*. 16 (4): 849–864.
19. Danielyan A, Wu Y W, Shih P Y, Dembitskaya Y, and Semyanov A (2014) Denoising of two-photon fluorescence images with block-matching 3D filtering. *Methods* 68 (2): 308–316.
20. Boulanger J, Kervrann C, Bouthemy P, Elbau P, Sibarita J B, and Salamero J (2010) Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans. Med. Imaging* 29 (2): 442–454.
21. Weigert M, Schmidt U, Boothe T, Muller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S, Rocha Martins M, Segovia-Miranda F, Norden C, Henriques R, Zerial M, Solimena M, Rink J, Tomancak P, Royer L, Jug F, and Myers E W. (2018) Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat Methods* 15 (12): 1090–1097.
22. Lee S, Negishi M, Urakubo H, Kasai H, and Ishii S (2020) Mu-net: Multi-scale u-net for two-photon microscopy image denoising and restoration. *Neural Networks*. 125: 92–103.
23. Ronneberger O, Fischer P, and Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *Int. Conf. Med. Image Comput. Comput.Assist. Interv*.: pp. 234–241.
24. Wang Z, Bovik A C, Sheikh H R, and Simoncelli E P (2004) Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process*. 13 (4): 600–612.
25. Feng L, Zhao T, and Kim J (2015) neuTube 1.0: A new design for efficient neuron reconstruction software based on the SWC format. *eNeuro* 2 (1): e0049–14.2014.
26. Sethian J A (1999) *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science, 3rd ed* (Cambridge University Press, Cambridge, UK).
27. Arganda-Carreras I, Turaga S C, Berger D R, Cires,an D, Giusti A, Gambardella L M, Schmidhuber J, Laptev D, Dwivedi S, Buhmann J M, Liu T, Seyedhosseini M, Tasdizen T, Kamentsky L, Burget R, Uher V, Tan X, Sun C, Pham T D, Bas E, Uzunbas M G, Cardona A, Schindelin J, and Seung H S (2015) Crowdsourcing the creation of image segmentation algorithms for connectomics. *Front. Neuroanat*. 9: 142.
28. Lee K, Zung J, Li P, Jain V, and Seung H S (2017) Superhuman accuracy on the SNEMI3D connectomics challenge arXiv. 1706–00120.
29. Januszewski M, Kornfeld J, Li P H, Pope A, Blakely T, Lindsey L, MaitinShepard J, Tyka M, Denk W, and Jain V (2018) High-precision automated reconstruction of neurons with flood-filling networks. *Nat. Methods* 15 (8): 605–610.
30. Urakubo H, Bullmann T, Kubota Y, Oba S, and Ishii S (2019) UNI-EM: An environment for deep neural network-based automated segmentation of neuronal electron microscopic images. *Sci. Rep*. 9: 19413.
31. Vangindertael J, Camacho R, Sempels W, Mizuno H, Dedecker P, and Janssen K (2018) An introduction to optical super-resolution microscopy for the adventurous biologist. *Methods Appl. Fluoresc*. 6 (2): 022003.
32. Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O, and Drummen G P (2019) Super-resolution microscopy demystified. *Nat. Cell Biol*. 21 (1): 72–84.
33. SNEMI3D. http://brainiac2.mit.edu/SNEMI3D/ accessed: 2019-11-10.
34. Kasthuri N, Hayworth K J, Berger D R, Schalek R L, Conchello J A, KnowlesBarley S, Lee D, Vazquez-Reina A, Kaynig V, Jones T R, Roberts M, Morgan J L, Tapia J C, Seung H S, Roncal W G, Vogelstein J T, Burns R, Sussman D L, Priebe C E, Pfister H, and Lichtman J W (2015) Saturated´ reconstruction of a volume of neocortex. *Cell* 162 (3): 648–661.
35. Kourkoutis L F, Plitzko J M, and Baumeister W (2012) Electron microscopy of biological materials at the nanometer scale. *Annu. Rev. Mater. Res*. 42: 33–58.
36. Blow N (2007) Following the wires. *Nat. Methods* 4 (11): 975–981.
37. Jain V, Murray J F, Roth F, Turaga S, Zhigulin V, Briggman K L, Helmstaedter M N, Denk W, and Seung H S (2007) Supervised learning of image restoration with convolutional networks. In: *2007 IEEE 11th Int. Conf. Comput. Vis*.: pp. 1–8.
38. Kaynig V, Vazquez-Reina A, Knowles-Barley S, Roberts M, Jones T R, Kasthuri N, Miller E, Lichtman J, and Pfister H (2015) Large-scale automatic reconstruction of neuronal processes from electron microscopy images. *Med. Image Anal*. 22 (1): 77–88.
39. ISBI2012. http://brainiac2.mit.edu/isbi.challenge/ accessed: 2019-11-10.
40. Ciresan D, Giusti A, Gambardella L M, and Schmidhuber J (2012) Deep neural networks segment neuronal membranes in electron microscopy images. In: *Adv. Neural Inf. Process. Syst*.: pp. 2843–2851.
41. Quan T M, Hildebrand D G, and Jeong W K (2016) FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics arXiv. 1612.05360.
42. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, and Pal C (2016) The importance of skip connections in biomedical image segmentation. In: *Deep Learning and Data Labeling for Medical Applications*: pp. 179–187 (Springer International Publishing).
43. Shen W, Wang B, Jiang Y, Wang Y, and Yuille A (2017) Multi-stage multirecursive-input fully convolutional networks for neuronal boundary detection. In: *Proc. IEEE Int. Conf. Comput. Vis*.: pp. 2391–2400.
44. Haehn D, Hoffer J, Matejek B, Suissa-Peleg A, Al-Awami A, Kamentsky L, Gonda F, Meng E, Zhang W, Schalek R, Wilson A, Parag T, Beyer J, Kaynig V, Jones T R, Tompkin J, Hadwiger M, Lichtman J W and Pfister H. (2017) Scalable interactive visualization for connectomics. *Informatics* 4 (3): 29.
45. Meirovitch Y, Mi L, Saribekyan H, Matveev A, Rolnick D, and Shavit N (2019) Cross-classification clustering: An efficient multi-object tracking technique for 3-D instance segmentation in connectomics. In: *Proc. IEEE Conf. Comput. Vis. Patt. Recog*.: pp. 8425–8435.
46. Hascoet T, Metge B, Takiguchi T, and Ariki Y (2019) Entropy policy for supervoxel agglomeration of neurite segmentation. In: *Int. Workshop Front. Comput. Vis*.: O3–4.
47. Matejek B, Haehn D, Zhu H, Wei D, Parag T, and Pfister H (2019) Biologicallyconstrained graphs for global connectomics reconstruction. In: *Proc. IEEE Conf. Comput. Vis. Patt. Recog*.: pp. 2089–2098.
48. Nunez-Iglesias J, Kennedy R, Plaza S M, Chakraborty A, and Katz W T (2014) Graph-based active learning of agglomeration (GALA): A python library to segment 2D and 3D neuroimages. *Front. Neuroinform*. 8: 34.
49. Funke J, Tschopp F, Grisaitis W, Sheridan A, Singh C, Saalfeld S, and Turaga S C (2018) Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Trans. Pattern Anal. Mach. Intel*. 41 (7): 1669–1680.
50. Takemura S, Xu C S, Lu Z, Rivlin P K, Parag T, Olbris D J, Plaza S, Zhao T, Katz W T, Umayam L, Weaver C, Hess H F, Horne J A, Nunez-Iglesias J, Aniceto R, Chang L A, Lauchie S, Nasca A, Ogundeyi O, Sigmund C, Takemura S, Tran J, Langille C, Le Lacheur K, McLin S, Shinomiya A, Chklovskii D B, Meinertzhagen I A, and Scheffer L K (2015) Synaptic circuits and their variations within different columns in the visual system of drosophila. *Proc. Natl. Acad. Sci. USA*. 112 (44): 13711–13716.
51. Li P H, Lindsey L F, Januszewski M, Zheng Z, Bates A S, Taisz I, Tyka M, Nichols M, Li F, Perlman E, Maitin-Shepard J, Blakely T, Leavitt L, Jefferis G S X E, Bock D, and Jain V (2019) Automated reconstruction of a serialsection EM drosophila brain with flood-filling networks and local realignment bioRxiv. 605634.

52. Zheng Z, Lauritzen J S, Perlman E, Robinson C G, Nichols M, Milkie D, Torrens O, Price J, Fisher C B, Sharifi N, Calle-Schuler S A, Kmecova L, Ali I J, Karsh B, Trautman E T, Bogovic J A, Hanslovsky P, Jefferis G S X E, Kazhdan M, Khairy K, Saalfeld S, Fetter R D, and Bock D D (2018) A complete electron microscopy volume of the brain of adult Drosophila melanogaster. *Cell* 174 (3): 730–743.
53. Nunez-Iglesias J, Kennedy R, Parag T, Shi J, and Chklovskii D B (2013) Machine learning of hierarchical clustering to segment 2D and 3D images. *PLoS One* 8 (8): e71715.
54. Andres B, Kroeger T, Briggman K L, Denk W, Korogod N, Knott G, Koethe U, and Hamprecht F A (2012) Globally optimal closed-surface segmentation for connectomics. In: *Eur. Conf. Comput. Vis.*: pp. 778–791.
55. Januszewski M, and Jain V (2019) Segmentation-enhanced CycleGAN bioRxiv. 548081.
56. Roels J, Hennies J, Saeys Y, Philips W, and Kreshuk A (2019) Domain adaptive segmentation in volume electron microscopy imaging. In: *2019 IEEE 16th Int. Symp. Biomed. Image*: pp. 1519–1522.
57. Bermúdez-Chacón R, Márquez-Neila P, Salzmann M, and Fua P (2018) A´ domain-adaptive two-stream U-net for electron microscopy image segmentation. In: *2018 IEEE 15th Int. Symp. Biomed. Image*: pp. 400–404.
58. Bae J A, Mu S, Kim J S, Turner N L, Tartavull I, Kemnitz N, Jordan C S, Norton A D, Silversmith W M, Prentki R, Sorek M, David C, Jones D L, Bland D, Sterling A L R, Park J, Briggman K L, Seung H S;Eyewirers (2018) Digital museum of retinal ganglion cells with dense anatomy and physiology. *Cell* 173 (5): 1293–1306.
59. Motta A, Berning M, Boergens K M, Staffler B, Beining M, Loomba S, Schramm C, Hennig P, Wissler H, and Helmstaedter M (2018) Dense connectomic reconstruction in layer 4 of the somatosensory cortex bioRxiv. 460618.
60. Lee K, Turner N, Macrina T, Wu J, Lu R, and Seung H S (2019) Convolutional nets for reconstructing neural circuits from brain images acquired by serial section electron microscopy. *Curr. Opin. Neurobiol.* 55: 188–198.
61. Motta A, Schurr M, Staffler B, and Helmstaedter M (2019) Big data in nanoscale connectomics, and the greed for training labels. *Curr. Opin. Neurobiol.* 55: 180–187.
62. Fiala J C (2005) Reconstruct: A free editor for serial section microscopy. *J. Microsc.* 218 (1): 52–61.
63. Sommer C, Straehle C, Koethe U, and Hamprecht F A (2011) Ilastik: Interactive learning and segmentation toolkit. In: *2011 IEEE Int. Symp. Biomed. Image*: pp. 230–233.
64. Helmstaedter M, Briggman K L, and Denk W (2011) High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nat. Neurosci.* 14 (8): 1081–1088.
65. Belevich I, Joensuu M, Kumar D, Vihinen H, and Jokitalo E (2016) Microscopy image browser: A platform for segmentation and analysis of multidimensional datasets. *PLoS Biol.* 14 (1): e1002340.
66. Berger D R, Seung H S, and Lichtman J W (2018) VAST (volume annotation and segmentation tool): Efficient manual and semi-automatic labeling of large 3D image stacks. *Front. Neural Circuits* 12: 88.
67. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jackel Z, Seiwald K, Dovzhenko A, Tietz O, Dal Bosco C, Walsh S, Saltukoglu D, Tay T L, Prinz M, Palme K, Simons M, Diester I, Brox T, and Ronneberger O (2019) U-net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16 (1): 67–70.
68. Gómez-de-Mariscal E, García-López-de-Haro C, Donati L, Unser M, Muñoz-Barrutia A, and Sage D (2019) DeepImageJ: A user-friendly plugin to run deep learning models in ImageJ bioRxiv. 799270.
69. Ito R, Nakae K, Hata J, Okano H, and Ishii S (2019) Semi-supervised deep learning of brain tissue segmentation. *Neural Netw.* 116: 25–34.