SparseSSP: 3D Subcellular Structure Prediction from Sparse-View Transmitted Light Images

Jintu Zheng[1,2], Yi Ding[3], Qizhe Liu[1], Yuehui Chen[3], Yi Cao[3], Ying Hu[1], and Zenan Wang[1]

[1] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
{jt.zheng,zn.wang1,ying.hu}@siat.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Jinan University, Shandong, China
dingyi@stu.ujn.edu.cn,ise_caoy@ujn.edu.cn

Abstract. Traditional fluorescence staining is phototoxic to live cells, slow, and expensive; thus, the subcellular structure prediction (SSP) from transmitted light (TL) images is emerging as a label-free, faster, low-cost alternative. However, existing approaches utilize 3D networks for one-to-one voxel level dense prediction, which necessitates a frequent and time-consuming Z-axis imaging process. Moreover, 3D convolutions inevitably lead to significant computation and GPU memory overhead. Therefore, we propose an efficient framework, SparseSSP, predicting fluorescent intensities within the target voxel grid in an efficient paradigm instead of relying entirely on 3D topologies. In particular, SparseSSP makes two pivotal improvements to prior works. First, SparseSSP introduces a one-to-many voxel mapping paradigm, which permits the sparse TL slices to reconstruct the subcellular structure. Secondly, we propose a hybrid dimensions topology, which folds the Z-axis information into channel features, enabling the 2D network layers to tackle SSP under low computational cost. We conduct extensive experiments to validate the effectiveness and advantages of SparseSSP on diverse sparse imaging ratios, and our approach achieves a leading performance compared to pure 3D topologies. SparseSSP reduces imaging frequencies compared to previous dense-view SSP (i.e., the number of imaging is reduced up to 87.5% at most), which is significant in visualizing rapid biological dynamics on low-cost devices and samples.

Keywords: Subcellular Structure Prediction · Hybrid Dimensions Topology · Microscopy Image

1 Introduction

Understanding information at the subcellular level is pivotal for biological and medical studies [1,2,5,7,8,24,25]. Fluorescence microscopy revolutionized modern biology by permitting labeled imaging and quantifying subcellular structures of

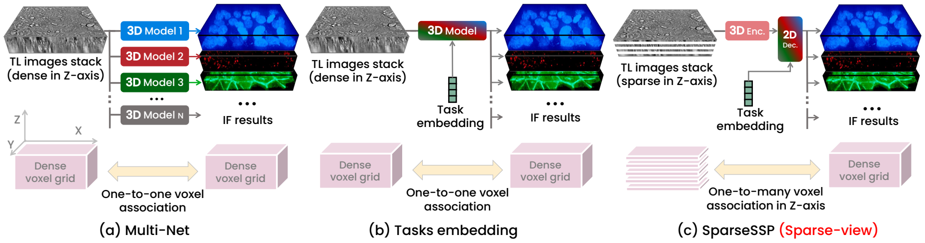


Fig. 1: Comparisons on SSP implementations. (a) Multi-Net adopts a set of independent 3D models for SSP. (b) Task embedding is an improved strategy which can learn all tasks relied on the one-hot encoding, however, this is still require a dense imaging procedure. Prolonged imaging time and high computational cost are unfriendly. (c) SparseSSP implements an efficient hybrid dimensions topology which only requires less TL imaging slices to achieve the excellent performance.

interest. This technology employs antibodies conjugated with various fluorescent dyes to cells, enabling the visualization of rich structural information in 3D through distinct fluorescent signals (i.e., immunofluorescence images).

Fluorescence staining requires expensive and advanced instrumentation and time-consuming preparation of materials [3, 12, 22]. Moreover, significant phototoxicity and photobleaching also damage the live cells [11]. An emerging technology, namely Subcellular Structure Prediction (SSP) [20, 30], enables direct prediction of 3D immunofluorescence (IF) from transmitted light (TL) images via 3D vision networks, significantly reduces sample preparation costs and is low phototoxic to live cells [25]. SSP predicts the fluorescent intensities scalar in a predefined 3D voxel grid, i.e., predicting an occupancy value for each voxel, which can be seen as a dense regression problem. Existing SSP approaches require the dense imaging process (i.e., a motor is required to drive the lens to scan layer by layer on the z-axis) for better data quality, which would cost a prolonged imaging time (e.g., in AllenCell collection [20], each subcellular type is imaged for up to 2.5 hours on a Zeiss spinning disk microscope). Prolonged imaging time is unfriendly to capturing the biodynamic process; the physiological motion, such as cell respiration, introduces the scanning position offset. Additionally, the nutrients produce wakes and motion artifacts, making observing live cell kinematics accurately challenging. Therefore, developing an approach for sparse-view SSP is desired, which makes it significant in visualizing rapid biological dynamics on low-cost devices and samples.

The essence of the SSP is learning the distribution of the fluorescence signal within the target voxel grid according to the subcellular structure characteristics in TL images. A logical approach is to train a one-to-one voxel mapping between the input and output 3D spaces via pure 3D topologies. FNet [20] employs multiple 3D models to learn diverse subcellular structures, as shown in Fig. 1(a). RepMode [30] introduces a task embedding solution for SSP, enabling multiple tasks to be embedded into a single model, reducing computational costs significantly, as shown in Fig. 1(b). Pure 3D modules (e.g., 3D convolutions), owing to their overhead on GPU memory and computational cost, pose substantial challenges in biological imaging applications.

Motivated by the aforementioned observations, we propose an efficient framework, SparseSSP, for predicting fluorescent intensities on less TL images (i.e., sparse-view) with a hybrid dimensions paradigm, as shown in Fig. 1(c). SparseSSP mainly makes two pivotal improvements to prior SSP works. First, SparseSSP introduces a one-to-many voxel mapping paradigm, enabling more minor microscopy imaging frequencies to capture the subcellular structure. We interpolate the sparse-view space into the pseudo voxel grid and aim to learn an implicit upsample optimization on the Z-axis. SparseSSP embeds the Z-axis information into channel features, enabling efficient 2D network layers to address this 3D prediction task. Recently, FlashOcc [27] proposed a pure 2D manner for 3D occupancy prediction [18, 19], which validates that the 2D networks are feasible to learn stereo representation in absolute 3D space. In addition, channel rearrangement [21] in super-resolution demonstrates the potentiality of reconstructing information missing in spatial dimensions via a channel-to-spatial manner. Inspired by the above works, we implement a Z-axis spatial information to channel cross-transformation, aiming to model the Z-axis sparse-view via efficient 2D decoding layers. We conduct extensive experiments to explore diverse hybrid dimensions topologies for sparse-view SSP. This substantiates that the 2D decoding layers are superior in modeling the sparse-view voxel grid after depth-to-channel transformation, demonstrating more excellent performance than pure 3D topologies.

Our contributions can be summarized as follows:

- We propose an efficient hybrid dimensions topology for SSP, enabling the 2D network layers to tackle the 3D voxel level prediction, which simultaneously trades off computational cost and precision, demonstrating its potential for more biological researches.
- We conduct extensive experiments to validate the effectiveness and advantages of SparseSSP on diverse sparse imaging ratios.
- To the best of the author's knowledge, we are the pioneers to study SSP from sparse-view TL images. Our approach significantly reduces the number of required imaging frequencies by up to 87.5% while maintaining leading performance compared to pure 3D topologies.

2 Related Work

2.1 Partially Labeling Issue in SSP

Diverse subcellular structures are annotated in different images because of the high laborious cost [30] and limitations in simultaneous multiple staining. It is a partial labeling issue similar to many medically dense prediction tasks, e.g., multi-organ segmentation. Fortunately, many mature solutions [6,20,23,26,28,30] are available for multi-task learning.

Multi-Net. Multi-Net adopts independent networks to learn tasks [4,13,14,20]. FNet [20] is the typical approach to employ multiple 3D UNets to predict diverse subcellular structures, which is inefficient and inflexible in tuning the network parameters.

Task Embedding. Some approaches [6,23,26,28] attempt to encode each task as the one-hot embedding feature in the network inference. Dmitriev et al. [6] propose a conditional framework CondNet on multi-task learning, while Zhang et al. [28] propose the DoDNet with the task controller and a dynamic head to formulate the multi-task problem as a single class task. Tgnet [26] improves DoDNet [28], embedding more task-guiding functional blocks in the network. These multi-head solutions share the feature extractor and employ different task embedding approaches to achieve varied output. Meanwhile, Zhou et al. [30] propose the RepMode to learn gating re-parameterize to generate specialized parameters for each subcellular structure, achieving state-of-the-art performance on dense-view TL images. However, all task embedding methods are focused on the dense-view SSP. Furthermore, especially for 3D prediction, these models heavily rely on computationally expensive pure 3D convolutions, which demand significant GPU memory and computational resources.

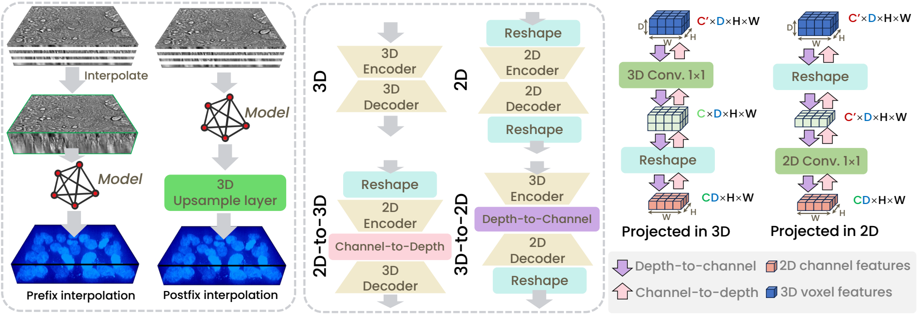
2.2 Sparse-View Techniques

Sparse-view techniques have emerged as a prominent research area in biological and medical imaging [9,10,15,29], garnering significant attention and interest. The sparse-view impact significantly brings up more physical gain than just the software advantages. For example, sparse-view techniques can reduce radiation dose in CT reconstruction with fewer projection times [16]. Similarly, less imaging time in SSP also reduces the phototoxicity of live cells [20]. More importantly, reducing imaging times to capture the subcellular structures allows biologists to observe the rapid biological dynamics in a low-cost preparation to understand subcellular-level activities better.

3 SparseSSP

3.1 Problem Definition

View Definition. Let $\mathbf{S}$ denote the target 3D voxel grid of size $D_s \times H \times W$, and we would like to obtain the subcellular structures in $\mathbf{S}$. The microscope requires TL imaging layer by layer in the Z-axis to reconstruct the structure information in this voxel grid. Let $\mathcal{I}$ denote imaging 3D voxel grid of size $D_i \times H \times W$. Assume the most minor interval of micromotor is the same as the voxel Z-axis resolution limitation. In a dense view, the motor must walk at least D_i steps to cover the voxel grid $\mathbf{S}$, i.e., $D_s = D_i$; this guarantees maximum imaging quality. However, this would slow down the imaging dramatically because the frequent acceleration and deceleration movements in a very short Z-axis interval require high fineness of the micromotor. To meet such high-precision imaging requirements, micromotors are usually mechanically designed to vary slowly to ensure accuracy. The goal of sparse-view SSP is to predict the voxel grid from $\mathbf{I} \rightarrow \mathbf{S}$ under the condition $D_s < D_i$, which can reduce the imaging frequencies. In order to make it more suitable for the actual situation of the imaging scenes, let $\mathbf{r}$ as the sparsity ratio, which is satisfying that the D_i is the multiple of $\mathbf{r}$ with D_s (which can be easily controlled by the micromotor walking interval), i.e., $\mathbf{r} \times D_i = D_s$, especially when $\mathbf{r} = 1$ it is the dense-view.



(a) Interpolation strategies (b) Topology strategies (c)Dimension transformation

Fig. 2: Optional strategies in SparseSSP. (a) demonstrates two strategies for one-to-many mapping in the Z-axis. (b) shows four combinations for different dimension encoders and decoders. (c) demonstrates diverse approaches to implementing the dimension transformation.

Data Definition. Following prior works [20,30], we assume that each TL input contains only one fluorescent label. Let $\mathcal{P} = \{(\mathbf{x}_n, \mathbf{y}_n, l_n)\}_{n=1}^{N}$ denotes the SSP dataset in the sparse-view setting with N samples. The n-th image $\mathbf{x}_n \in \mathcal{P}_n$ is associated with the label $\mathbf{y}_n \in \mathcal{P}_n$, where $\mathbf{y}_n = \mathbb{R}^{D_n \times H_n \times W_n}$ denotes the fluorescent label, and the $\mathbf{x}_n = \mathbb{R}^{(D_n/\mathbf{r}) \times H_n \times W_n}$ denotes the TL input ($D_n/\mathbf{r}$ images). The $\mathbf{y}_n$ satisfies the view definition $\mathbf{S}$, and the $\mathbf{x}_n$ satisfies the view definition $\mathbf{I}$. We define the T as the number of tasks, i.e., the number of subcellular structures that require learning, and the $l_n \in \mathcal{L} = \{1, 2, ..., T\}$ represents that $\mathbf{y}_n$ is the label of the l_n-th subcellular structure.

3.2 One-to-Many Mapping in Z-axis

Sparse-view is sparse on the Z-axis which means that the Z-axis information is incomplete in describing the target voxel grid S. In contrast to dense-view, which solely focuses on prediction, sparse-view necessitates tackling both prediction and upsampling within a network. Here, we propose two strategies, prefix and postfix interpolation, as shown in Fig. 2(a). The prefix strategy involves performing interpolation prior to the network, with the aim of mapping into a pseudo voxel grid. The network requires to learn a mapping, $M_{pre} : \mathbf{S}' \rightarrow \mathbf{S}$, where $\mathbf{S}'$ denotes this pseudo voxel grid of size $D_s \times H \times W$. The postfix strategy is firstly to learn a one-to-one voxel grid mapping, $M_{post} : \mathbf{I} \rightarrow \mathbf{I}' \rightarrow \mathbf{S}$, then interpolate it to S, where $\mathbf{I}'$ denotes the network output voxel grid of size $D_i \times H \times W$. The prefix interpolation is the traditional algorithm implementation, e.g., nearest or trilinear, while the postfix interpolation is a learnable module that adopts deconvolution as the upsampling layer. The prefix strategy generates the pseudo voxel grid before the model input; in this strategy, the Z-axis information is implicitly restored through learning the fluorescence prediction. In contrast, the postfix strategy, learning the restore procedure through an explicit upsampling layer, separates two processing purposes. We argue that implicitly learning the spatial context restoration benefits the subcellular prediction because the prefix strategy provides more prior structure knowledge with interpolation.

3.3 Hybrid Dimensions Topology

All prior works employ pure 3D topologies for prediction because the input and output is the one-to-one voxel associated to 3D space. Despite their accuracy, convolutions or deconvolutions implemented in the 3D manner are cost more GPU memory and FLOPs than 2D. So it is natural to reduce the 3D operations. To this end, we propose a hybrid dimensions topology eschewing the use of costing 3D convolutions in encoding and decoding. In Fig. 2(b), we demonstrate four topology strategies (i.e., 2D, 3D, 2-to-3D and 3-to-2D). Here we adopt the native UNet as the baseline (i.e., contains 5 layers in encoder and decoder), and the UNet layers follow the residual-style convolution architecture.

Dimension transformation. This core implementation of SparseSSP enables the 3-to-2D or 2-to-3D feature transformation. Transforming to a 2D feature permits all existing advanced modules or technologies achieved on 2D features to migrate into our framework. It is an encouraging improvement, enabling more extensive development based on SparseSSP in the future. As depicted in Fig. 2(c), there are two transformation routes, i.e., depth-to-channel and channel-to-depth. We introduce the route of depth-to-channel first. In simple terms, we compress the 3D feature into a 2D space, and it looks like we are folding the dimensions from the Z-axis into channels. Here, we discuss two projection implementations, as shown in the right part of Fig. 2(c). The embedding in 2D means the feature is projected in 2D space after the channel arrangement. Embedding in 3D means that the feature is projected in 3D space (as shown in the left part of Fig. 2(c)), which keeps relatively complete 3D structural representation information before the channel arrangement. Dimension transformation via channel arrangement would break the structural expression of 3D space; thus, it is more rational to embed in 3D, though it costs a little more computation than embedding in 2D. The route of channel-to-depth is similar to the above implementations, and the most significant difference is to unsqueeze the 2D feature to a 3D representation.

Pure 2D Topology. This 2D network for 3D SSP is utilized to compare with the following hybrid dimensional topologies. Inspired by FlashOcc's [27] fully 2D operations for 3D occupancy prediction, we would like to verify the effectiveness of this implementation in SSP. We first reshape the input voxel into 2D space. Specifically, the Z-axis dimension is squeezed to the channel dimension, as formulated like $\mathbf{x}_n \in \mathbb{R}^{1\times D_n \times H_n \times W_n}$ transforms to $\mathbf{x'}_n \in \mathbb{R}^{D_n \times H_n \times W_n}$, where $\mathbf{x}_n$

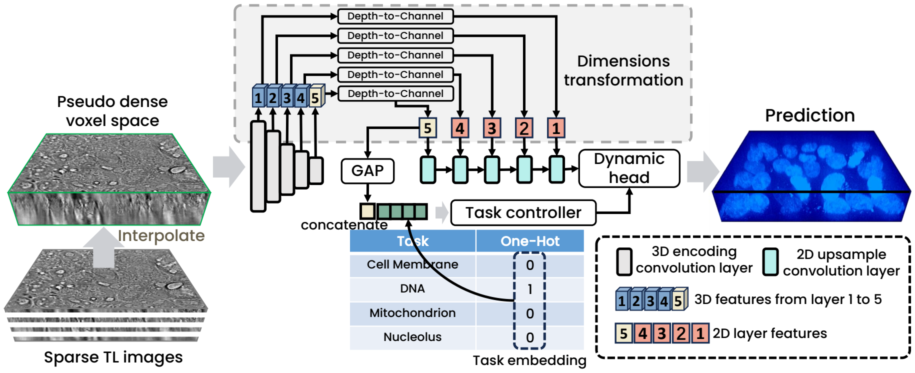


Fig. 3: Overview of SparseSSP with 3-to-2D topology and prefix interpolation. It is an example following the DoDNet [28] style that utilizes the task controller and dynamic head. The framework's task embedding approach can be conveniently changed to other technologies.

denotes the TL input. For the network output, we reshape the prediction from $\mathbf{y'}_n \in \mathbb{R}^{D_n \times H_n \times W_n}$ to $\mathbf{y}_n \in \mathbb{R}^{1 \times D_n \times H_n \times W_n}$. In this topology, the depth D_n of the 3D feature is treated as the channel dimension during the encoding and decoding.

Hybrid of 3-to-2D. The hybrid of 3-to-2D topology contains a 3D encoder and a 2D decoder. After the 3D feature extraction, a dimension transformation in a depth-to-channel manner is required to convert the 3D features into 2D space. There are 5 layers in the encoder, following the RepMode layer settings, let $\mathbf{C}_e = \{32, 64, 128, 256, 256\}$ denotes as the output channels of 3D space. Note that the last layer is the bottleneck layer. We perform depth-to-channel transformation for each layer. Let $k \in [1, 5]$ denote the layer index, and the resolution of 3D features can be formatted as $x_k \in \mathbb{R}^{C_k \times D_k \times H_k \times W_k}, C_k \in \mathbf{C}_e$. Let U denote the upsample embedding uniform dimension (we fixed the setting as 256 in the experiments; see more comparisons about this hyper-parameter in supplementary materials), which can trade off the model's performance and scale. Here, we discuss two projection implementations in formulation.

1) If we assume that the projection is embedded in 3D space. Define the $\hat{x_k} = \varGamma(\lambda, \mu, x_k)$ as the projection function, where the λ and μ are the input and output channels, where $\hat{x_k}$ denotes the projected feature; here, it must satisfy that $\lambda = C_k$ and $\mu = C_u / D_k$ (e.g., in the output of bottleneck layer, $U = 256, C_k = 256, D_k = 2, H_k = 8, W_k = 8$, and the λ and μ must be defined as 256 and 128). We employ a 3D convolution on kernel size of 1 as the projecting function. Then reshape the $\hat{x_k} \in \mathbb{R}^{\mu \times D_k \times H_k \times W_k}$ to $\mathbb{R}^{\mu D_k \times H_k \times W_k}$.

2) If we assume that the projection is embedded in 2D space. We firstly reshape $\hat{x_k} \in \mathbb{R}^{C_k \times D_k \times H_k \times W_k}$ to $\mathbb{R}^{C_k D_k \times H_k \times W_k}$. For the projection function $\varGamma$, the parameters λ and μ are required to be formulated as $\lambda = C_k D_k$ and $\mu = U$. We employ a 2D convolution as the projecting function.

After the dimension transformation, we employ the 2D decoding layers to upsample these features. The output channel of the final convolution head equals the length of the Z-axis in the target voxel grid, e.g., an inference patch is $32 \times 128 \times 128$, and the final output channel is 32.

Hybrid of 2-to-3D. The hybrid 3-to-2D topology contains a 2D encoder and a 3D decoder. We first reshaped the 3D TL input into a 2D shape like the pure 2D topology. Different from 3-to-2D, we employ the channel-to-depth transformation, as shown in Fig. 2(b) and (c). The 2D encoder can be changed to other backbones, and the decoder is the same as the typical 3D topology.

3.4 Task Eembeding Mechanisms

SparseSSP is designed as a plug-and-play paradigm, meaning all components can be conveniently changed to existing state-of-the-art technologies. For example, the task embedding approach, e.g., task controller (DoDNet) [28], task guiding enhancement (Tgnet) [26], task switching (TSNs) [23], and reparameterize diverse expert (RepMode) [17,30]. We demonstrate an example of a task controller and dynamic head following DoDNet in Fig. 3.

4 Experiments

4.1 Experimental setup

Benchmark. We conduct the experiments based on AllenCell collection [20], containing 12 subcellular structures (i.e., ActinFilament, ActomBundle, CellMembrane, Desmosome, DNA, EndopReticulum, GolgiApparatus, Microtubule, Mitochondria, NuclearEnvelope, Nucleolus and TightJunction). We follow the data distribution (i.e., for each structure, select 25% of samples for evaluation and 10% of the rest for the test) and preprocessing (e.g., z-scored normalization for voxels and data augmentation) as identical to the RepMode [30] (we directly adopt the prefabricated data splitting from the RepMode [30]). The fluorescence labels are resized to make the voxel correspond to $0.29 \times 0.29 \times 0.29 \mu m^3$, while the TL input images are resized to make the voxel correspond to $(0.29\mathbf{r}) \times 0.29 \times 0.29 \mu m^3$ ($\mathbf{r}$ is the sparsity rate). We mainly compare the sparsity ratios of 2, 4, and 8, and the sparsity ratio 8 means that it can already reduce 87.5% imaging frequencies compared to the source microscopy session.

Evaluation Metrics. Following the RepMode, we adopt three metrics for SSP, i.e., the mean absolute error (MAE), mean square error (MSE), and the coefficient of determination (R^2) to measure the quality of prediction. MSE and MAE are in arbitrary ranges, and the smaller value is better, while the R^2 ranges from 0 to 1, and the larger value is better performance. Moreover, unless otherwise specified, we compute the mean value from 12 datasets as the overall performance.

Training Settings. For 3D topologies, we use 1000 training epochs, the Adam optimizer, MSELoss function, batch size of 8, and evaluate every 20 epochs, following RepMode. For 3D and 2D blending networks, we extend training to 3000

			r = 2			**r** = 4			**r** = 8		
Approach	Interp.	Topology	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
RepMode	None	2D	.6055	.4420	.3632	.6096	.4513	.3542	.6109	.4523	.3493
RepMode	None	3-to-2D	.5668	.4376	.4032	.5769	.4376	.3813	.6041	.4423	.3690
DoDNet	None	2D	.6103	.4509	.3524	.6238	.4623	.3243	.6298	.4602	.3103
DoDNet	None	3-to-2D	.5781	.4368	.3813	.5860	.4421	.3794	.6013	.4469	.3638
RepMode [30]	post	3D	.5455	.4234	.4193	.5539	.4360	.4137	.5886	.4424	.3736
RepMode	pre	3D	.5210	.4248	.4453	.5320	.4197	.4359	.5687	.4359	.3984
RepMode	post	2D	.6043	.4471	.3613	.6154	.4500	.3460	.6189	.4562	.3324
RepMode	pre	2D	.5704	.4532	.3984	.5812	.4383	.3857	.5896	.4484	.3849
RepMode	post	2-to-3D	.5458	.4236	.4250	.5523	.4312	.4153	.5896	.4447	.3790
RepMode	pre	2-to-3D	.5135	.4173	.4543	.5299	.4194	.4397	.5624	.4286	.3971
RepMode	post	3-to-2D	.5234	.4132	.4589	.5356	.4232	.4264	.5780	.4313	.3862
RepMode	pre	3-to-2D	**.5069**	.4140	**.4616**	**.5159**	.4136	**.4523**	.5468	.4222	.4207
DoDNet [28]	post	3D	.5343	.4241	.4243	.5541	.4313	.4117	.5768	.4381	.3861
DoDNet	pre	3D	.5173	.4257	.4512	.5392	.4318	.4288	.5572	.4316	.4103
DoDNet	post	2D	.6012	.4487	.3643	.6045	.4412	.3632	.6123	.4561	.3520
DoDNet	pre	2D	.5751	.4372	.3883	.5793	.4342	.3734	.5823	.4413	.3699
DoDNet	post	2-to-3D	.5486	.4329	.4232	.5554	.4382	.4143	.5774	.4367	.3903
DoDNet	pre	2-to-3D	.5244	.4185	.4463	.5367	.4150	.4221	.5535	.4234	.4145
DoDNet	post	3-to-2D	.5354	.4213	.4201	.5475	.4335	.4172	.5634	.4318	.4032
DoDNet	pre	3-to-2D	.5128	**.4118**	.4516	.5229	**.4133**	.4452	**.5440**	**.4209**	**.4236**

Table 1: Impact of the interpolation strategies. The prefix strategies have a clear lead compared with no interpolation and the postfix. The **r** is the sparsity ratio of the input voxel grid. No interpolation can only be implemented in a 2D decoding manner. Gray-back areas are the better settings.

		r = 2			**r** = 4			**r** = 8		
Proj. space	Topology	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
3D	3-to-2D	.5128	.4118	.4516	.5229	.4133	.4452	.5440	.4209	.4236
2D	3-to-2D	.5351	.4149	.4336	.5405	.4221	.4143	.5660	.4389	.3969
3D	2-to-3D	.5244	.4185	.4463	.5367	.4150	.4221	.5535	.4234	.4145
2D	2-to-3D	.5313	.4142	.4377	.5437	.4265	.4246	.5718	.4382	.3913

Table 2: Impact of the projection space. We compare the different implementations in dimension transformation based on the DoDNet [28]. The projection in 3D space performed better than the projection in 2D space. It can maintain more spatial context for 3-to-2D topologies, while for 2-to-3D, it can pre-reconstruct a 3D space and refine it during the layer-by-layer upsampling.

epochs. Pure 2D topologies undergo 7000 epochs of training with consistent settings. Despite needing more epochs to converge, the hybrid dimensions topology trains faster than pure 3D layers. Detailed resource costs are in Sec. 4.3. Using the Gaussian sliding inference strategy from RepMode [30], the window size is $32 \times 128 \times 128$. All experiments were conducted on an NVIDIA RTX 3090 with 24GB memory, and our experiments are easily replicable on devices with at least 18GB of memory.

4.2 Ablation studies

We conduct ablative experiments to evaluate each proposal in our framework. We implement the SOTA multi-task methods (i.e., DoDNet [28], Tgnet [26], TSNs [23], CondNet [6], and RepMode [30]) on our topologies strategies (i.e.,3D, 2D, 2-to-3D and 3-to-2D) with different interpolation strategies. We employ the 3D channel projection for hybrid dimensions topologies in both channel-to-depth and depth-to-channel unless stated otherwise.

		$\mathbf{r} = 2$			$\mathbf{r} = 4$			$\mathbf{r} = 8$		
Approach	Topology	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
TSNs [23]	3D	.5540	.4325	.4135	.5663	.4358	.4009	.5871	.4463	.3796
TSNs	2D	.6024	.4474	.3640	.6046	.4488	.3617	.6111	.4513	.3551
TSNs	2-to-3D	.5343	.4198	.4323	.5620	.4322	.4025	.5846	.4396	.3897
TSNs	3-to-2D	.5164	.4101	.4519	.5328	.4174	.4351	.5542	.4262	.4132
CondNet [6]	3D	.5388	.4279	.4225	.5425	.4334	.4113	.5633	.4446	.3928
CondNet	2D	.5864	.4419	.3821	.5878	.4517	.3814	.5943	.4528	.3683
CondNet	2-to-3D	.5267	.4134	.4416	.5413	.4221	.4094	.5612	.4148	.4001
CondNet	3-to-2D	.5211	.4123	.4501	.5326	.4195	.4281	.5498	.4202	.4205
DoDNet [28]	3D	.5173	.4257	.4512	.5392	.4318	.4288	.5572	.4316	.4103
DoDNet	2D	.5751	.4372	.3883	.5793	.4342	.3734	.5823	.4413	.3699
DoDNet	2-to-3D	.5244	.4185	.4463	.5367	.4150	.4221	.5535	.4234	.4145
DoDNet	3-to-2D	.5128	.4118	.4516	.5229	**.4133**	.4452	.5440	**.4209**	.4236
RepMode [30]	3D	.5210	.4248	.4453	.5320	.4197	.4359	.5687	.4359	.3984
RepMode	2D	.5704	.4532	.3984	.5812	.4383	.3857	.5896	.4484	.3849
RepMode	2-to-3D	.5135	.4173	.4543	.5299	.4194	.4397	.5624	.4286	.3971
RepMode	3-to-2D	.5069	.4140	.4616	.5159	.4136	.4523	.5468	.4222	.4207
Tgnet [26]	3D	.5224	.4259	.4459	.5421	.4315	.4257	.5691	.4437	.3981
Tgnet	2D	.5690	.4327	.4034	.5713	.4368	.3947	.5773	.4351	.3896
Tgnet	2-to-3D	.5034	.4122	.4519	.5231	.4143	.4503	.5498	.4291	.4164
Tgnet	3-to-2D	**.4995**	**.4113**	**.4693**	**.5158**	.4166	**.4526**	**.5386**	.4231	**.4292**

Table 3: Topologies comparisons for sparse-view SSP. We compare the performance of existing mainstream multi-task approaches implemented in diverse topologies. For hybrid dimensions topologies, i.e., the 3-to-2D and 2-to-3D, employ the 3D space projection in the dimension transformation. All of these experiments are prefix interpolations. The red-bold row has less MSE and MAE and the largest R^2, and the rows in gray-back are the winning strategy with a group.

Interpolation Strategies. We implement the active interpolation in SparseSSP for one-to-many voxel mapping in Z-axis. We compare the performance on two approaches, i.e., prefix and postfix, and the details are provided in Tab. 1(row5-20). For a fairness, we utilize the nearest algorithm in all interpolations. In addition, there is a naive alternative to interpolation, but only for the topologies with the 2D decoding manner. That is, directly define the output channels as same as the Z-axis length of voxel grid (e.g., the Z-length of target grid is 32, and set the output channels of final head as 32). We demonstrate the performance on this alternative in Tab. 1(row1-4), which are described as "None". The results indicate that, when employing the prefix interpolation technique, both DoDNet and RepMode representational models outperform the postfix and none interpolation. Prefix interpolation enables the incorporation of prior spatial knowledge into the model learning process, thereby enhancing the performance of sparse-view SSP; therefore, we perform the prefix strategy on the following experiments.

Projection Operation. Projection is an essential operation in hybrid dimensions topology, and the space projected in which dimension means maintaining the spatial context in where. We compare two implementations in Tab. 2, i.e., projected in 3D or 2D, on DoDNet, and it can be observed that projection in 3D outperforms 2D space. Projection in 3D maintains more spatial context for 3-to-2D topologies, while for 2-to-3D, it can pre-reconstruct a 3D space and refine it during the layer-by-layer upsampling. Maintaining a relatively complete feature transformation is crucial for hybrid dimensions topology.

Topology Strategies on Diverse Methodologies. We compare 5 SOTA multi-task methodologies, and the performance of diverse topology strategies are as listed in Tab. 3. We demonstrate a trend of R^2 values as the sparsity

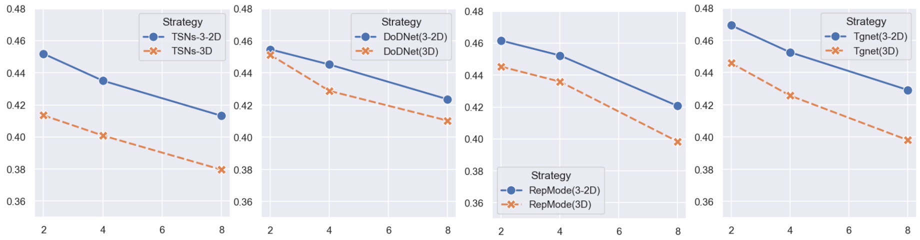


Fig. 4: Trend of R^2 value as sparsity ratio increased from 2 to 8. Hybrid dimensions topology 3-to-2D (i.e., the blue lines in the figure) shows a slower decay and higher global scores than pure 3D topology (i.e., the orange lines).

ratio increased from 2 to 8 in Fig. 4, and it can be obverse that the quality of the prediction keeps getting worse as the sparsity ratio increases; however, the 3-to-2D topology shows a slower decay than pure 3D networks. From Tab. 3, it can be observed that the hybrids, i.e., 3-to-2D and 2-to-3D, outperform pure 2D and 3D on diverse sparsity ratios, and the pure 2D strategies are much poorer than other implementations. Notably, the 3-to-2D strategy based on Tgnet [26] achieves the best score in all metrics, and RepMode [30], which is the SOTA method in dense-view SSP, fell to Tgnet by a narrow margin. Hybrid dimensions topologies have the leading advantages on sparse-view SSP, especially the 3-to-2D strategy, which encodes in 3D space and decodes the 2D features. It means that the 3D feature extractor is essential for a hybrid strategy. Meanwhile, the Z-axis dimension is folded into the channel features after the transformation of dimensions, and it permits the sparse Z-axis information to be refined during the 2D layer-by-layer upsample process. In contrast, 2D features from 2D extractors are more difficult for downstream 3D decoders to understand, and the cases involved in 2D encoding would lead to a longer convergence process. Therefore, we will discuss more potential factors in the following sections.

More Training Epochs. We save the best checkpoint on validation data during training. For 3D and 3-to-2D topologies, increasing training epochs to 7000 does not produce a better checkpoint; however, epochs increasing benefits the topologies with 2D encoders. As shown in Tab. 4, more training epochs would lead to a little better performance, which means that there is still room for optimization.

		r = 2		
epochs	Topology	MSE	MAE	R^2
7000	2D	.5751	.4372	.3883
14000	2D	.5623	.4252	.4176
3000	2-to-3D	.5244	.4185	.4463
7000	2-to-3D	.5231	.4183	.4485

Table 4: We increase the training epochs on DoDNet [28] to study whether the 2D encoder requires a longer convergence time.

Impact of 2D Backbone in 2-to-3D Topology. The 2D feature extractor involved topologies, i.e., the pure 2D and 2-to-3D, which enables the replacement of the 2D modules for its components. We found that increasing the epochs would improve its results, which means it is possible to increase the volume of 2D network parameters. We conducted an ablation study on TSNs, which demonstrated inferior performance

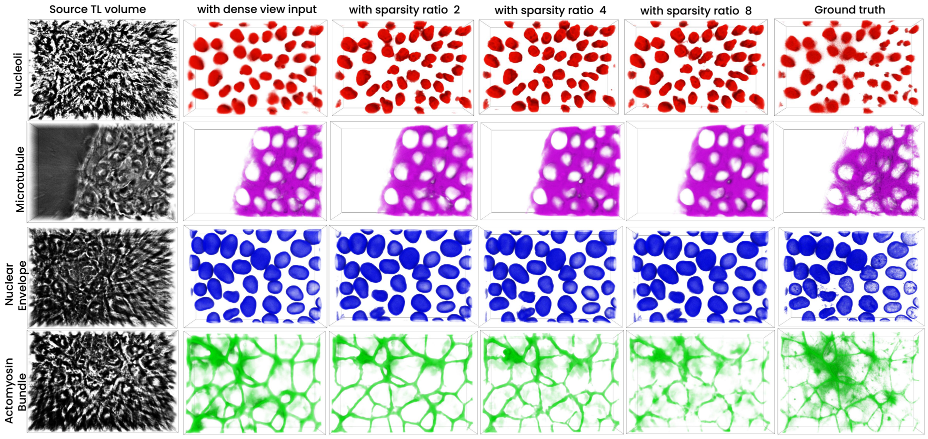


Fig. 5: Visualization of subcellular structure prediction. A stereoscopic volume display from a top-down Z-axis view. We compare results of four structures using dense-view and sparse-view TL images at different sparsity ratios. The second column shows results from RepMode (best for dense-view SSP), and columns three to six show results from Tgnet with the best 3-to-2D topology for sparse-view.

compared to other implementations, to assess the impact of backbone replacement. We employ a SOTA transformer-based backbone Swin-Transformer in its base architecture (Swin-B) for replacement. Swin-B is much larger than the native backbone ResNet18 of TSNs on parameters. In Tab. 5, we can observe a better performance after increasing the parameters, which means there is currently a correlation between the topologies' performance in 2D feature extractors and the design of their encoders. It shows the potential to introduce more powerful technologies from 2D network design, e.g., some of the latest transformer designs.

Summarized for Ablations. We demonstrate the performances of two interpolation strategies and four topologies' combinations in sparse-view SSP. We

		r = 2			**r = 8**		
Backbone	Topology	MSE	MAE	R^2	MSE	MAE	R^2
RestNet18	2D	.6024	.4474	.3640	.6111	.4513	.3551
Swin-Base	2D	.5867	.4412	.3800	.5977	.4490	.3687
ViT-Base	2D	.5889	.4420	.3812	.5813	.4323	.3846
DINOv2-Base	2D	.6143	.4563	.3507	.6324	.4682	.3340
ViT-Large	2D	.5674	.4425	.3995	.5713	.4402	.3893
RestNet18	2-to-3D	.5343	.4198	.4323	.5846	.4396	.3897
Swin-Base	2-to-3D	.5266	.4132	.4501	.5623	.4323	.4133
ViT-Base	2-to-3D	.5300	.4157	.4385	.5641	.4428	.3970
DINOv2-Base	2-to-3D	.5543	.4305	.4124	.5952	.4561	.3628
ViT-Large	2-to-3D	**.5155**	**.4098**	**.4559**	**.5524**	**.4240**	**.4194**

Table 5: Ablations on 2D backbones, based on the TSNs [23], generally exhibit poorer performance than other implementations; thus, we replaced the 2D backbone for it. It is encouraging to see that the results are improved with the transformer backbones, which means that there is potential to migrate more powerful 2D network technologies into our framework, e.g., some of the latest transformer designs.

found that the combination of prefix interpolation and 3-to-2D strategies demonstrated significantly better performance than others. Moreover, the projection in 3D space during the dimension transformation of hybrid dimensions topologies is better than in 2D space. In addition, training epochs and backbone types would also affect the topologies of 2D and 2-to-3D. The effectiveness and advantages of our proposals in this paper are validated in the above ablation studies.

4.3 Analyzation of Resource Consumption

SSP is a 3D dense regression problem, and it is logical to utilize the 3D networks to solve it. The proposed hybrid dimensions topologies incorporate dimension transformation, facilitating the replacement of 3D operations and significantly reducing computational costs (including the training and inference procedures). We demonstrate details of resource consumption on different topologies based on the experiments under sparsity ratio 2. In Tab. 6 it can be seen that pure 2D topology is the most efficient. However, its evaluation metrics are poorer than those of other topologies. Hybrid dimension topologies cost less resources and achieve excellent performance compared to 3D. The pure 3D topology costs larger MACs and memory than others. The 2-to-3D surpasses the 3-to-2D in resource efficiency due to the replacement of the feature extractor with 2D layers. Although 2-to-3D is slightly behind 3-to-2D in performance, it trades off performance and efficiency, which is the most cost-effective solution.

4.4 Different Subcellular Structures

Visualization. We compare the prediction results under dense-view and sparse-view input in Fig. 5. Without staining or using SSP, it is difficult for biologists to directly observe subcellular structures from TL images, as shown in column 1. As shown in Fig. 5, the sparse-view SSP can show the position and shape of organelles and fluorescent staining. From the visualization results, SSP is a reliable alternative to fluorescent staining. It can be observed that with the

		GPU Infer.		GPU Train.		Computation
Approach	Topology	time(s/iter)	Mem.(MiB)	Speed(iter/s)	Mem.(MiB)	MACs
RepMode	3D	4.47	9122	0.89	17843	66.29G
RepMode	2D	0.66	2472	3.31	3548	2.33G
RepMode	2-to-3D	1.86	3666	1.86	8984	30.11G
RepMode	3-to-2D	2.55	5428	1.29	15521	43.47G
DoDNet	3D	2.11	4710	2.64	16054	113.86G
DoDNet	2D	0.35	1982	4.70	2692	1.82G
DoDNet	2-to-3D	1.04	2897	3.31	4384	41.02G
DoDNet	3-to-2D	1.35	4124	4.53	14362	76.61G
TSNs	3D	1.09	6862	7.19	10458	55.70G
TSNs	2D	0.37	1720	31.12	2793	2.05G
TSNs	2-to-3D	0.86	2028	10.2	5729	13.94G
TSNs	3-to-2D	0.91	3554	8.91	7522	46.77G

Table 6: Comparisons on resource consumption. We compare the TSNs [23], DoDNet [28] and RepMode [30] in diverse topologies. Hybrid dimensions topologies demonstrate less resource consumption than pure 3D, especially in MACs. The number of iterations in training is the number of the loss backward operations.

Methods	Actin Filament			Actom. Bundle			Cell Membrane			Desmosome			DNA			Endop. Reticulum			Golgi Apparatus		
	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
Multi-Net($\mathbf{r} = 1$) [20]	.4241	.4716	.5695	.7247	.4443	.2606	.5940	.4351	.3930	.8393	.5640	.0162	.5806	.5033	.3822	.4635	.4914	.5262	.8023	.5732	.0801
Tgnet-3-to-2D($\mathbf{r} = 2$) [26]	.3896	.4476	.6044	.6670	.4080	.3196	.5079	.4160	.4810	.8379	.5646	.0175	.4873	.4639	.4808	.4219	.4648	.5688	.7822	.5754	.1028
Tgnet-3-to-2D($\mathbf{r} = 4$)	.3928	.4492	.6011	.6726	.4084	.3138	.5335	.4223	.4548	.8385	.5633	.0168	.4797	.4587	.4888	.4372	.4736	.5532	.7931	.5790	.0905
Tgnet-3-to-2D($\mathbf{r} = 8$)	.4107	.4570	.5829	.6973	.4166	.2886	.5529	.4298	.4350	.8408	.5637	.0141	.5087	.4720	.4578	.4449	.4748	.5453	.7975	.5777	.0853

Methods	Microtubule			Mitochondria			Nuclear Envelope			Nucleolus			Tight Junction			All			Δ_{Imp} (%)		
	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2	MSE	MAE	R^2
Multi-Net($\mathbf{r} = 1$) [20]	.3682	.4348	.6296	.4684	.3921	.5172	.3014	.3006	.6954	.2164	.1789	.7826	.6474	.3369	.3370	.5341	.4269	.4337	0.000	0.000	0.000
Tgnet-3-to-2D($\mathbf{r} = 2$) [26]	.3230	.4082	.6750	.4726	.4058	.5129	.2783	.2892	.7186	.1995	.1679	.7996	.6295	.3262	.3554	.4995	.4113	.4693	6.4782	3.6543	8.2084
Tgnet-3-to-2D($\mathbf{r} = 4$)	.3546	.4292	.6432	.5011	.4132	.4835	.3044	.2968	.6923	.2101	.1708	.7890	.6699	.3322	.7926	.5158	.4166	.4526	3.4263	2.4127	4.3579
Tgnet-3-to-2D($\mathbf{r} = 8$)	.3401	.4165	.6578	.5403	.4249	.4431	.3471	.3149	.6491	.2356	.1825	.7633	.7520	.3491	.2301	.5386	.4231	.4292	-0.8425	.8901	-1.037

Table 7: Comparisons of different structures on diverse sparsity ratios. We utilize the FNet as the reliable baseline to metric our proposals. The Tgnet-3-to-2D is the best setting in SparseSSP, and its performance on different subcellular structures achieves a reliable level for actual biological research. In sparsity ratio 8, i.e., reducing 87.5% imaging times, SparseSSP only shows a minor performance decay.

sparse view in higher ratios, the prediction results maintain the quality well with dense views, e.g., the Nucleoli, Microtubule, and NuclearEnvelope. While for the slender striped structures, e.g., ActomyosinBundle, sparse-view would cause a few deficiencies. This is primarily due to the electronic noise of the fluorescence label adversely affecting the learning of ActomyosinBundle.

Quantitative Comparison. Tab. 7 reports the detailed results of 12 datasets. FNet [20] (i.e., the Multi-Net in the table) is sufficient to assist with biological research; thus, we set it as the reliable baseline for application use on different subcellular structures. Following RepMode, we utilize a relative performance improvement metric over Multi-Net (i.e., Δ_{Imp}). We compare the best solution in hybrid dimensions topology, i.e., Tgnet-3-to-2D, on diverse sparsity ratios. Our framework proposal (i.e., hybrid dimensions topology) demonstrates keeping the small MSE and MAE with reduced imaging frequency.

5 Conclusion

We propose a novel framework, SparseSSP, featuring hybrid dimensions topology to achieve fast and efficient SSP using sparse transmitted light images, thus reducing the need for extensive microscopy imaging and computational costs. SparseSSP leverages a one-to-many voxel mapping paradigm to implicitly refine sparse Z-axis information. Unlike prior works that rely solely on pure 3D topologies, our approach integrates hybrid 3D and 2D network layers to enhance efficiency, striking a balance between computational cost and prediction quality. By incorporating dimension transformation into the network, we convert Z-axis information into channel features, allowing advanced 2D modules to effectively address 3D SSP challenges. Furthermore, we explore the potential of migrating 2D backbones to SparseSSP. Extensive experiments with SOTA multi-task methods across various sparse ratios demonstrate the effectiveness and advantages of our approach. To the best of our knowledge, this work is pioneering in utilizing sparse-view TL images for SSP, significantly advancing the visualization of rapid biological dynamics on low-cost devices and samples.

Acknowledgments. We thank all authors, reviewers and the chair for the excellent contributions. Q. Liu and Y. Ding contributed equally to this paper. This work was supported in part by a grant from Key Area Research and Development Program of Guangdong Province (2022B1111010003), in part by the National Key Research and Development Plan of China (2023YFB3210400), in part by a grant from the Innovative Team Project of Jinan (No. 202333009), in part by a grant from National Natural Science Foundation of China (62273329), in part by a grant from Key Fundamental Research Program of Shenzhen (JCYJ20210324115601004), in part by a grant from Natural Science Foundation of Guangdong Province (2022A1515011399), and in part by a grant from the Major Scientific and Technological Innovation Project of Shandong Province (2021CXGC010603).

References

1. Bock, F.J., Tait, S.W.: Mitochondria as multifaceted regulators of cell death. Nature reviews Molecular cell biology **21**(2), 85–100 (2020) 1
2. Carlton, J.G., Jones, H., Eggert, U.S.: Membrane and organelle dynamics during cell division. Nature Reviews Molecular Cell Biology **21**(3), 151–166 (2020) 1
3. Chen, B.C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E., Davidson, M.W., Janetopoulos, C., Wu, X.S., Hammer III, J.A., Liu, Z., et al.: Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. Science **346**(6208), 1257998 (2014) 2
4. Cheng, S., Fu, S., Kim, Y.M., Song, W., Li, Y., Xue, Y., Yi, J., Tian, L.: Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy. Science advances **7**(3), eabe0431 (2021) 4
5. Christopher, J.A., Stadler, C., Martin, C.E., Morgenstern, M., Pan, Y., Betsinger, C.N., Rattray, D.G., Mahdessian, D., Gingras, A.C., Warscheid, B., et al.: Subcellular proteomics. Nature Reviews Methods Primers **1**(1), 1–24 (2021) 1
6. Dmitriev, K., Kaufman, A.E.: Learning multi-class segmentations from single-class datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9501–9511 (2019) 3, 4, 9, 10
7. Guo, Y., Li, D., Zhang, S., Yang, Y., Liu, J.J., Wang, X., Liu, C., Milkie, D.E., Moore, R.P., Tulu, U.S., et al.: Visualizing intracellular organelle and cytoskeletal interactions at nanoscale resolution on millisecond timescales. Cell **175**(5), 1430–1442 (2018) 1
8. Gut, G., Herrmann, M.D., Pelkmans, L.: Multiplexed protein maps link subcellular organization to cellular states. Science **361**(6401), eaar7042 (2018) 1
9. Han, Y., Ye, J.C.: Framing u-net via deep convolutional framelets: Application to sparse-view ct. IEEE transactions on medical imaging **37**(6), 1418–1429 (2018) 4
10. Hu, D., Liu, J., Lv, T., Zhao, Q., Zhang, Y., Quan, G., Feng, J., Chen, Y., Luo, L.: Hybrid-domain neural network processing for sparse-view ct reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(1), 88–98 (2020) 4
11. Icha, J., Weber, M., Waters, J.C., Norden, C.: Phototoxicity in live fluorescence microscopy, and how to avoid it. BioEssays **39**(8), 1700003 (2017) 2
12. Im, K., Mareninov, S., Diaz, M., Yong, W.H.: An introduction to performing immunofluorescence staining. Biobanking pp. 299–311 (2019) 2
13. Jo, Y., Cho, H., Park, W.S., Kim, G., Ryu, D., Kim, Y.S., Lee, M., Park, S., Lee, M.J., Joo, H., et al.: Label-free multiplexed microtomography of endogenous subcellular dynamics using generalizable deep learning. Nature Cell Biology **23**(12), 1329–1337 (2021) 4

14. Kandel, M.E., He, Y.R., Lee, Y.J., Chen, T.H.Y., Sullivan, K.M., Aydin, O., Saif, M.T.A., Kong, H., Sobh, N., Popescu, G.: Phase imaging with computational specificity (pics) for measuring dry mass changes in sub-cellular compartments. Nature communications **11**(1), 1–10 (2020) 4
15. Lee, H., Lee, J., Kim, H., Cho, B., Cho, S.: Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences **3**(2), 109–119 (2018) 4
16. Li, Z., Ma, C., Chen, J., Zhang, J., Shan, H.: Learning to distill global representation for sparse-view ct. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21196–21207 (2023) 4
17. Liu, P., Gu, C., Wu, B., Liao, X., Qian, Y., Chen, G.: 3d multi-organ and tumor segmentation based on re-parameterize diverse experts. Mathematics **11**(23), 4868 (2023) 8
18. Mescheder, L.M., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. CoRR **abs/1812.03828** (2018), http://arxiv.org/abs/1812.03828 3
19. Min, C., Xiao, L., Zhao, D., Nie, Y., Dai, B.: Uniscene: Multi-camera unified pre-training via 3d scene reconstruction (2024) 3
20. Ounkomol, C., Seshamani, S., Maleckar, M.M., Collman, F., Johnson, G.R.: Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. Nature methods **15**(11), 917–920 (2018) 2, 3, 4, 5, 8, 14
21. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016) 3
22. Skylaki, S., Hilsenbeck, O., Schroeder, T.: Challenges in long-term imaging and quantification of single-cell dynamics. Nature Biotechnology **34**(11), 1137–1144 (2016) 2
23. Sun, G., Probst, T., Paudel, D.P., Popović, N., Kanakis, M., Patel, J., Dai, D., Van Gool, L.: Task switching network for multi-task learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8291–8300 (2021) 3, 4, 8, 9, 10, 12, 13
24. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al.: A subcellular map of the human proteome. Science **356**(6340), eaal3321 (2017) 1
25. Wolff, G., Limpens, R.W., Zevenhoven-Dobbe, J.C., Laugks, U., Zheng, S., de Jong, A.W., Koning, R.I., Agard, D.A., Grünewald, K., Koster, A.J., et al.: A molecular pore spans the double membrane of the coronavirus replication organelle. Science **369**(6509), 1395–1398 (2020) 1, 2
26. Wu, H., Pang, S., Sowmya, A.: Tgnet: A task-guided network architecture for multi-organ and tumour segmentation from partially labelled datasets. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022) 3, 4, 8, 9, 10, 11, 14
27. Yu, Z., Shu, C., Deng, J., Lu, K., Liu, Z., Yu, J., Yang, D., Li, H., Chen, Y.: Flashocc: Fast and memory-efficient occupancy prediction via channel-to-height plugin. arXiv preprint arXiv:2311.12058 (2023) 3, 6
28. Zhang, J., Xie, Y., Xia, Y., Shen, C.: Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1195–1204 (2021) 3, 4, 7, 8, 9, 10, 11, 13

29. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view ct reconstruction method based on combination of densenet and deconvolution. IEEE transactions on medical imaging **37**(6), 1407–1417 (2018) 4
30. Zhou, D., Gu, C., Xu, J., Liu, F., Wang, Q., Chen, G., Heng, P.A.: Repmode: Learning to re-parameterize diverse experts for subcellular structure prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3312–3322 (2023) 2, 3, 4, 5, 8, 9, 10, 11, 13