

<https://doi.org/10.1038/s44387-025-00004-y>

Hitchlearning: a general free-lunch paradigm for single-image enhancement by unifying inference and training

Jiaxin Liu^{1,2,3,7}, Yuchen Guo^{1,3,4,7}, Lu Fang^{1,3,5} ✉ & Qionghai Dai^{1,3,6} ✉

Deep learning (DL) has ushered in a suite of promising tools for image processing, including denoising (DN), deblurring (DB), and super-resolution (SR). However, traditional DL methods assume independent and identically distributed (i.i.d.) data for model training and inference, which does not hold in practice due to factors such as sample variation, variability in imaging conditions, and temporal effects in living cells. This discrepancy prevents models, even when trained on extensive datasets, from reaching their full performance potential during inference. This practical issue, unfortunately, is still unexplored. To address this issue, drawing inspiration from the way biological intelligence adapts through past experiences to shape future learning, we introduce HitchLearning—a revolutionary paradigm that breaks away from the conventional separation of training and inference. In HitchLearning, we leverage a single inference image to unsupervisedly optimize the model by aligning the training images with it. Subsequently, we employ this optimized model for processing individual inference images. This approach allows the model to adapt to the specific characteristics of each inference image, leading to improved results in a manner reminiscent of a “free lunch.” We conducted a thorough evaluation of our method across three distinct tasks within both supervised and unsupervised frameworks, utilizing four diverse datasets. Compared to conventional training methods, HitchLearning demonstrated average performance increases of 4.34 dB, 4.08 dB, and 0.54 dB for the DN, DN, and SR tasks, respectively. The experimental results unequivocally demonstrate that our algorithm offers a universally applicable and cost-free optimization solution for processing image and can be used in other fields as well.

Since the time of Galileo, imaging has been the “eyes of science”¹. The acquisition of high-quality biological images has assumed a pivotal role in biological science research, as advancements in imaging technology have provided essential tools for direct investigation^{2,3}. Traditional techniques like widefield fluorescence and laser scanning confocal microscopy are crucial in biology^{4,5}, and provide structural data⁶ but face resolution limits⁷. Advanced methods like confocal microscopy^{7–11}, multiphoton microscopy¹², structured illumination^{4,13,14}, light-sheet imaging^{6,15–18}, and STED microscopy^{19–22} aim to increase resolution, aiding neuroscience research. Nonetheless, these methods involve inherent physical resolution constraints⁷, necessitating the continued development of advanced

techniques to enhance image quality through denoising, deblurring, and resolution improvement.

DL excels in and provides a promising tool to image enhancement and restoration^{23,24}. With large datasets, given in a supervised way^{25–30} like low-high quality image pairs, or unsupervised way^{31–36} like corrupted pairs, DL may have good performance in many cases^{28–30,32,34,37–39}. However, it should be noted that most DL algorithms assume independently and identically distributed⁴⁰ (i.i.d.) training and inference data. Regrettably, this assumption is frequently violated in real-world scenarios due to factors such as sample variation (biological samples can exhibit significant differences even within the same species or cell line, stemming from genetic variations,

¹Beijing National Research Center for Information Science and Technology, Tsinghua University, 100084 Beijing, China. ²Tsinghua Shenzhen International Graduate School, Tsinghua University, 518071 Shenzhen, China. ³Institute for Brain and Cognitive Sciences, Tsinghua University, 100084 Beijing, China.

⁴Hangzhou Zhuoxi Institute of Brain and Intelligence, 311121 Hangzhou, China. ⁵Department of Electronic Engineering, Tsinghua University, 100084 Beijing, China.

⁶Department of Automation, Tsinghua University, 100084 Beijing, China. ⁷These authors contributed equally: Jiaxin Liu, Yuchen Guo.

✉ e-mail: fanglu@tsinghua.edu.cn; qhchai@tsinghua.edu.cn

environmental factors, or developmental stages), variability in imaging conditions (variations in lighting, exposure time, and focus can yield non-i.i.d data, and different microscopes or separate sessions with the same microscope can introduce variability in captured images), and temporal effects in living cells (time-lapse live-cell imaging can introduce non-i.i.d data due to changes in cells or imaging conditions during the experiment), ultimately resulting in less-than-optimal outcomes. In a simple experiment, we found that the PSNR may drop by approximately 0.5 to 6 dB when a DL model is applied to a different dataset. We require a novel approach to address the non-i.i.d issue between training and inference and unlock the model's performance potential, developing effective single-image restoration methods is pivotal to surmount these data and non-i.i.d -related challenges.

The detrimental effects of non-i.i.d. data stem from the segregation of training and inference (as shown in Fig. 1a, traditional paradigm is a two-stage framework), hindering the model's ability to adjust to the distinct attributes of individual inference images. To improve its performance, the central challenge lies in effectively utilizing a single inference image to optimize the model for the specific characteristics of the image. Inspired by biological intelligence, where past experiences sculpt neural circuits for future learning⁴¹, we introduce the Hitchlearning paradigm, which unifies training and inference (as shown in Fig. 1a, HitchLearning paradigm is a one-stage framework), by instantaneously adapting the model with any new inference data and the existing training data. Our paradigm (Fig. 1a, b) utilizes historical non-i.i.d. image data (Source) as prior experience, merging this prior experience with a single new inference image (Target) to optimize the model by aligning their distribution. In this way, the prior data provide general information for the task, while the inference image gives detailed features for the specific scenario. Traditional domain alignment strategies need multiple target data and source data to achieve the goal using deep neural network, while in our paradigm we only use single target data and source to realize domain-alignment in Fourier domain (Fig. 1b, c). With the adapted model, better performance can be achieved. Experiments demonstrate HitchLearning is versatile, as it can be applied to both supervised and unsupervised frameworks for deblurring, denoising, and super-resolution across various datasets (Fig. 1b).

Results

Hitchlearning facilitates distribution consistency

Training with single target image offers important but limited information, whereas historical images are abundant. However, due to the non-i.i.d. of historical images and single target image, domain adaptation⁴² (DA) is needed to bridge this gap. Here, we propose a simple yet effective strategy based on Fourier domain adaptation⁴³ (FDA) for HitchLearning. HitchLearning aligns source and target image distributions in the Fourier domain by replacing the source's amplitude spectrum with that of target (Fig. 1c), without requiring complex networks, convolution operations, or extensive data. This process enhances source-target distribution similarity, providing consistent data for further analysis.

HitchLearning minimizes discrepancies between training and inference images. We employ Kullback-Leibler divergence (KLD) to quantify the distribution shift from source image data before and after HitchLearning to target image data^{44,45}. Figure 2a depicts the fluctuations in Golgi vs. Golgi transferred by F-actin and Microtubule vs. Microtubule transferred by ER, indicating a reduction in KLD following the HitchLearning process. Additionally, the processed images became visually similar to the target images (Fig. 2b). The KLD results and the visual evidence show that HitchLearning enables consistent distributions, which indicates that HitchLearning can solve non-i.i.d. problems.

Hitchlearning provides advanced one-image denoising techniques for enhancing live-cell imaging

We employed a state-of-the-art model, Neighbor2Neighbor³³, as a foundation for unsupervised denoising of live-cell images. We adapted the Neighbor2Neighbor framework to include source and single target image

data, expanding its applications. Our study comprised seven distinct experimental settings (Supplementary Table 1), each utilizing a differently composed training dataset as described in Methods. We analyzed four datasets: endoplasmic reticulum (ER), F-actin filaments (F-actin)⁴⁶ Golgi, and Microtubule³⁸. ER and F-actin were imaged using multimodal SIM, while the Golgi apparatus and microtubules were examined via confocal microscopy. These datasets vary in structure, shape, and photographic style, impacting noise levels and resolution. In our approach, Golgi was the source of F-actin and Microtubule was the source of ER. We divided the F-actin and ER datasets into training and inference segments (details in Methods) using only inference ground truth data for metric evaluations. All the datasets were normalized from 0 to 1 to reduce scale differences and improve model performance. The model underwent 10 trial runs to obtain average results across the inference dataset, facilitating a comparison of the seven experimental settings. Here, Src embodies the existing traditional deep learning paradigm, while AllTrg symbolizes the theoretical maximum achievable under this paradigm, albeit requiring extensive additional data collection.

To quantitatively evaluate our HitchLearning method, we used four metrics: the mean-square error (MSE), mean-absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). For the F-actin dataset, HitchLearning notably exceeded the theoretical performance ceiling, AllTrg, our benchmark for theoretical maximum performance, across all metrics. These results, averaged over ten runs, are detailed in Fig. 3c and Supplementary Table 2. The Trg setting, which used only a single inference image for training, failed dramatically, achieving a PSNR of 11.64 dB and an SSIM of 0.15. In contrast, the SrcTrg setting performed much better (35.63 dB PSNR and 0.97 SSIM), but still did not match the AllTrg's performance (35.95 dB PSNR and 0.97 SSIM) that requires the collection of a large amount of data and is akin to training and testing on the dataset from the same domain, underscoring a significant gap in challenging scenarios. Remarkably, HitchLearning achieved 36.03 dB PSNR and 0.98 SSIM, demonstrating its superior ability to address the non-i.i.d problem in denoising tasks. Similarly, in the ER dataset, HitchLearning also yielded the best results, as shown in Supplementary Fig. 1b and Table 3. In addition to quantitative measures, visual comparisons (Fig. 3a and Supplementary Fig. 1a, 4) further confirmed HitchLearning's superiority over the other six settings. To directly visualize the improvements in these denoised images, we computed the pixelwise absolute difference between the denoised and corresponding ground truth images shown under each restored image.

Compared to the raw input inference images, the denoised images revealed a multitude of finer structures, including F-actin filaments. Moreover, we demonstrate that HitchLearning more precisely resolves the densely crisscross regions of F-actin cytoskeleton (Fig. 3b) and the line-scan profiles of F-actins inferred from HitchLearning are closer to the profiles of GT than those from Src (Fig. 3d). The statistical analysis qualified that HitchLearning typically achieved better DN imaging performance in terms of four metrics (Fig. 3e). This comparison clearly demonstrated that HitchLearning effectively facilitates the denoising of live-cell images, as evidenced by the reduced discrepancies in the fine structural details. In summary, HitchLearning presents a highly effective solution for addressing non-i.i.d challenges in image denoising tasks.

Hitchlearning significantly enhances the performance of bio-image deblurring

When photographing living cells, the movements of cells caught motion blur, out of focus caught defocus blur, etc. To better analyze cell growth and morphology, it is essential to obtain deblurred cell images. We chose MPRNet⁴⁷ as the base model for this task, when MPRNet was combined with HitchLearning, deblurring with only one image became reality. MPRNet is a supervised method, here, we use the deblurred image obtained from Source model as the pseudo ground truth to train models in the remaining settings. The data for the deblurring are generated by adding motion blur to the raw images³⁸. We use different blur kernels for the source

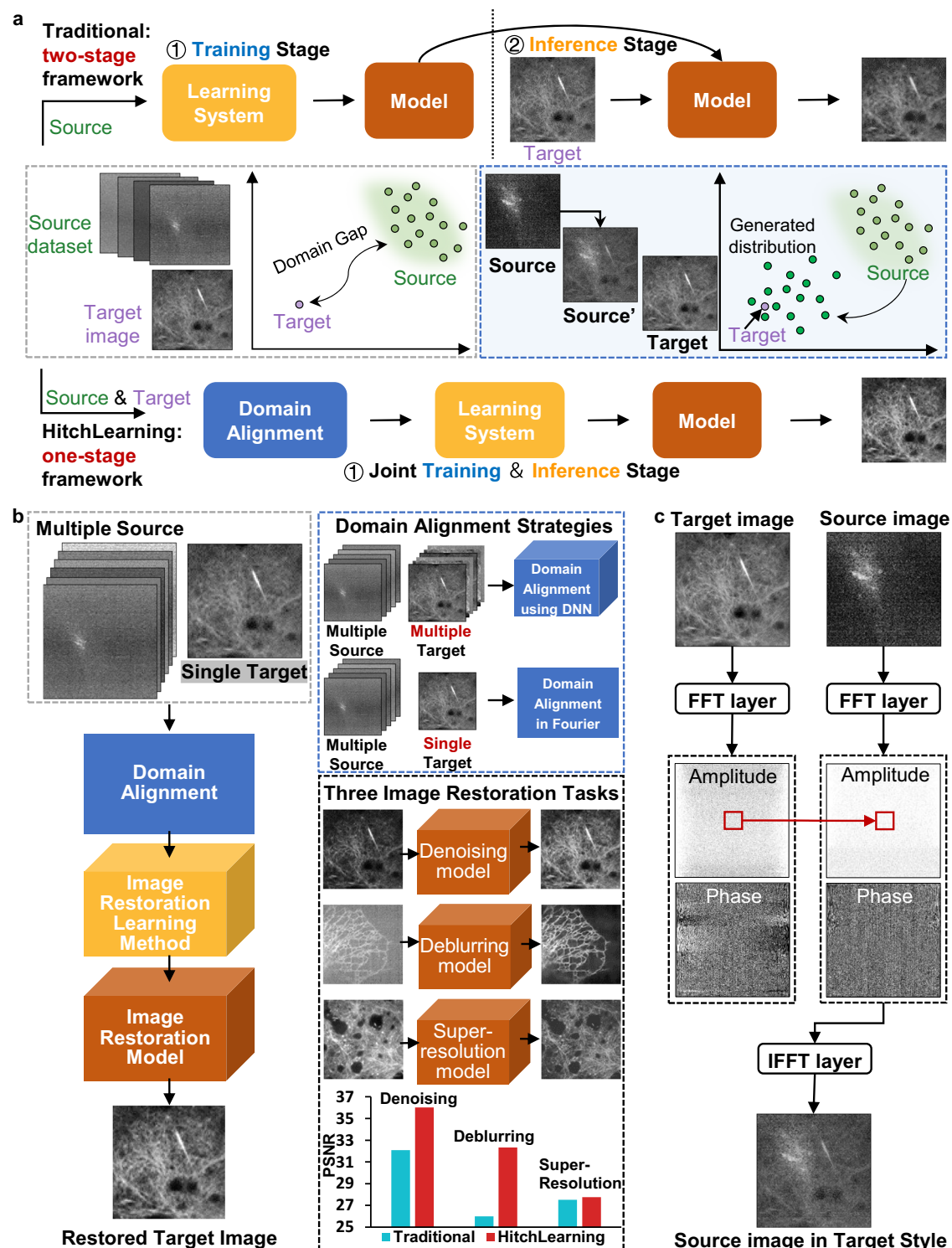


Fig. 1 | HitchLearning paradigm mechanism. **a** Comparison of the traditional and HitchLearning paradigms. The traditional machine learning paradigm has two stages: the first stage (upper left) uses a large amount of training data to create a well-performing model through various learning methods, and the second stage (upper right) uses this trained model to infer a restored output. Our new paradigm, shown at the bottom, unifies these stages by incorporating inference data during training. The two-stage framework faces a non-i.i.d. problem due to the domain gap between source and target data (middle left). In HitchLearning, this problem is alleviated by using inference data to generate new source data during model training. **b** The architecture of the HitchLearning paradigm. Domain alignment (blue) can utilize any method to align the distribution of source and target images. The image restoration learning method can be any approach (supervised, self-supervised, or

unsupervised) suitable for tasks such as image denoising, deblurring, and super-resolution. The image restoration model is a well-trained model with strong performance on a single target image. The blue dotted box illustrates our method's domain alignment in the Fourier domain, which requires only one target image, increasing costs. The black dotted box displays results of three tasks using our paradigm, with the bar chart at the bottom showing the PSNR comparison of two main training paradigms against HitchLearning across these tasks. **c** A simple but effective distribution alignment method. Here, we use Fourier Domain Adaption (FDA) method to obtain domain alignment, by doing this, the source image has a similar style to the target and can be verified by the data shown in the Results.

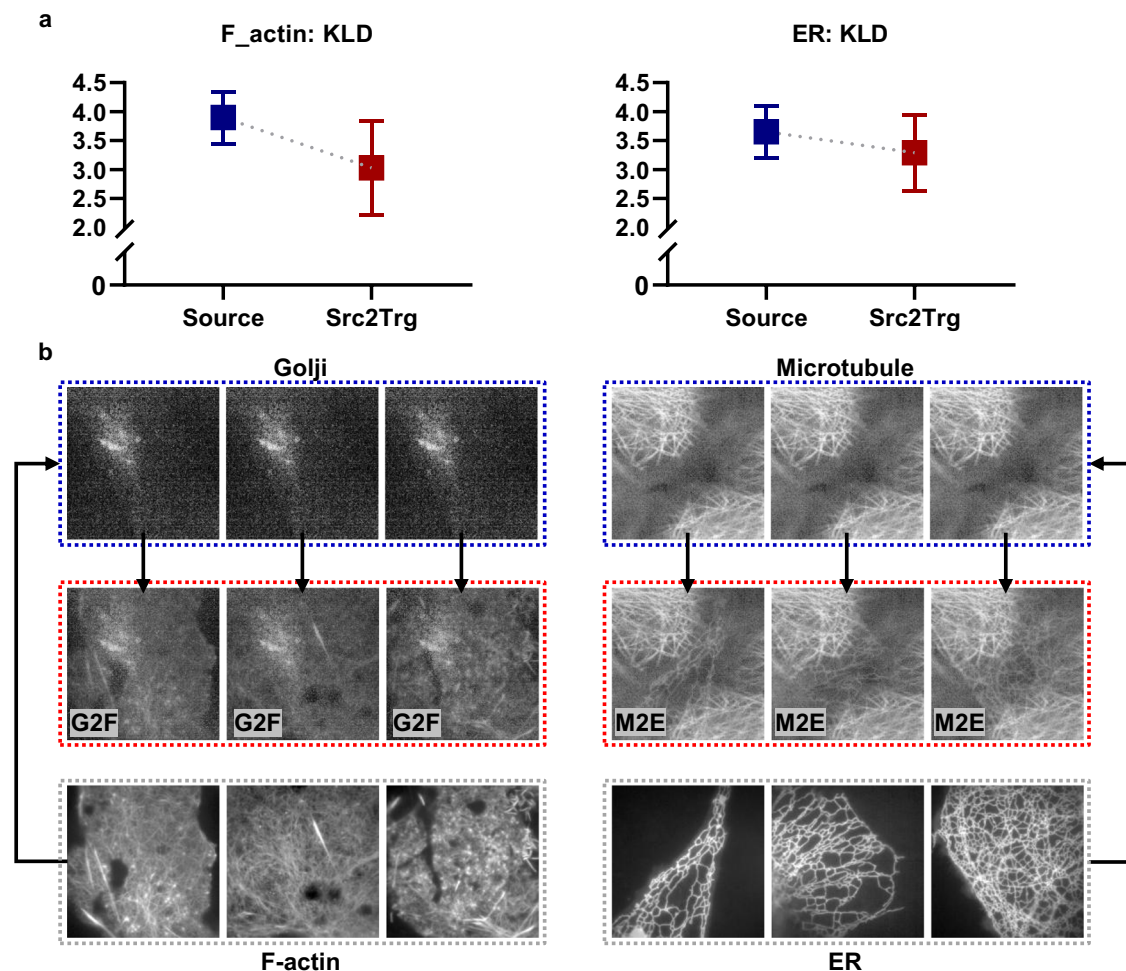


Fig. 2 | Evaluation of FDA mechanism. **a** Quantitative Analysis of Kullback–Leibler Divergence (KLD) Reduction. A side-by-side comparison of the KLD values before and after applying our proposed transformation method. On the left panel, the analysis focused on F-actin as the target, while the right panel addresses the endoplasmic reticulum (ER) as the target. The data clearly illustrate a significant decrease in KLD following the transformation process. This reduction is indicative of the enhanced alignment between the source and target distributions, thereby affirming the efficacy of our method in mitigating discrepancies between disparate datasets. The graphical representation provides a compelling visual proof of concept for the effectiveness of our approach in harmonizing data distributions in distinct biological contexts. **b** Comparative Visualization of Domain Adaptation Algorithm

Effectiveness. This figure demonstrates the outcomes (indicated by the red box) of employing Frequency Domain Adaptation (FDA) for domain adaptation on a single source image (highlighted in the blue box) across multiple target images (depicted in the gray box). Following the transformation, the source image shows significant visual alignment with each target image, affirming the adaptability and efficacy of the FDA algorithm for accessing various target images. The figure's details emphasize subtle shifts in visual characteristics, highlighting the transformative capacity and flexibility of the approach. Specifically, the left panel depicts the transformation of the Golgi apparatus to F-actin, while the right panel illustrates the transformation of Microtubule to the endoplasmic reticulum (ER) attributes.

datasets (Golgi and Microtubule) and target datasets (F-actin and ER) to generate blurry datasets.

As previously discussed, we assessed the performance of seven training settings for the deblurring task on both the F-actin and ER datasets. In this comparison, HitchLearning outperformed all the other settings across the four metrics for both the F-actin and ER datasets, with the exception of a slight decrease in the PSNR metric for the ER dataset under the Finetune setting, as shown Supplementary Tables 4, 5. Notably, HitchLearning exhibited significant improvements over Src and AllTrg in the F-actin dataset, with average PSNR enhancements of 6.34 dB and 1.02 dB, respectively. Similarly, on the ER dataset, HitchLearning improved the PSNR by 1.82 dB and 0.06 dB compared to Src and AllTrg, respectively. Although Finetune slightly outperformed HitchLearning in terms of the PSNR by 0.13 dB, HitchLearning achieved a notably greater SSIM, indicating a more substantial improvement. The visual results of the F-actin and ER datasets, which vary in structural complexity, are displayed in Fig. 4a and Supplementary Fig. 2a, 5. Despite a minor reduction in deblurring effectiveness with increasing complexity, HitchLearning consistently delivered the best

visual outcomes among all the settings. These results, encompassing both quantitative metrics and visual analysis, demonstrate that HitchLearning can surpass models trained with extensive i.i.d data, even with the use of just a single inference image.

To test whether HitchLearning outperforms the traditional Src in the deblurring task, we compared the restored images obtain from these two paradigms. The visual results shown in Fig. 4b demonstrated that HitchLearning achieved more precise cellular structures than Src. Moreover, the four metrics also reveal that our HitchLearning paradigm has a better performance.

Hitchlearning empowers the single-image super-resolution of live-cell images

To address the single-image super-resolution (SISR) challenge^{30,37,46,48–52}, we employed SwinIR⁵³ for SISR tasks on the F-actin and ER datasets, using Golgi and Microtubule as the respective sources. SwinIR, noted for its shallow and deep feature extraction capabilities coupled with high-quality image super-resolution, leverages the Swin Transformer for improved performance in

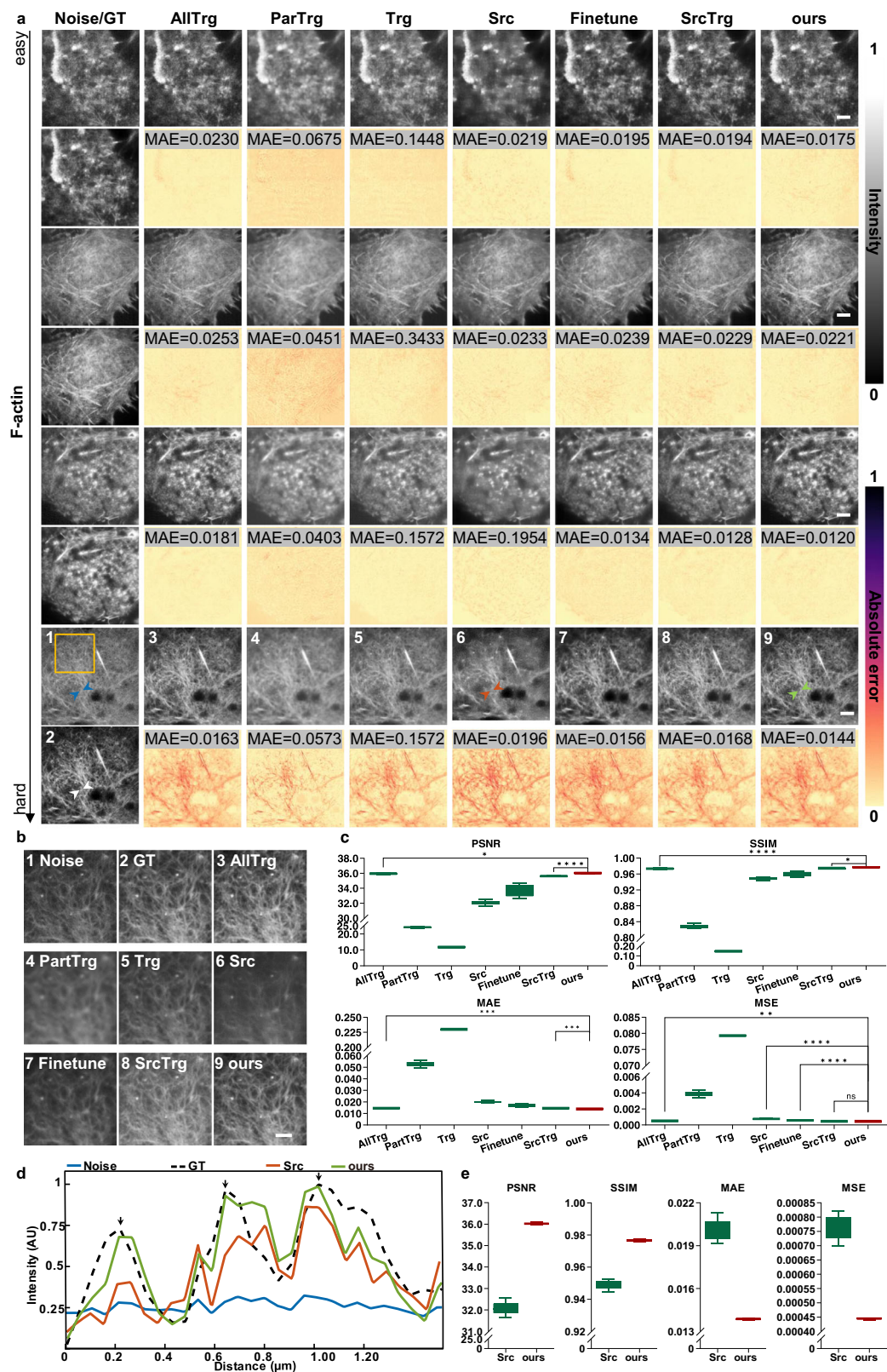


image restoration tasks. In our experiments, we tailored the input module of SwinIR to facilitate source-to-target domain transfer, enhancing the model's efficacy in single-image super-resolution tasks. Since SwinIR is a supervised model, we first generated low-resolution (LR) images of Golgi and Microtubule by downsampling the raw Golgi and Microtubule images and subsequently used Golgi and Microtubule as source to train the model to

generate pseudo super-resolution (SR) images of F-actin and ER, respectively, finally, these pseudo-SR images were used to train the model under the remaining settings.

Metric comparisons reveal that HitchLearning achieves optimal performance across all the metrics in both the F-actin and ER experiments, as detailed in Supplementary Tables 6, 7. In the F-actin experiment,

Fig. 3 | Systematic denoising analysis of F-actin. **a** Representative denoised (DN) images inferred from noisy images of F-actin. Following each noisy image, seven DN renditions produced under different settings are presented. Directly beneath each DN image is its corresponding difference image, juxtaposed against the ground truth (GT). The Mean Absolute Error (MAE) is quantified and annotated at the top left corner of each difference image. **b** Higher-magnification views of F-actin (marked in yellow rectangular regions in (a)). This sequence of images demonstrates the comparative effectiveness of the denoising process, with insets providing a close-up view of the enhanced structural details after denoising. **c** Statistical comparison of the seven settings in terms of the PSNR, SSIM, MSE and MAE for F-actin. This analysis comprehensively contrasted the performance of seven distinct settings, and Tukey

box-and-whisker plots with significance analysis are shown (Methods). **d** Intensity profiles along the lines indicated by the arrowheads in the images of Noise (blue), GT (white), Src (orange) and ours (green). **e** The comparison of four metrics between the traditional (Src) and HitchLearning (ours). These side-by-side comparisons facilitate the assessment of denoising algorithm performance in preserving structural integrity while reducing noise. This set of images underscores the advanced capability of our paradigm to maintain critical biological structures amidst substantial noise reduction. Through a combination of visual results and quantitative analysis across seven settings, our innovative paradigm, HitchLearning, demonstrated superior performance in single-image denoising tailored for live-cell imaging. Scale bar, 4 μm (a), 2 μm (b) for magnified images.

HitchLearning not only outperformed Src and AllTrg by 0.84 dB and 1.99 dB, respectively, but also surpassed the second-best model by an average of 0.05 dB in PSNR, signifying substantial improvements. The visual results, depicted in Fig. 5a–c and Supplementary Fig. 6, show that HitchLearning produces superior SR images for both the F-actin and ER datasets compared to Src, however, performance slightly diminishes as complexity increases. The SRs obtained via HitchLearning demonstrated a marked enhancement in image details (Fig. 5d). In the Trg setting, the performance is notably poor, which suggests that relying on a single inference image is insufficient for effective model training. However, in our HitchLearning paradigm, this problem can be mitigated by utilizing another dataset. These findings underscore the efficacy of our method in addressing the non-i.i.d nature of training and inference datasets.

As exemplified in Fig. 5e, the line-scan profiles from Src are closer to those of GT. However, as the complexity of the structure increases, the profiles obtained by HitchLearning and Src decrease, while HitchLearning consistently outperforms Src. Representative colored frames (Fig. 5f and Supplementary Fig. 3) show that HitchLearning resolves the crisscrossing regions of F-actin and ER cytoskeleton well, and it goes beyond the traditional Src paradigm.

Discussion

The field of image restoration has undergone a significant transformation with the advent of deep learning. The introduction of convolutional neural networks⁵⁴ (CNNs) and advanced architectures has led to substantial advancements in denoising, super-resolution, and deblurring. Deep neural networks, especially CNNs, have excelled in capturing complex image patterns and learning restoration processes directly from training data. The ability of these methods to automatically learn complex mappings has notably enhanced image restoration, often surpassing traditional methods. However, these advancements have come with challenges. Supervised methods^{41,55} demand large quantities of low-quality and high-quality image pairs, while self-supervised^{31–33,56,57} and unsupervised methods^{34,35,58} aim to reduce this requirement. However, a common limitation across these methods is the separation of training and inference phases, leading to the non-i.i.d problem when the inference dataset distribution differs from the training dataset, adversely affecting model performance. Moreover, conventional methods typically require numerous training images from the same distribution, a requirement not always feasible, especially in biology. To address these issues, we introduce a novel deep learning paradigm that combines the target image with any available data for training, demonstrating effective performance, as shown in our results.

At its core, HitchLearning is a transfer learning paradigm, yet traditional frameworks demonstrate inadequacy for our specific challenges. Current transfer learning methods predominantly rely on deep learning models employing complex operations, such as convolution^{59,60} and generative adversarial networks^{61,62}. Recent studies have shown that such models require extensive image data for training, a demand often impractical in real-world applications. Our paradigm offers a solution to this limitation. By utilizing any existing image dataset as source data and transforming it to match the distribution of a single target image, we can train the model using both the transformed source images and the target

image, effectively restoring the image with a single inference. This approach ensures that only one image is sufficient for successful outcomes. Our results demonstrate that HitchLearning is effective in aligning the distributions of different datasets and in accomplishing transfer learning with a single inference image.

In this study, we evaluated the efficacy of our method across three tasks—denoising, deblurring, and super-resolution—and within both supervised and unsupervised frameworks on four datasets. Our findings demonstrate that our approach consistently yields the best overall results. Specifically, in denoising and super-resolution tasks, HitchLearning achieves superior performance both quantitatively and visually. In deblurring, while Finetune for ER marginally outperforms HitchLearning, the difference is minimal, and visually, HitchLearning yields excellent results. These outcomes illustrate the broad applicability of our paradigm across various tasks and datasets, regardless of whether supervised or unsupervised models are employed. We further simulated variations in feature extraction capacity by altering the model's depth and architectural design. Experimental results (Supplementary Fig. 7 and Tables 8–10) demonstrate that HitchLearning consistently performs well across models with different feature extraction capabilities, further indicating its robustness and broad applicability.

We further provide a theoretical analysis of HitchLearning by examining its generalization behavior through the lens of the target-domain error bound proposed by Ben-David et al.⁶³. Let h denote the true restoration function, and f the hypothesis learned by HitchLearning. The expected risks of f on the source domain \mathcal{D}^s and the target domain \mathcal{D}^t are defined respectively as:

$$\epsilon_s(f) = \mathbb{E}_{x \sim p_s} [|h(x) - f(x)|]$$

$$\epsilon_t(f) = \mathbb{E}_{x \sim p_t} [|h(x) - f(x)|]$$

where x represents the raw input data. According to Theorem 1 in ref. 63, if the hypothesis space \mathcal{H} containing f is of VC-dimension \bar{d} , then with probability at least $1 - \delta$ for every $f \in \mathcal{H}$, the expected target domain error $\epsilon_t(f)$ in \mathcal{D}^t is upper-bounded by:

$$\epsilon_t(f) \leq \hat{\epsilon}_s(f) + \sqrt{\frac{4}{n} \left(\bar{d} \log \frac{2en}{\bar{d}} + \log \frac{4}{\delta} \right)} + d_{\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t) + \lambda$$

where $\hat{\epsilon}_s(f)$ is the empirical risk on the source domain, $d_{\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t)$ denotes the divergence between source and target distributions with respect to \mathcal{H} and λ is the minimum joint error achievable by any hypothesis in \mathcal{H} across both domains. This decomposition highlights three critical factors that influence generalization to unseen domains: (1) the empirical risk on the source domain, (2) the distributional divergence between domains, and (3) the inherent difficulty of the task, captured by the joint error term. HitchLearning is specifically designed to address the second component, domain discrepancy, by aligning the source and target distributions at the input level prior to training. This alignment is accomplished through Frequency Domain Adaptation

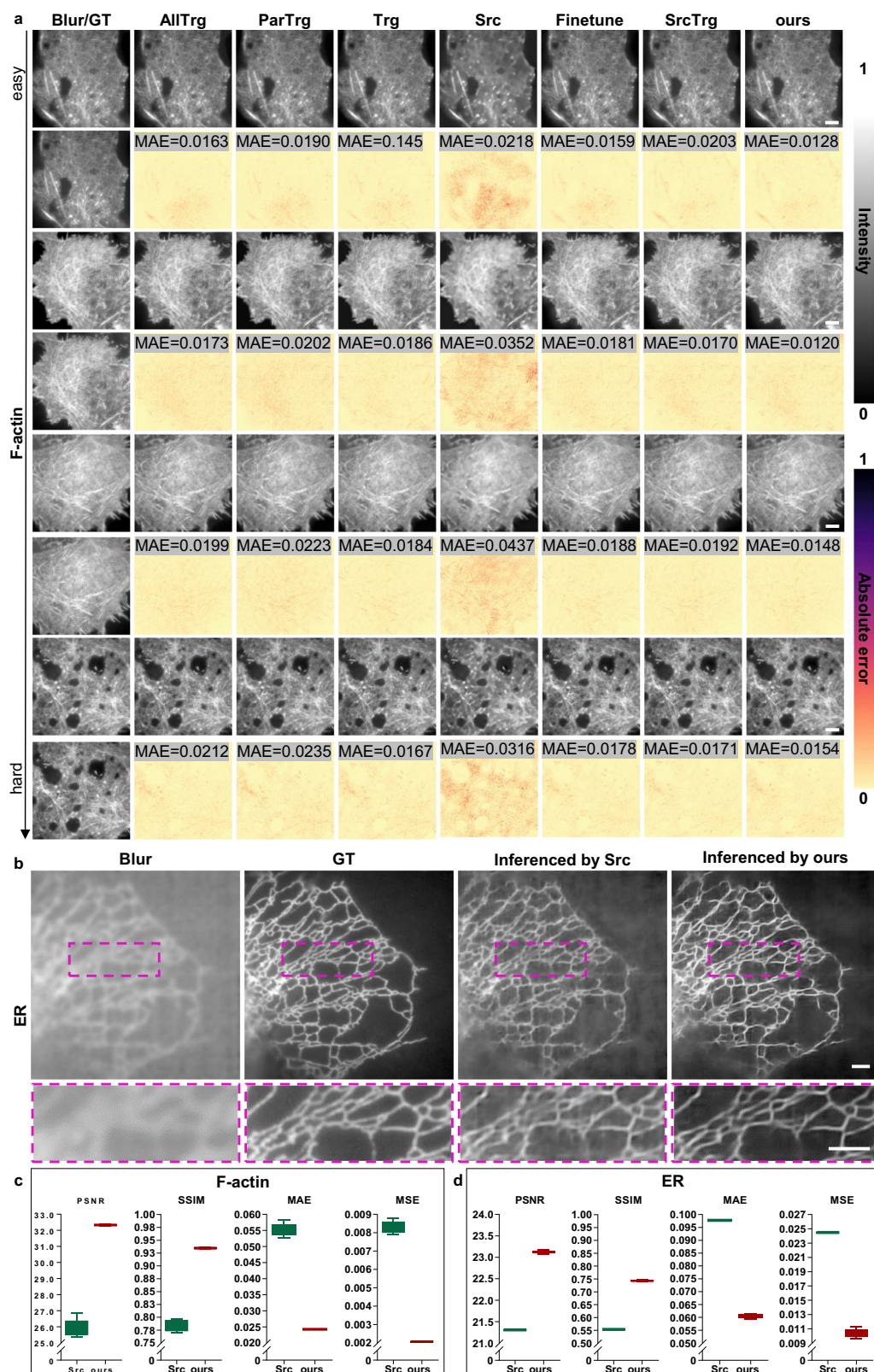


Fig. 4 | Systematic deblurring analysis of F-actin and ER. a Representative deblurred (DB) images derived from blurry F-actin. Subsequent to each blurry image, seven distinct Deblurred (DB) renditions, realized through varied settings, are exhibited. Beneath each DB image, the corresponding difference image is presented, strategically aligned against the ground truth (GT) for direct comparison. The Mean Absolute Error (MAE) for each difference image was meticulously quantified and annotated at the upper left corner, facilitating a precise evaluation of the deblurring efficacy. Scale bar, 4 μ m. **b** Comparison of DB images of ERs inferred

by Src and ours. Blur and GT images are shown for reference. Lower row shows the magnified images of the boxed regions in the upper images. Scale bar 1 μ m.

c, d Statistical comparison of the traditional and HitchLearning in terms of PSNR, SSIM, MSE and MAE for F-actin (**c**) and ER (**d**). Combining visual results with a comprehensive quantitative analysis across seven distinct settings, our novel paradigm, HitchLearning, achieves unparalleled performance in single-image deblurring specifically for bioimaging applications.

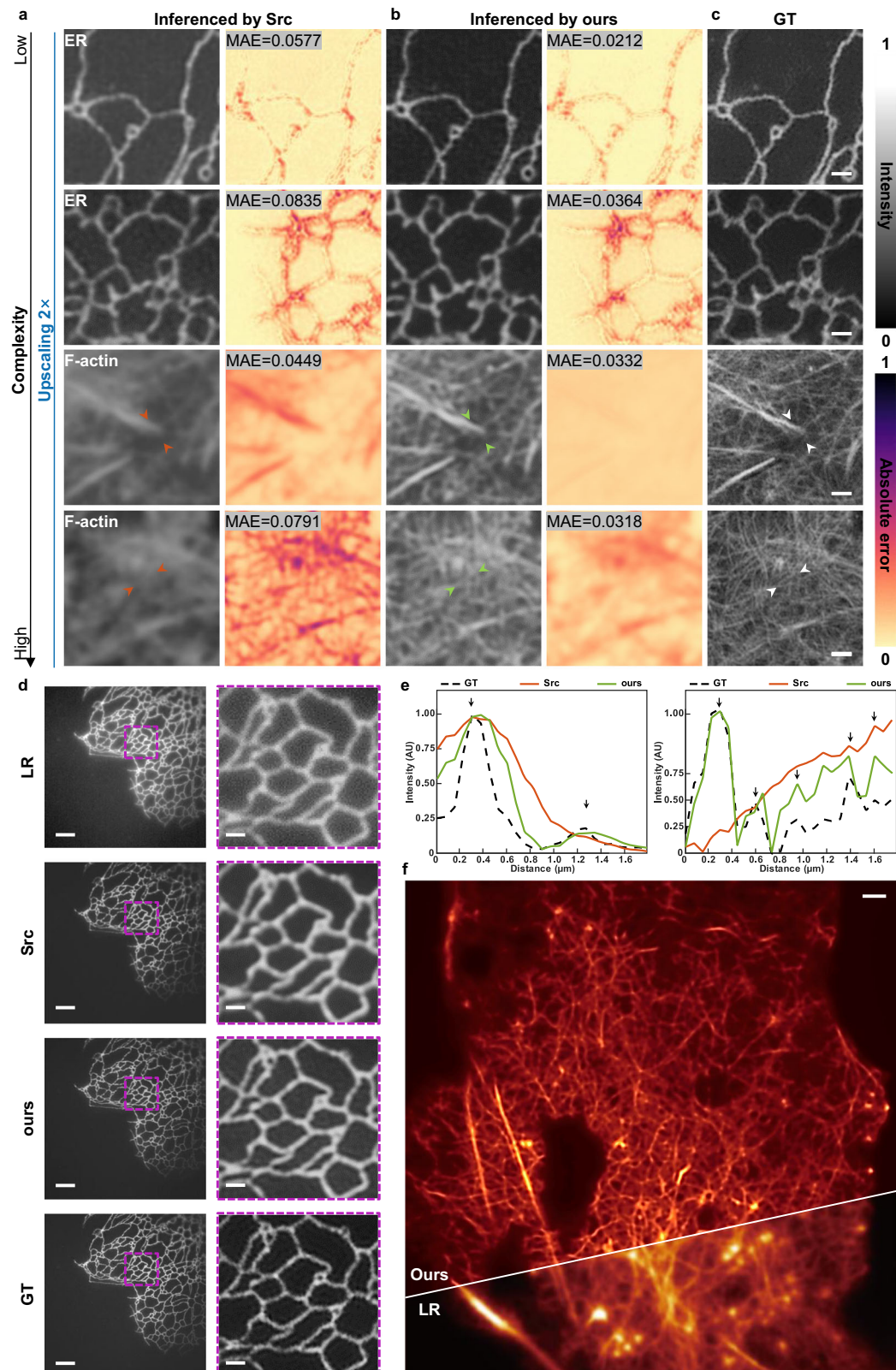


Fig. 5 | Systematic super-resolution analysis for F-actin and ER. **a, b** Representative SR images reconstructed by Src (**a**) and ours (**b**) from raw images from ER and F-actin. **c** The corresponding GT images. Scale bar, 1 μm for ER and F-actin. **d** Comparison of SR images of ERs inferred by Src and ours. LR and GT images are shown for reference. right row shows the magnified images of the boxed regions in the left images. Scale bar 5 μm; 1 μm for magnified images. **e** Intensity profiles of F-actin (2× upscaling) along the lines indicated by the two arrowheads in the images of (**a**) Src (orange), (**b**) ours (green) and (**c**) GT (white). The arrows indicate that the structures

reconstructed by ours are closer to GT than those reconstructed by Src. **f** Representative SR images of F-actin generated by HitchLearning. Bottom: a fraction of the corresponding LR image. Scale bar, 2 μm. Through a combination of visual assessments and quantitative analysis across seven distinct settings, our innovative paradigm, HitchLearning, was demonstrated to achieve the most effective performance in Single-Image Super-Resolution (SISR) specifically for live-cell imaging. For comparison purposes, the low-resolution (LR) images were resized to match the dimensions of the super-resolution (SR) images.

(FDA), which reduces degradation-specific discrepancies such as noise patterns, blur kernels, and spectral distortions.

The core idea behind HitchLearning is to approximate the i.i.d. assumption fundamental to most learning-theoretic guarantees. By narrowing the distributional gap between source and target data, HitchLearning facilitates more stable training and improved generalization to unseen domains. In the ideal case, when source and target domains are fully aligned, the distributional distance, measured using metrics such as the Kullback–Leibler Divergence (KLD), will be very small, thus minimizing the upper bound of the target-domain error. It is also worth noting that the performance of the underlying base model f plays a significant role in the final outcomes. Differences in model architecture or learning capacity can affect both the empirical risk and the alignment quality, which may in part explain the variability in performance observed across different models.

Although our paradigm has performed well in low-vision tasks such as image denoising, image deblurring and image super-resolution, there are still several limitations. First, performance varies across datasets due to differences in raw data quality, with some exhibiting higher degradation and greater restoration difficulty, which inherently impacts results. Despite this, HitchLearning achieves notable PSNR improvements on F-actin and ER, indicating its robustness under challenging conditions. This suggests that although absolute performance may fluctuate with dataset difficulty, relative gains remain substantial. Moreover, experiments show that HitchLearning more effectively reduces distributional discrepancies—reflected by a greater reduction in Kullback–Leibler divergence—between F-actin and Golgi than between ER and Microtubule. The performance gain for F-actin was also more pronounced than that for ER, suggesting that uneven improvements may hinder downstream tasks. Addressing how to consistently minimize inter-dataset differences remains an important direction for future research. Second, the compatibility of our paradigm with various image restoration models varies. Essentially, the better the inherent performance of a restoration model is, the greater its compatibility with our paradigm. This is primarily due to the differing abilities of the models.

We recognize that employing KLD as a measure of distributional consistency entails certain limitations, particularly in capturing complex, high-dimensional alignment characteristics. In future work, we plan to investigate the theoretical properties of the feature alignment process in HitchLearning, particularly how the implicit constraints introduced during alignment affect feature across different model architectures. Additionally, we aim to explore more principled metrics for distribution compatibility beyond KLD, and consider model-agnostic alignment strategies that can better adapt to architectures with differing feature representations.

Methods

HitchLearning

The central dilemma of HitchLearning is how to utilize a single test image to construct a high-performance model with thousands of parameters. The fundamental approach is to use a single test image to finetune a well pre-trained model, despite the several advantages of fine-tuning, these drawbacks, such as loss of original knowledge and training time and resource requirements, cannot be ignored in fine-tuning. Moreover, when the distribution of the inference image is quite far from the training dataset, the fine-tuning method may lose effect. Here, we propose a simple new paradigm that involves utilizing old image resources to provide additional information for new test images to address these issues. The focus of HitchLearning is how to take advantage of the amount of information already available to pave the way for a new task to achieve a better result; since the gap is the non-i.i.d. problem, we first came up with the idea of trying to align the distributions of two datasets. There are many domain adaptation methods available; however, by analyzing the restoration problem carefully, we find that the core objective is to align the distribution of the noise and blur, which can be easier to separate in the Fourier domain. Then, we harness the Fourier Domain Adaptation method in our work to make the distributions of the two datasets consistent. The main idea of the FDA is to Fourier transform the image and replace the amplitude of the high-

frequency region of the source domain image with that of the target domain image with the phase invariant.

Neighbor2Neighbor model for denoising

The Neighbor2Neighbor methodology is a self-supervised approach designed to train Convolutional Neural Network (CNN) denoisers using a single observation of noisy imagery. The proposed training strategy encompasses two components. The initial phase involves generating pairs of noisy images through the utilization of random neighbor subsampling techniques. In the subsequent phase, these sub-sampled image pairs are employed for self-supervised training, complemented by a regularized loss introduced to address the nonzero ground-truth gap existing between the paired, subsampled, noisy images. This regularized loss algorithm integrates a reconstruction segment with a regularization segment as follows:

$$\mathcal{L}_{\text{Neighbor2Neighbor}} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{reg}}$$

where

$$\mathcal{L}_{\text{rec}} = \|f_{\theta}(g_1(y)) - g_2(y)\|$$

$$\mathcal{L}_{\text{reg}} = \gamma \cdot \|f_{\theta}(g_1(y)) - g_2(y) - (g_1(f_{\theta}(y)) - g_2(f_{\theta}(y)))\|$$

Here, $f_{\theta}(\cdot)$ is the denoising network with an arbitrary network design (here we use UNet), y is the inference image, $g_1(\cdot)$ and $g_2(\cdot)$ represent the neighbor subsampler, and γ is a hyperparameter controlling the strength of the regularization term. To ensure learning stability, we halt the gradients of $g_1(f_{\theta}(y))$ and $g_2(f_{\theta}(y))$, and incrementally increase the parameter γ to its predetermined value throughout the training process.

The MPRNet model for deblurring

The proposed image restoration framework is a tri-stage process designed for progressive restoration. The first two stages rely on encoder-decoder subnetworks to learn broad contextual information. The final stage, in recognition of the position-sensitive nature of image restoration, employs a subnetwork that operates on the original input image resolution, ensuring preservation of fine textures. This is not a simple cascade; supervised attention modules are integrated between each stage to rescale feature maps, and a cross-stage feature fusion mechanism is used to consolidate intermediate features. In each stage, MPRNet has access to the input image, which is divided into nonoverlapping patches for different stages. In each stage, a residual image is predicted, which when added to the degraded input, yields the final restored image. MPRNet is optimized end-to-end with the following loss function:

$$\mathcal{L}_{\text{MPRNet}} = \sum_{S=1}^3 [\mathcal{L}_{\text{char}}(X_S, Y) + \lambda \mathcal{L}_{\text{edge}}(X_S, Y)]$$

S represents the stage, X_S is the restored image that is obtained by adding the residual image to the input in each stage, Y is the ground-truth, and λ is set to 0.5 as a parameter controls that the relative importance of Chabonier loss $\mathcal{L}_{\text{char}}(\cdot)$ and edge loss $\mathcal{L}_{\text{edge}}(\cdot)$ defined as follows:

$$\mathcal{L}_{\text{char}} = \sqrt{\|X_S - Y\|^2 + \varepsilon^2}$$

$$\mathcal{L}_{\text{edge}} = \sqrt{\|\Delta(X_S) - \Delta(Y)\|^2 + \varepsilon^2}$$

Δ denotes the Laplacian operator, and ε is a constant empirically set to 10^{-3} .

SwinIR model for super-resolution

SwinIR consists of three modules: shallow feature extraction, deep feature extraction and high-quality (HQ) image reconstruction modules. SwinIR employ the same feature extraction modules for all restoration tasks but uses different reconstruction modules for different tasks. Given a low-quality input, SwinIR use a convolution layer to extract shallow feature F_0 . Then, a deep feature extraction module consisting of residual Swin Transformer blocks and a convolution layer is used to extract the deep feature F_{DF} . Specifically, intermediate features F_1, F_2, \dots, F_K and the output deep feature F_{DF} are extracted block by block. Then, the high-quality image I_{RHQ} is reconstructed by aggregating shallow and deep features to a reconstruction module with the loss as follows:

$$\mathcal{L}_{SwinIR} = ||X - Y||_1$$

The loss is simply a \mathcal{L}_1 pixel loss, and X, Y represent the restored image and ground-truth respectively.

The shallow feature mainly contains low frequencies, while the deep feature focuses on recovering lost high frequencies. With a long skip connection, SwinIR can transmit low-frequency information directly to the reconstruction module, which can help the deep feature extraction module focus on high-frequency information and stabilize training.

Image preprocessing and training

Here, we used F-actin and ER from the figshare repository as the target dataset, Golgi and Microtubule from the Zenodo repository are used as the sources for F-actin and ER, respectively. Although we have the ground truths of all four datasets, we only use the ground truths of source datasets in super-resolution task for training, in other conditions, we only use the ground truths to compute the evaluation metrics. There are 51 cells in the F-actin datasets with 12 shots per cell, and we randomly choose 10 cells of the data volume (approximately 20%) as the test dataset, which means that 120 images are used as the test dataset and 492 images are used as training dataset. For ER, we used the same strategy. ER contains 68 cells with 6 shots per cell, we choose 13 cells (approximately 20%) as test set and 55 cells as train set, that is, 78 images were used as the test set and 330 images were used as the training set. For the source, 388 images of Golgi and 618 images of Microtubule were used.

In the denoising, deblurring and super-resolution tasks, Golgi was set as the source of F-actin, and Microtubule were set as the source of ER. In these tasks, we remove only the background of images by merely subtracting the 10% minimum value of the whole image, setting the negative values to 0 and then normalizing the image to the range 0 to 1. Specifically, in the deblurring task, we do not have the real blurred images of these datasets; then, we generate blurred images by adding motion blur to each image at different angles and values.

In our training protocol, as detailed in Supplementary Table 1, we employ seven distinct configurations. AllTrg, PartTrg, and Trg involve training the model with datasets matching the distribution of the inference dataset called the target. These settings differ in terms of the volume of imagery utilized: AllTrg employs the entire training set, PartTrg uses a subset of 10 images, and Trg leverages the inference image itself. In contrast, Src, the conventional approach, trains the model with datasets that typically do not match the inference dataset's distribution, called the source. Finetune refines the model trained under the Src setting using the inference image. Finally, SrcTrg and HitchLearning integrate both the source training data and the inference image in the training process.

Assessment metrics calculation

In this work, we utilized four key metrics for image quality evaluation: the mean absolute error (MAE), mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM). These metrics jointly allowed for robust and comprehensive assessment of the quality of image reconstruction methodologies.

The MAE is a widely used statistical metric for evaluating the accuracy of predictions in regression analysis. The average of the absolute differences between the predicted values and observed (actual) values was calculated. The MAE provides a straightforward interpretation of the average magnitude of errors in predictions without considering their direction and is calculated as

$$MAE = \frac{1}{mn} \sum_{i,j=1,1}^{m,n} |I(i,j) - K(i,j)|$$

where I and K are the original and compared images, respectively, m and n are the dimensions of the images; and i and j are the pixel positions.

The MSE quantifies the average of the squares of the error between the original and compared images and is a common metric used in statistical analysis, particularly in the context of regression analysis and forecasting. The method for evaluating the performance of an estimator involves calculating the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value, calculated as

$$MSE = \left(\frac{1}{mn} \right) \sum_{i,j=1,1}^{m,n} [I(i,j) - K(i,j)]^2$$

where I and K are the original and compared images, respectively; m and n are the dimensions of the images; and i and j are the pixel positions.

The PSNR is commonly employed in image processing to measure the quality of reconstructed images, particularly for assessing the quality of reconstructed or compressed images in comparison to their original, uncompressed versions. The PSNR is based on the mean squared error (MSE) between the original image and a compressed or reconstructed image and provides a measure of reconstruction quality. It is defined as

$$PSNR = 20 \times \log_{10}(MAX_I) - 10 \times \log_{10}(MSE)$$

where MAX_I is the maximum possible pixel value of the image.

The SSIM quantifies image quality by comparing image luminance, contrast, and structure; this metric has a decimal value between -1 and 1, and a value of 1 is only reachable when two sets of data are identical. The SSIM was developed to provide a more accurate and consistent assessment of the perceived quality of an image. The SSIM is based on the understanding that the human visual system detects structural information in a scene; thus, a measure of structural information can provide an indication of perceived image quality. A higher SSIM index indicates greater structural similarity between two images. It is calculated using the following formula:

$$SSIM = \frac{(2\mu_x\mu_y + c_1) \times (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) \times (\sigma_x^2 + \sigma_y^2 + c_2)}$$

where x and y are the two images being compared; μ_x and μ_y are the averages of x and y , respectively; σ_x^2 and σ_y^2 are the variances of x and y , respectively; and σ_{xy} is the covariance of x and y .

Data availability

Training and test datasets for denoising, deblurring and super-resolution are publicly available at figshare repository (<https://doi.org/10.6084/m9.figshare.13264793>) and the Zenodo repository (<https://zenodo.org/records/4624364#.YF3jsa9Kibg>) for target data and source data respectively.

Code availability

The pytorch codes of HitchLearning, three tasks model, as well as some example images for are publicly available at <https://github.com/LiuJiaxin-1/HitchLearning>.

Received: 28 October 2024; Accepted: 5 April 2025;
Published online: 14 May 2025

References

1. Kherlopian, A. R. et al. A review of imaging techniques for systems biology. *BMC Syst. Biol.* **2**, 74 (2008).
2. Elliott, R. Executive functions and their disorders. *Br. Med. Bull.* **65**, 49–59 (2003).
3. Li, B., Wu, C., Wang, M., Charan, K. & Xu, C. An adaptive excitation source for high-speed multiphoton microscopy. *Nat. Methods* **17**, 163–166 (2020).
4. Gustafsson, M. G. L. et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957–4970 (2008).
5. Fadero, T. C. et al. LITE microscopy: tilted light-sheet excitation of model organisms offers high resolution and low photobleaching. *J. Cell Biol.* **217**, 1869–1882 (2018).
6. Olarte, O. E., Andilla, J., Gualda, E. J. & Loza-Alvarez, P. Light-sheet microscopy: a tutorial. *Adv. Opt. Photonics* **10**, 111 (2018).
7. Gustafsson, M. G. Extended resolution fluorescence microscopy. *Curr. Opin. Struct. Biol.* **9**, 627–628 (1999).
8. Wilson, T. Optical sectioning in confocal fluorescent microscopes. *J. Microsc.* **154**, 143–156 (1989).
9. Wilson, T. in *Handbook of Biological Confocal Microscopy* (ed. Pawley, J. B.) 167–182 (Springer US, 1995). https://doi.org/10.1007/978-1-4757-5348-6_11.
10. Pacheco, S. et al. High resolution, high speed, long working distance, large field of view confocal fluorescence microscope. *Sci. Rep.* **7**, 13349 (2017).
11. Horstmeyer, R., Ruan, H. & Yang, C. Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue. *Nat. Photonics* **9**, 563–571 (2015).
12. König, K. Multiphoton microscopy in life sciences. *J. Microsc.* **200**, 83–104 (2000).
13. Neil, M. A., Juškaitis, R. & Wilson, T. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Opt. Lett.* **22**, 1905–1907 (1997).
14. Ströhl, F. & Kaminski, C. F. Frontiers in structured illumination microscopy[J]. *Optica*. **3**, 667–677 (2016).
15. Stelzer, E. H. K. Light sheet fluorescence microscopy. *Nat. Rev. Methods Primer* **1**, 74 (2021).
16. Schueth, A. et al. Efficient 3D light-sheet imaging of very large-scale optically cleared human brain and prostate tissue samples. *Commun. Biol.* **6**, 1–15 (2023).
17. Hahn, V. et al. Light-sheet 3D microprinting via two-colour two-step absorption. *Nat. Photonics* **16**, 784–791 (2022).
18. Dvinskikh, L. et al. Remote-refocusing light-sheet fluorescence microscopy enables 3D imaging of electromechanical coupling of hiPSC-derived and adult cardiomyocytes in co-culture. *Sci. Rep.* **13**, 3342 (2023).
19. Dyba, M., Jakobs, S. & Hell, S. W. Immunofluorescence stimulated emission depletion microscopy. *Nat. Biotechnol.* **21**, 1303–1304 (2003).
20. Gonzalez Pisfil, M. et al. Stimulated emission depletion microscopy with a single depletion laser using five fluorochromes and fluorescence lifetime phasor separation. *Sci. Rep.* **12**, 14027 (2022).
21. Liang, L. et al. Continuous-wave near-infrared stimulated-emission depletion microscopy using downshifting lanthanide nanoparticles. *Nat. Nanotechnol.* **16**, 975–980 (2021).
22. Sarmiento, M. J. et al. Exploiting the tunability of stimulated emission depletion microscopy for super-resolution imaging of nuclear structures. *Nat. Commun.* **9**, 3415 (2018).
23. Wang, F. et al. Phase imaging with an untrained neural network. *Light Sci. Appl.* **9**, 77 (2020).
24. Wu, Y. et al. Multiview confocal super-resolution microscopy. *Nature* **600**, 279–284 (2021).
25. Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**, 3142–3155 (2017).
26. Zhang, K., Zuo, W. & Zhang, L. FFDNet: toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **27**, 4608–4622 (2018).
27. Sun, J. et al. Learning a convolutional neural network for non-uniform motion blur removal. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 769–777 (IEEE, 2015). <https://doi.org/10.1109/CVPR.2015.7298677>.
28. Tao, X. et al. Scale-recurrent network for deep image deblurring[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition* 27, 8174–8182 (2018).
29. Zhang, K., Zuo, W., Zhang, L. Deep plug-and-play super-resolution for arbitrary blur kernels[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1671–1681 (2019).
30. Dong, C., Loy, C. C., He, K. & Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2016).
31. Lehtinen J. et al. Noise2Noise: Learning Image Restoration without Clean Data[C]. *International Conference on Machine Learning*. PMLR, 2965–2974 (2018).
32. Batson, J. & Royer, L. Noise2self: Blind denoising by self-supervision[C]. *International conference on machine learning*. PMLR, **97**, 524–533 (2019).
33. Huang, T., Li, S., Jia, X. et al. Neighbor2neighbor: Self-supervised denoising from single noisy images[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14781–14790 (2021).
34. Lu, B., Chen, J.-C. & Chellappa, R. Unsupervised domain-specific deblurring via disentangled representations. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 10217–10226 (IEEE, 2019). <https://doi.org/10.1109/CVPR.2019.01047>.
35. Yuan, Y. et al. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 701–710 (2018).
36. Zamir, S. W. et al. Cycleisp: Real image restoration via improved data synthesis[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2696–2705 (2020).
37. de Leeuw den Bouter, M. L. et al. Deep learning-based single image super-resolution for low-field MR brain images. *Sci. Rep.* **12**, 6362 (2022).
38. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat. Methods* **18**, 678–687 (2021).
39. Sinha, A., Lee, J., Li, S. & Barbastathis, G. Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).
40. Tan, C. et al. A survey on deep transfer learning. In: *Artificial Neural Networks and Machine Learning—ICANN 2018* (eds. Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L. & Maglogiannis, I.) 270–279 (Springer International Publishing, 2018). https://doi.org/10.1007/978-3-030-01424-7_27.
41. Sharpe, M. J., Batchelor, H. M., Mueller, L. E., Gardner, M. P. H. & Schoenbaum, G. Past experience shapes the neural circuits recruited for future learning. *Nat. Neurosci.* **24**, 391–400 (2021).
42. Ben-David, S. et al. A theory of learning from different domains. *Mach. Learn.* **79**, 151–175 (2010).
43. Yang, Y. & Soatto, S. FDA: Fourier domain adaptation for semantic segmentation. 11 <https://doi.org/10.1109/CVPR42600.2020.00414>.

44. Joyce, J. M. in *International Encyclopedia of Statistical Science* (ed. Lovric, M.) 720–722 (Springer, 2011). https://doi.org/10.1007/978-3-642-04898-2_327.
45. Belov, D. I. & Armstrong, R. D. Distributions of the Kullback–Leibler divergence with applications. *Br. J. Math. Stat. Psychol.* **64**, 291–309 (2011).
46. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
47. Zamir, S. W. et al. Multi-stage progressive image restoration. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 14816–14826 (IEEE, 2021). <https://doi.org/10.1109/CVPR46437.2021.01458>.
48. Dong, C., Loy, C. C., He, K. & Tang, X. Learning a deep convolutional network for image super-resolution. In: *Computer Vision—ECCV 2014* (eds. Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 184–199 (Springer International Publishing, 2014). https://doi.org/10.1007/978-3-319-10593-2_13.
49. Yang, W. et al. Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multimed.* **21**, 3106–3121 (2019).
50. Zhang, K., Van Gool, L. & Timofte, R. Deep unfolding network for image super-resolution. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 3214–3223 (IEEE, 2020). <https://doi.org/10.1109/CVPR42600.2020.00328>.
51. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).
52. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
53. Liang, J. et al. SwinIR: image restoration using swin transformer. In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* 1833–1844 (IEEE, 2021). <https://doi.org/10.1109/ICCVW54120.2021.00210>.
54. Chua, L. O. & Roska, T. The CNN paradigm. *IEEE Trans. Circuits Syst. Fundam. Theory Appl.* **40**, 147–156 (1993).
55. Gu, S., Li, Y., Gool, L. V. & Timofte, R. Self-guided network for fast image denoising. <https://doi.org/10.1109/ICCV.2019.00260>.
56. Liu, P., Janai, J., Pollefeys, M., Sattler, T. & Geiger, A. Self-supervised linear motion deblurring. *IEEE Robot. Autom. Lett.* **5**, 2475–2482 (2020).
57. Nguyen, N. L. et al. Self-supervised super-resolution for multi-exposure push-frame satellites[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1858–1868 (2022).
58. Kim, K., Soltanayev, S. & Chun, S. Y. Unsupervised Training of Denoisers for Low-Dose CT Reconstruction Without Full-Dose Ground Truth. *IEEE J. Sel. Top. Signal Process.* **14**, 1112–1125 (2020).
59. Tammina, S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ. (IJSRP)* **9** (2019).
60. Raghu, S., Sriraam, N., Temel, Y., Rao, S. V. & Kubben, P. L. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Netw.* **124**, 202–212 (2020).
61. Majurski, M. et al. Cell image segmentation using generative adversarial networks, transfer learning, and augmentations[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, (eds Larry, D., Philip, T., Song-Chun, Z.) (2019).
62. Soleimani, E. & Nazerfard, E. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* **426**, 26–34 (2021).
63. Ben-David, S., Blitzer, J., Crammer, K. & Pereira, F. in *Advances in Neural Information Processing Systems* Vol. 19 (MIT Press, 2006).

Acknowledgements

This work was supported by National Science and Technology Major Project (2023ZD0506304), National Natural Science Foundation of China (82441013, 62125106), and the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No. 2024C01142), the Beijing Outstanding Young Scientist Program under contract No. JWZQ20240101009, Tsinghua University-Migu Xinkong Culture Technology (Xiamen) Co.Ltd. Joint Research Center for Intelligent Light Field and Interaction Technology, Tsinghua University-Joint research and development project (R24119F0).

Author contributions

Jiaxin Liu completed the experiments described in the paper, wrote the manuscript, and created the figures; Yuchen Guo conducted the experimental design, wrote and revised the manuscript; Lu Fang and Qionghai Dai supervised the experimental design and manuscript writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44387-025-00004-y>.

Correspondence and requests for materials should be addressed to Lu Fang or Qionghai Dai.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025