



RESEARCH ARTICLE **OPEN ACCESS**

Uncertainty-Guided Selective Adaptation Enables Cross-Platform Predictive Fluorescence Microscopy

Kai-Wen K. Yang¹ | Andrew Bai² | Alexandra Bermudez^{3,4} | Yunqi Hong² | Zoe Latham⁴ | Iris Sloan⁴ | Michael Liu² | Vishrut Goyal² | Cho-Jui Hsieh²  | Neil Y. C. Lin^{3,4,5,6} 

¹Independent Researcher, San Francisco, CA, USA | ²Computer Science Department, University of California, Los Angeles, CA, USA | ³Mechanical and Aerospace Engineering Department, University of California, Los Angeles, CA, USA | ⁴Bioengineering Department, University of California, Los Angeles, CA, USA | ⁵Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA | ⁶Institute for Quantitative and Computational Biosciences, University of California, Los Angeles, CA, USA

Correspondence: Cho-Jui Hsieh (chohsieh@cs.ucla.edu) | Neil Y. C. Lin (neillin@g.ucla.edu)

Received: 25 March 2026 | **Accepted:** 27 March 2026

Keywords: domain adaptation | fluorescent microscopy | image-image translation

ABSTRACT

Deep learning is transforming microscopy, yet models often fail when applied to images from new instruments or acquisition settings. Conventional adversarial discriminative domain adaptation (ADDA) retrains entire networks, often disrupting learned semantic representations. Here, we overturn this paradigm by showing that adapting only the earliest convolutional layers, while freezing deeper layers, yields reliable transfer. Building on this principle, we introduce subnetwork image translation ADDA with automatic depth selection (SIT-ADDA-Auto), a self-configuring framework that integrates shallow-layer adversarial alignment with predictive uncertainty to automatically select adaptation depth without target labels. We demonstrate robustness via multi-metric evaluation, blinded expert assessment, and uncertainty-depth ablations. Across exposure and illumination shifts, cross-instrument transfer, and multiple stains, SIT-ADDA improves prediction fidelity and downstream segmentation over full-encoder adaptation and nonadversarial baselines, with reduced drift of semantic features. Our results provide a design rule for label-free adaptation in microscopy and a recipe for field settings; the code is publicly available.

1 | Introduction

Deep learning is reshaping optical bioimaging by redefining what microscopes can observe, resolve, and interpret. Neural networks increasingly act as algorithmic optics, inferring biological structures from morphological context, blurred by diffraction, or invisible to conventional contrast [1–4]. This transformation spans a continuum: from inferring lost signal, to transcending optical limits, to synthesizing entirely new forms of contrast. Deep learning networks now infer dynamic processes from photon-starved data [5–7]; in light-sheet imaging of zebrafish embryos, content-aware models generate crisp subcellular time-lapses in which mitotic spindles can be tracked deep within tissue [8]. Super-resolution networks extend beyond diffraction, predicting actin filaments at ~120 nm while avoiding the phototoxicity of structured

illumination [9–12]. Generative models now computationally synthesize biological contrast: in cryo-electron tomograms, self-supervised denoisers reveal membrane-protein complexes otherwise lost in noise [13–15]. Together, these advances illustrate a shift: microscopy is evolving into an inference engine where data and algorithms converge to drive discovery [16, 17].

Despite these breakthroughs, a persistent barrier limits the broad adoption of AI in microscopy: trained models often fail to generalize across imaging domains [18, 19]. Domain shifts, arising from differences in sample preparation, staining protocols, hardware configurations, or acquisition settings, cause substantial changes in image distributions. Even minor shifts make models brittle [20, 21], blocking reproducibility and broad use. As such, multiple strategies have attempted to bridge these gaps. Pixel-

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Advanced Intelligent Discovery* published by Wiley-VCH GmbH.

level style transfer approaches harmonize image appearance across domains [22, 23], while data-centric methods such as stain normalization and augmentation simulate variability [24–26]. Recent innovations like ContriMix disentangle biological features from acquisition-specific artifacts to improve training robustness [27]. Despite tangible progress, these strategies often depend on assumptions about the form of domain variation or require representative labeled data from anticipated use cases [28], leaving generalization gaps unresolved [29, 30].

Here, we introduce subnetwork image translation ADDA with automatic depth selection (SIT-ADDA-Auto), a self-configuring scheme for unsupervised adaptation of trained predictors to new acquisition domains. SIT-ADDA-Auto freezes high-level blocks, performs adversarial alignment only in early layers, and uses predictive uncertainty to automatically set adaptation depth on unlabeled target streams. We emphasize that these outputs are predictive proxies designed to facilitate segmentation and alignment, rather than quantitative molecular measurements. Because our

regime is unpaired and label-free, we ground the method in ADDA (Figure 1A) [31] and compare against full-encoder adaptation [27, 32–35].

Our design rests on the principle that most microscopy domain shifts perturb low-level optical statistics while leaving the label function (biological semantics) largely stable [18]. This suggests a simple rule: adapt shallow features, freeze deep semantics. We instantiate this rule as a microscopy-optimized ADDA (Figure 1A) [31]; the uncertainty signal then chooses how far to adapt, avoiding per-site heuristics. Across cross-modality and cross-platform transfers, including low-cost and 3D-printed microscopes (Figure 1B), SIT-ADDA-Auto improves prediction and downstream performance over conventional ADDA and related methods while preserving high-level semantics under optical shifts. By reducing technical and data demands, SIT-ADDA-Auto offers a lightweight, plug-and-play path to adapt trained models to new microscopes and conditions, broadening access in both advanced facilities and resource-limited settings.

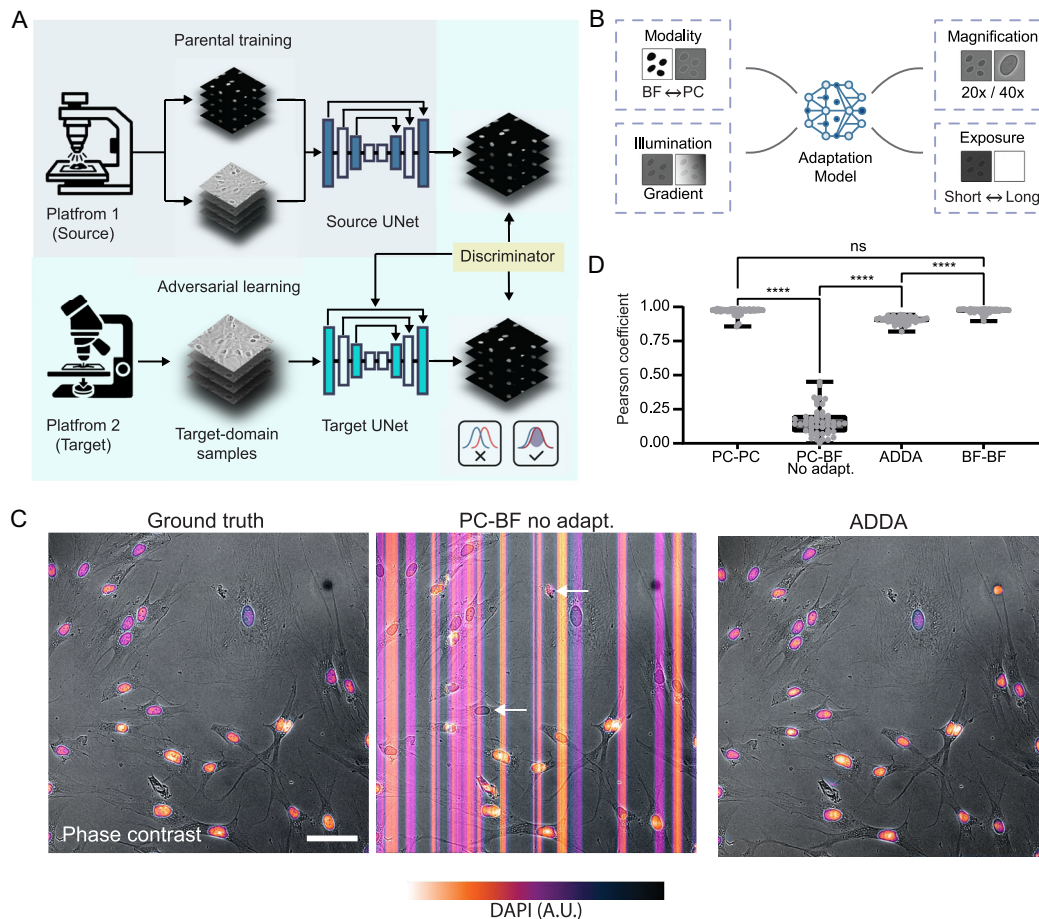


FIGURE 1 | Adversarial discriminative domain adaptation (ADDA) enables robust cross-domain image translation in microscopy.(A) Schematic of the ADDA workflow. Stage 1 (Parental training): train an encoder and predictor (e.g., U-Net) on labeled source images. Stage 2 (Adversarial learning): adapt a target encoder to match the source using unlabeled target data and a discriminator. Stage 3: perform inference on target images with the adapted encoder and fixed predictor. Our variant, SIT-ADDA, selectively fine-tunes a subset of layers to improve robustness across microscopy platforms. (B) This framework generalizes across diverse imaging domain shifts, including differences in magnification, imaging modality, illumination conditions, and exposure settings. (C) Example application to cross-modality nuclear fluorescence prediction. A U-Net trained on phase-contrast data fails to generalize to brightfield inputs (middle), introducing artifacts (white arrows) and missing nuclei. In contrast, ADDA adaptation (right) preserves nuclear morphology and internuclear variability in the target domain. (D) Quantitative evaluation using pixel-wise Pearson correlation shows that direct transfer from phase contrast to brightfield causes accuracy to collapse from ~ 0.98 to ~ 0.16 , while ADDA adaptation restores performance to 0.91, closely matching source-domain baselines. Scale bar: 100 μm . “ns” and “****” refer to $p \geq 0.05$ and < 0.0001 , respectively.

2 | Results

2.1 | Adversarial Domain Adaptation Enables Cross-Modality Image Translation in Microscopy

To evaluate the impact of domain adaptation on cross-modality translation in microscopy, we first tested whether a U-Net [36], pre-trained to predict DAPI fluorescence (nucleus) from phase-contrast images, could generalize to brightfield data (Figure 1). Without adaptation, performance degraded markedly under domain shift, producing stripe artifacts, spurious elongated structures, false-positive nuclei in background regions, and missed detections in dense fields of view (Figure 1C, middle; white arrows). Applying conventional ADDA (Figure 1A) restored reliable cross-modality translation: most nuclei were accurately delineated, predicted fluorescence intensities closely matched those of the source domain, and both nuclear morphology and internuclear variability were preserved (Figure 1C, right).

To quantify the impact of domain shift on image translation, we computed the pixel-wise Pearson correlation coefficient between predicted and ground-truth images. Applying a phase-contrast-trained model directly to brightfield inputs caused the correlation to plummet from ~ 0.98 to ~ 0.16 , reflecting substantial loss of predictive accuracy. Following adaptation with conventional ADDA, this performance deficit was largely mitigated: the Pearson coefficient recovered to 0.91 (Figure 1D). Additionally, Cellpose-based nucleus segmentation of these ADDA predictions showed agreement with ground truth in nuclear centroid positions (Pearson $r \approx 1.0$) and nuclear areas ($r \approx 0.674$ after outlier removal) (Figure S1). First, despite lower apparent contrast, brightfield images contain sufficient structure for accurate nuclear labeling. Second, aligning feature distributions across modalities restores prediction accuracy without requiring target-domain annotations.

2.2 | SIT-ADDA: Selective Layer-Freezing Enhances Domain Adaptation in Microscopy

Conventional ADDA adapts the full encoder but can struggle when domain shifts primarily affect low-level image statistics (illumination, exposure, contrast, sensor noise). Adapting the entire network risks altering higher-level semantic features learned on the source domain and can reduce prediction quality. Targeting adaptation to early layers that encode contrast and texture, therefore, may more effectively correct acquisition-specific differences while better preserving domain-invariant representations.

Motivated by this rationale, we introduce selective image-translation ADDA (SIT-ADDA), a transfer-learning-inspired variant that freezes deeper, semantically rich U-Net blocks while adapting the earlier convolutional layers. This reversal of the traditional classification-oriented freezing schedule leverages the empirical fact that deeper layers tend to capture object-level features, whereas early layers encode microscope-specific details [37]. We conducted systematic freezing schedule experiments by testing early layers, deep layers, and dropping individual layers (Figure S2). These results validated our design choice, confirming that the strategy preserves generalizable structural representations while enabling microscope-specific corrections.

To evaluate the effectiveness of our approach, we tested it on mesenchymal stem cells (MSCs) stained for CD29 (integrin $\beta 1$), where perinuclear signal enrichment reflects intracellular

trafficking or signaling activity [38, 39]. CD29 produces a spatially subtle, texture-dependent signal that does not coincide with obvious, high-contrast organelles. As a result, faithfully predicting CD29 localization requires preservation of fine-scale intensity and structural detail. In contrast, nuclei are easily visualized in transmitted-light microscopy and thus serve as a simpler test case. These features make CD29 a particularly stringent and informative target for evaluating biological validity.

To characterize SIT-ADDA, we conducted an ablation study in which different combinations of encoder and decoder layers were frozen. We evaluated predictions with complementary metrics: Pearson correlation for linear agreement with ground truth, peak signal-to-noise ratio (PSNR) for prediction fidelity, and the structural similarity index (SSIM) for perceptual quality and structural integrity. As shown in Figure 2A–C, fine-tuning only the first three convolutional layers yielded the best adaptation, improving Pearson correlation (0.84 vs. 0.75), PSNR (20.8 vs. 19.0), and SSIM (0.61 vs. 0.48) compared with ADDA. These improvements across all three metrics demonstrate that SIT-ADDA enhances the quantitative accuracy of predicted fluorescence signals.

We also observed that SIT-ADDA outperforms conventional ADDA in capturing fine-scale CD29 localization. While conventional ADDA accurately captured the overall cellular distribution of CD29, it failed to resolve the perinuclear signal enrichment evident in the ground truth, instead introducing cross-shaped artifacts (middle panel in Figure 2D). In contrast, SIT-ADDA markedly improved subcellular prediction, visualizing perinuclear halo-like structures with minimal hallucinated features. It also better preserved signal at the cell periphery, resulting in sharper boundary delineation. These improvements further reflect the importance of selectively fine-tuning early layers that govern fine-grained morphological details.

Our quantitative summary of the Pearson coefficient (Figure 2E) showed that traditional ADDA improved prediction accuracy over the no-adaptation baseline, raising the correlation from 0.73 to 0.77. SIT-ADDA further increased it to 0.85, representing an overall improvement of 17% compared with the baseline. Lastly, we assessed the effectiveness of adaptation strategies by visualizing the distribution of predicted images in a 2D embedding space (uniform manifold approximation and projection (UMAP), where overlap between the source and target domains reflects successful alignment. As shown in Figure 2F, when brightfield input images were processed by a network trained on phase-contrast data, the resulting predictions (PC-BF) exhibited a pronounced shift in distribution, indicating a degradation in image-to-image translation performance caused by the domain mismatch. While traditional ADDA partially corrected this drift, SIT-ADDA effectively aligned the predicted image distribution (SIT-ADDA) with the original target domain cluster. Quantitative analysis using Euclidean distance between cluster centroids calculated in the high-dimensional feature space confirmed that the SIT-ADDA predictions were consistently closer to the ground truth and to direct same-modality predictions (e.g., BF-BF, PC-PC) than those from ADDA or the no-adaptation baseline (Figure 2G). This closer correspondence suggests a more successful domain adaptation.

In this work, our comparative analysis primarily focuses on the conventional ADDA framework. To the best of our knowledge, ADDA remains the only method that preserves a fixed predictor

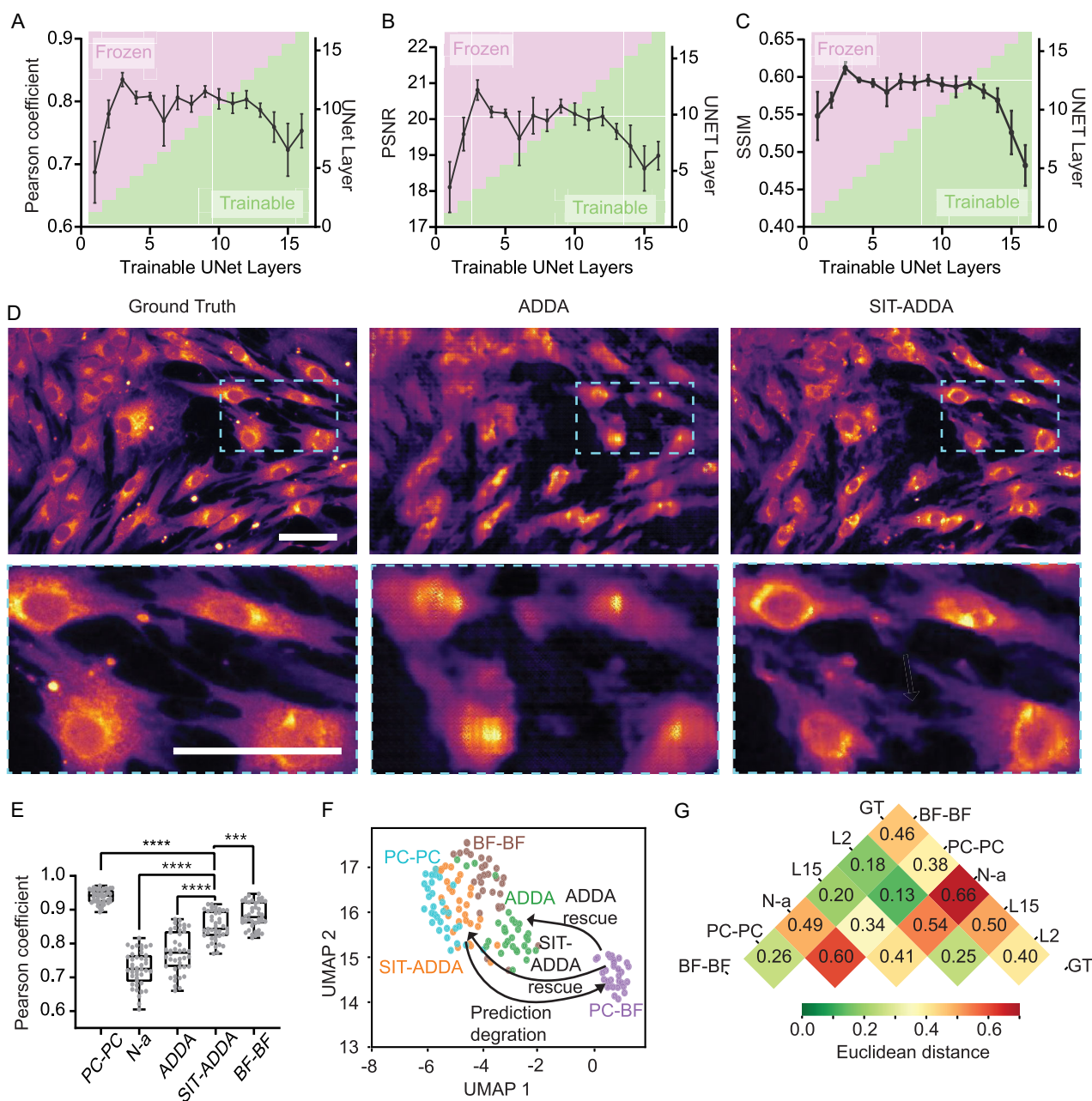


FIGURE 2 | SIT-ADDA improves cross-modal prediction and preserves subcellular detail.(A–C) Ablation results indicate that keeping deeper, semantically meaningful blocks frozen and fine-tuning only the first three convolutional layers (SIT-ADDA) gives the best balance of stability and adaptability—outperforming conventional ADDA (all 16 layers trainable) across Pearson correlation (A), PSNR (B), and SSIM (C). (D) Representative examples of CD29 fluorescence prediction show that conventional ADDA can introduce cross-shaped and other hallucinated artifacts and fail to capture perinuclear enrichment, whereas SIT-ADDA more faithfully reconstructs perinuclear halo-like signal and preserves peripheral intensity. Scale bars: 100 μ m. (E) Aggregate prediction accuracy across the test set: conventional ADDA raised Pearson correlation from a no-adaptation (N-a) baseline of 0.73–0.77, while SIT-ADDA further increased it to 0.85, reflecting superior inference of fine-scale subcellular localization. (F) 2D UMAP embeddings of predicted images were computed for four conditions: source-trained (PC-PC and BF-BF), no adaptation (PC-BF), ADDA-adapted (ADDA), and SIT-ADDA-adapted (SIT-ADDA). SIT-ADDA predictions align more closely with the target-domain clusters (PC-PC and BF-BF), indicating improved correction of domain-induced distributional shifts. (G) Quantification via Euclidean distance between cluster centroids (calculated in the high-dimensional feature space) confirms this visual observation: SIT-ADDA predictions are consistently closer to ground-truth and same-modality outputs than those produced by conventional ADDA or by the no-adaptation baseline. *** and **** refer to $p < 0.001$ and < 0.0001 , respectively.

while exclusively adapting the encoder using unlabeled target-domain data. However, to contextualize the prediction accuracy of SIT-ADDA within the broader landscape of unpaired image-to-image translation, we also benchmarked our model against contrastive unpaired translation (CUT) [40, 41]. While we found

that CUT successfully inferred the high-level morphological features and stylistic attributes of the immunofluorescent images from brightfield inputs, its predicted pixel-level intensities deviated substantially from the ground truth (Figures S3A and S3B). Consequently, the quantitative prediction accuracy of CUT

underperformed even conventional ADDA for both the DAPI (Figure S3C) and CD29 (Figure S3D) channels, highlighting its limitations for quantitative biological imaging.

2.3 | SIT-ADDA Enables Domain Adaptation Across Microscopy Variabilities

When pretrained models are applied to new microscopy datasets, domain shifts emerge from variations in illumination, contrast, or modality-specific characteristics. To systematically evaluate the capacity of SIT-ADDA to generalize across different shifts, we simulated three canonical sources of domain variability (Figure 3A): (1) variations in exposure levels, driven by changes in illumination

intensity and exposure time (Figures 3B–D); (2) spatial illumination gradients introduced by optical misalignment (Figures 3E–G); and (3) changes in magnification, reflecting differences in objective lenses and pixel resolution (Figures 3H–J); We synthetically applied these perturbations to controlled image sets, enabling a reproducible assessment of domain adaptation strategies under realistic microscopy conditions. We quantified the extent of these domain shifts by characterizing the perturbed images using cosine similarity, UMAP projections (Figure S4), and Shannon entropy (Figure S5). Overexposure and scaling produced the greatest divergence, whereas gradient perturbations caused the least.

We then applied both ADDA and SIT-ADDA to these perturbed datasets, systematically varying the number of frozen layers to

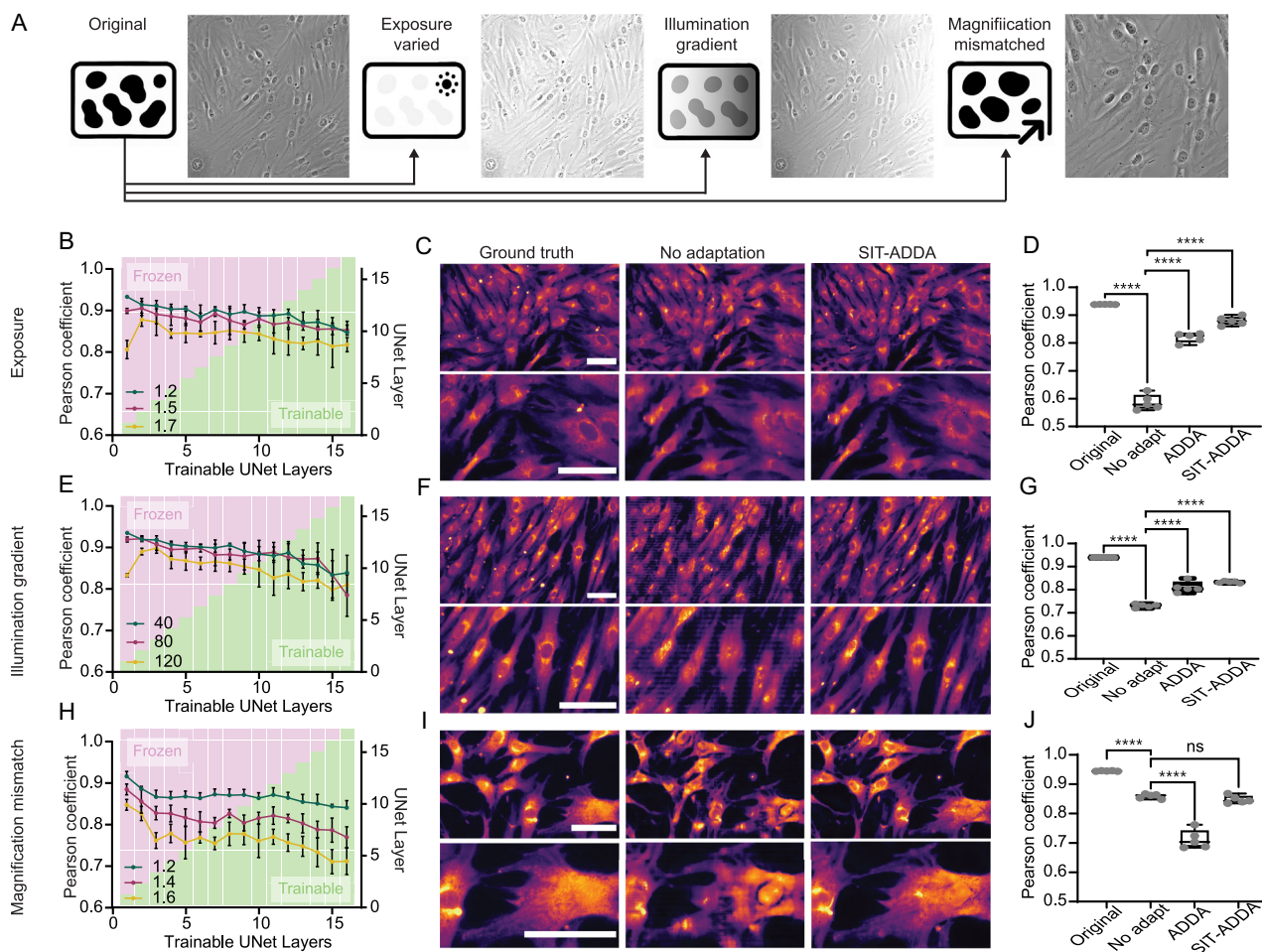


FIGURE 3 | SIT-ADDA produces prediction improvements across common microscopy domain shifts.(A) Experimental design: we simulated three canonical domain perturbations, namely overexposure, spatial illumination gradient, and magnification (scaling), and evaluated adaptation performance by measuring prediction fidelity (Pearson correlation) across different perturbation strengths. (B–D) Overexposure results. (B) Summary curves show Pearson correlation as a function of increasing overexposure: SIT-ADDA consistently outperforms conventional ADDA and the no-adaptation baseline. (C) Representative image examples: extreme overexposure that produces saturation and clipping in baseline predictions is corrected by SIT-ADDA, restoring signal delineation and reducing hallucinated features. (D) Quantification: SIT-ADDA increased Pearson correlation by 7.8% compared to conventional ADDA, representing a statistically significant improvement (1.7× overexposure). (E–G) Illumination-gradient results. (E) Summary curves for increasing gradient strength indicate greater prediction accuracy with SIT-ADDA. (F) SIT-ADDA normalizes spatial bias introduced by misaligned illumination, producing more uniform intensity profiles. (G) SIT-ADDA exceeded conventional ADDA by 13.2% in Pearson correlation (gradient strength = 120). (H–J) Magnification (scaling) results. (H) Pearson versus scaling factor shows that both ADDA and SIT-ADDA achieve smaller gains compared to the other perturbation types. (I) SIT-ADDA can capture global morphology and some subcellular detail (e.g., perinuclear CD29 enrichment) in moderate scaling. (J) Although SIT-ADDA outperforms conventional ADDA by 14.3% (1.6× scaling), the source-domain U-Net already shows strong robustness to magnification shifts, yielding stable and reliable predictions. Scale bars: 100 μm. “ns” and “****” refer to $p \geq 0.05$ and < 0.0001 , respectively.

identify configurations that best balance generalization and fine-tuning. Across all scenarios, SIT-ADDA outperformed conventional ADDA by roughly 10% on average relative to the ADDA baseline. As shown in Figure 3D,G, and J, prediction accuracy improvements were consistent across benchmarking metrics, with gains in Pearson correlation ($11.8\% \pm 4.5\%$ [6.0%, 19.1%]), PSNR ($14.0\% \pm 3.7\%$ [8.4%, 17.7%]), and SSIM ($23.4\% \pm 4.1\%$ [18.3%, 32.2%]). Counterintuitively, fine-tuning only the first layer often maximized performance. Crucially, this single-layer adaptation significantly outperformed the ‘no-adaptation’ baseline (Figure 3D,G). These data confirm that while deep retraining induces semantic drift, shallow adaptation remains indispensable for correcting acquisition shifts. In cases of more severe perturbations, such as 1.7-fold overexposure (Figure 3B, yellow curve) or a 120% illumination gradient (Figure 3E, yellow curve), this was not sufficient. In these conditions, incorporating the second layer into training was necessary to achieve robust adaptation. In contrast, magnification shifts (Figure 3J) yielded minimal gains, suggesting the baseline U-Net is inherently scale-tolerant. Notably, conventional ADDA reduced performance below the baseline in this regime. This degradation is consistent with negative transfer [42], where aggressive adversarial alignment disrupts the encoder’s intrinsic scale-tolerance to satisfy a strict distribution-matching objective.

Closer inspection of the microscopy images revealed in detail how adaptation improved predictions. Overexposed inputs that produced saturation artifacts in baseline predictions were corrected by SIT-ADDA, yielding improved signal delineation (Figure 3C). Illumination gradients that introduced intensity biases were effectively normalized, producing outputs with uniform intensity profiles (Figure 3F). In scaled images, SIT-ADDA restored the model’s ability to predict both cellular morphology and perinuclear CD29 enrichment (Figure 3I).

To compare AI-based generalization with parametric image restoration, we evaluated histogram normalization, a standard method that mitigates overexposure by rescaling intensities to approximate the original distribution. We found that this method provided only modest improvements in prediction accuracy on overexposed images (Figure S6). However, it did not correct saturation-induced intensity clipping, resulting in persistent artifacts and performance inferior to ADDA and SIT-ADDA. Together, these results show that selectively fine-tuning early U-Net layers allows SIT-ADDA to outperform parametric methods and robustly accommodate diverse domain shifts.

2.4 | Independent Model Ensembles Enable Unsupervised Optimization of SIT-ADDA

The effectiveness of SIT-ADDA depends on selecting the appropriate number of layers to fine-tune. In practical adaptation scenarios, only unlabeled target-domain images are available, while ground-truth annotations used in our experiments above are not. To address this challenge, we established an epistemic uncertainty framework that quantifies prediction variability across $K = 5$ independently trained U-Net models (Figure 4A). For each input x , we calculated the set of predictions $f_k(x)_{k=1}^K$; The sample mean of these predictions provided the point estimate, while their sample variance yielded a per-pixel uncertainty score. This variance provides a Monte Carlo approximation of the

epistemic component of the predictive variance, serving as a surrogate for full Bayesian integration [43, 44]. Variability across independently trained models has been shown to yield well-calibrated uncertainty estimates, to track true error in dense prediction tasks such as optical flow and medical image segmentation [45–47], and to detect distributional shifts [48]. Moreover, because independent stochastic gradient descent runs converge to different connected minima, their predictions approximate posterior samples [49].

To illustrate how our uncertainty framework guides adaptation, we first focused on the cross-modal scenario (Figure 2). In this setting, fine-tuning three layers with SIT-ADDA had previously outperformed full-model adaptation with fifteen trainable layers in traditional ADDA. This case provided a benchmark for evaluating whether the distribution of standard deviation (Std) values, computed for individual inputs, could reliably capture differences in model performance (Figure 4B). The analysis revealed a clear separation between the two conditions: SIT-ADDA exhibited consistently lower Std values and higher Pearson correlation coefficients relative to ADDA. This pattern was robust across domain shifts induced by overexposure (Figure 4C), illumination gradients (Figure 4D), and scaling perturbations (Figure 4E). To further examine the link between uncertainty and predictive accuracy, we compared Std values directly with Pearson correlation coefficients (Figure 4F–I), revealing an inverse relationship in which lower Std values coincided with higher Pearson correlations. Accordingly, the ensemble-derived standard deviation provides a practical criterion for selecting adaptation depth, enabling unsupervised optimization of SIT-ADDA across the tested domain shifts.

2.5 | SIT-ADDA Enables Robust Model Transfer Among Diverse Imaging Platforms

To evaluate real-world utility, we next tested whether SIT-ADDA can support cross-platform adaptation, a challenging scenario where multiple domain shifts arise simultaneously (Figure 5A). To this end, we employed mouse kidney tissue sections (Figure 5B) labeled with DAPI (nucleus) and phalloidin (F-actin) stains. As depicted in Figure 5A, training data for the baseline models were acquired on a high-end epifluorescence imaging system (Echo). For cross-platform testing, we introduced images from four microscopes representing a broad spectrum of optical performance: laboratory-grade optics (Etaluma), educational-grade systems (Olympus paired with an iPhone, as used in K-12 environments; Olympus-P), advanced hobbyist platforms (OpenFlexure 3D-printed microscope with pseudo-phase contrast modifications; 3DP-PPC), and citizen-science tools (OpenFlexure in a standard configuration; 3DP-BF). See Figure S7 for 3D-printed microscope configurations. This experimental design also demonstrates a unique application of SIT-ADDA, enabling compensation for the absence of fluorescence imaging capabilities in the latter three platforms. Representative images from the five microscopy systems and their quantitative characterizations can be found in Figures S8 and S9, respectively. Before testing domain shifts, we verified that the Echo dataset enabled accurate U-Net predictions of actin and nuclei (Figure 5C). Actin predictions preserved filamentous structures with luminal enrichment, while DAPI predictions matched nuclear morphology and distribution. These Echo-trained predictions serve as a stable reference for all downstream assessments,

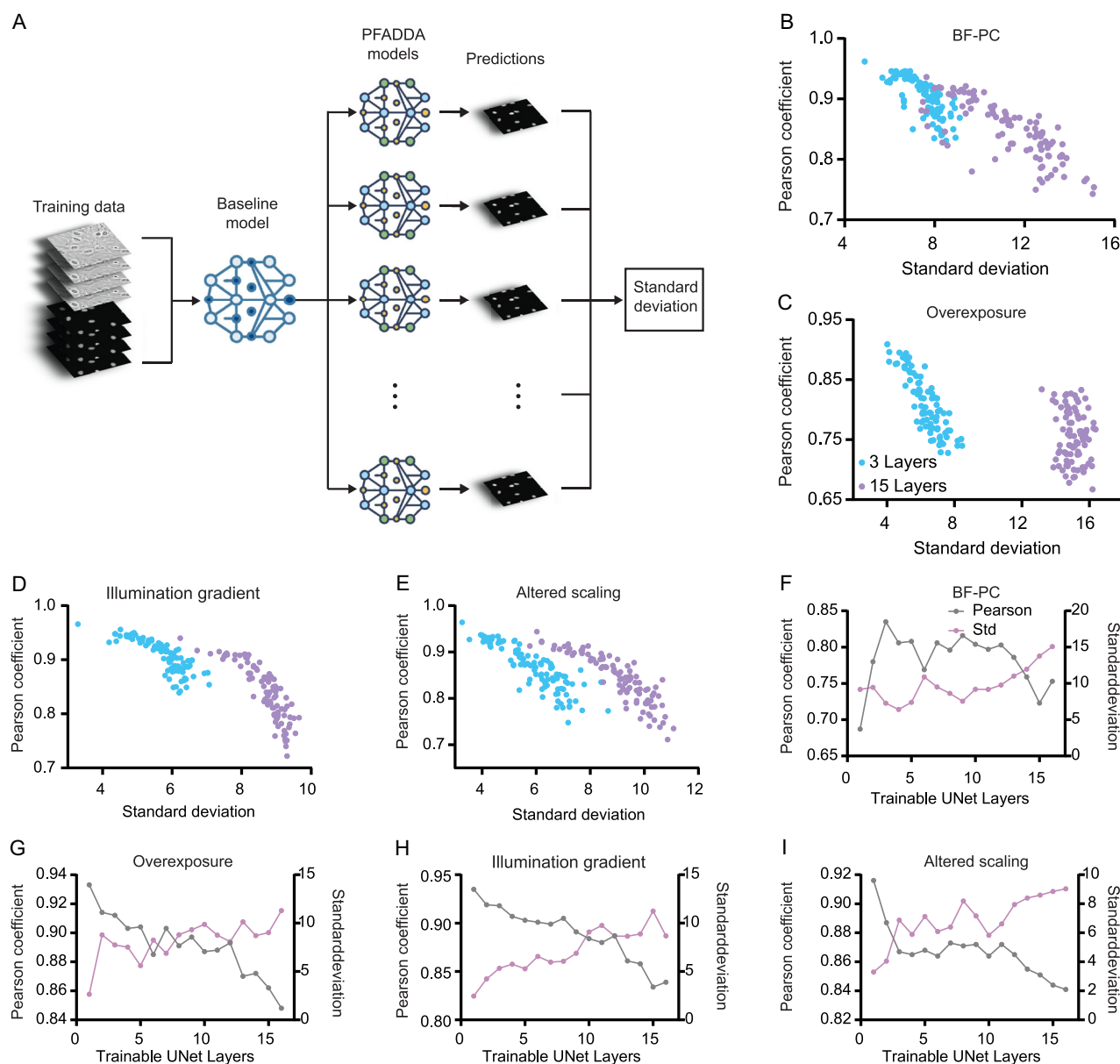


FIGURE 4 | Ensemble-based uncertainty enables unsupervised optimization of SIT-ADDA.(A) Workflow for estimating epistemic uncertainty. Five independently trained U-Net models were used to form an ensemble, with per-pixel variance across predictions serving as an approximation of predictive uncertainty. The mean yields the point estimate, while variance quantifies uncertainty without requiring ground truth. (B–E) Across all tested domain shifts, the scatter plots show a distinct separation between methods: SIT-ADDA (Layer 3, blue dots) predictions cluster tightly in the low-variance, high-accuracy region, whereas ADDA predictions (Layer 15, purple dots) spread into a regime of higher variance and lower accuracy. (F–I) Uncertainty inversely correlates with prediction accuracy. As standard deviation (Std) decreases, Pearson correlation systematically increases, confirming that predictive uncertainty is a reliable surrogate for adaptation performance.

allowing paired comparisons of adapted versus nonadapted outputs per field of view.

Across all four tested target domains, which formed distinct clusters in feature space (Figure 5D), SIT-ADDA substantially improved virtual staining relative to the no-adaptation baseline (Figure 5E–H). This effect was consistent across microscopes of markedly different optical quality, indicating that the gains are not idiosyncratic to a single device but reflect robustness to the combined shifts of contrast, illumination, and aberrations encountered in routine imaging. On the Etaluma (Figure 5E) and Olympus-P (Figure 5F) datasets in particular, SIT-ADDA produced sharper nuclear and cytoskeletal predictions with

reduced background artifacts, whereas baseline models yielded distorted outputs lacking biologically meaningful features. In the more challenging OpenFlexure domains, both 3DP-PPC (Figure 5G) and 3DP-BF (Figure 5H) inputs presented significant obstacles to accurate prediction, owing to their inherently low contrast, uneven illumination, and pronounced optical aberrations. Nevertheless, SIT-ADDA successfully adapted the model to infer actin-rich cell boundaries and discrete nuclear features that were barely discernible in the raw inputs. Notably, improvement on the lowest-grade platform (3DP-BF) was most evident for nuclei, consistent with the method inferring features that are reliably present despite degraded inputs.

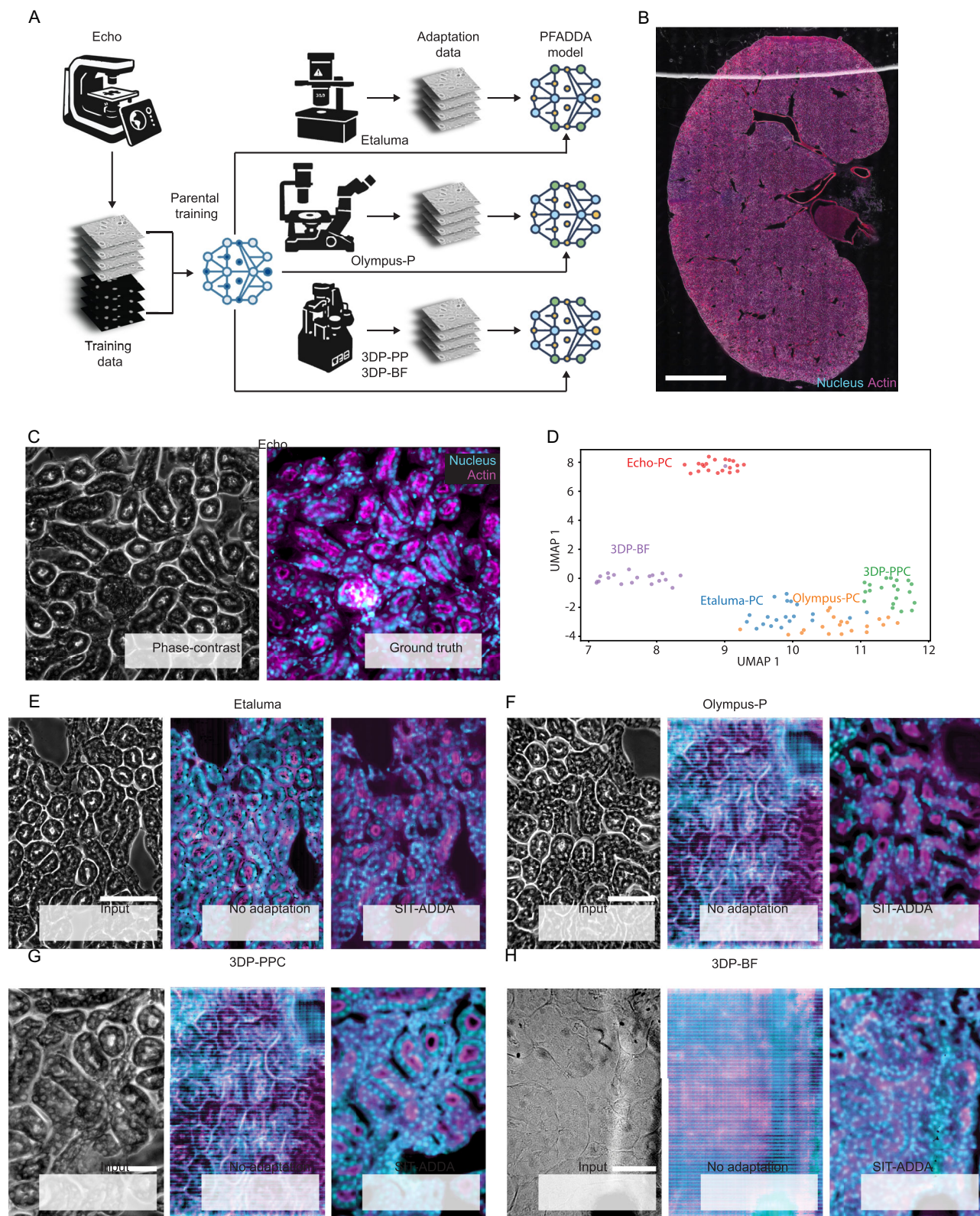


FIGURE 5 | SIT-ADDA enables robust model transfer across diverse microscopy systems.(A) Experimental setup: models trained on Echo microscope data were tested on four distinct target platforms spanning a spectrum of optical performance, from lab-grade instruments to citizen-science tools, to evaluate the ability of SIT-ADDA to infer fluorescence-like predictions. (B) Mouse kidney tissue labeled with DAPI (nucleus, cyan) and phalloidin (actin, magenta) was used to assess adaptation performance. Scale bar: 500 μm . (C) Source domain reference: predictions from the Echo-trained model closely matched ground truth, accurately predicting actin filaments and nuclear morphology. (D) Feature-space analysis revealed five distinct and well-separated clusters, corresponding to the Echo source images and the four target platforms, validating pronounced distributional shifts. (E–H) Representative inputs and predictions for each target platform. Without adaptation, predictions failed to capture meaningful nuclear or cytoskeletal structure. In contrast, SIT-ADDA consistently generated biologically plausible features. Scale bars: 100 μm .

Using UMAP (Figure 6A), we verified that SIT-ADDA shifts unadapted images closer to the Echo prediction cluster across all microscopy platforms. To assess adaptation effectiveness without ground truth, we used three complementary metrics: (1) image quality and cosine similarity to Echo predictions, measuring resemblance to reference images; (2) expert evaluation, benchmarking biological realism and interpretability; and (3) nucleus segmentation performance, testing utility for downstream analysis. These readouts were prespecified to reflect the biology in our samples. Nucleus-centric endpoints capture cell counts and enable density and morphology analyses, while lumen clarity reflects essential renal tubule architecture

characteristics. On lower-grade platforms without target-domain fluorescence ground truth, this combination provides orthogonal evidence: perceptual quality, blinded biological interpretability, and quantitative downstream performance. Also, because per-field fluorescence cannot be acquired on these devices, direct prediction-to-ground-truth comparisons are not possible. Our evaluation instead separates a distribution-alignment proxy (similarity to Echo) from independent endpoints not anchored to Echo (expert judgment and Cellpose accuracy). Consistent gains across these views would indicate that SIT-ADDA improves biologically meaningful signal rather than fitting only to Echo-specific appearance.

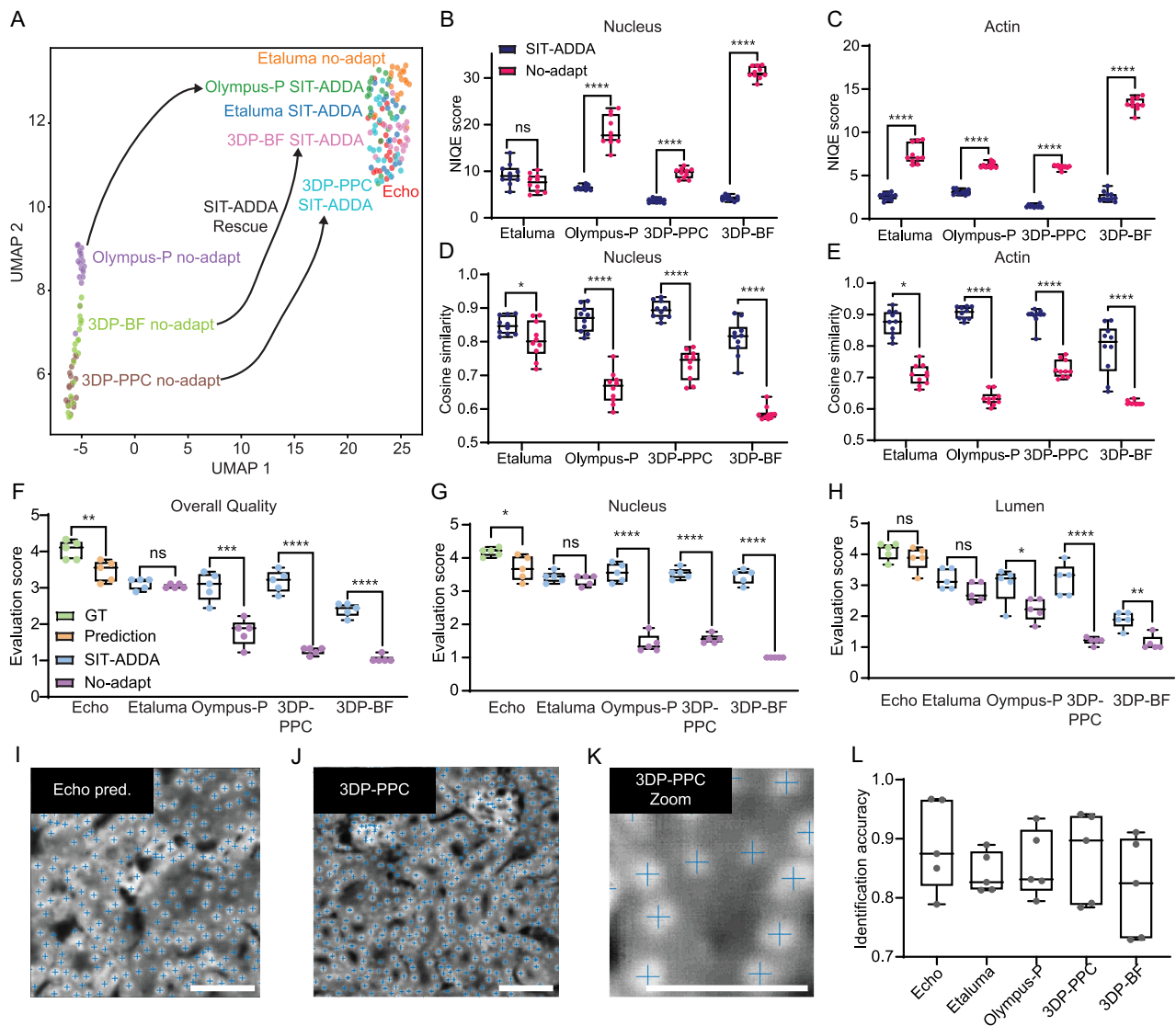


FIGURE 6 | SIT-ADDA improves cross-platform prediction quality and preserves downstream analytical utility. (A) UMAP embedding of predicted images demonstrates that SIT-ADDA shifts the nonadapted outputs toward the original Echo prediction cluster, indicating effective cross-domain alignment. Here, Echo predictions are used as a platform-consistent reference because co-registered fluorescence is unavailable on target devices. (B,C) NIQE scores, where lower values correspond to sharper and more natural images, demonstrate that SIT-ADDA consistently improves perceptual quality for both DAPI (nucleus, B) and phalloidin (actin, C) predictions relative to no adaptation. (D,E) Cosine similarity analysis further confirms that adapted predictions are more closely aligned with Echo references than their nonadapted counterparts, for both DAPI (nucleus, D) and phalloidin (actin, E). (F–H) Results from a blinded expert questionnaire show that SIT-ADDA-adapted images received higher ratings for (F) overall image quality, (G) nuclear clarity, and (H) lumen clarity. Adapted Etaluma predictions approached the quality of Echo outputs, while even lower-grade systems such as 3DP-BF benefited in terms of nuclear identification (G). (I–L) Cellpose-based segmentation analysis reveals that adapted predictions achieve nucleus identification accuracy comparable to Echo outputs, with near 90% centroid detection accuracy across all platforms. Scale bars: 50 μ m. ‘ns’, *, **, ***, and **** refer to $p \geq 0.05$, <0.05 , <0.01 , <0.001 , and <0.0001 , respectively.

We first evaluated image quality using the Naturalness Image Quality Evaluator (NIQE), where lower scores indicate sharper structures and reduced distortion. SIT-ADDA consistently lowered NIQE for both DAPI (Figure 6B) and actin predictions (Figure 6C). Adapted images also exhibited higher cosine similarity to the original predictions (Figure 6D,E). Finally, Laplacian variance, which reflects high-frequency artifacts, was reduced, while root mean square contrast increased, indicating improved structural clarity (Figure S10). For each platform we performed paired comparisons across matched fields of view between adapted and nonadapted outputs; significance indicators in Figure 6B–E reflect two-sided nonparametric paired tests with multiplicity control across metrics and platforms, with exact procedures detailed in Methods.

Beyond quantitative image quality metrics, we assessed biological interpretability through expert evaluation. A blinded questionnaire (example shown in Figure S11) of nine experienced microscopists asked participants to rate overall image quality (Figure 6F), clarity of nuclear visualization (Figure 6G), and definition of luminal structures (Figure 6H). Adapted images consistently received higher scores across all three metrics (Figure S12). Specifically, SIT-ADDA successfully adapted inputs from Etaluma, Olympus-P, and 3DP-PPC, yielding overall image quality and nuclear and lumen interpretability comparable to Echo predictions. This pattern held for both high- and low-grade instruments, reinforcing that the perceived gains are not device-specific but reflect improvements relevant to end-users' qualitative judgments. 3DP-BF adaptations showed improvement primarily in nuclear identification (Figure 6G) but not in lumen morphology (Figure 6H), likely reflecting the substantially lower quality of the raw inputs. The consistent improvement in nuclear identification likely reflects the visibility of nuclei across imaging conditions, enabling robust inference by SIT-ADDA.

Lastly, we assessed a critical utility of SIT-ADDA-generated images by performing a standard downstream analysis with Cellpose nucleus segmentation (Figure 6I–K). We focused on nucleus centroid labeling (blue crosses), a key metric for cell counting and subsequent packing or morphological analyses. We prioritize centroid detection because it is the decision-making endpoint in regular cell-based assays (confluence estimation, regional cell counts) and is less confounded by ambiguous nuclear boundaries across devices. From our analysis, labeling accuracy, defined as the fraction of true positives among all labels, showed no significant differences between Echo predictions and adapted conditions, with all reaching nearly 90% accuracy (Figure 6L). Across $n = 10$ manually annotated fields per platform, adapted predictions achieved nucleus centroid detection comparable to Echo references, indicating that perceived improvements translate into routine quantitative workflows. Visual inspection confirmed that the centroids identified by Cellpose coincided with the corresponding intensity centers (Figure 6K). Together, these results show that SIT-ADDA extends deep learning-based virtual staining across diverse imaging platforms, from commercial microscopes to fully 3D-printed systems, while maintaining the accuracy needed for reliable biological interpretation. It enables robust adaptation from high-end fluorescence instruments to resource-limited settings and provides a practical framework for making advanced virtual staining accessible in both scientific and educational contexts.

3 | Discussion

Our results indicate that full-network adaptation is often unnecessary and present a simple, stable recipe for handling domain shifts in microscopy. Conceptually, our finding is consistent with transfer-learning observations that early layers encode task-agnostic, low-level statistics, whereas deeper layers capture task-specific semantic structure that should be preserved [37, 50, 51]. In our setting, constraining updates to early filters absorbs acquisition variability (contrast, illumination, noise, optics) without disturbing downstream semantic representations of subcellular morphology. This functional separation is distinctly observable in the prediction artifacts, which offer a qualitative heuristic for optimizing adaptation strategy. For example, the observed high-frequency striations (Figure 1C) generally signal insufficient shallow alignment, whereas macroscopic, block-like hallucinations (Figure 2D) suggest semantic drift in deeper layers. These visual signatures provide a diagnostic cue for calibrating the freezing schedule, allowing the extent of adaptation to be tuned to the specific nature of the domain shift. To rigorously validate these observations, a mechanistic next step is to quantify layer-wise drift between source and target using centered-kernel alignment or related similarity measures to test this hypothesis directly [52, 53].

Operationally, we show that predictive uncertainty can act as a principled, label-free signal to choose adaptation depth in real deployments. While uncertainty has most often been used to prioritize labeling or flag low-confidence inputs, recent work has leveraged it for model optimization and hyperparameter selection [54, 55]. Building on that perspective, we repurpose prediction variance as a configuration selector for how many early layers to fine-tune, providing an actionable control knob when paired targets are unavailable. This converts a potentially brittle design choice into a data-dependent decision, aiding nonexpert users facing new microscopes or acquisition settings.

SIT-ADDA does have limitations that point to immediate refinements. First, magnification (scale) shifts remain challenging: the nonadapted model shows partial robustness, and early-layer adaptation alone does not reliably close the gap. In practice, this can be mitigated with scale-aware modules (e.g., spatial transformers [56] or multiscale front ends such as feature-pyramid networks [57]) and simple predeployment diagnostics (e.g., resampling target images to match source pixel size) to distinguish cases where preprocessing rather than adversarial alignment is appropriate [53]. Second, our formulation implicitly assumes the target prediction distribution resembles the source (e.g., similar specimens). This can be relaxed by (i) broadening pretraining with diverse, multisource microscopy and self-supervised contrastive objectives to bias encoders toward content-centric features [58, 59], and (ii) incorporating test-time, self-supervised adjustment that updates only normalization statistics or lightweight adapters via invariance objectives, stabilizing predictions without forcing source-like outputs [60–62].

In summary, by freezing high-level semantics and adapting low-level, domain-variant filters, SIT-ADDA offers a robust, easily deployed solution for virtual-staining domain shifts. Across modality, exposure, and illumination changes it outperforms conventional ADDA. The accompanying unsupervised ensemble procedure uses predictive variance to select the number of trainable layers, enabling no-ground-truth deployment. These

properties make SIT-ADDA well suited for cross-platform transfer, including to compact 3D-printed microscopes lacking fluorescence, and offer a practical template for domain adaptation in scientific imaging. Beyond microscopy, the idea of subnetwork-level adaptation informed by uncertainty may translate to medical imaging, materials characterization, and remote sensing, where labeled targets are scarce and distribution shifts are common.

4 | Methods

4.1 | Microscopy Datasets and Imaging Protocols

Human bone marrow-derived MSCs (ATCC, PCS-500-012) were cultured following the manufacturer's instructions and established protocols [4, 47]. For immunofluorescence, cells were washed with PBS+/+ and fixed in 4% paraformaldehyde (Thermo Fisher Scientific, 28908) in PBS+/+ (Gibco), followed by PBS+/+ washes. Blocking was performed with 2% donkey serum (Sigma-Aldrich, D9663-10ML) and 0.5% Triton X-100 (Sigma-Aldrich, T8787-50ML). After two PBS+/+ washes, cells were incubated with primary antibody solution (0.5% BSA, 0.25% Triton X-100, and primary antibody) for 30 min and washed twice with PBS+/+. Secondary antibody staining was performed for 30 min, followed by two PBS+/+ washes. Imaging was conducted using phase-contrast, brightfield, and fluorescence microscopy (Etaluma LS720, Lumaview 720/600 software) with a $20\times/0.45\text{NA}$ phase-contrast objective (Olympus, LCACHN 20XIPC).

For the mouse kidney dataset, a pre-labeled cryosection (FluoCells Prepared Slide #3, Invitrogen, F24630; $16\ \mu\text{m}$) stained with Alexa Fluor 568 phalloidin and DAPI was imaged. Fluorescence images were acquired in the DAPI and Alexa Fluor 568 channels using both an Echo Revolve microscope and an Etaluma LS720 microscope equipped with a $20\times/0.45\text{NA}$ phase-contrast objective (Olympus, LCACHN 20XIPC). Transmitted light images were additionally acquired using an Olympus CKX41 microscope with a Plan N $20\times\text{Ph1}$ phase-contrast objective coupled to an iPhone 4 via the eyepiece, and a 3D-printed OpenFlexure microscope fitted with an Olympus PlanC N $40\times\text{Ph2}$ phase-contrast objective. The illumination module of the OpenFlexure system was modified with a 30-mm LED ring to produce pseudo-phase-contrast images.

4.2 | Image Translation Model Architecture

The generative model employed in this work is based on a U-Net architecture with eight levels of downsampling and upsampling, designed for image-to-image translation tasks at a resolution of 1056×1056 pixels. The encoder and decoder components consist of convolutional and transposed convolutional layers, respectively, both utilizing 4×4 kernels with a stride of 2 and padding of 1 to appropriately preserve spatial dimensions during downsampling and upsampling operations. Skip connections are incorporated between corresponding encoder and decoder layers to facilitate gradient flow and retain fine-grained spatial information throughout the network. Normalization layers are applied after each convolutional layer except the outermost output layer. Activation functions include Leaky rectified linear unit (ReLU)

in the encoder and ReLU in the decoder, with a final Tanh activation to produce output images normalized between -1 and 1 . The discriminator model follows a PatchGAN design, structured as a three-layer convolutional neural network. Each convolution uses 4×4 kernels with a stride of 2 and padding of 1, progressively increasing the number of feature channels while reducing spatial resolution. Leaky ReLU activations and normalization layers are employed after each convolution except the last, which outputs a single-channel spatial map indicating the discriminator's real/fake classification confidence for each image patch. This patch-level discrimination allows the model to focus on local structure, encouraging high-frequency detail consistency in the generated images.

4.3 | Adaptation Model Implementation

4.3.1 | Traditional ADDA

We implemented ADDA to enable unsupervised domain adaptation. Let $\mathcal{D}_S = \{(x_S, y_S)\}$ denote the labeled source domain, and $\mathcal{D}_T = \{x_T\}$ the unlabeled target domain. The goal of the adaptation is to learn a target feature representation that aligns with the source feature space such that a classifier trained on \mathcal{D}_S generalizes effectively to \mathcal{D}_T .

Step 1: Source model training. A source encoder $F_{\theta_S} : \mathcal{R}^{H\times W\times C} \rightarrow \mathcal{R}^{H\times W\times C}$ is trained using the mean squared error (MSE) loss: $\mathcal{L}_{\text{MSE}}(y, \hat{y}) = \mathbb{E}\|y - \hat{y}\|_2^2$, where y and \hat{y} denote the ground truth and predicted pixel values, respectively:

$$\min_{\theta_S} \mathcal{L}_{\text{MSE}}(F_{\theta_S}(x_S), y_S).$$

Step 2: ADDA. Let F_{θ_S} denote the source encoder, F_{θ_T} the target encoder with parameters θ_T , and D_ϕ the discriminator with parameters ϕ . The discriminator maps feature representations to a binary domain label, $D_\phi : \mathcal{R}^{H\times W\times C} \rightarrow \{0, 1\}$.

The target encoder and discriminator are jointly trained by iteratively optimizing discriminative and encoding losses in an adversarial manner. Specifically, the discriminator is trained to classify whether the features originate from the source or target domain, while the encoder is trained to fool the discriminator

$$\min_{\theta_T} \max_{\phi} \mathbb{E}_{x_S} [\log D_\phi(F_{\theta_S}(x_S))] + \mathbb{E}_{x_T} [\log (1 - D_\phi(F_{\theta_T}(x_T)))] .$$

The source encoder F_{θ_S} is frozen during adaptation, and the target encoder F_{θ_T} is initialized from F_{θ_S} . During the iterative optimization process, D_ϕ is frozen while F_{θ_T} is optimized to align feature distributions, and F_{θ_T} is frozen while D_ϕ is optimized to discriminate between the outputs from the source and target encoders.

4.3.2 | Subnetwork Image Translation ADDA

Due to the limited amount of image data, properly limiting the model class size is essential to achieve proper generalization. To this end, we explored freezing the model weights of some layers in the target encoder while performing domain adaptation. Let $F_T = [f_1, f_2, \dots, f_n]$ denote the encoder's n convolutional layers. To determine the most effective weight-freezing strategy for domain adaptation in microscopy, we empirically evaluated three

different configurations: (1) freezing the first k layers while training f_{k+1} to f_n , (2) freezing the last k layers while updating the earlier f_1 to f_{k-1} ones, and (3) freezing all layers while only training f_k .

In doing this, we adopted a selective freezing scheme in SIT-ADDA, where only the first k convolutional blocks of the target encoder/decoder F_{θ_r} are trainable, and the remaining blocks are frozen: $\theta_{f_1}, \dots, \theta_{f_k}$ are updated; $\theta_{f_{k+1}}, \dots, \theta_{f_n}$ are fixed. We empirically tested values of $k \in \{1, \dots, 16\}$ and found that tuning just the first 1–3 blocks ($k \in \{1, 2, 3\}$) consistently yielded the most robust adaptation performance across diverse domain shifts in the images tested (see Figures 2A–C and 3B,E,H).

Baseline rationale. To rigorously isolate the algorithmic contribution of SIT-ADDA, we prioritized methods that resolve domain shifts within the latent feature space rather than through synthetic image generation. We primarily benchmarked against conventional ADDA, as its asymmetric architecture uniquely aligns with our strictly unsupervised setting (fixed source-trained predictor; unlabeled target images). Excluding methods that require target labels or predictor retraining prevents confounding comparisons and directly isolates the effect of subnetwork freezing. Additionally, to provide broader context against state-of-the-art image-to-image translation, we included CUT as a benchmarking baseline.

4.4 | Comparison with Contrastive Unpaired Translation

Unpaired image-to-image translation was performed using the CUT architecture [40]. The model utilized a 9-block ResNet generator and a basic PatchGAN discriminator. For training, input images were scaled to a resolution of 1024 pixels and randomly cropped to 512 by 512 pixel patches. The network was optimized using a least-squares GAN objective function with a batch size of 2. Optimization was performed with an initial learning rate of 0.0002 and momentum beta parameters of 0.5 and 0.999. The training process spanned a total of 200 epochs, where the learning rate was kept constant for the first 100 epochs and linearly decayed over the remaining 100 epochs.

4.5 | Image Data Perturbation

To systematically evaluate the robustness of our model under acquisition variability, we synthetically perturbed MSC CD29 images using three canonical transformations: scaling, overexposure, and illumination gradient. Image scaling was performed by simulating a digital zoom effect. Each image was cropped at the center by a factor inversely proportional to the zoom factor ($\times 1.2$, $\times 1.4$, and $\times 1.6$), followed by resizing to the original resolution. This mimics optical zoom or focal shift artifacts that may alter cellular context or spatial resolution. Overexposure was simulated by increasing the image brightness using a linear multiplicative factor ($\times 1.2$, $\times 1.5$, and $\times 1.7$) applied to pixel intensities via brightness enhancement. This perturbation emulates sensor saturation or improper illumination calibration often encountered in high-throughput imaging setups. Illumination gradient was introduced by superimposing a left-to-right horizontal intensity gradient across the image. The gradient values linearly increased from 0 to a maximum value (40, 80, or 120), then added channel-wise to the original image before clipping to the valid dynamic

range (0–255). This simulates uneven illumination across the field of view due to shading or vignetting artifacts in microscopy. Each perturbation was applied independently to the full set of CD29 images, resulting in a set of systematically distorted inputs used to assess domain generalization performance.

4.6 | Training & Evaluation Setup

Each dataset was split into training, validation, and test subsets (7:1.5:1.5). Models were implemented in PyTorch [63], with images resized to 1024×1024 pixels and processed in batches of 8. A U-Net backbone was pretrained on the source domain, after which selected convolutional blocks were frozen to various depths and adapted using our SIT-ADDA approach with domain-adversarial training. Models were trained for up to 200 epochs with the Adam optimizer, using a linear learning-rate decay and checkpointing based on validation Pearson correlation. Grid sweeps were performed across discriminator learning rates (10^{-3} – 10^{-6}) and 16 weight-freezing configurations. For evaluation, predictions were compared against ground truth using Pearson correlation, PSNR, and SSIM; predictions were histogram-matched before computing PSNR and SSIM. For each configuration, five independent models were trained, and performance was reported as the mean and standard deviation across runs, with unstable or nonconvergent models excluded. For unsupervised optimization of SIT-ADDA, five models were trained per configuration, and pixel-wise prediction standard deviations were averaged to quantify uncertainty. This was validated on cross-platform and domain-shifted CD29 datasets, where higher prediction accuracy consistently corresponded to lower predictive uncertainty. The optimal learning rate and freezing configuration were therefore selected as those minimizing uncertainty.

4.7 | Metrics of Prediction Accuracy Evaluation

To evaluate how the predicted images \hat{y} match the ground truth y , we used three complementary metrics. We quantify pixel-wise correlation using the Pearson correlation coefficient, which measures global linear intensity agreement between the reference image y

and the prediction \hat{y} : $\rho(y, \hat{y}) = \frac{\sum_{p \in \Omega} (y_p - \bar{y})(\hat{y}_p - \bar{\hat{y}})}{\sqrt{\sum_{p \in \Omega} (y_p - \bar{y})^2} \sqrt{\sum_{p \in \Omega} (\hat{y}_p - \bar{\hat{y}})^2}}$.

Here, y_p and \hat{y}_p are the intensities at pixel $p \in \Omega$, Ω is the set of all pixels with cardinality $|\Omega|$, $\bar{y} = \frac{1}{|\Omega|} \sum_{p \in \Omega} y_p$ and $\bar{\hat{y}} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \hat{y}_p$ are the corresponding sample means. Second, the PSNR $\text{PSNR}(y, \hat{y}) = 10 \log_{10} \left(\frac{L^2}{\mathbb{E}[(y - \hat{y})^2]} \right)$, where L denotes the maximum pixel value, quantifying prediction quality. Third, the SSIM $\text{SSIM}(y, \hat{y}) = \frac{(2\mu_y \mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + c_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)}$, with μ , σ , and $\sigma_{y\hat{y}}$ representing patch-wise means, variances, and cross-covariances, and c_1, c_2 stabilization constants, capturing structural and perceptual consistency.

4.8 | UMAP Visualization, Quality Metrics, and Segmentation Evaluation

To visualize image distributions and quantify domain shifts, each image was embedded using the pretrained BiomedCLIP [64]

image encoder to produce 512-D feature vectors. For qualitative visualization, these vectors were projected into two dimensions using UMAP (umap-learn v0.5.3 [65]) configured with $n_{\text{neighbors}} = 15$, $\text{min}_{\text{dist}} = 0.1$, and $\text{metric} = \text{'euclidean'}$, with all other parameters kept at their defaults. The resulting 2D embedding was used solely to illustrate the separation of domains in feature space. For quantitative analysis, pairwise Euclidean distances between cluster centroids were computed directly in the original 512-dimensional feature space. This approach avoids the topological distortions inherent to dimensionality reduction and ensures that reported distances accurately reflect structural similarity.

NIQE scores were computed by first fitting a custom model to the Echo prediction images as the reference set using MATLAB's `fitniqe` function. Following model fitting, each test image was similarly preprocessed and the NIQE score was calculated using the `niqe` function together with the fitted model parameters. Finally, per-image NIQE values were aggregated by calculating the mean score for each test set, yielding a comparison of image quality metrics across all experimental conditions.

Image similarity was quantified by extracting a 512-D feature vector from ResNet-18's global average-pooled convolutional output. Each vector was L2-normalized and the ResNet-18 model was pretrained on ImageNet (ILSVRC2012). Images were resized to 224×224 px and grayscale images were replicated across the three input channels to match the network's RGB input. Cosine similarities between each test vector and all reference embeddings were then computed. For each image, the maximum similarity (clamped to $[-1, +1]$) was retained, and these maxima were then averaged across samples to yield the mean similarity score per condition.

Nuclear segmentation was carried out in MATLAB (R2024a; The MathWorks) using the Cellpose algorithm (v1.0) [66]. Raw fluorescence images were input directly to Cellpose with the 'nuclei' pretrained model and default diameter settings. Segmentation masks were then compared against manually curated ground-truth annotations on 10 fields of view. Overall segmentation accuracy was reported as the complement of the mean false-positive and false-negative rates across all test images.

4.9 | Assessment of Image Interpretability and Prediction Quality

To quantitatively assess the interpretability and prediction quality of images produced by our method and baseline comparisons, a blinded questionnaire study was conducted with nine expert microscopists. Each participant was shown 50 fluorescent images and asked to rate three criteria: nuclear visibility (delineation of nuclear boundaries in DAPI), lumen visibility (perinuclear CD29 enrichment in MSCs), and overall image quality (artifact suppression under illumination perturbations). All ratings were provided on a 5-point Likert scale (1 = poor, 5 = excellent). Images were presented in a randomized and blinded fashion, with no information regarding imaging source, adaptation status, or prediction type disclosed. A brief anatomical orientation and visual reference were provided prior to the evaluation to standardize rating criteria. For each image, scores across participants were averaged to produce a consensus rating per criterion. Group-level

comparisons were performed using paired t-tests to identify conditions yielding significantly different interpretability scores.

Author Contributions

K.W.K.Y., A.Bai, C.J.H., and N.Y.C.L. conceived the experiments, A.Bermudez, Z.L., I.S., and M.L. acquired and processed the microscopy data, M.L. and V.G. constructed the early version of code., K.W.K.Y. and A.Bai finalized the coded and conducted the image adaptation. All authors analyzed the results. All authors reviewed the manuscript.

Acknowledgments

This work was partially supported by NSF (CBET-2244760, DBI-2325121, IIS-2048280) and NIH NIGMS (R35GM146735).

Funding

This study was supported by Division of Chemical, Bioengineering, Environmental, and Transport Systems (2244760), Division of Biological Infrastructure (2325121), Division of Information and Intelligent Systems (2048280), National Institute of General Medical Sciences (R35GM146735).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All code required to reproduce training and adaptation, including instructions and configuration files, is openly available on GitHub. <https://github.com/kaiwentw1018/SIT-ADDA>. Raw and processed imaging data, including MSC DAPI and CD29 images as well as cross-platform datasets, are deposited in Zenodo. <https://doi.org/10.5281/zenodo.17066331>.

References

1. M. Guo, Y. Wu, C. M. Hobson, et al., "Deep Learning-Based Aberration Compensation Improves Contrast and Resolution in Fluorescence Microscopy," *Nature Communications* 16 (2025): 313.
2. C. Lu, K. Chen, H. Qiu, et al., "Diffusion-Based Deep Learning Method for Augmenting Ultrastructural Imaging and Volume Electron Microscopy," *Nature Communications* 15 (2024): 4677.
3. M. Priessner, D. C. A. Gaboriau, A. Sheridan, et al., "Content-Aware Frame Interpolation (CAFI): Deep Learning-Based Temporal Super-Resolution for Fast Bioimaging," *Nature Methods* 21 (2024): 322–330.
4. S. Imboden, X. Liu, B. S. Lee, M. C. Payne, C.-J. Hsieh, and N. Y. C. Lin, "Investigating Heterogeneities of Live Mesenchymal Stromal Cells Using AI-Based Label-Free Imaging," *Scientific Reports* 11 (2021): 6728.
5. L. F. Morales-Curiel, A. C. Gonzalez, G. Castro-Olvera, et al., "Volumetric Imaging of Fast Cellular Dynamics with Deep Learning Enhanced Bioluminescence Microscopy," *Communications Biology* 5 (2022): 1330.
6. R. Cao, N. S. Divekar, J. K. Nunez, S. Upadhyayula, and L. Waller, "Neural Space-time Model for Dynamic Multi-Shot Imaging," *Nature Methods* 21 (2024): 2336–2341.
7. C. Qiao, S. Liu, Y. Wang, et al., "A Neural Network for Long-Term Super-Resolution Imaging of Live Cells with Reliable Confidence Quantification," *Nature Biotechnology* 44 (2025): 1–10.
8. M. Zhang, R. Li, S. Fu, et al., "Deep Learning Enhanced Light Sheet Fluorescence Microscopy for In Vivo 4d Imaging of Zebrafish Heart Beating," *Light: Science and Applications* 14 (2025): 92.

9. H. Wang, Y. Rivenson, Y. Jin, et al., “Deep Learning Enables Cross-Modality Super-Resolution in Fluorescence Microscopy,” *Nature Methods* 16 (2019): 103–110.
10. C. Qiao, Z. Li, Z. Wang, et al., “Fast-Adaptive Super-Resolution Lattice Light-Sheet Microscopy for Rapid, Long-Term, Near-Isotropic Subcellular Imaging,” *Nature Methods* 22 (2025): 1–11.
11. L. Jin, B. Liu, F. Zhao, et al., “Deep Learning Enables Structured Illumination Microscopy with Low Light Levels and Enhanced Speed,” *Nature Communications* 11 (1934): 2020.
12. C. Qiao, D. Li, Y. Guo, et al., “Evaluation and Development of Deep Neural Networks for Image Super-Resolution in Optical Microscopy,” *Nature Methods* 18 (2021): 194–202.
13. X. Liu, M. Kassab, M. Xu, and Q. Ho, “J-Invariant Volume Shuffle for Self-Supervised Cryo-Electron Tomogram Denoising on Single Noisy Volume,” *arXiv Preprint arXiv: 2411.15248* (2024).
14. J. Zhang, T. Zhao, S. Hu, and X. Zhao, “Robust single-particle cryo-em image denoising and restoration,” in ICASSP. 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE, 2024), 2995–2999.
15. R. Jiang, G. Zhang, Y. Yang, et al., *Cryoccd: Conditional Cycle-Consistent Diffusion with Biophysical Modeling for Cryo-Em Synthesis* (2025), 2505.23444.
16. I. Cunha, E. Latron, S. Bauer, D. Sage, and J. Griffié, “Machine Learning in Microscopy—insights, Opportunities and Challenges,” *Journal of Cell Science* 137 (2024): jcs262095.
17. L. Morgado, E. Gómez-de Mariscal, H. S. Heil, and R. Henriques, “The Rise of Data-Driven Microscopy Powered by Machine Learning,” *Journal of Microscopy* 295 (2024): 85–92.
18. H. Guanand M. Liu, “Domain Adaptation for Medical Image Analysis: A Survey,” *IEEE Transactions on Biomedical Engineering* 69 (2022): 1173–1185.
19. J. S. Yoon, K. Oh, Y. Shin, M. A. Mazurowski, and H.-I. Suk, “Domain Generalization for Medical Image Analysis: A Review,” *Proceedings of the IEEE* 112 (2024): 1583.
20. M. W. Lafarge, J. P. Plum, K. A. Eppenhof, P. Moeskops, and M. Veta, “Domain-Adversarial Neural Networks to Address the Appearance Variability of Histopathology Images,” in *International Workshop on Deep Learning in Medical Image Analysis* (Springer, 2017), 83–91.
21. K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, “A Closer Look at Domain Shift for Deep Learning in Histopathology,” *arXiv preprint arXiv:1909.11575* (2019).
22. F. Xing, T. C. Cornish, T. D. Bennett, and D. Ghosh, “Bidirectional Mapping-Based Domain Adaptation for Nucleus Detection in Cross-Modality Microscopy Images,” *IEEE Transactions on Medical Imaging* 40 (2021): 2880–2896.
23. Y.-C. Chen, Y.-Y. Lin, M.-H. Yang, and J.-B. Huang, “Crdoco: Pixel-level domain transfer with cross-domain consistency,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019), 1791–1800.
24. A. Mahbod, G. Dorffner, I. Ellinger, R. Woitek, and S. Hatamikia, “Improving Generalization Capability of Deep Learning-Based Nuclei Instance Segmentation by Non-Deterministic Train Time and Deterministic Test Time Stain Normalization,” *Computational and Structural Biotechnology Journal* 23 (2024): 669–678.
25. J. C. Gutiérrez Pérez, D. Otero Bager, and P. Maass, “Staincut: Stain Normalization with Contrastive Learning,” *Journal of Imaging* 8 (2022): 202.
26. Y. Zhong, W. Zhou, and Z. Wang, “A Survey of Data Augmentation in Domain Generalization,” *Neural Processing Letters* 57 (2025): 34.
27. T. H. Nguyen, D. Juyal, J. Li, et al., “Contrimix: Scalable Stain Color Augmentation for Domain Generalization without Domain Labels in Digital Pathology,” *arXiv preprint arXiv:2306.04527* (2023).
28. K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, “Domain Generalization: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2022): 4396–4415.
29. S. B. David, T. Lu, T. Luu, and D. Pál, “Impossibility theorems for domain adaptation,” in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, (JMLR Workshop and Conference Proceedings), (2010), 129–136.
30. K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, “Measuring Domain Shift for Deep Learning in Histopathology,” *IEEE Journal of Biomedical and Health Informatics* 25 (2020): 325–336.
31. E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), 7167–7176.
32. S. Bhattacharya, A. Wasit, J. M. Earles, N. Nitin and J. Yi, “Enhancing AI Microscopy for Foodborne Bacterial Classification via Adversarial Domain Adaptation across Optical and Biological Variability,” *arXiv preprint arXiv:2411.19514* (2024).
33. W. M. Pernice, M. Doron, A. Quach, et al., “Out of distribution generalization via interventional style transfer in single-cell microscopy,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2023), 4326–4335.
34. S.-Y. Guo, L. Huang, Y.-H. Mu, and T. Bai, “Unsupervised Adversarial Domain Adaptation with Hierarchical Semantic Consistency for Cross-Modal Nuclei Detection,” *Journal of Computer Science and Technology* 40 (2025): 1–12.
35. X. Ju, J. Su, S. Dai, X. Wu, M. Li, and D. Hu, “Domain Adversarial Neural Network with Reliable Pseudo-Labels Iteration for Cross-Subject EEG Emotion Recognition,” *Knowledge-Based Systems* 316 (2025): 113368.
36. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-assisted Intervention (Springer, 2015), 234–241.
37. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How Transferable Are Features in Deep Neural Networks?,” *Advances in Neural Information Processing Systems* 27 (2014).
38. N. De Franceschi, H. Hamidi, J. Alanko, P. Sahgal, and J. Ivaska, “Integrin Traffic—the Update,” *Journal of Cell Science* 128 (2015): 839–852.
39. P. P. Togarrati, N. Dinglasan, S. Desai, W. R. Ryan, and M. O. Muench, “CD 29 is Highly Expressed on Epithelial, Myoepithelial, and Mesenchymal Stromal Cells of Human Salivary Glands,” *Oral Diseases* 24 (2018): 561–572.
40. T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, “Contrastive learning for unpaired image-to-image translation,” in European Conference on Computer Vision (Springer, 2020), 319–345.
41. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Proceedings of the IEEE International Conference on Computer Vision, (2017), 2223–2232.
42. S. J. Panand Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering* 22 (2010): 1345–1359.
43. A. Kendall and Y. Gal, “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?,” *Advances in Neural Information Processing Systems* 30 (2017).
44. B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles,” *Advances in Neural Information Processing Systems* 30 (2017).
45. E. Ilg, Ö. Çiçek, S. Galesso, et al., “Uncertainty estimates and multi-hypotheses networks for optical flow,” in Proceedings of the European Conference on Computer Vision (ECCV), (2018), 652–667.
46. J. Park and G. Doğan, “Probabilistic U-Net with Kendall Shape Spaces for Geometry-Aware Segmentations of Images,” *arXiv preprint arXiv:2410.14017* (2024).

