COSE474-2024F: Final Project
Modifying UNIFMIR for domain-general isotropic reconstruction of fluorescence microscopy

Jungwoo Park

1. Introduction

Motivation Anisotropic resolution in three-dimensional (3D) imaging remains a significant challenge in fluorescence microscopy. The axial resolution (along the z-axis) is often significantly poorer compared to the lateral resolution (along the x- and y-axes), due to the inherent limitations of optical systems. This resolution imbalance creates difficulties in accurately reconstructing 3D structures, which are critical for detailed cellular and subcellular studies. Overcoming this challenge has been a persistent focus of research in the field.

In recent years, deep learning has emerged as a transformative solution to address the limitations of traditional methods. By leveraging neural networks, deep learning approaches have demonstrated the ability to learn complex patterns directly from the data, enabling significant improvements in 3D reconstruction accuracy. Regression-based reconstruction methods train networks to directly predict high-resolution 3D structures from low-resolution inputs. Adversarial learning, through the use of generative adversarial networks (GANs), has shown promise in enhancing image quality by generating realistic high-resolution outputs. More recently, diffusion models have been explored as a probabilistic framework for reconstruction, offering robust performance in capturing fine details and generating uncertainty estimates.

Despite these advancements, a critical challenge remains: the frequent need for retraining these models for new imaging conditions, datasets, or tasks. This issue arises because most existing deep learning approaches are designed for task-specific applications and lack the generalization capabilities necessary to adapt to diverse datasets without extensive fine-tuning. The reliance on retraining not only increases computational cost and time but also limits the accessibility of these methods for broader scientific applications.

This limitation underscores the growing need for foundational models in fluorescence microscopy and 3D imaging. Foundational models, trained on diverse datasets and tasks, offer the potential to generalize across a wide range of imaging scenarios. By leveraging transfer learning or zero-shot learning paradigms, such models could significantly reduce the dependency on retraining and accelerate the adoption of advanced computational methods in microscopy. Developing such foundational models would represent a paradigm shift, enabling robust and efficient solutions to overcome the long-standing challenges of anisotropic 3D resolution in fluorescence microscopy.

Problem Definition This project addresses the challenge of isotropic reconstruction in fluorescence microscopy, focusing on restoring anisotropic volume images effectively. Existing methods often require extensive fine-tuning, limiting their generalization capabilities across different imaging domains and degradation types. To overcome these limitations, we propose leveraging the pre-trained, non-generative UNIFMIR model, which incorporates xy-slice content to reconstruct anisotropic data without full model fine-tuning. By integrating a prompt and an additional patch attention layer, our approach enables task-specific adaptability, extending the model's capability to handle diverse degradation scenarios.

Contribution Existing methods for isotropic restoration often require fully training the entire model architecture for each specific imaging condition, including different image domains and degradation processes. In contrast, our approach leverages a pre-trained model without full fine-tuning, enabling parameter-efficient adaptation to various imaging conditions. Furthermore, our method introduces a novel "prompt tuning"-based framework for isotropic restoration. While some prior research explores all-in-one image restoration using contrastive learning or prompt-based approaches, these are primarily designed for natural images and do not incorporate reference images, such as xy slices, during training.

2. Methods

Overview of Our Methods and its Novelty UNIFMIR is a pre-trained model based on the Swin Transformer, inspired by SWINIR. It is designed for diverse tasks in fluorescence imaging, and features a multi-head, multi-tail architecture including an intermediate Swin Transformer module for en-

hanced feature representation. Trained under a multitask learning scheme, this intermediate module has acquired a broad range of features from various imaging domains, making it highly versatile for fluorescence image analysis.
However, UNIFMIR is a generalist for imaging task, not a specialist for restoring anisotropic images. Being necessary to fine-tune this model for down-stream task, our method integrates an additional patch-attention module after the frozen UNIFMIR model to enhance performance. The key idea is that the attention mechanism, with the ability of prompt to adaptively augment and refine features, selectively enhances high-frequency feature representations without the need for full model fine-tuning. In the module, prompt might work as the ”guidance” for the additional refinement for fine-tuning. Below we explain the functionality of patch attention module in specific.
(i) The module incorporates learnable 4×4 prompt patches, which serve as reference templates for the attention mechanism. These prompts are pre-initialized and learnable during training, allowing the model to adapt to various imaging conditions and degradation levels. Prompts act as a condensed representation of essential patterns, enabling efficient feature refinement without introducing excessive computational overhead.
(ii) Input features are divided into non-overlapping 4×4 patches, which are independently processed. This patch-based approach allows the model to focus on local details while preserving global context. Key Insight: Dividing images into patches helps localize attention mechanisms, ensuring that high-frequency details are appropriately handled.
(iii) The module employs a Query-Key-Value (QKV) attention mechanism, where Query is generated from the input patches to capture their unique characteristics, and Key, values are generated from the learnable prompts, encoding domain-specific knowledge. Attention scores are computed by the dot product of Query and Key, followed by a softmax normalization. These scores determine the relative importance of each prompt in refining the input patches. The adaptive nature of attention enables the module to selectively enhance regions with higher degradation.
(iv) Attention scores are applied to the Value to produce refined patch features, which are then reassembled into the original image structure using a folding operation. This step ensures that all enhanced patch features contribute to the final output, preserving both local and global consistency.
Model Architecture

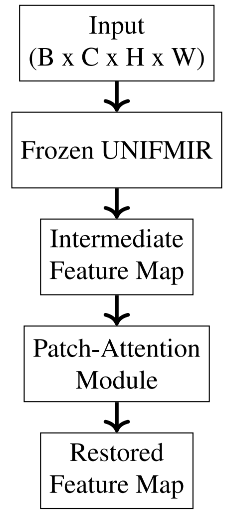

Figure 1. Model architecture overview, showing the process of input processing through the frozen UNIFMIR model, intermediate feature map extraction, and the application of the Patch-Attention module for feature refinement. Only Patch-attention module is learnable

Training Strategy Adam optimizer was used with learning rate 0.05, updating only the parameter of patch attention module. For the loss, mixture of MS-SSIM loss and L1 loss was used, since MSE itself would be a suboptimal loss in image restoration task. Batch size was set as 8, and the mixture rate of two loss was (0.5, 0.5) for MS-SSIM and L1 loss.
The degradation process used during training is average downsampling, and the rate of downsampling varies per dataset.

Algorithm 1 Training epoch

Initialize XY (XY slice of Image volume), model (class), Loss (class), max_epoch, optimizer
Function degrad() : Y-direction image degradation function

input_batch = XY
target_batch = XY
for i in range(max_epoch) **do**
    for j in range(input_batch.shape[0]) **do**
        input_tensor = degrad(input_batch)
        target_tensor = target_batch[j]
        output = model(input_tensor)
        loss = Loss(output, target_tensor)
        optimizer.step

3. Experiments

Dataset and Computer Resources FIB Dataset, and Liver Dataset from UNIFMIR model are used for checking performance. All of them is a 3D volumetric representation of fluorescence microscopy image, having a size of (256, 256, 256). No ground truth image was available for those images. All the training and further experiments are executed under Colab Free environment, with T4 GPU.

Pytorch framework was used for managing the model.

Experimental Design The model is trained using XY slices, which allow us to create pairs of (target, degradation) for training. We train the model separately for each dataset and evaluate its performance by providing ZX slices as input during testing. Since ground truth (GT) data is required for quantitative evaluation, quantitative comparisons are made using only the training images (XY slices). For the ZX slices, Only qualitative evaluation is performed, and comparison is performed against the original UNIFMIR model without any fine-tuning.

Experiment 1 (Liver Dataset) ZX slices of CREMI is downsampled 8-fold. Though we do not know the exact degradation process, we averaged the XY slices per 8th slice.

As shown in the figure 2, image restoration network can adapt to the unseen image with only prompts and attention. A vertical blur in the input image has been successfully restored, having sharper edge. PSNR was 59.42 dB.

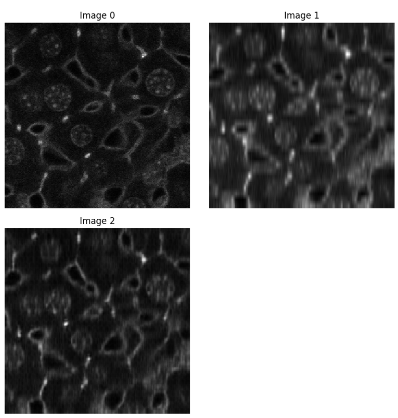

Figure 2. Iteration of final epoch. top left is target XY slice, top right is degraded slice, and bottom left is the result.

Experiment 2 (FIB) ZX slices of FIB is downsampled 8-fold. Though we do not know the exact degradation process, we averaged the XY slices per 8th slice.

The edge became sharper and the blurry effect has been alleviated. However, there are some stripe artifacts in the result, and PSNR was 78.97 dB.

Discussion and Analysis

The table shows that using prompt and patch attention layer for adaptation yields improved result. Since prompts work as basic template for revision, the model is capable of revising the high-frequency detail. However, if the downsampling rate is over 8, not only high-frequency detail but also the shape and structure of image has been distorted. In order to restore this, it is required to inject the prompt in the intermediate layer.

PSNR	UNIFMIR + prompt	UNIFMIR
LIVER	59.42	54.67
FIB	78.97	70.44

Table 1. PSNR Comparison against pretrained UNIFMIR

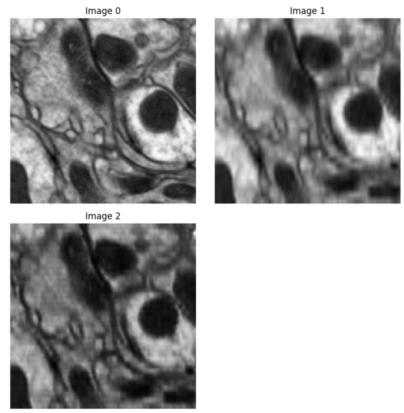

Figure 3. Iteration of final epoch. top left is target XY slice, top right is degraded slice, and bottom left is the result.

4. Future work

In this project, we tested the capability of model to adapt to various restoration scenario, with prompt given. Still the model is trained per each scenario, but this might be not favorable to real world application. Recent study for all-in-one image restoration aims to deal with generalization problem in end-to-end manner, without any further training process. This prompt-based approach might be able to be integrated with those methods, and thus will be our future goals.

5. History

Github link : `https://github.com/pianistoscar/20242R0136COSE47402/tree/920dfa323a3ca4a083169745b63742e8b191aa38/FinalProject`

Github commit history : `https://github.com/pianistoscar/20242R0136COSE47402/commits/main/`

Overleaf view link : `https://www.overleaf.com/read/fbmxkkpyfwxh#b747a4`

Overleaf history :

(i) Dec 8th, 2024 : Overall draft for the report

(ii) Dec 9th, 2024 : Adding and editing the experiment, method part

References

He, J. and et al. Isovem: Isotropic reconstruction for volume electron microscopy based on transformer. https://www.biorxiv.org/content/10.1101/2023.11.22.567807v3, 2023. bioRxiv Preprint.

Korkmaz, Y., Dar, S., Yurt, M., Özbey, M., and Çukur, T. Unsupervised mri reconstruction via zero-shot learned adversarial transformers. *IEEE Transactions on Medical Imaging*, 41:1747–1763, 2021. doi: 10.1109/TMI.2022.3147426.

Lee, K. and Jeong, W. K. Reference-free Axial Super-resolution of 3D Microscopy Images using Implicit Neural Representation with a 2D Diffusion Prior . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15007. Springer Nature Switzerland, October 2024.

Lumentut, J. S., Marchellus, M., Santoso, J., Kim, T. H., Chang, J. Y., and Park, I. K. Universal framework for joint image restoration and 3d body reconstruction. *IEEE Access*, 9:162543–162552, 2021. doi: 10.1109/ACCESS.2021.3132148.

Ma, C., Tan, W., He, R., and Yan, B. Pretraining a foundation model for generalizable fluorescence microscopy-based image restoration. *Nature Methods*, 21:1558–1567, 2024. doi: 10.1038/s41592-024-02244-3.

Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., Jug, F., and Myers, E. W. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat methods*, 15:1090–1097, 2018.