Step-Calibrated Diffusion for Biomedical Optical Image Restoration

Yiwei Lyu[1*], Sung Jik Cha[2*], Cheng Jiang[1], Asadur Zaman Chowdury[1], Xinhai Hou[1]
Edward S. Harake[1], Akhil Kondepudi[1], Christian Freudiger[3], Honglak Lee[14], Todd C. Hollon[1]
[1]University of Michigan
[2]Western Michigan University
[3] Invenio Imaging
[4] LG AI Research
yiweilyu@umich.edu

Abstract

High-quality, high-resolution medical imaging is essential for clinical care. Raman-based biomedical optical imaging uses non-ionizing infrared radiation to evaluate human tissues in real time and is used for early cancer detection, brain tumor diagnosis, and intraoperative tissue analysis. Unfortunately, optical imaging is vulnerable to image degradation due to laser scattering and absorption, which can result in diagnostic errors and misguided treatment. Restoration of optical images is a challenging computer vision task because the sources of image degradation are multi-factorial, stochastic, and tissue-dependent, preventing a straightforward method to obtain paired low-quality/high-quality data. Here, we present Restorative Step-Calibrated Diffusion (RSCD), an unpaired diffusion-based image restoration method that uses a step calibrator model to dynamically determine the number of steps required to complete the reverse diffusion process for image restoration. RSCD outperforms other widely used unpaired image restoration methods on both image quality and perceptual evaluation metrics for restoring optical images. Medical imaging experts consistently prefer images restored using RSCD in blinded comparison experiments and report minimal to no hallucinations. Finally, we show that RSCD improves performance on downstream clinical imaging tasks, including automated brain tumor diagnosis and deep tissue imaging.

Code — https://github.com/MLNeurosurg/restorative_step-calibrated_diffusion

Extended version — https://arxiv.org/pdf/2403.13680

Introduction

Medical imaging plays a major role in clinical medicine. Computed tomography, radiography, magnetic resonance imaging, and optical imaging are examples of common and indispensable medical imaging modalities used for diagnosis, guiding treatment decisions, and monitoring treatment response. Raman-based biomedical optical imaging uses Raman scattering to non-invasively evaluate human tissues for diagnostic purposes. As an advanced medical imaging modality, it has an increasing role in patient care and is now being

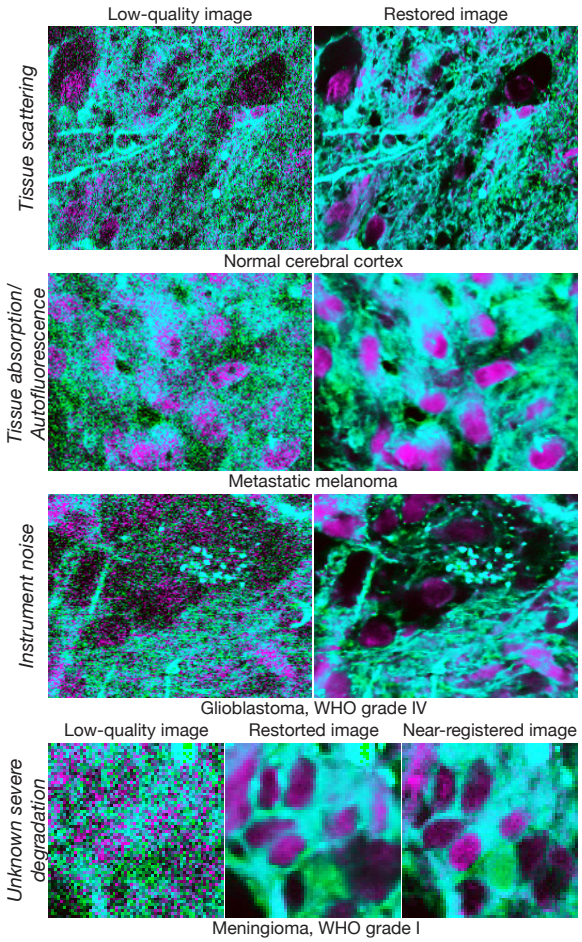

Figure 1: Examples of biomedical optical images restored using our proposed method, ***RSCD***. Across a range of known and unknown sources of image degradation, RSCD provides high-quality image restoration of fresh, surgical specimens imaged during brain tumor surgery. Our method can restore optical images with severe image degradation such that, after restoration, they can be used for downstream clinical tasks, including automated brain tumor diagnosis and deep tissue imaging during surgery.

*These authors contributed equally.

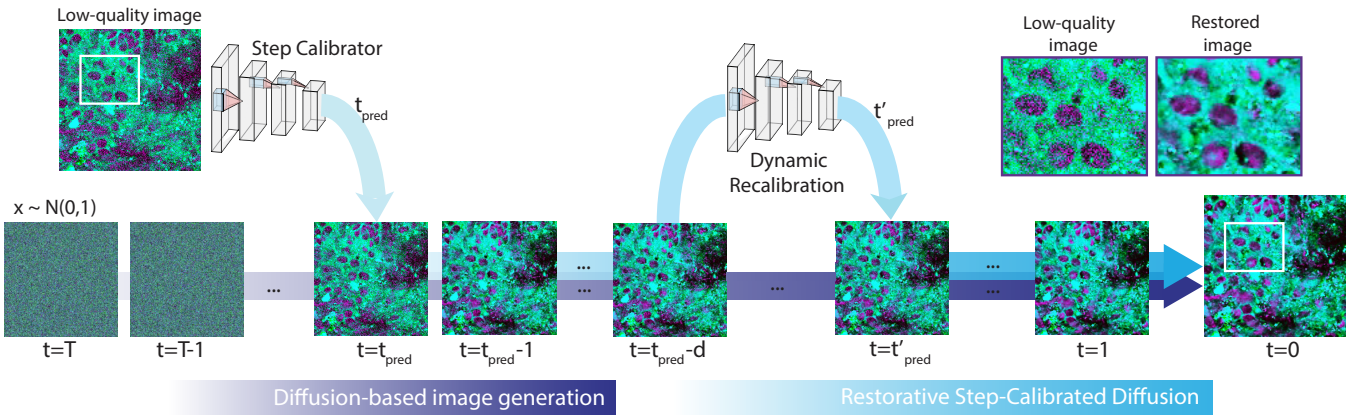

Figure 2: An overview of ***Restorative Step-Calibrated Diffusion (RSCD)***. We view the low-quality image as the output of an incomplete diffusion generation process that starts from Gaussian noise ($t = T$) and performs T steps of denoising (reverse diffusion) to generate a restored image at $t = 0$. We use a step calibrator model to predict t_{pred}, the number of steps of diffusion model denoising needed for image restoration, and we perform the reverse diffusion starting from t_{pred}. In addition, we use dynamic recalibration to dynamically adjust the number of steps required for optimal image restoration, t'_{pred}. The dynamic recalibration process and subsequent d steps of reverse diffusion denoising can be repeated until the restoration process runs to completion, obtaining the restored image at $t = 0$.

used for non-invasive cancer detection (Waterhouse et al. 2019; Lui et al. 2012), brain tumor diagnosis (Hollon et al. 2020, 2023), and surgical specimen analysis (Orringer et al. 2017; Mannas et al. 2023; Hoesli et al. 2017). Raman-based optical imaging has several advantages over conventional medical imaging as it does not require ionizing radiation, does not cause tissue damage, and can acquire images rapidly (within seconds) at the patient's bedside. Optical imaging uses light in the infrared electromagnetic spectrum to visualize biological tissues. Unfortunately, imaging within this spectral region causes optical imaging to be vulnerable to image degradation. Moreover, the sources of image degradation are multifaceted (Waterhouse et al. 2019; Manifold et al. 2019) (see Figure 1):

- *Tissue scattering*: scattering of the incident light can result in noisy images, especially when imaging at depth.
- *Tissue absorption*: biologically tissues contain chromophores that absorb light at specific wavelengths, reducing signal strength.
- *Auto-fluorescence*: some biological tissues have intrinsic fluorescence that degrades optical images by reducing signal-to-noise ratios.
- *Instrument noise*: image noise can be introduced from the detectors, electronic components, and external sources of interference.

The frequency and degree of image degradation are inherently unpredictable, making widespread clinical integration of Raman-based optical imaging challenging (Hollon et al. 2016; Waterhouse et al. 2019). Moreover, while small-scale imperfectly-paired image data has been generated in controlled laboratory settings (Manifold et al. 2019; Weigert et al. 2018), it is impossible to collect a large-scale paired dataset of perfectly aligned low-quality/high-quality clinical optical imaging that includes the full range of possible image degradations (Min et al. 2011; Audier et al. 2020; Moester, Ariese, and De Boer 2015).

Unfortunately, previous unpaired restoration methods suffer from hallucinations and perceptually poor reconstructions (Belthangady and Royer 2019), which can have detrimental downstream effects in medical imaging, increasing the risk of nondiagnostic images or diagnostic errors. An ideal method for Raman-based optical image restoration would (1) restore images degraded from a wide range of corruption sources, (2) only require unpaired data, (3) avoid hallucinations or perceptual artifacts, and (4) be time-efficient to allow for real-time, intraoperative image restoration.

We present ***Restorative Step-Calibrated Diffusion (RSCD)***, a novel diffusion-based unpaired image restoration method that efficiently restores low-quality Raman-based optical images with minimal to no perceptual artifacts or hallucinations. The rationale behind RSCD is that (1) Gaussian-based DDPMs have demonstrated generalization capacities in restoring non-Gaussian degradation (Chung, Sim, and Ye 2022; Chung, Lee, and Ye 2023), which makes them the ideal restoration model for unpredictable and unknown degradations; (2) DDPM restoration by directly performing reverse diffusion steps on degraded images requires much fewer steps than the full generative reverse diffusion process; and finally (3) when the degradation is unpredictable, the number of reverse diffusion steps required should vary for each image depending on the severity and pattern of the degradation.

Thus, RSCD includes a step calibrator model that determines the number of restoration steps required and a generative diffusion model that completes the restorative steps. Both models can be trained using unpaired high-quality images. RSCD is hallucination-resistant because it only involves editing the image via noise removal rather than full image generation from a random prior as is conventionally done in generative diffusion-based image restoration methods (Kawar

et al. 2022). Moreover, RSCD is time-efficient because it does not require the full reverse diffusion process. To further improve restoration quality and stability on unpredictable degradations, we designed a dynamic recalibration process that dynamically adjusts the number of remaining steps *during* restoration. The well-trained step calibrator model and dynamic recalibration enable RSCD to consistently restore biomedical optical images that have various degrees and distributions of image degradation, as shown in Figure 1.

We summarize our contributions as follows:

- We propose a novel unpaired image restoration method, RSCD, that is fast, reliable, and ideally suited for Raman-based optical imaging where noise is unpredictable with varying strength and pattern across and within images.
- RSCD is evaluated against other image restoration baselines, and outperforms them on image quality metrics. RSCD also achieves state-of-the-art performance on various unpaired perceptual metrics.
- Optical imaging experts consistently prefer images restored via RSCD over other methods, and report minimal to no hallucinations during human evaluations.
- RSCD can improve performance on downstream clinical computer vision tasks, including automated brain tumor diagnosis and deep tissue imaging.

Background

Intraoperative Raman-based Optical Imaging

In this paper, we focus on the restoration of stimulated Raman histology (SRH), a rapid and label-free optical imaging method based on Raman spectroscopy (Freudiger et al. 2008). SRH is used for a wide range of biomedical imaging tasks and has been clinically validated for imaging fresh, unprocessed human tissues and surgical specimens (Orringer et al. 2017). SRH begins with imaging a tissue specimen at two Raman shifts, 2845cm^{-1} and 2930cm^{-1}, which highlights the optical image features generated by the lipid and protein concentrations, respectively, to generate image contrast. SRH can capture high-resolution, diagnostic-quality images across multiple organs and tissues (Orringer et al. 2017; Hollon et al. 2018). A major advantage of SRH is that it can be performed rapidly ($\sim$1 minute) without the need for tissue processing or staining, making it ideally suited for intraoperative tissue evaluation and diagnosis during surgery.

Generative Diffusion Models

Denoising diffusion probabilistic models are used to generate high-quality and diverse images (Ho, Jain, and Abbeel 2020; Dhariwal and Nichol 2021). Training diffusion models consist of two processes. The first is the forward diffusion process, which gradually adds Gaussian noise to an image x_0 over T steps to obtain $x_1, x_2, \ldots, x_T$, where $x_t \sim N(\sqrt{1-\beta_t}x_{t-1}; \beta_t I)$ for each $1 \leq t \leq T$ and $\beta_1, \ldots, \beta_T$ follows a noise schedule. The total noise added over the T steps should be strong enough to reduce the image to Gaussian noise, such that $x_T \sim N(0, I)$. Since combining multiple steps of Gaussian noise results in Gaussian noise, we see that for each t, $x_t \sim N(\sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$ where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \Pi_{i=1}^{t}\alpha_i$. The second process is the restoration phase, where we train a model ϵ_θ that takes in a noised image ($x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ where $\epsilon \sim N(0, I)$) and the step number t, and tries to predict the noise (ϵ), and the training loss is $||\epsilon - \epsilon_\theta(x_t, t)||$. When we use the model to generate an image, we start by randomly sampling $x_T \sim N(0, 1)$, and we gradually remove noise by sampling $x_{t-1} \sim N\left(\frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)), \beta_t I\right)$, and we repeat this T times until we reach clean image x_0.

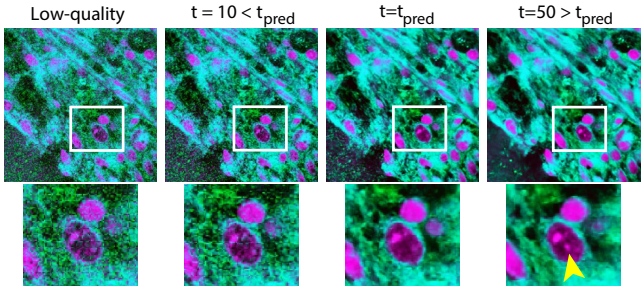

Figure 3: Importance of the **Step calibrator**. If we perform less than the optimal number of diffusion steps ($t < t_{pred}$), the image remains degraded. If we perform more than the needed steps ($t > t_{pred}$), the output image is excessively smooth, fine details are removed, and contains hallucinations (yellow arrow). ($t_{pred} = 34$ for this image.)

Methodology

The key idea behind RSCD is that when using the last t steps of a T-step generative diffusion process to perform image restoration, the number of steps t required should differ for each image due to variances in degradation severity and pattern. RSCD uses a trainable step calibrator model that predicts the value of t for each low-quality image. The step calibration is crucial for restoration quality because, as illustrated in Figure 3, if t is too small, then the reverse diffusion process cannot complete a sufficient number of steps to restore the low-quality image. If t is too large, then the reverse diffusion process generates excessively smooth and homogenized images with hallucinations. Thus, setting a fixed t value as a hyperparameter results in suboptimal image quality and perceptual features. To further improve restoration quality, we dynamically recalibrate the number of steps needed to restore the image during the reverse diffusion process. An overview of our method is shown in Figure 2.

Training Data

Training data was generated from approximately 2500 patients who underwent intraoperative SRH imaging to evaluate tissue during surgery (Orringer et al. 2017). Whole slide SRH images are approximately 6000×6000 pixels, which are then divided into 256×256 pixel patches, resulting in $\sim$1 million total patches. To obtain high-quality SRH images, optical imaging experts manually selected 4.5K high-quality patches, and then we automatically filtered through the remaining patches to obtain 840K relatively high-quality patches, using

the 4.5K patches as guidance.

Step Calibrator

We used a ResNet-50 model with MLP prediction head as the step calibrator. During training, a step number $t \sim \mathcal{U}(0, T)$ is sampled and t steps of Gaussian noise are added to a high-quality image according to a cosine schedule (Ho, Jain, and Abbeel 2020). The calibrator is trained to predict t using an L_2 loss between the prediction t_{pred} and t. A challenge we identified with this naive implementation is that the severity of image degradation varies *within* an optical image, and the trained model tends to predict steps based on the region with the least amount of degradation. To address this issue, we perform the following augmentation when training the step calibrator: we randomly divide the image into two regions, and we add t steps of noise to one of the regions; we then sample a second, smaller t' such that $0 \leq t' \leq t$, and add t' steps of noise to the other region. We train the model to predict the larger noise t to ensure the step calibrator favors calibrations that will restore the most degraded regions in the image.

In practice, we set $T = 1000$. After training, the step calibrator was used to predict noise level on approximately 63K low-quality images that had the lowest score from section . We found that no low-quality image required more than 200 steps for image restoration.

Diffusion Model

The diffusion model is trained as a generative diffusion model that generates high-quality SRH images unconditionally with a cosine noise schedule. We follow a similar training objective and procedure as described in (Ho, Jain, and Abbeel 2020), except that we use a shortcut for training efficiency: RSCD only requires training the model with noise level randomly sampled between 1 and T' steps (where T' just need to be larger than the maximum t_{pred} of low-quality images) instead of sampling from the full range 1 to T when training diffusion-based generation models. In practice, we set $T' = 200$ due to the distribution in of t_{pred} as discussed in the previous section. This allows us to train the model more efficiently than conventional diffusion-based image restoration methods (Kawar et al. 2022). We first train the model for one pass through all 840K images, then fine-tune the model on the 4.5K high-quality images for 20 epochs.

Dynamic Recalibration

The step calibrator is trained to predict Gaussian noise levels; however, Raman-based optical image degradation is not limited to Gaussian noise. When image degradation deviates significantly from Gaussian noise, the step calibrator is more likely to under or overestimate the required number of diffusion steps, which results in poor image restorations. Therefore, to better calibrate the number of steps *during* the image restoration process, we perform dynamic recalibration: after we predict t_{pred} for an input low-quality image, instead of directly performing all t_{pred} steps of denoising, we only apply d steps of denoising, i.e. $x_{t_{pred}} \rightarrow x_{t_{pred}-1} \rightarrow ... \rightarrow x_{t_{pred}-d}$. Then, we will use the step calibrator to predict the remaining steps of denoising needed for $x_{t_{pred}-d}$, and continue denoising starting from the updated predicted number of steps. Another d steps of denoising are performed before additional calibration. We repeat the process until the predicted number of steps remaining is less than d, and then we complete the remaining steps without additional recalibration. The process is illustrated in Algorithm 1.

Algorithm 1: Restorative Step-Calibrated Diffusion with Step Calibration and Dynamic Recalibration: Sampling

Requires: Low-quality image x, step calibrator S, diffusion model ϵ_θ, recalibration interval d, total steps T, hyperparameters based on noise schedule α_1,...,α_T,$\bar{\alpha}_1$,...,$\bar{\alpha}_T$,β_1,...,β_T
$t \leftarrow S(x)$ *# Initial step calibration*
$t_{end} \leftarrow 0$
while $t > 0$ **do**
  $x_t \leftarrow x$
  $t_{end} = max(t - d, 0)$ *# d-steps of reverse diffusion*
  for k **in** $[t, t-1, ..., t_{end}+1]$ **do**
    $z \sim N(0, 1)$
    $x_{k-1} \leftarrow \frac{1}{\sqrt{\alpha_k}}(x_k - \frac{1-\alpha_k}{\sqrt{1-\bar{\alpha}_k}}\epsilon_\theta(x_k, k)) + \sqrt{\beta_k}z$
  end for
  $x \leftarrow x_{t_{end}}$
  $t \leftarrow S(x)$ *# Dynamic recalibration*
end while
return x

Note that even though the observed noise from SRH images is not Gaussian, we can train our step calibrator and diffusion model with Gaussian noise because (1) Gaussian diffusion models are known to have generalization capacities for restoring degradations that deviates from Gaussian distributions (Chung, Lee, and Ye 2023; Chung and Ye 2022), as small Gaussian noise is added during each DDPM step that makes the resulting degradation distribution more Gaussian-like; and (2) dynamic recalibration can mitigate possible step prediction inaccuracies due to different noise patterns by dynamically adjusting the remaining number of steps.

Experiments

Baselines and Ablations

In the following experiments, we compare RSCD to the several unpaired image restoration baselines: CycleGAN(Zhu et al. 2017), synthetic noise/noise2noise (Lehtinen et al. 2018; Manifold et al. 2019), conditional diffusion (Kawar et al. 2022; Saharia et al. 2022), CCDF (Chung, Sim, and Ye 2022), regularized reverse diffusion (RRD) (Chung, Lee, and Ye 2023), deep image prior (Ulyanov, Vedaldi, and Lempitsky 2016), and median blur. In addition, we also conduct ablation studies over the following design choices: dynamic recalibration, step calibrator, and cosine noise schedule. We compare our method against no dynamic recalibration, no step calibrator (either using a non-parametric noise estimator (Chen, Zhu, and Heng 2015) as replacement or denoise for a fixed number of steps), evaluating the commonly-used linear noise schedule versus the cosine schedule, and training our step calibrator without augmentation.

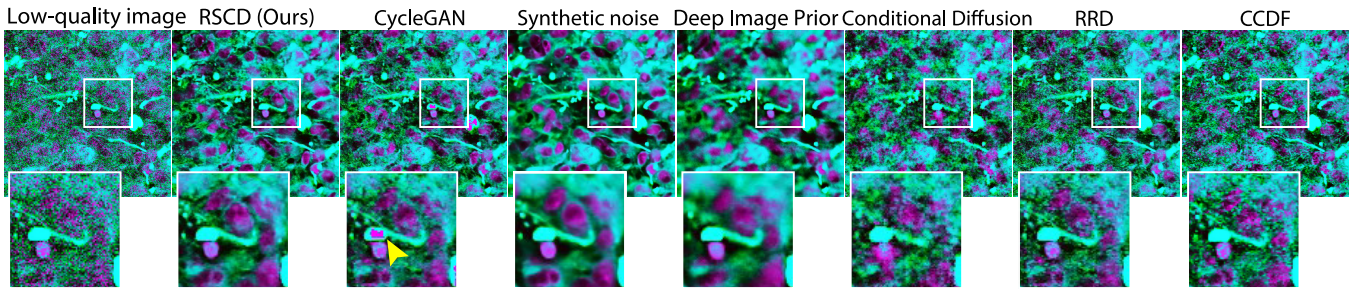

Figure 4: Visual comparison of unpaired image restoration methods. CycleGAN hallucinates/inpaints nuclei within non-cellular structures (yellow arrow). Synthetic noise and Deep Image Prior (DIP) produce overly smoothed, unrealistic images. Conditional diffusion and Regularized Reverse Diffusion (RRD) generally perform insufficient image restoration.

		FID ↓	CMMD ↓
	Original unrestored LQ images	53.22	0.566
Baselines	Median Blur	95.28	1.099
	Deep Image Prior	57.56	0.570
	Synthetic noise	58.24	0.356
	CycleGAN	37.26	0.196
	Conditional Diffusion	47.66	0.581
	CCDF	43.21	0.216
	Regularized Reverse Diffusion	38.37	0.337
Ours	RSCD	**32.02**	**0.128**
Ablations	No Dynamic Recalibration	34.62	0.137
	Nonparametric Noise Estimator	35.53	0.270
	No Step Calibration, 10 steps fixed	36.41	0.174
	No Step Calibration, 50 steps fixed	42.06	0.224
	Linear noise schedule	38.71	0.150
	Unaug. Step Calibrator Training	33.14	0.133

Table 1: Results restoring low-quality open-source SRH images from OpenSRH (Jiang et al. 2022). FID score and CMMD score are measured between all restored images and the 4.5K expert-selected high-quality SRH images.

Evaluation on Unpaired Images

We evaluate the perceptual quality of image restoration on 12K unpaired low-quality images sampled from the largest public SRH dataset, OpenSRH (Jiang et al. 2022). We evaluate RSCD against the above baseline methods and ablations. We use Frechet Inception Distance (FID) (Heusel et al. 2017) and CLIP Maximum Mean Discrepancy (CMMD) (Jayasumana et al. 2024) between the restored images and the 4.5K expert-selected high-quality images as metrics.

We report all results in Table 1. RSCD outperformed all baselines and ablations, indicating that our method produces high-quality and realistic image restorations of low-quality Raman-based optical imaging. We present an example of qualitative comparison in Figure 4.

Evaluation on Near-registered Images

Next, we aim to evaluate the quality and fidelity of image restoration using an approximately paired low-quality/high-quality SRH imaging dataset. As mentioned before, perfectly paired images are challenging to obtain because optical image degradation is conditional on several (stochastic) factors. However, it is possible to obtain noisy/clean images that are nearly paired, by scanning the same specimen with a cold and warm laser (cold laser is known to produce more noisy images) and then re-align the two images. Through this process, we obtained 2,135 pairs of near-registered low-quality/high-quality SRH patches. These pairs are called 'near-registered' because non-affine deformation and optical tissue sectioning make perfect spatial alignment impossible, therefore these images cannot be used as paired training data. Nevertheless, the near-registered SRH dataset is a useful evaluation benchmark for restoration quality and fidelity, as paired data allows additional evaluation metrics (PSNR, SSIM and LPIPS).

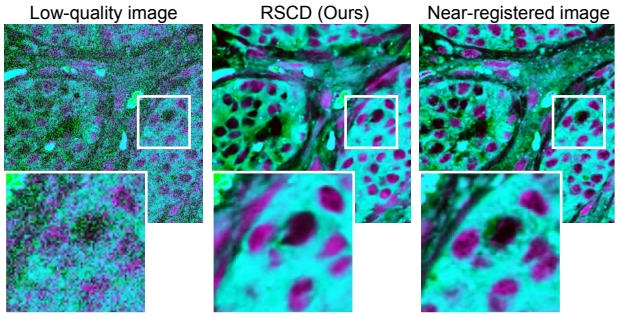

Figure 5: Examples of paired low-quality/near-registered SRH images, and the restored image via RSCD.

We report image quality and restoration metrics comparing RSCD and above baselines in Table 2. Our method achieves the best FID and CMMD, which means our restored images are perceptually the closest to real high-quality images, while most baselines produce blurry or unrealistic images that result in significantly worse FID and CMMD scores. RSCD also has the second highest PSNR and SSIM scores, only behind deep image prior. This indicates that the images restored by our method are close to the ground truth high-quality images while avoiding being excessively smooth and unrealistic (excessively smooth or blurry images, such as the ones produced by median blur or deep image prior, are known to have inflated scores on metrics based on pixel-wise MSE (Wang and Bovik 2009), including PSNR and SSIM). RSCD was the only method to achieve good metrics on both pixel-level and perceptual metrics. Examples of near-registered image pairs restored using RSCD are shown in Figure 5.

Our method also achieved top performance among all

		FID ↓	CMMD ↓	PSNR ↑	SSIM ↑	LPIPS ↓
	Paired HQ Images	0.00	0.000	100.00	1.000	0.000
	Paired LQ Images	47.64	0.506	26.97	0.604	0.598
Baselines	Median Blur	72.71	1.047	28.87	0.791	0.336
	Deep Image Prior	41.51	0.385	**29.27**	**0.806**	0.284
	Synthetic noise	77.43	1.987	22.94	0.662	0.464
	CycleGAN	22.03	0.185	28.39	0.779	0.667
	Conditional Diffusion	34.58	0.332	28.07	0.740	0.377
	CCDF	30.80	0.238	28.04	0.704	0.359
	Regularized Reverse Diffusion	23.37	0.270	27.89	0.682	0.459
Ours	RSCD	**21.05**	**0.104**	29.06	0.791	**0.280**
Ablations	No Dynamic Recalibration	21.44	0.111	28.99	0.781	0.281
	Nonparametric Noise Estimator	22.77	0.215	27.90	0.691	0.418
	No Step Calibrator, 10 steps fixed	22.58	0.163	28.29	0.721	0.357
	No Step Calibrator, 50 steps fixed	29.72	0.136	29.05	0.788	0.335
	Linear noise schedule	32.71	0.386	29.06	0.789	0.283
	Unaugmented Step Calibrator Training	21.15	0.105	29.03	0.786	0.284

Table 2: Results of restoring paired low-quality/near-registered high-quality SRH images. FID score is computed between all restored low-quality images and all near-registered images. Bolded numbers are best in each metric, and underlined are second best. RSCD achieved the best FID and LPIPS scores and the second-best PSNR and SSIM across all baselines and ablations.

5 metrics compared to all ablations, justifying our design choices of using the step calibrator, cosine noise schedule, and dynamic recalibration.

Human Expert Preference

The primary application of RSCD is to restore low-quality SRH images for intraoperative surgical specimen analysis and diagnosis by clinicians and domain experts, so their opinions are **the most important metric** for evaluating SRH image restoration quality. We recruited three clinicians and three optical imaging experts to evaluate the quality of the restored images and provide preference ratings. Each rater was asked to give their preferences between RSCD restoration and the five baseline methods' restorations. For every preference task, the experts are given the original low-quality image and the two restored images, one from our method and one from the baselines, in random and blinded order. The experts then selected their preferred restored image based on restoration quality and fidelity. In addition, raters were asked if either restoration contained hallucinations. We obtained expert preferences on restorations of 100 randomly selected low-quality images, and we gave every preference task to two raters to measure inter-rater agreement.

We report the results in Figure 6. Human experts preferred RSCD restoration more often than baselines. Raters also reported the least number of hallucinations in our method, less than deterministic methods such as median blur that can create artifacts with severely degraded images. Inter-rater agreement was over 80% for both quality and hallucination assessments. Our method generates high-quality and reliable restorations of low-quality SRH images according to clinicians and domain experts.

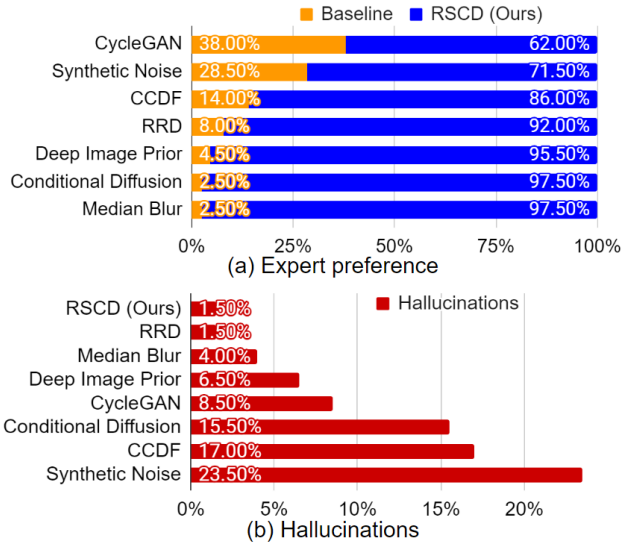

Figure 6: Results of human expert evaluations. **(a)** The human expert preferences between RSCD and baselines. **(b)** Percentage of restored images by each method that the experts indicated hallucinated. As shown, the experts overall preferred the restorations from RSCD over the baselines.

Downstream Clinical Tasks

In the previous section, we showed that RSCD can restore low-quality SRH images to higher quality, more realistic, and more preferable images compared to other existing unpaired image restoration methods. In this section, we demonstrate the downstream potential of RSCD in restoring SRH images on two clinical tasks: automated deep learning-based tumor diagnostics and z-stack restoration in deep tissue imaging.

Deep Learning-based Tumor Diagnostics

Image degradation can decrease the performance of computer vision systems and result in decreased diagnostic accuracy for automated brain tumor classification (Hollon et al. 2020). In this experiment, we test whether RSCD can be used to facilitate more accurate deep learning-based diagnosis of SRH images by restoring model inputs.

We use the SRH tumor classification model from (Hollon et al. 2020), the most widely accepted study on deep learning-based brain tumor diagnosis using SRH. The model

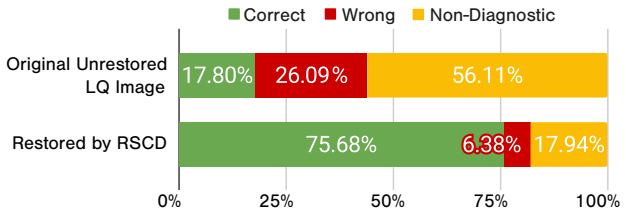

Figure 7: Classification accuracy of the SRH tumor classifier from (Hollon et al. 2020) on low-quality images and restored images by RSCD.

classifies SRH images into one of 3 classes: normal tissue, tumor tissue, or non-diagnostic. The model classifies images as non-diagnostic if the image quality is sufficiently poor that human experts are not able to determine the underlying tissue diagnosis. For each unpaired low-quality image, we pass both the original low-quality SRH image and the restored image through the SRH tumor classification model. We report the classification results in Figure 7.

Without restoration, over 50% of the SRH images were classified as non-diagnostic and another 26% were incorrectly classified. After restoration with RSCD, the classifier correctly diagnosed over 75% low-quality SRH images, and only miss-classified 6% of images. Restoring SRH images via RSCD can significantly improve deep learning-based automated diagnostic accuracy, and can drastically reduce non-diagnostic predictions caused by low image quality. Importantly, SRH restoration significantly reduces the risk of a wrong diagnosis, which can have a severe detrimental effect on patient care and surgical treatment. RSCD can make existing automated diagnostic tools safer and more reliable.

Z-stack Image Restoration

A known limitation of Raman-based optical imaging is that signal-to-noise ratios decrease as imaging depth increases due to laser scattering and absorption by tissue above the scanned depth. In this experiment, we show that RSCD can restore z-stack data, which contains SRH images acquired at sequentially deeper spatial locations in the tissue. z-stack images are volumetric and capture 3-D biological structures by imaging in all three spatial dimensions.

RSCD was used to restore z-stacked SRH images taken from 425 surgical specimens. FID scores were computed at each z-depth level for original and restored images with respect to the 4.5K high-quality SRH images. We show the results in Figure 8. FID scores for the original low-quality images steadily increase as imaging depth increases, indicating worse image quality with imaging depth. RSCD consistently reduces the FID score for deep SRH images.

Related Works

Several previous works share the similarity of using only the last part of the reverse diffusion generative process for image generation or restoration purposes (instead of requiring the full reverse diffusion) with our work. TDPM (Zheng et al. 2022) increases the efficiency of diffusion-based image generation by replacing the majority of the reverse diffusion

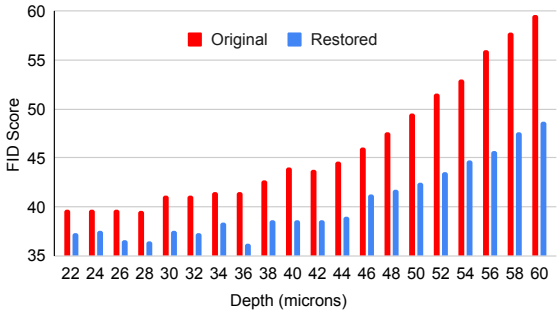

Figure 8: FID scores of original and restored z-stack images. Lower FID scores indicate better perceptual quality. RSCD consistently improves image quality at all depths.

steps with a GAN and only performing the last part of the reverse diffusion process. SDEdit (Meng et al. 2021) and CCDF (Chung, Sim, and Ye 2022) perform image editing and restoration by first adding a strong Gaussian noise to the input image, then performing a few reverse diffusion steps. Regularized Reverse Diffusion (RRD) (Chung, Lee, and Ye 2023) builds upon CCDF: it does not require noise addition, uses a non-parametric Gaussian noise level estimator (Chen, Zhu, and Heng 2015) to determine the number of reverse diffusion steps to perform, and regularizes the restoration process with Fourier features. In comparison, RSCD has the following advantages: (1) we designed and trained a highly effective step calibrator to determine the number of steps required, while TDPM, SDEdit, and CCDF set this as a fixed hyperparameter and RRD determines this with a non-parametric Gaussian noise level estimator, all of which does not work well when the input image has non-uniform noise of varying strengths (as evident by our ablation studies on fixed steps and nonparametric noise estimator); (2) Adding a large amount of noise at once to the input image (like SDEdit and CCDF) increases the risk of hallucination, which is unacceptable for medical images; instead, RSCD mitigates the noise distribution difference between real noise and Gaussian noise by gradually adding back a small amount of Gaussian noise during each DDPM step and dynamic recalibration.

Conclusion

We present Restorative Step-Calibrated Diffusion (RSCD), a reliable and efficient method to restore biomedical optical imaging without requiring paired high-quality data. RSCD outperforms other widely used unpaired image restoration methods on quantitative metrics; more importantly, experts in biomedical optical imaging consistently prefer images restored using RSCD in blinded comparison experiments and report minimal to no hallucinations. RSCD can improve model performance on downstream clinical tasks, including automated brain tumor diagnosis and deep tissue imaging. Our method reduces diagnostic errors that can have detrimental impacts on clinical care. This study demonstrates the potential of AI in improving automated clinical diagnostics and patient care in today's precision medicine landscape.

Ethical Statement

The SRH datasets used in this study include specimens from patients who underwent brain tumor biopsy or tumor resection. Patients were consecutively and prospectively enrolled at University of Michigan for intraoperative SRH imaging, and this study was approved by the Institutional Review Board (HUM#00083059). Informed consent was obtained for each patient prior to SRH imaging and the use of tumor specimens for research and development was approved.

Acknowledgements

We would like to thank Karen Eddy, Lin Wang, Hubert Zhang for their administrative support and data collection.

This work was supported by the following National Institute of Health (NIH) funding sources: K12NS080223 (T.C.H.), T32GM141746 (C.J.), F31NS135973 (C.J.). This work was support by the Chan Zuckerberg Foundation (CZI) Advancing Imaging Through Collaborative Project grant, Cook Family Brain Tumor Research Fund (T.C.H), the Mark Trauner Brain Research Fund, the Zenkel Family Foundation (T.C.H.), Ian's Friends Foundation (T.C.H.) and the UM Precision Health Investigators Awards grant program (T.C.H.).

This research was also supported, in part, through computational resources and services provided by Advanced Research Computing, a division of Information and Technology Services at the University of Michigan.

T.C.H. and C.F. are shareholders of Invenio Imaging, Inc., a company developing SRH microscopy systems.

References

Audier, X.; Heuke, S.; Volz, P.; Rimke, I.; and Rigneault, H. 2020. Noise in stimulated Raman scattering measurement: From basics to practice. *Apl Photonics*, 5(1).

Belthangady, C.; and Royer, L. A. 2019. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods*, 16(12): 1215–1225.

Chen, G.; Zhu, F.; and Heng, P. A. 2015. An Efficient Statistical Method for Image Noise Level Estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, 477–485.

Chung, H.; Lee, E. S.; and Ye, J. C. 2023. MR Image Denoising and Super-Resolution Using Regularized Reverse Diffusion. *IEEE Trans. Med. Imaging*, 42(4): 922–934.

Chung, H.; Sim, B.; and Ye, J. C. 2022. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12413–12422.

Chung, H.; and Ye, J. C. 2022. Score-based diffusion models for accelerated MRI. arXiv:2110.05243.

Dhariwal, P.; and Nichol, A. 2021. Diffusion Models Beat GANs on Image Synthesis.

Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; and Xie, X. S. 2008. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science*, 322(5909): 1857–1861.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6629–6640. Red Hook, NY, USA: Curran Associates Inc.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, number Article 574 in NIPS'20, 6840–6851. Red Hook, NY, USA: Curran Associates Inc.

Hoesli, R. C.; Orringer, D. A.; McHugh, J. B.; and Spector, M. E. 2017. Coherent Raman Scattering Microscopy for Evaluation of Head and Neck Carcinoma. *Otolaryngol. Head Neck Surg.*, 157(3): 448–453.

Hollon, T.; Jiang, C.; Chowdury, A.; Nasir-Moin, M.; Kondepudi, A.; Aabedi, A.; Adapa, A.; Al-Holou, W.; Heth, J.; Sagher, O.; Lowenstein, P.; Castro, M.; Wadiura, L. I.; Widhalm, G.; Neuschmelting, V.; Reinecke, D.; von Spreckelsen, N.; Berger, M. S.; Hervey-Jumper, S. L.; Golfinos, J. G.; Snuderl, M.; Camelo-Piragua, S.; Freudiger, C.; Lee, H.; and Orringer, D. A. 2023. Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. *Nat. Med.*

Hollon, T.; Lewis, S.; Freudiger, C. W.; Sunney Xie, X.; and Orringer, D. A. 2016. Improving the accuracy of brain tumor surgery via Raman-based technology. *Neurosurg. Focus*, 40(3): E9.

Hollon, T. C.; Lewis, S.; Pandian, B.; Niknafs, Y. S.; Garrard, M. R.; Garton, H.; Maher, C. O.; McFadden, K.; Snuderl, M.; Lieberman, A. P.; Muraszko, K.; Camelo-Piragua, S.; and Orringer, D. A. 2018. Rapid Intraoperative Diagnosis of Pediatric Brain Tumors Using Stimulated Raman Histology. *Cancer Res.*, 78(1): 278–289.

Hollon, T. C.; Pandian, B.; Adapa, A. R.; Urias, E.; Save, A. V.; Khalsa, S. S. S.; Eichberg, D. G.; D'Amico, R. S.; Farooq, Z. U.; Lewis, S.; Petridis, P. D.; Marie, T.; Shah, A. H.; Garton, H. J. L.; Maher, C. O.; Heth, J. A.; McKean, E. L.; Sullivan, S. E.; Hervey-Jumper, S. L.; Patil, P. G.; Thompson, B. G.; Sagher, O.; McKhann, G. M., 2nd; Komotar, R. J.; Ivan, M. E.; Snuderl, M.; Otten, M. L.; Johnson, T. D.; Sisti, M. B.; Bruce, J. N.; Muraszko, K. M.; Trautman, J.; Freudiger, C. W.; Canoll, P.; Lee, H.; Camelo-Piragua, S.; and Orringer, D. A. 2020. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.*

Jayasumana, S.; Ramalingam, S.; Veit, A.; Glasner, D.; Chakrabarti, A.; and Kumar, S. 2024. Rethinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9307–9315.

Jiang, C.; Chowdury, A.; Hou, X.; Kondepudi, A.; Freudiger, C. W.; Conway, K.; Camelo-Piragua, S.; Orringer, D. A.; Lee, H.; and Hollon, T. C. 2022. OpenSRH: optimizing brain tumor surgery using intraoperative stimulated Raman histology. *Adv. Neural Inf. Process. Syst.*, 35(DB): 28502–28516.

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising Diffusion Restoration Models.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2Noise: Learning Image Restoration without Clean Data. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 2965–2974. PMLR.

Lui, H.; Zhao, J.; McLean, D.; and Zeng, H. 2012. Real-time Raman spectroscopy for in vivo skin cancer diagnosis. *Cancer Res.*, 72(10): 2491–2500.

Manifold, B.; Thomas, E.; Francis, A. T.; Hill, A. H.; and Fu, D. 2019. Denoising of stimulated Raman scattering microscopy images via deep learning. *Biomed. Opt. Express*, 10(8): 3860–3874.

Mannas, M. P.; Jones, D.; Deng, F.-M.; Hoskoppal, D.; Melamed, J.; Orringer, D. A.; and Taneja, S. S. 2023. Stimulated Raman histology, a novel method to allow for rapid pathologic examination of unprocessed, fresh prostate biopsies. *Prostate*, 83(11): 1060–1067.

Meng, C.; He, Y.; Song, Y.; Song, J.; Wu, J.; Zhu, J.-Y.; and Ermon, S. 2021. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*.

Min, W.; Freudiger, C. W.; Lu, S.; and Xie, X. S. 2011. Coherent nonlinear optical imaging: beyond fluorescence microscopy. *Annual review of physical chemistry*, 62(1): 507–530.

Moester, M.; Ariese, F.; and De Boer, J. 2015. Optimized signal-to-noise ratio with shot noise limited detection in stimulated Raman scattering microscopy. *Journal of the European Optical Society-Rapid publications*, 10: 15022.

Orringer, D. A.; Pandian, B.; Niknafs, Y. S.; Hollon, T. C.; Boyle, J.; Lewis, S.; Garrard, M.; Hervey-Jumper, S. L.; Garton, H. J. L.; Maher, C. O.; Heth, J. A.; Sagher, O.; Wilkinson, D. A.; Snuderl, M.; Venneti, S.; Ramkissoon, S. H.; McFadden, K. A.; Fisher-Hubbard, A.; Lieberman, A. P.; Johnson, T. D.; Xie, X. S.; Trautman, J. K.; Freudiger, C. W.; and Camelo-Piragua, S. 2017. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat Biomed Eng*, 1.

Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.; Ghasemipour, S. K. S.; Gontijo-Lopes, R.; Ayan, B. K.; Salimans, T.; Ho, J.; Fleet, D. J.; and Norouzi, M. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization.

Wang, Z.; and Bovik, A. C. 2009. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, 26(1): 98–117.

Waterhouse, D. J.; Fitzpatrick, C. R. M.; Pogue, B. W.; O'Connor, J. P. B.; and Bohndiek, S. E. 2019. A roadmap for the clinical implementation of optical-imaging biomarkers. *Nat Biomed Eng*, 3(5): 339–353.

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. 2018. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods*, 15(12): 1090–1097.

Zheng, H.; He, P.; Chen, W.; and Zhou, M. 2022. Truncated diffusion probabilistic models and diffusion-based adversarial auto-encoders. *arXiv preprint arXiv:2202.09671*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.