UCLA
UCLA Previously Published Works

Title

Data-Efficient Prostate Micro-Ultrasound Image Analysis using a Modality-specific Foundation Model.

Permalink

https://escholarship.org/uc/item/1xd9d517

Journal

medRxiv

Authors

Pang, Kaifeng
Fan, Junxin
Hung, Alex Ling Yu
et al.

Publication Date

2026-10-05

DOI

10.64898/2026.09.29.26364330

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Highlights

Micro-Ultrasound Super-Resolution with Geometry-Driven Consistency Models

Kaifeng Pang,Kai Zhao,Qi Miao,Alex Ling Yu Hung,Changsuk Oh,Raymi Ramirez,Qiudi He,Jordan Klein,Wei Shao,Wayne Brisbane,Kyunghyun Sung

- We propose UltraCCM, a geometry-driven consistency model for efficient super-resolution of reformatted micro-ultrasound images.
- UltraCCM performs angular super-resolution in the polar coordinate system with self-supervised training using simulated data, eliminating the need for high-resolution target-plane supervision.
- The method achieves superior image quality with single-step inference, validated on *in vivo*, *ex vivo*, and external datasets, as well as downstream prostate segmentation and expert reader studies.

Micro-Ultrasound Super-Resolution with Geometry-Driven Consistency Models

Kaifeng Pang[a,b], Kai Zhao[b,f,*], Qi Miao[b], Alex Ling Yu Hung[b,c], Changsuk Oh[b], Raymi Ramirez[b], Qiudi He[a,b], Jordan Klein[g], Wei Shao[h], Wayne Brisbane[e] and Kyunghyun Sung[b,d]

[a]Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095, United States
[b]Department of Radiological Sciences, University of California, Los Angeles, CA, 90095, United States
[c]Department of Computer Science, University of California, Los Angeles, CA, 90095, United States
[d]Department of Bioengineering, University of California, Los Angeles, CA, 90095, United States
[e]Department of Urology, University of California, Los Angeles, CA, 90095, United States
[f]School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China
[g]University of Wisconsin-Madison, WI, 53706, United States
[h]Department of Medicine, University of Florida, FL, 32611, United States

ARTICLE INFO

Keywords:
Micro-ultrasound
Consistency models
Super-resolution
Multi-planar reformation
Prostate

ABSTRACT

Prostate cancer diagnosis typically relies on image-guided biopsy. Micro-ultrasound (MicroUS) has recently emerged as a promising imaging modality, offering high spatial resolution with comparable diagnostic performance to multi-parametric MRI (mpMRI) at a lower cost. However, unlike MRI, MicroUS volumes are acquired in a fan-shaped, angular geometry, making direct alignment and comparison across imaging modalities challenging. High-quality multi-planar reformation (MPR) is therefore expected in treatment planning, yet conventional MPR produces blurred images with slice discontinuities. In this paper, we present **UltraCCM**, a fully self-supervised, geometry-driven conditional consistency model for super-resolution of reformatted MicroUS images. UltraCCM explicitly incorporates the geometric characteristics of MicroUS acquisition and formulates MPR as an angular super-resolution problem, enabling fast, single-step recovery of fine anatomical details without requiring high-resolution target-plane supervision. Extensive experiments on *in vivo* and *ex vivo* datasets, including quantitative evaluation, expert reader studies, and a downstream prostate segmentation task, demonstrate that UltraCCM improves perceptual quality and fine-detail visualization while preserving anatomy-relevant information for downstream image analysis. The code of UltraCCM is available at: https://github.com/Calvin-Pang/UltraCCM.

1. Introduction

Prostate imaging has undergone a renaissance over the last decade with the wide, standard implementation of multi-parametric MRI (Stabile et al., 2020), the introduction of PSMA PET scans (Maurer et al., 2016), and the leverage of artificial intelligence to interpret imaging (Zheng et al., 2024). Micro-ultrasound (MicroUS), a 29-MHz ultrasound technology, is an emerging, cost-effective prostate imaging technology that provides 3-4 times higher spatial resolution than conventional ultrasound (Pensa et al., 2024). Recent studies demonstrate that MicroUS has the potential to achieve diagnostic accuracy comparable to that of mpMRI, with a significantly lower cost (Pensa et al., 2024; Klotz et al., 2020; Kinnaird et al., 2025; Miao et al., 2026). Its high resolution and real-time imaging capabilities make it well-suited for guiding prostate biopsy and early diagnosis, reducing a common source of registration errors in traditional MRI-ultrasound fusion-guided biopsy (Williams et al., 2022). However, MicroUS differs fundamentally from MRI, which is typically based on multi-slice volumetric acquisitions. As shown in Fig 1(a), MicroUS slices are acquired in a fan-shaped (sector) geometry through manual probe rotation, resulting in non-uniform angular spacing between frames. To enable clinical comparison with MRI, multi-planar reformation (MPR) can be applied to reformat MicroUS volumes into different imaging orientations (Pang et al., 2025a). Commercial MicroUS systems, such as ExactVu (Exact Imaging, n.d.a), provide elastic MRI-MicroUS fusion at the millimeter scale with the built-in MPR function ("Transverse Mode") (Exact Imaging, n.d.b), which is conceptually similar to interpolation-based reformation (Pensa et al., 2024). However, accurate pixel-level comparison remains challenging because MicroUS acquisition is limited by sparse angular sampling and the fan-shaped geometry. Sparse angular sampling produces large gaps between acquired slices, leading to inter-slice discontinuities, while the fan-shaped geometry results in progressively lower anterior spatial resolution and interpolation-induced blurring (Fig 1(b)) (Pang et al., 2025a). Recent studies have

*Corresponding author
kaifengpang@mednet.ucla.edu (K. Pang); kz@kaizhao.net (K. Zhao); qmiao@mednet.ucla.edu (Q. Miao); alexhung96@g.ucla.edu (A.L.Y. Hung); changsukoh@mednet.ucla.edu (C. Oh); raymiramirez@mednet.ucla.edu (R. Ramirez); qiudihe@mednet.ucla.edu (Q. He); jordanklein1009@gmail.com (J. Klein); wei.shao@medicine.ufl.edu (W. Shao); wbrisbane@mednet.ucla.edu (W. Brisbane); ksung@mednet.ucla.edu (K. Sung)
ORCID(s): 0009-0001-7613-6559 (K. Pang); 0000-0002-2496-0829 (K. Zhao); 0000-0003-2235-8997 (Q. Miao); 0000-0002-2664-7430 (A.L.Y. Hung); 0000-0002-7665-0451 (C. Oh); 0000-0001-7458-6006 (R. Ramirez); 0000-0003-4931-4839 (W. Shao); 0000-0003-0470-5262 (W. Brisbane); 0000-0003-4175-5322 (K. Sung)

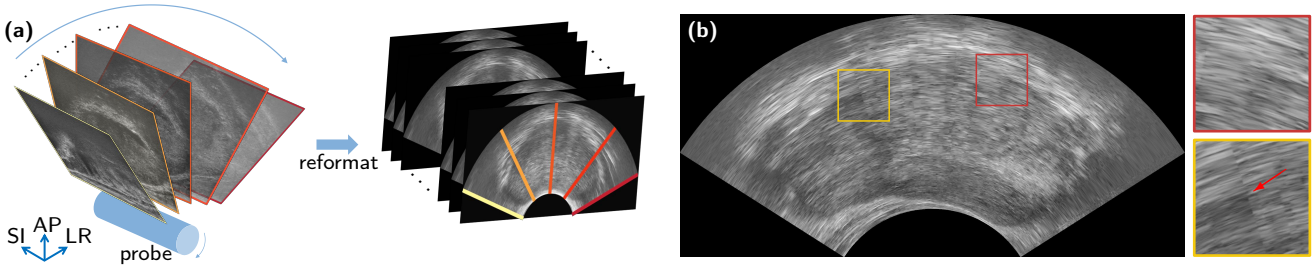


Figure 1: Example of MicroUS acquisition, reformation, and interpolation artifacts. (a) MicroUS acquires pseudo-sagittal slices using a rotating probe, which are assembled into sector-shaped matrices and reformatted into the axial plane. Anatomical labels (SI: superior–inferior, AP: anterior–posterior, LR: left–right) indicate physical directions and are unrelated to the low-resolution (LR) notation used elsewhere in the paper. The schematic represents the simplified acquisition geometry rather than the exact mechanical geometry of the ExactVu probe. (b) Conventional interpolation-based MPR introduces blurring (red box) and slice discontinuity artifacts (yellow box).

explored post-MPR processing to enhance MicroUS volumes by predicting intermediate slices along the acquisition trajectory to reconstruct continuous MicroUS volumes, e.g., using implicit neural representations (Chojnowski et al., 2026). Nevertheless, recovering high-resolution reformatted images with faithful fine-grained anatomical details remains challenging.

Deep learning (DL)-based super-resolution (SR) has been widely explored for enhancing the spatial resolution of medical images (Zhao et al., 2024; Pang et al., 2025b), demonstrating strong potential for improving the image quality of MicroUS images. Among them, generative models, such as Generative Adversarial Networks (GANs) (Ledig et al., 2017) have demonstrated a strong capability in recovering fine-grained image details, while often suffering from training instability (Arjovsky et al., 2017). Diffusion-based models have emerged as a powerful approach due to their high-fidelity generation and stable optimization (Saharia et al., 2022; Zhao et al., 2024), and have been previously applied to post-MPR MicroUS SR (Pang et al., 2025a). However, diffusion model inference relies on many iterative denoising steps, leading to high computational cost and limiting fast volumetric deployment (Yue et al., 2023; Zhao et al., 2024). Consistency Models (CMs) offer a more efficient solution by enforcing generation consistency across noise scales, enabling high-quality image generation in a single inference step (Song et al., 2023).

Despite this progress, applying DL-based SR to MicroUS images remains challenging due to three fundamental factors (Pang et al., 2025a). First, each MicroUS study contains hundreds of slices, making iterative generative models (e.g., diffusion models) computationally prohibitive for efficient volumetric enhancement. Second, the non-uniform angular sampling violates the regular-grid assumptions of standard neural networks. Third, high-quality target-plane images are unavailable, making conventional supervised SR training infeasible.

To address these challenges, we propose **UltraCCM**, a geometry-driven conditional consistency model for accelerated super-resolution of multi-planar reformatted MicroUS images. To leverage the geometric characteristics of MicroUS acquisition, i.e., the non-uniform sector pixel layout, UltraCCM formulates the task as an angular SR problem in polar space and incorporates polar-coordinate positional encoding to provide geometry-aware positional guidance for handling non-uniform sampling. A fully self-supervised training strategy is introduced by extracting paired LR–HR sector patches from the acquisition plane, eliminating the need for target-plane supervision. Moreover, since diffusion-style generative objectives do not explicitly constrain global intensity statistics and may introduce intensity shifts (Song and Ermon, 2020), we incorporate an intensity distribution alignment regularization. This alignment is implemented via kernel density estimation (KDE) to enable differentiable comparison of discrete-valued intensity distributions. Once trained, UltraCCM enables SR in under one second per image, achieving significant acceleration over existing generative methods. To the best of our knowledge, our method is the first to enable single-step, high-quality MicroUS SR.

The contributions of this work are summarized as follows:

- We introduce a novel SR framework for post-MPR MicroUS images based on consistency models, which explicitly incorporates the geometric characteristics of MicroUS acquisitions by conducting angular super-resolution directly in the polar coordinate space.
- We propose a self-supervised training strategy that eliminates the need for HR target plane images, enabling effective learning.
- We introduce an intensity distribution alignment regularization to mitigate brightness and contrast drift of consistency models.
- We demonstrate superior image quality and fast inference capability through comprehensive *in vivo*, *ex vivo*, and downstream task evaluations.

2. Related Work

2.1. Micro-Ultrasound Imaging

MicroUS has recently emerged as a novel imaging modality for prostate biopsy guidance and prostate cancer diagnosis, offering substantially higher spatial resolution (∼70 μm at 29 MHz) than conventional transrectal ultrasound (Pensa et al., 2024). Multiple preliminary studies have shown its potential diagnostic value (Klotz et al., 2020; Pensa et al., 2021; Dias et al., 2022; Pensa et al., 2024). These efforts motivated the international multi-institutional OPTIMUM trial (Kinnaird et al., 2025), which randomized 802 patients to conventional MRI/US fusion versus MicroUS-guided biopsy. OPTIMUM demonstrated non-inferiority of MicroUS for detecting Grade Group $\geq$ 2 prostate cancer, consolidating its role as a potential alternative to MRI guidance.

Recent efforts have also applied deep learning to MicroUS image analysis. Jiang et al. (Jiang et al., 2024) proposed a multiscale annotation-guided transformer UNet model designed for MicroUS prostate segmentation. Gilany et al. (Gilany et al., 2023) designed TRUSformer, a context-aware model for prostate cancer detection in MicroUS images. Wilson et al. (Wilson et al., 2023) leveraged self-supervised representation learning on unlabeled MicroUS data to improve prostate cancer tissue detection. Despite its clinical potential, MicroUS image quality remains limited due to sparse angular sampling and degraded anterior resolution, while relatively little work has focused on MicroUS image quality enhancement. Recent learning-based approaches have begun to address this problem, including implicit neural representations (INR) for intermediate slice prediction (Chojnowski et al., 2026) and diffusion-based models for post-MPR super-resolution (Pang et al., 2025a). Nevertheless, INR-based methods require computationally expensive volumetric inference (Chojnowski et al., 2026), and may have limited ability to recover fine-grained anatomical details under sparse and non-uniform angular sampling. Diffusion-based approaches often rely on expensive iterative sampling (Pang et al., 2025a).

2.2. Image Super-Resolution

In image processing, a variety of classical image SR methods have been developed, including edge-based approaches (Sun et al., 2008; Freedman and Fattal, 2011) and sparse-coding-based approaches (Yang et al., 2008, 2010). With the development of deep learning, CNN- and transformer-based SR models (Liang et al., 2021; Li et al., 2024b) have become widely used. In medical imaging, learning-based SR methods have been increasingly explored to enhance image resolution (Li et al., 2025; Pang et al., 2025b).

Recently, deep generative models, e.g., Generative Adversarial Networks (GANs) (Goodfellow et al., 2020), Variational Autoencoders (VAEs) (Kingma and Welling, 2013), Diffusion Models (Ho et al., 2020; Song et al., 2020), Consistency Models (Song et al., 2023) have shown remarkable capabilities in producing realistic images, subsequently leading to their application in SR tasks. Ledig et al. (Ledig et al., 2017) proposed a GAN-based model to restore photo-realistic details for SR results. Li et al. (Li et al., 2024a) incorporated semantic priors from pretrained vision foundation models to enhance GAN-based SR with fine-grained, semantic-aware details. Diffusion-based approaches, including SR3 (Saharia et al., 2022), progressively generate SR outputs conditioned on LR inputs. To accelerate the iterative sampling process of diffusion models, Yue et al. introduced ResShift (Yue et al., 2023), an efficient SR diffusion model that reduces the number of diffusion steps by constructing a residual shifting Markov chain between HR images and LR images. In medical imaging, Hung et al.(Hung et al., 2023) developed a conditional diffusion framework for various medical image generation tasks, including SR. Zhao et al. (Zhao et al., 2024) proposed a partial diffusion model using LR latents to approximate the HR latents during the diffusion trajectory for fast MRI image SR. Liu et al. (Liu et al., 2024) introduced a wavelet-based diffusion model with multi-layer self-attention for ultrasound SR.

Although diffusion models have achieved state-of-the-art performance in SR and recent efforts such as ResShift (Yue et al., 2023) and PDM (Zhao et al., 2024) improve efficiency compared with conventional diffusion-based SR, they still require multiple iterative steps and hinder the volumetric real-time SR for MicroUS images. In contrast, consistency models (Song et al., 2023) generate high-quality results in a single inference step, making them particularly well-suited for our framework.

3. Method

3.1. Overview

As illustrated in Fig 2, UltraCCM is trained using paired HR-LR patches extracted from the acquired MicroUS slices and, once trained, is applied to enhance reformatted target-plane images during inference.

In clinical practice, a MicroUS volume is acquired through a rotational sweep of the prostate, although the actual trajectory may be affected by residual probe translation or tilt, pressure-induced tissue deformation, and patient motion. Following previous work (Pensa et al., 2024; Jiang et al., 2024; Zhou et al., 2025), we adopt a simplified fan-shaped model of the dominant rotational geometry and its associated angular sampling, while not explicitly accounting for these pose and deformation effects. Each volume consists of N image slices of size $H \times W$, denoted as $\{\boldsymbol{S}^p \in \mathbb{R}^{H\times W}, p = 1, ..., N\}$, where each image is associated with an acquisition angle α_p. Given the sector acquisition geometry, the volume can be initially reformatted to polar coordinate space as a set of W sector matrices: $\{\boldsymbol{M}^q \in \mathbb{R}^{H\times N}, q = 1, ..., W\}$. Each matrix $\boldsymbol{M}^q$ corresponds to the q-th column across all N slices, forming N radial lines at angular positions $\{\alpha_p\}$, each of length H. The pixel correspondence in this initial reformation can be expressed as $\boldsymbol{M}^q(i, j) = \boldsymbol{S}^j(i, q)$, where (i, j) represents the radial and angular pixel position in polar space.

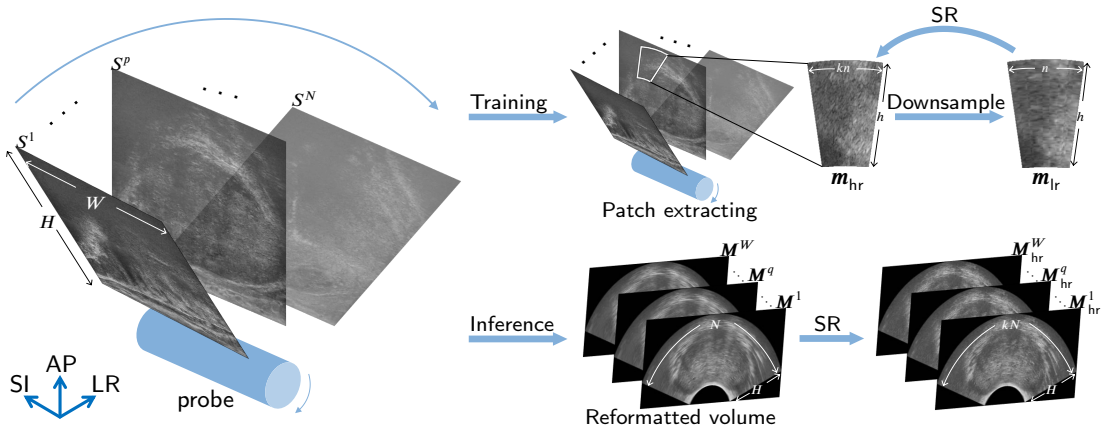


Figure 2: Training and inference workflow of UltraCCM. The schematic follows the simplified acquisition geometry described in Section 3.1. *Training:* Paired HR–LR patches are extracted from the acquired slices to supervise the model training. *Inference:* Given reformatted MicroUS slices, the trained model performs SR to enhance the resolution.

The traditional MicroUS MPR approach performs direct bilinear interpolation on $\boldsymbol{M}^q$ to obtain the corresponding Cartesian-grid image $\boldsymbol{I}^q_{\text{lr}}$ (Pensa et al., 2024). However, due to the sparsity and non-uniform nature of the angular dimension, this process often introduces severe artifacts in the reformatted images (Fig 1). Thus, we propose an angular super-resolution approach based on a conditional consistency model to enhance the image quality of the reformatted images. Specifically, given an upsampling factor k, the model first generates a high-resolution sector matrix $\boldsymbol{M}^q_{\text{hr}} \in \mathbb{R}^{H\times kN}$, where the columns correspond to linearly spaced angular positions. This high-resolution representation is subsequently mapped into the Cartesian grid to obtain the final reformatted image $\boldsymbol{I}^q_{\text{hr}}$.

Throughout the remainder of this paper, we refer to matrices like $\boldsymbol{M}^q$ as **polar images**, where pixels are arranged according to angular and radial locations in a polar coordinate system. The corresponding **Cartesian images**, $\boldsymbol{I}^q_{\text{lr}}$, refer to the version of polar images reprojected onto a Cartesian equidistant (x, y) grid through coordinate transformation and interpolation. **For clarity, in all figures in this paper, polar images are interpolated as Cartesian images for better visualization.**

3.2. Self-Supervised Training

Due to the lack of ground-truth HR polar images, it is not feasible to perform supervised training with explicit LR and HR pairs. However, MicroUS image characteristics are direction-dependent, with differences in spatial resolution and filtering across imaging directions. The acquired images $\{\boldsymbol{S}\}$ in the original acquisition plane have relatively high native spatial resolution, whereas the resolution in the reformatted target plane is primarily limited by sparse sampling along the angular direction and the fan-shaped geometry. We exploit this directional disparity by synthesizing LR-HR polar image patch pairs from the original plane, thereby transferring their high-resolution image statistics to the lower-resolution target plane through self-supervised training, as shown in Fig 2. We design a polar grid generation scheme with random angular sampling, detailed in Algorithm 1, which approximates the dominant sector pixel layout in real reformatted images $\{\boldsymbol{M}\}$ and angular sampling under the simplified geometry described in Sec 3.1. In this scheme, we also account for the radial offset between the assumed rotational axis and the first image row.

Algorithm 1 Polar Patch Sampling

Require: Acquired HR slices set $\{\boldsymbol{S}\}$; full polar image height H; LR patch size $h \times n$; angular density range $[N_{\min}, N_{\max}]$; upsampling factor k; rotational axis-to-image radial offset r_{offset}; pixel spacing ρ

repeat

Sample an image $\boldsymbol{S} \sim \{\boldsymbol{S}\}$, and number of LR angular coordinates $N \sim [N_{min}, N_{max}]$

Random angular sampling: Sample N angular coordinates $\{\alpha_j\}_{j=1}^{N} \sim \mathcal{U}\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, sorted in ascending order

Construct radial coordinates $\{r_i\}_{i=1}^{H}$ as:

$$r_i = \frac{\frac{r_{\text{offset}}}{\rho} + i}{\frac{r_{\text{offset}}}{\rho} + H}, \quad i = 1, \ldots, H$$

Sample $v \sim [1, N - n + 1]$ and $u \sim [1, H - h + 1]$; extract n angular coordinates $\{\alpha_j\}_{j=v}^{v+n-1}$ and h radial coordinates $\{r_i\}_{i=u}^{u+h-1}$

Construct LR coordinate map $\boldsymbol{\Pi}_{\text{lr}} \in \mathbb{R}^{h\times n\times 2}$ via meshgrid over $\{r_i\}_{i=u}^{u+h-1}$ and $\{\alpha_j\}_{j=v}^{v+n-1}$

Construct HR angular coordinates $\{\tilde{\alpha}_j\}_{j=1}^{kn}$ as kn linearly spaced values between α_v and α_{v+n-1};

Construct HR coordinate map $\boldsymbol{\Pi}_{\text{hr}} \in \mathbb{R}^{h\times kn\times 2}$ via meshgrid over $\{r_i\}_{i=u}^{u+h-1}$ and $\{\tilde{\alpha}_j\}_{j=1}^{kn}$

Extract LR patch $\boldsymbol{m}_{\text{lr}} \in \mathbb{R}^{h\times n}$ and HR patch $\boldsymbol{m}_{\text{hr}} \in \mathbb{R}^{h\times kn}$ from $\boldsymbol{S}$ using interpolation with $\boldsymbol{\Pi}_{\text{lr}}$ and $\boldsymbol{\Pi}_{\text{hr}}$

until Training convergence

The procedure begins by constructing normalized radial and angular coordinates for the full LR polar image layout. A variable number of angular slices, $N \sim [N_{min}, N_{max}]$, is sampled to reflect acquisition density variability across different scans. Within this coordinate domain, LR radial and angular coordinates are randomly extracted to form an LR coordinate map of size $h \times n \times 2$. The corresponding HR angular coordinates are then generated as kn linearly spaced values between the LR angular endpoints, forming an HR coordinate map of size $h \times kn \times 2$. The resulting coordinate maps are used to extract paired LR and HR polar image patches from the original HR images $\{\boldsymbol{S}\}$ via interpolation.

During training, the extracted LR and HR polar patches, $(\boldsymbol{m}_{\text{lr}}, \boldsymbol{m}_{\text{hr}})$ are used to supervise the model. During inference, the model directly produces a super-resolved HR polar image $\boldsymbol{M}_{\text{hr}} \in \mathbb{R}^{H \times kN}$ from the full LR polar image $\boldsymbol{M}_{\text{lr}} \in \mathbb{R}^{H \times N}$.

3.3. Geometry-driven Consistency Models

3.3.1. Background of Consistency Models

Consistency models (CMs) (Song et al., 2023) build on the theoretical foundations of diffusion models (Ho et al., 2020), i.e., score-based generative models (Song et al., 2020). In diffusion models, a clean image $\boldsymbol{x}_0 \sim p_{\text{data}}$ is gradually corrupted into noisy states $\boldsymbol{x}_t \sim p_t$ by Gaussian noise. The reverse process learns a denoising dynamics that iteratively maps $\boldsymbol{x}_t$ back to $\boldsymbol{x}_0$. Specifically, under the Variance Exploding (VE) formulation (Karras et al., 2022), this reverse process follows an empirical probability-flow ordinary differential equation (PF-ODE): $\frac{d\boldsymbol{x}_t}{dt} = -t\boldsymbol{s}_\phi(\boldsymbol{x}_t, t)$, where $\boldsymbol{s}_\phi$ is a neural network trained to approximate the gradient of the log-density of p_t.

A consistency model (CM) (Song et al., 2023) replaces the iterative generation of the PF-ODE with a single-step function $\boldsymbol{f}_\theta(\boldsymbol{x}_t, t)$ that directly predicts the clean image $\boldsymbol{x}_0$ from any intermediate noisy sample $\boldsymbol{x}_t$:

$$\boldsymbol{f}_\theta(\boldsymbol{x}_t, t) = \boldsymbol{x}_0, \quad t \in [0, T] \tag{1}$$

Once trained, CM enables single-step generation of clean images from pure Gaussian noise inputs $\boldsymbol{x}_T$. To ensure valid generative behavior along the PF-ODE trajectory, CM must satisfy two properties: i) *self-consistency*, meaning predictions from any two points on the trajectory must agree, $\boldsymbol{f}_\theta(\boldsymbol{x}_t, t) = \boldsymbol{f}_\theta(\boldsymbol{x}_{t'}, t')$; and ii) the *boundary condition*, $\boldsymbol{f}_\theta(\boldsymbol{x}_0, 0) = \boldsymbol{x}_0$.

3.3.2. Conditional Formulation

Fig 3(a) illustrates the UltraCCM framework. To recover the HR image $\boldsymbol{M}_{\text{hr}}$ from its LR counterpart $\boldsymbol{M}_{\text{lr}}$, we propose a conditional CM by incorporating two conditions to guide the generation.

(i) **Anatomical condition.** Motivated by prior work (Saharia et al., 2022; Hung et al., 2023), we introduce the LR polar image as a condition to preserve the anatomical structure. Specifically, $\boldsymbol{M}_{\text{lr}}$ is first interpolated to HR resolution, denoted as $\tilde{\boldsymbol{M}}_{\text{lr}}$. $\tilde{\boldsymbol{M}}_{\text{lr}}$ is then concatenated with the noisy input $\boldsymbol{x}_t$ and fed into the model.

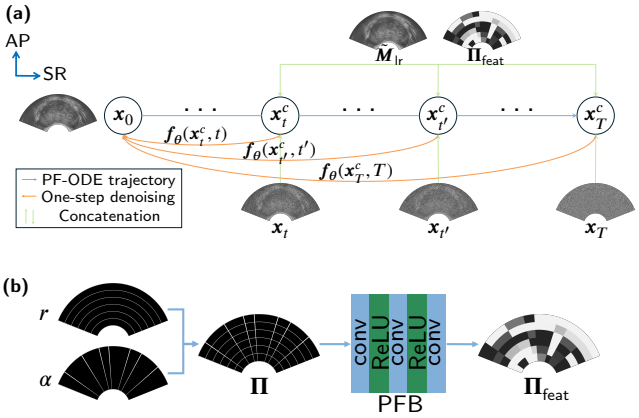


Figure 3: Overview of UltraCCM and positional conditioning. (a) UltraCCM learns a consistency mapping along the PF-ODE trajectory to generate an SR image $\boldsymbol{x}_0$ in a single step, conditioned on the interpolated LR image $\tilde{\boldsymbol{M}}_{\text{lr}}$ and positional features $\boldsymbol{\Pi}_{\text{feat}}$. (b) The Positional Feature Block (PFB) encodes polar coordinates $\boldsymbol{\Pi}$ into a compact positional feature map $\boldsymbol{\Pi}_{\text{feat}}$ for geometry-aware conditioning.

(ii) **Positional encoding.** As discussed in Sec 1, a key geometric characteristic of MicroUS is its non-uniform sector pixel layout, where pixel spacing increases with depth and angular sampling varies due to manual probe sweep of clinicians. Such irregularity violates the regular grid assumptions of standard neural networks. To address this, we design a geometry-aware positional encoding scheme that injects pixel-wise spatial information as an additional conditioning input to the model, as illustrated in Fig 3(b).

Specifically, given an LR polar image $\boldsymbol{M}_{\text{lr}} \in \mathbb{R}^{H \times N}$, each column $\boldsymbol{M}_{\text{lr}}(*, j)$ corresponds to an acquired slice with angular position α_j and each row $\boldsymbol{M}_{\text{lr}}(i, *)$, shares a radial distance r_i. Inspired by relative positional encoding in sequence models (Shaw et al., 2018), we represent angular positions using relative offsets rather than absolute angles to avoid overfitting to scan-specific angle values and to better capture local angular sampling density. To enable the computation of relative offsets at the boundaries, we pad the angular sequence $\{\alpha\}$ such that the first and last angles have well-defined left and right neighbors. Specifically, for each angular index j, we compute the left and right angular residuals as:

$$\Delta\alpha_j^- = \alpha_j - \alpha_{j-1}, \quad \Delta\alpha_j^+ = \alpha_{j+1} - \alpha_j. \tag{2}$$

This results in a 3-channel polar coordinate map $\boldsymbol{\Pi} \in \mathbb{R}^{H \times N \times 3}$, defined as:

$$\boldsymbol{\Pi}(i, j) = \begin{bmatrix} r_i & \Delta\alpha_j^- & \Delta\alpha_j^+ \end{bmatrix}^\top, i \in [1, H],\ j \in [1, N], \tag{3}$$

which encodes radial positions and angular contexts. During the model feedforward, $\boldsymbol{\Pi}$ is first interpolated to HR resolution, yielding $\tilde{\boldsymbol{\Pi}} \in \mathbb{R}^{H \times kN \times 3}$. The latter is passed through a lightweight positional feature block (PFB) with three 3×3 convolutions and ReLU activations, producing a positional feature map $\boldsymbol{\Pi}_{\text{feat}}$.

The final model input concatenates the noisy sample $\boldsymbol{x}_t$, anatomical condition $\tilde{\boldsymbol{M}}_{\text{lr}}$, and positional feature $\boldsymbol{\Pi}_{\text{feat}}$:

$$\boldsymbol{x}_t^c = [\boldsymbol{x}_t,\ \tilde{\boldsymbol{M}}_{\text{lr}},\ \boldsymbol{\Pi}_{\text{feat}}], \tag{4}$$

and the geometry-aware conditional CM is defined as

$$\boldsymbol{f}(\boldsymbol{x}_t^c, t) = \boldsymbol{x}_0, \quad t \in [0, T]. \tag{5}$$

3.4. Loss Function

3.4.1. Consistency Loss

As proposed in (Song et al., 2023), CM can be trained from scratch by enforcing self-consistency between two adjacent diffusion steps. A student network $\boldsymbol{\theta}$ predicts the output at t_{n+1}, while a teacher network $\boldsymbol{\theta}^-$, updated via exponential moving average (EMA), provides the prediction output at t_n. The consistency loss can be formulated as:

$$\mathcal{L}_{\text{CM}} = \mathrm{d}\big(\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}^c_{t_{n+1}}, t_{n+1}),\ \boldsymbol{f}_{\boldsymbol{\theta}^-}(\boldsymbol{x}^c_{t_n}, t_n)\big), \tag{6}$$

where $\mathrm{d}(\cdot,\cdot)$ denotes the perceptual distance measure.

3.4.2. Intensity Distribution Alignment

In medical imaging, preserving the intensity distribution is essential, as intensity distributions often reflect clinically relevant tissue characteristics (Masoudi et al., 2021). To compensate for the lack of explicit intensity constraints in the diffusion-style objective (Song and Ermon, 2020), we introduce a differentiable kernel-density-estimation (KDE) based Intensity Distribution Alignment (IDA) regularization that aligns the pixel intensity distributions of the SR output and the LR input.

Specifically, given an image $\boldsymbol{I} \in \mathbb{R}^{H\times W}$, we treat its pixel values as samples and define the KDE-based marginal intensity distribution as:

$$\hat{p}_{\boldsymbol{I}}(y) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} K_h(y - \boldsymbol{I}_{ij}), \tag{7}$$

where y denotes an intensity query point, $I_{i,j}$ is the pixel intensity at location (i, j), and $K_h(\cdot)$ is a Gaussian kernel with bandwidth h, defined as:

$$K_h(\delta) = \frac{1}{\sqrt{2\pi}h}\exp\left(-\frac{\delta^2}{2h^2}\right). \tag{8}$$

The regularization term is then calculated with the Kullback–Leibler (KL) divergence between the KDE-estimated distributions of the SR output and LR input:

$$\mathcal{L}_{\text{IDA}} = \mathrm{KL}\left(\hat{p}_{\boldsymbol{M}_{\text{sr}}}(\boldsymbol{y}) \parallel \hat{p}_{\tilde{\boldsymbol{M}}_{\text{lr}}}(\boldsymbol{y})\right), \tag{9}$$

where $\boldsymbol{M}_{\text{sr}}$ denotes the output generated by the student network in Eq (6), i.e., $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}^c_{t_{n+1}}, t_{n+1})$.

The final loss function used to optimize UltraCCM combines the CM loss $\mathcal{L}_{\text{CM}}$ and the IDA regularization term $\mathcal{L}_{\text{IDA}}$:

$$\mathcal{L} = \mathcal{L}_{\text{CM}} + \gamma\mathcal{L}_{\text{IDA}} \tag{10}$$

where γ is a tunable hyperparameter that balances the two objectives.

4. Experiments

In this section, we describe the implementation details of our experiments and present the results. Our experimental study was performed in compliance with the United States Health Insurance Portability and Accountability Act (HIPAA) of 1996 and was approved by the institutional review board (IRB) with a waiver of the requirement for informed consent.

4.1. Dataset

Our experiments were conducted on both *in vivo* and *ex vivo* prostate MicroUS datasets. The *in vivo* dataset reflects realistic clinical scenarios, while the *ex vivo* dataset includes HR reference images for paired quantitative evaluation. The internal *in vivo* and *ex vivo* datasets were collected at a single academic institution using an ExactVu MicroUS system (Exact Imaging, Markham, Ontario). In all experiments, we used the standard vendor-reconstructed B-mode images without additional probe-specific geometric correction.

4.1.1. in vivo Dataset

A total of 37 clinical MicroUS studies (February 2023 – September 2024) were retrospectively collected. Seven studies (1,716 slices) were used for training, and the remaining 30 for SR inference in the reformatted plane. Each scan contains 150–300 slices spanning 109.3°–185.4°, stored as DICOM images with 70 μm in-plane resolution and an original size of 962 × 1372 pixels. A fixed radial offset of 15 mm exists between the rotational axis of the probe and the first row of the MicroUS image.

4.1.2. ex vivo Dataset

Fourteen *ex vivo* prostate specimens (October 2021 – May 2022) were scanned using a custom MicroUS scanning system described in (Pensa et al., 2021), with 7 of them assigned for training and 7 for testing. An ExactVu probe mounted on a motorized translation stage (X-USBDC, Zaber Technologies Inc., British Columbia, Canada) translated laterally at 0.65 mm/s inside a sealed 3D-printed imaging tank (Lulzbot Taz 6, Fargo Additive Manufacturing Equipment 3D LLC, USA) designed for patient-specific molds. Each specimen was placed in a mold to preserve axial anatomical orientation. Images were acquired every 70 μm, producing densely sampled stacks of 902–960 slices per case (original size 962 × 1372). To simulate LR volumes, we uniformly downsampled all cases by selecting every k-th slice. The selected slices from the 7 training specimens were used for training, resulting in a total of 638 slices. For the test cases, the original densely sampled scans are reformatted as the HR target-plane images, while the downsampled scans are similarly reformatted as the corresponding LR images.

4.2. Experimental Setup

Our UltraCCM model is implemented with a 2D U-Net-based encoder-decoder backbone (Ronneberger et al., 2015), with additional inputs for time-step embeddings, LR conditioning, and positional features. The encoder comprises four

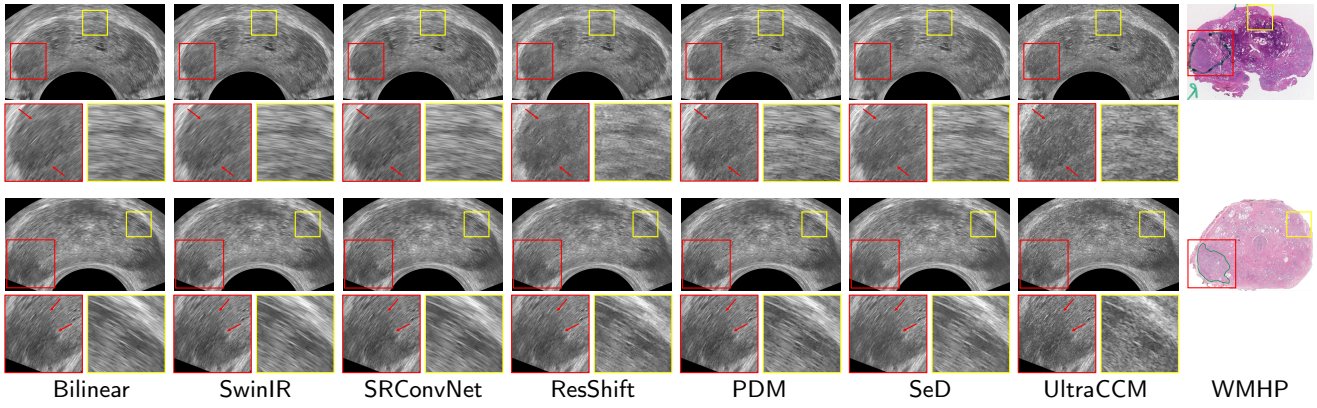


Figure 4: Qualitative comparison on the *in vivo* data. Lesions are delineated by green contours in the whole-mount histopathology (WMH) images and highlighted with red boxes in the MicroUS images. Key differences in lesion boundaries are further indicated by red arrows. Yellow boxes highlight regions with pronounced slice discontinuities and fine-textural degradation in the original LR images, where differences in SR quality among different methods are compared.

downsampling stages with a base feature channel size of 64 and channel multipliers of (1, 2, 4, 8, 16), respectively. Time-step embeddings are computed using sinusoidal encoding followed by two fully connected layers with SiLU activation (Elfwing et al., 2018) and are injected into each encoder and decoder stage via concatenation.

We compare our method against several SR approaches. These include bilinear interpolation (Pensa et al., 2024); two non-generative methods, SwinIR (Liang et al., 2021) and SRConvNet (Li et al., 2024b); and three generative SR approaches, SeD (Li et al., 2024a), ResShift (Yue et al., 2023), and PartialDiffusion (PDM) (Zhao et al., 2024). Among the diffusion-based methods, ResShift and PDM require 15 and 50 iterative steps, respectively. All baseline methods are trained using the same self-supervised strategy as UltraCCM. For stochastic approaches, including ResShift, PDM, and UltraCCM, inference was repeated using five different random seeds, and results are reported as mean ± standard deviation across runs. For all slice-wise metrics, values were first averaged within each patient for the *in vivo* dataset or within each specimen for the *ex vivo* dataset. The reported summary statistics and significance tests were computed using these patient- or specimen-level values. For all results in tables, different background colors indicate the best and second best results.

We conducted all experiments on a single NVIDIA Quadro RTX8000 GPU. Training was performed for 30 epochs using Adam optimizer (Kingma, 2014) with a learning rate of 5×10^{-6}, batch size 2. The upsampling factor was fixed at $k = 8$. Following (Song et al., 2023), LPIPS distance (Zhang et al., 2018) was used in $\mathcal{L}_{\text{CM}}$, and the IDA weight was set to $\gamma = 5$. During each epoch, 10 LR-HR patch pairs were randomly extracted from every slice, with LR and HR polar patch sizes of 256×32 and 256×256, respectively. Angular density range $[N_{\min}, N_{\max}]$ was set as [150, 300].

4.3. *in vivo* Evaluation

We present the evaluation results on the *in vivo* dataset. As illustrated in Fig 1, SR was performed in the reformatted axial plane for each MicroUS volume of size $H \times W \times N$. Because adjacent axial slices along the W direction exhibit minimal structural variation, we performed inference on every fourth slice to reduce computation while preserving spatial diversity, yielding 9,479 SR slices across the 30 test volumes.

4.3.1. *Qualitative Results*

Fig 4 presents representative qualitative comparisons, where lesions are highlighted and zoomed in with red boxes. For reference, the whole-mount histopathology (WMH) images with lesions outlined in green are provided. Compared with other methods, UltraCCM produces significantly sharper lesion boundaries and clearer anatomical structures, while introducing minimal artifacts. Furthermore, as marked by the yellow boxes, UltraCCM substantially reduces slice discontinuities and reconstructs realistic textural details that closely resemble true MicroUS patterns, whereas non-generative baselines fail to recover fine-grained structures and other generative methods tend to produce unrealistic artifacts.

4.3.2. *Expert Reader Study*

To assess expert-perceived image quality and clinical interpretability, we conducted a volumetric expert reader study on the SR *in vivo* MicroUS images. Since non-generative methods provided only marginal or negligible improvements, the reader study focused on comparing UltraCCM against generative baselines. For each test case, SR slices were reassembled into volumes, and the four volumes (one per method) corresponding to the same case were grouped together for evaluation. Two experts, a senior clinical urologist (W.B.) and an experienced radiologist specializing in urogenital imaging (M.Q.), independently assessed the image quality after training sessions. All patient

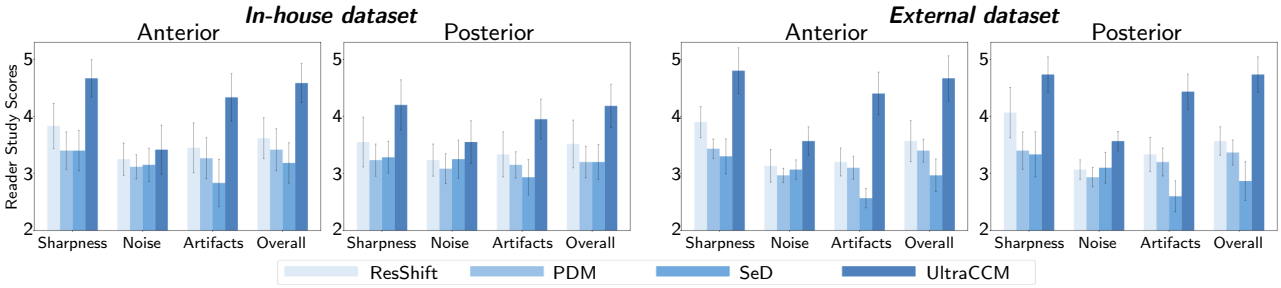


Figure 5: Results of the expert reader study on the *in vivo* data. Reader scores for anterior and posterior regions are shown for our dataset (left) and an external validation dataset from UF (right). Error bars denote standard deviations. All comparisons between UltraCCM and other methods show statistically significant differences ($p_{adj} < 0.05$, paired t-test with Holm correction).

Metrics	Bilinear	SwinIR	SRConvNet	ResShift	PDM	SeD	UltraCCM
Inference Steps	1	1	1	15	50	1	1
Sharpness ($\times 10^2$) ↑	6.252*	6.263*	6.271*	5.947* $\pm 9\times10^{-5}$	7.129* $\pm 1\times10^{-4}$	6.233*	7.514 $\pm 2\times10^{-4}$
TOPIQ (NR) ↑	0.5817*	0.5822*	0.5800*	0.6514 $\pm 6\times10^{-6}$	0.6314* $\pm 2\times10^{-5}$	0.6359*	0.6536 $\pm 1\times10^{-5}$
TRES ↑	61.14*	61.13*	60.93*	77.75 $\pm 4\times10^{-3}$	74.26* $\pm 1\times10^{-3}$	75.94*	77.33 $\pm 9\times10^{-4}$
CLIPIQA ↑	0.6343	0.6346	0.6287	0.6296* $\pm 5\times10^{-5}$	0.6091 $\pm 5\times10^{-5}$	0.5886*	0.6364 $\pm 4\times10^{-5}$
MANIQA ↑	0.3117*	0.3114*	0.3110*	0.3214* $\pm 1\times10^{-5}$	0.3108* $\pm 1\times10^{-5}$	0.3350	0.3359 $\pm 1\times10^{-5}$
CNNIQA ↑	0.7234*	0.7235*	0.7232*	0.7233* $\pm 1\times10^{-5}$	0.7228* $\pm 1\times10^{-5}$	0.7238*	0.7246 $\pm 2\times10^{-5}$
FID ↓	12.75	12.71	12.06	6.23 $\pm 6\times10^{-4}$	3.31 $\pm 4\times10^{-3}$	4.49	2.90 $\pm 6\times10^{-4}$
KID ($\times 10^{-3}$) ↓	10.85	10.85	11.04	7.15 $\pm 3\times10^{-3}$	8.24 $\pm 8\times10^{-3}$	9.14	6.23 $\pm 3\times10^{-3}$

Table 1
Patient-level quantitative comparison of UltraCCM with other methods on the *in vivo* dataset. Stochastic methods are reported as mean ± standard deviation over five random seeds, whereas deterministic methods use single-run results. * indicates a significant difference from UltraCCM based on patient-level paired *t*-tests with Holm correction ($p_{adj} < 0.01$). Statistical testing was not applied to FID and KID, which are dataset-level distribution metrics.

identifiers and SR method labels were anonymized, except that the bilinear-interpolated volumes were disclosed as a reference. Image quality was scored for sharpness, noise, artifacts, and overall quality on a 5-point Likert scale relative to the bilinear baseline (1: much worse, 2: worse, 3: fair, 4: better, 5: much better). Evaluations were performed separately for the anterior and posterior prostate regions.

To provide an initial assessment of generalizability, we conducted an external validation reader study on *in vivo* cases ($n = 15$) from the University of Florida (UF). All imaging protocols and acquisition configurations were consistent with our dataset, as described in Sec 4.1.1. The model trained exclusively on our dataset was directly applied to the UF cases without any fine-tuning. The same evaluation protocol was followed.

Fig 5 presents the average reader scores across cases for both the internal and external datasets. UltraCCM consistently achieved the highest ratings across all criteria and in both anatomical regions on both datasets, with particularly significant improvements in the anterior region where image degradation is most severe. These findings indicate that UltraCCM produces clinician-preferred enhanced images among the evaluated methods and maintains its performance on the external cohort, demonstrating the potential of cross-institutional generalizability.

4.3.3. Quantitative Results

We perform a non-reference image quality evaluation on the *in vivo* dataset using the sharpness measurement, non-reference image quality assessment (NR-IQA) metrics, and feature distribution measurements. Sharpness is quantified using the Tenengrad measure (Krotkov, 1988), which computes the average gradient magnitude obtained from Sobel filtering in both horizontal and vertical directions. NR-IQA metrics include the non-reference variant of TOPIQ (Chen et al., 2024), TRES (Golestaneh et al., 2022), CLIPIQA (Wang et al., 2023), MANIQA (Yang et al., 2022), and CNNIQA (Kang et al., 2014). We compute Fréchet Inception Distance (FID) (Heusel et al., 2017) and Kernel Inception Distance (KID) (Bińkowski et al., 2018) to compare the distribution of deep features extracted from the SR axial slices against those from the acquired HR slices in the original plane, thereby assessing whether the generated textures resemble those in real acquisitions. Quantitative results in Tab 1 show that non-generative baselines (SwinIR, SRConvNet) provide minimal improvement over bilinear interpolation. Among generative approaches, UltraCCM achieves the best sharpness score and ranks first or second across all NR-IQA metrics, while also obtaining the lowest FID and KID. These results indicate that UltraCCM achieves both superior perceptual quality and closer alignment with the distribution of real HR textures. Patient-level paired

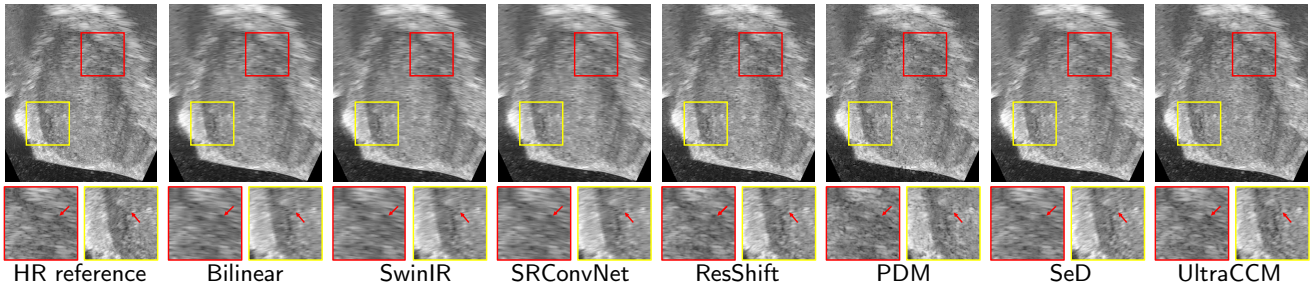


Figure 6: Qualitative comparison on the *ex vivo* data. Red arrows highlight noticeable differences in anatomical structures across the results of different methods. Zoom-in regions are provided to highlight differences in anatomical details.

Metrics	Bilinear	SwinIR	SRConvNet	ResShift	PDM	SeD	UltraCCM
Inference Steps	1	1	1	15	50	1	1
PSNR ↑	30.78*	30.79*	30.74*	29.88* $\pm 5\times10^{-4}$	28.16* $\pm 1\times10^{-3}$	29.83*	29.35 $\pm 8\times10^{-4}$
SSIM ↑	0.8569*	0.8565*	0.8561*	0.8411* $\pm 1\times10^{-5}$	0.7897* $\pm 3\times10^{-5}$	0.8477*	0.8243 $\pm 2\times10^{-5}$
LPIPS ↓	0.1835*	0.1829*	0.1799*	0.1416 $\pm 1\times10^{-5}$	0.1455 $\pm 7\times10^{-5}$	0.1417	0.1405 $\pm 2\times10^{-5}$
DISTS ↓	0.2032*	0.2019*	0.2006*	0.1262 $\pm 5\times10^{-5}$	0.1202 $\pm 1\times10^{-4}$	0.1454*	0.1231 $\pm 6\times10^{-5}$
TOPIQ (FR) ↑	0.5510*	0.5538*	0.5542*	0.6370 $\pm 6\times10^{-5}$	0.6500* $\pm 2\times10^{-4}$	0.6361*	0.6473 $\pm 5\times10^{-5}$
PieAPP ↓	1.4049*	1.3874*	1.3923*	1.0024 $\pm 6\times10^{-4}$	0.8381* $\pm 1\times10^{-3}$	1.2015*	0.9650 $\pm 7\times10^{-4}$

Table 2
Specimen-level quantitative comparison of UltraCCM with other methods on the *ex vivo* dataset. Stochastic methods are reported as mean ± standard deviation over five random seeds, whereas deterministic methods use single-run results. * indicates a significant difference from UltraCCM based on specimen-level paired t-tests with Holm correction ($p_{adj} < 0.01$).

t-tests with Holm correction further show significant differences between UltraCCM and other methods for most comparisons ($p_{adj} < 0.01$).

4.4. *ex vivo* Evaluation

We perform full-reference evaluation on the *ex vivo* dataset. Because the *ex vivo* data were acquired using linear translation, the densely sampled slices were repositioned at pseudo-angular locations to match the fan-shaped geometric representation used as input to UltraCCM. This computational representation is used for controlled sparse-to-dense evaluation and does not represent the physical geometry of the original *ex vivo* acquisition. HR reference volumes were obtained by performing MPR on all repositioned slices, whereas LR inputs were simulated by uniformly subsampling the slices (every k-th slice) before MPR. Following the *in vivo* protocol, inference was performed on every fourth slice along the W dimension, yielding 2,037 SR axial slices.

4.4.1. Qualitative Results

Fig 6 presents the *ex vivo* qualitative comparison. Compared to the HR reference, non-generative methods provide limited enhancement and yield blurred anatomical structures. Among generative methods, SeD produces noticeable artifacts, likely due to the inherent instability of GAN-based training. In contrast, UltraCCM performs comparably to ResShift and PDM, effectively recovering sharp edges and reducing artifacts, while requiring only a single inference step.

4.4.2. Quantitative Results

Tab 2 summarizes the quantitative comparison. We evaluate pixel-level fidelity using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), and perceptual quality using Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018), Deep Image Structure and Texture Similarity (DISTS) (Ding et al., 2020), full-reference TOPIQ score (Chen et al., 2024), and PieAPP (Prashnani et al., 2018).

Non-generative methods achieve high PSNR and SSIM, whereas generative methods generally perform better on perceptual metrics. PDM yields the best DISTS, TOPIQ, and PieAPP scores but requires 50 inference steps. In contrast, using only a single inference step, UltraCCM attains the best LPIPS score and the second-best performance across other perceptual metrics, while maintaining competitive distortion-based performance. Specimen-level paired t-tests with Holm correction show significant differences between UltraCCM and the competing methods for most comparisons ($p_{adj} < 0.01$), as indicated in Table 2. These results indicate that UltraCCM provides a favorable trade-off between perceptual quality and computational efficiency. Although UltraCCM achieves slightly lower PSNR and SSIM than the distortion-oriented non-generative baselines, it remains competitive on these pixel-level metrics while substantially improving perceptual quality, reflecting the expected perception–distortion trade-off. This suggests that the perceptual gains are not accompanied by large pixel- or structure-level distortions, although these metrics alone cannot exclude localized anatomically inconsistent details.

Metrics	Bilinear	SwinIR	SRConvNet	ResShift	PDM	SeD	UltraCCM
Dice (%) ↑	89.90**	89.88**	89.79**	90.45** ±0.08	90.64* ±0.07	90.20**	90.95 ±0.07
IoU (%) ↑	81.87**	81.84**	81.69**	82.78** ±0.13	83.09* ±0.11	82.37**	83.59 ±0.11
RAVD (%) ↓	19.34**	19.32**	19.97**	17.26** ±0.20	15.91** ±0.25	18.06**	14.18 ±0.29
ASSD (mm) ↓	2.8440**	2.8450**	2.8811**	2.6736** ±0.0391	2.5726 ±0.0251	2.7509**	2.5035 ±0.0181

Table 3
Patient-level prostate segmentation results on the *in vivo* dataset. Stochastic methods are reported as mean ± standard deviation over five random seeds. * and ** indicate significant differences from UltraCCM based on patient-level paired t-tests with Holm correction ($p_{adj} < 0.05$ and $p_{adj} < 0.01$, respectively).

Method	Steps	Time (s / image)
Bilinear	1	0.037
SwinIR	1	1.366
SRConvNet	1	0.177
SeD	1	5.394
ResShift	15	75.370
PDM	50	88.570
UltraCCM	1	0.523

Table 4
Inference efficiency comparison on the *in vivo* dataset.

PF	×	✓	✓
Random $\{\alpha\}$	✓	×	✓
Sharpness ($\times 10^2$) ↑	7.004 $\pm 1\times10^{-4}$	7.079 $\pm 6\times10^{-5}$	7.514 $\pm 2\times10^{-4}$
TOPIQ (NR) ↑	0.6502 $\pm 6\times10^{-6}$	0.6464 $\pm 1\times10^{-5}$	0.6536 $\pm 1\times10^{-5}$
TRES ↑	75.63 $\pm 1\times10^{-3}$	76.80 $\pm 1\times10^{-3}$	77.33 $\pm 9\times10^{-4}$
CLIPIQA ↑	0.6141 $\pm 5\times10^{-5}$	0.6209 $\pm 5\times10^{-5}$	0.6364 $\pm 4\times10^{-5}$
MANIQA ↑	0.3274 $\pm 3\times10^{-5}$	0.3565 $\pm 1\times10^{-5}$	0.3359 $\pm 1\times10^{-5}$
CNNIQA ↑	0.7247 $\pm 2\times10^{-5}$	0.7238 $\pm 1\times10^{-5}$	0.7246 $\pm 2\times10^{-5}$
FID ↓	3.75 $\pm 8\times10^{-4}$	3.01 $\pm 6\times10^{-4}$	2.90 $\pm 6\times10^{-4}$
KID ($\times 10^{-3}$) ↓	6.62 $\pm 7\times10^{-3}$	8.23 $\pm 7\times10^{-2}$	6.23 $\pm 3\times10^{-3}$

Table 5
Ablation results evaluating the effect of positional features (PF) and random angular sampling ($\{\alpha\}$).

Together with the expert reader study results, these findings provide complementary evidence of the practical benefits of UltraCCM. The relevance to downstream image analysis is further evaluated through the prostate segmentation task in the following section.

4.5. Computational Efficiency

We benchmarked inference speed on the *in vivo* dataset by measuring the average inference time per target-plane slice for all comparison methods. As presented in Tab 4, UltraCCM achieves a runtime comparable to non-generative approaches while delivering substantially better image quality (Sections 4.3 and 4.4). Compared with other generative competitors, UltraCCM provides **10**× to **169**× acceleration. This highlights the effectiveness of UltraCCM in achieving real-time SR of reformatted MicroUS images, thereby supporting practical and time-efficient deployment in clinical workflows.

4.6. Application to Prostate Segmentation

Prostate segmentation is a clinically important task that delineates the prostate capsule and supports multi-modality registration and disease diagnosis (Jiang et al., 2024). To evaluate the downstream utility of the enhanced images, we performed prostate segmentation on target-plane slices generated by different SR methods. A radiologist annotated 3–5 axial slices per test case in the *in vivo* dataset, resulting in 126 labeled images. We applied the pre-trained medical image segmentation foundation model, MedSAM (Ma et al., 2024) for zero-shot prostate segmentation. The bounding boxes of the annotated masks were extended by 50 pixels to simulate approximate prostate localization and were used as prompts to feed into the model. We used the Dice score, Intersection over Union (IoU) score, Relative Absolute Volume Difference (RAVD), and Average Symmetric Surface Distance (ASSD) to assess the segmentation performance. For each metric, slice-wise values were first averaged within each test case, and the resulting patient-level values were reported in Tab 3. UltraCCM consistently outperforms other methods across all metrics, demonstrating improved utility for downstream prostate segmentation and providing complementary evidence that the SR images may enhance the information relevant to anatomy delineation.

4.7. Ablation Study

We conducted ablation studies on the *in vivo* dataset to evaluate the contributions of key components in UltraCCM.

4.7.1. Positional Features and Random Angular Sampling

We investigate the contribution of the proposed positional features (PF, Sec 3.3.2) and the random angular sampling strategy ($\{\alpha\}$, Sec 3.2) by training ablation variants in which PF was removed or random angular sampling was replaced with uniform sampling. As summarized in Tab 5, removing either component results in worse performance across most of the metrics. These results highlight that PF provides essential geometric information for recovering fine-grained tissue structures, while random angular sampling improves generalization to realistic acquisition patterns. Qualitative comparisons in Fig 7 further show that both ablated models exhibit blurring artifacts, whereas the full UltraCCM restores clearer prostate boundaries and richer anatomical textures.

4.7.2. Intensity Distribution Alignment

To assess the effect of the proposed intensity distribution alignment (IDA) regularization in Sec 3.4.2, we trained an

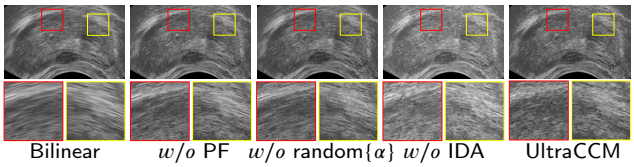


Figure 7: Qualitative comparisons in the ablation study. Zoom-in regions are provided to highlight differences in anatomical details.

Metrics	*w/o* IDA	*w/* IDA
JS ($\times 10^{-2}$) ↓	$4.92_{\pm 1\times 10^{-3}}$	$1.10_{\pm 4\times 10^{-3}}$
W1 ($\times 10^{-2}$) ↓	$3.58_{\pm 4\times 10^{-4}}$	$0.39_{\pm 1\times 10^{-3}}$
HI (%) ↑	$82.69_{\pm 2\times 10^{-3}}$	$95.86_{\pm 1\times 10^{-2}}$
Δ_μ ($\times 10^{-2}$) ↓	$3.57_{\pm 5\times 10^{-4}}$	$0.31_{\pm 1\times 10^{-3}}$
Δ_σ ($\times 10^{-2}$) ↓	$2.26_{\pm 4\times 10^{-4}}$	$0.27_{\pm 5\times 10^{-4}}$
Δ_{RMS} ($\times 10^{-2}$) ↓	$1.86_{\pm 2\times 10^{-4}}$	$0.40_{\pm 8\times 10^{-5}}$

Table 6
Ablation results with and without intensity distribution alignment (IDA). Results are averaged over five runs.

Metrics	Self-supervised	Fully supervised
PSNR ↑	$29.35_{\pm 8\times 10^{-4}}$	$28.63_{\pm 6\times 10^{-4}}$
SSIM ↑	$0.8243_{\pm 2\times 10^{-5}}$	$0.8042_{\pm 1\times 10^{-5}}$
LPIPS ↓	$0.1405_{\pm 2\times 10^{-5}}$	$0.0897_{\pm 4\times 10^{-5}}$
DISTS ↓	$0.1231_{\pm 6\times 10^{-5}}$	$0.0627_{\pm 3\times 10^{-5}}$
TOPIQ-FR ↑	$0.6473_{\pm 5\times 10^{-5}}$	$0.7848_{\pm 1\times 10^{-4}}$
PieAPP ↓	$0.9650_{\pm 7\times 10^{-4}}$	$0.6824_{\pm 8\times 10^{-4}}$

Table 7
Comparison of self-supervised and fully supervised UltraCCM training on the *ex vivo* dataset. Results are averaged over five runs.

ablation model without this term and compared its ability to preserve intensity statistics (e.g., contrast and brightness) between SR outputs and the original LR counterparts. We employed six metrics: Jensen–Shannon divergence (JS) (Lin, 1991), 1D Wasserstein distance (W1) (Villani et al., 2008), histogram intersection (HI) (Swain and Ballard, 1991), absolute differences of mean intensity ($\Delta\mu = |\mu_{LR} - \mu_{SR}|$), standard deviation ($\Delta\sigma = |\sigma_{LR} - \sigma_{SR}|$), and RMS contrast ($\Delta\text{RMS} = \left|\frac{\sigma_{LR}}{\mu_{LR}} - \frac{\sigma_{SR}}{\mu_{SR}}\right|$). Given normalized intensity histograms p and q of LR and SR images, JS quantifies the similarity between them, W1 measures the Earth Mover's distance, and HI evaluates their overlap.

As shown in Tab 6, incorporating IDA substantially improves all metrics, demonstrating its effectiveness in preserving the native intensity distribution. Qualitative results in Fig 7 further show that IDA prevents the brightness drift and contrast imbalance observed when the regularization is removed, producing SR images whose intensity profiles more closely match those of the LR inputs.

4.7.3. Self-Supervised Training Strategy

To assess how well the proposed simulated degradation approximates the sparse-sampling degradation in the target plane, we trained UltraCCM using paired LR-HR target-plane images from the densely acquired *ex vivo* dataset. As shown in Tab 7, self-supervised training achieved higher PSNR and SSIM, whereas fully supervised training yielded better perceptual metrics, benefiting from direct supervision from the same target-plane distribution as the inference data. These results suggest that the proposed simulated degradation provides a reasonable approximation of the dominant sparse-sampling characteristics in the target plane under this controlled setting, while minor discrepancies remain in fine-scale image details and textural patterns.

5. Discussion

We demonstrate that the proposed method can significantly improve the image quality and the potential clinical value of reformatted MicroUS images with a fast inference speed. We propose a self-supervised strategy that exploits the directional resolution disparity of MicroUS by learning HR image statistics from the original acquisition plane and transferring them to the LR target plane. By incorporating positional features as a conditional model input and applying random angular sampling during training, UltraCCM effectively leverages the geometric characteristics of MicroUS acquisition to enhance fine anatomical details. In addition, we introduced an intensity distribution alignment regularization to preserve the brightness and contrast of SR outputs. Unlike diffusion-based approaches, UltraCCM requires only a single inference step, achieving orders-of-magnitude faster inference while producing high-quality images. Extensive evaluations on both *in vivo* and *ex vivo* datasets, together with expert reader studies and a downstream prostate segmentation task, demonstrate that UltraCCM achieves state-of-the-art perceptual quality and provides practical benefits for MicroUS anatomical visualization.

Despite the advantages, several limitations remain. First, the proposed SR formulation and simulated degradation focus on the rotational acquisition geometry and sparse angular sampling, but do not explicitly model residual probe translation or tilt, pressure-induced tissue deformation, or patient motion. Consequently, robustness to individual pose and deformation effects has not been comprehensively evaluated. Second, UltraCCM is currently implemented as a 2D super-resolution framework, which may not fully capture continuity across adjacent slices in the reformatted volume or the broader volumetric anatomical context. Third, although expert reader studies were conducted, the number of readers was relatively small, which may limit the robustness of subjective evaluation. Fourth, the limited number and diversity of training subjects may not fully capture patient-, operator-, and anatomy-level variability. Future work will extend UltraCCM to 3D modeling and explore other cross-plane ultrasound SR applications while incorporating pose-, motion-, and deformation-aware reconstruction. We will also perform larger-scale, multi-operator, and multi-institutional evaluation with additional readers and lesion-level or task-based validation.

6. Conclusion

In this paper, we presented UltraCCM, a geometry-driven, self-supervised consistency model for super-resolution of reformatted MicroUS images. By explicitly leveraging the geometric information of MicroUS acquisition through positional features and random angular sampling, and by introducing an intensity distribution alignment regularization, UltraCCM generates high-quality SR images that preserve both anatomical details and intensity distribution characteristics. Compared with previous methods, UltraCCM performs single-step inference, enabling faster super-resolution processing while maintaining superior image quality. Comprehensive evaluations on internal and external *in vivo* cohorts and *ex vivo* data, along with expert reader studies and a downstream segmentation task, demonstrate that UltraCCM substantially improves perceptual quality, reduces artifacts, and supports downstream image analysis. These results highlight the potential of UltraCCM to improve prostate cancer imaging workflows by improving reformatted MicroUS visualization and supporting both diagnostic and downstream computational applications.

Funding sources

This research was funded in part by the National Institutes of Health under grants R01-CA248506 and R01-CA272702, and by the Integrated Diagnostics Program of the Departments of Radiological Sciences and Pathology in the David Geffen School of Medicine, University of California at Los Angeles (UCLA).

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Grammarly to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability

The data used in this study are not publicly available due to patient privacy and institutional restrictions, but may be available from the corresponding author upon reasonable request and with permission from the Institutional Review Board (IRB).

Declaration of competing interest

Dr. Wayne Brisbane is a consultant for Exact Imaging. Dr. Kyunghyun Sung has a financial interest in Siemens Healthineers. All other authors declare no competing interests.

CRediT authorship contribution statement

Kaifeng Pang: Conceptualization, Methodology, Data curation, Software, Validation, Writing – original draft. **Kai Zhao:** Writing - review and editing, Methodology, Project administration, Supervision, Data curation, Conceptualization, Investigation. **Qi Miao:** Conceptualization, Investigation, Formal analysis, Validation, Data curation. **Alex Ling Yu Hung:** Methodology, Writing - review and editing. **Changsuk Oh:** Writing - review and editing, Formal analysis. **Raymi Ramirez:** Writing - review and editing, Formal analysis. **Qiudi He:** Writing - review and editing, Formal analysis. **Jordan Klein:** Writing - review and editing. **Wei Shao:** Writing - review and editing, Resources. **Wayne Brisbane:** Data curation, Supervision, Resources, Project administration, Conceptualization, Investigation, Writing - review and editing, Writing - original draft, Funding acquisition. **Kyunghyun Sung:** Writing - original draft, Writing - review and editing, Supervision, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Data curation, Formal analysis.

References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks, in: International conference on machine learning, PMLR. pp. 214–223.

Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A., 2018. Demystifying mmd gans. arXiv preprint arXiv:1801.01401 .

Chen, C., Mo, J., Hou, J., Wu, H., Liao, L., Sun, W., Yan, Q., Lin, W., 2024. Topiq: A top-down approach from semantics to distortions for image quality assessment. IEEE Transactions on Image Processing 33, 2404–2418.

Chojnowski, H., Ates, G.C., Brisbane, W.G., Shao, W., 2026. Geometry-aware implicit neural reconstruction of oblique micro-ultrasound scans .

Dias, A.B., O'Brien, C., Correas, J.M., Ghai, S., 2022. Multiparametric ultrasound and micro-ultrasound in prostate cancer: a comprehensive review. The British Journal of Radiology 95, 20210633.

Ding, K., Ma, K., Wang, S., Simoncelli, E.P., 2020. Image quality assessment: Unifying structure and texture similarity. IEEE transactions on pattern analysis and machine intelligence 44, 2567–2581.

Elfwing, S., Uchibe, E., Doya, K., 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural networks 107, 3–11.

Exact Imaging, n.d.a. Fusionvu: A simplified fusion approach. https://www.exactimaging.com/fusionvu. Accessed July 26, 2026.

Exact Imaging, n.d.b. Operation and Safety Manual for the ExactVu High Resolution Micro-Ultrasound System. revision 3.7 ed. Exact Imaging. Markham, Ontario, Canada. Accessed July 26, 2026.

Freedman, G., Fattal, R., 2011. Image and video upscaling from local self-examples. ACM Transactions on Graphics (TOG) 30, 1–11.

Gilany, M., Wilson, P., Perera-Ortega, A., Jamzad, A., To, M.N.N., Fooladgar, F., Wodlinger, B., Abolmaesumi, P., Mousavi, P., 2023. Trusformer: improving prostate cancer detection from micro-ultrasound using attention and self-supervision. International Journal of Computer Assisted Radiology and Surgery 18, 1193–1200.

Golestaneh, S.A., Dadsetan, S., Kitani, K.M., 2022. No-reference image quality assessment via transformers, relative ranking, and self-consistency, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1220–1230.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Communications of the ACM 63, 139–144.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems 30.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Advances in neural information processing systems 33, 6840–6851.

Hung, A.L.Y., Zhao, K., Zheng, H., Yan, R., Raman, S.S., Terzopoulos, D., Sung, K., 2023. Med-cdiff: Conditional medical image generation with diffusion models. Bioengineering 10, 1258.

Jiang, H., Imran, M., Muralidharan, P., Patel, A., Pensa, J., Liang, M., Benidir, T., Grajo, J.R., Joseph, J.P., Terry, R., et al., 2024. Microsegnet: A deep learning approach for prostate segmentation on micro-ultrasound images. Computerized Medical Imaging and Graphics 112, 102326.

Kang, L., Ye, P., Li, Y., Doermann, D., 2014. Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1733–1740.

Karras, T., Aittala, M., Aila, T., Laine, S., 2022. Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems 35, 26565–26577.

Kingma, D.P., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 .

Kinnaird, A., Luger, F., Cash, H., Ghai, S., Urdaneta-Salegui, L.F., Pavlovich, C.P., Brito, J., Shore, N.D., Struck, J.P., Schostak, M., et al., 2025. Microultrasonography-guided vs mri-guided biopsy for prostate cancer diagnosis: the optimum randomized clinical trial. JAMA 333, 1679–1687.

Klotz, L., Lughezzani, G., Maffei, D., Sánchez, A., Pereira, J.G., Staerman, F., Cash, H., Luger, F., Lopez, L., Sanchez-Salas, R., et al., 2020. Comparison of micro-ultrasound and multiparametric magnetic resonance imaging for prostate cancer: A multicenter, prospective analysis. Canadian Urological Association Journal 15, E11.

Krotkov, E., 1988. Focusing. International Journal of Computer Vision 1, 223–237.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681–4690.

Li, B., Li, X., Zhu, H., Jin, Y., Feng, R., Zhang, Z., Chen, Z., 2024a. Sed: Semantic-aware discriminator for image super-resolution, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 25784–25795.

Li, F., Cong, R., Wu, J., Bai, H., Wang, M., Zhao, Y., 2024b. Srconvnet: A transformer-style convnet for lightweight image super-resolution. International Journal of Computer Vision , 1–17.

Li, J., Yang, H., Yi, Q., Lu, M., Shi, J., Zeng, T., 2025. High-frequency modulated transformer for multi-contrast mri super-resolution. IEEE Transactions on Medical Imaging .

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R., 2021. Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 1833–1844.

Lin, J., 1991. Divergence measures based on the shannon entropy. IEEE Transactions on Information theory 37, 145–151.

Liu, T., Han, S., Xie, L., Xing, W., Liu, C., Li, B., Ta, D., 2024. Super-resolution reconstruction of ultrasound image using a modified diffusion model. Physics in Medicine & Biology 69, 125026.

Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., 2024. Segment anything in medical images. Nature Communications 15, 654.

Masoudi, S., Harmon, S.A., Mehralivand, S., Walker, S.M., Raviprakash, H., Bagci, U., Choyke, P.L., Turkbey, B., 2021. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. Journal of Medical Imaging 8, 010901–010901.

Maurer, T., Eiber, M., Schwaiger, M., Gschwend, J.E., 2016. Current use of psma–pet in prostate cancer management. Nature Reviews Urology 13, 226–235.

Miao, Q., Sisk, A., Sonni, I., Grunden, K., Ludwig, V., Pang, K., Walsh, K., Klein, J., Goldar, Z.M., Reiter, R.E., et al., 2026. Head-to-head comparison of multimodal imaging for prostate cancer detection and local staging with whole-mount histopathology reference. European Urology Oncology .

Pang, K., Miao, Q., Hung, A.L.Y., Zhao, K., Oh, E., Ramirez, R., Brisbane, W., Sung, K., 2025a. Mpr-diff: A self-supervised diffusion model for multi-planar reformation in prostate micro-ultrasound imaging, in: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1–5.

Pang, K., Zhao, K., Hung, A.L.Y., Zheng, H., Yan, R., Sung, K., 2025b. Nexpr: Neural explicit representation for fast arbitrary-scale medical image super-resolution. Computers in biology and medicine 184, 109354.

Pensa, J., Brisbane, W., Kinnaird, A., Kuppermann, D., Hughes, G., Ushko, D., Priester, A., Gonzalez, S., Reiter, R., Chin, A., et al., 2024. Evaluation of prostate cancer detection using micro-ultrasound versus mri through co-registration to whole-mount pathology. Scientific Reports 14, 18910.

Pensa, J., Brisbane, W., Priester, A., Sisk, A., Marks, L., Geoghegan, R., 2021. A system for co-registration of high-resolution ultrasound, magnetic resonance imaging, and whole-mount pathology for prostate cancer, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE. pp. 3890–3893.

Prashnani, E., Cai, H., Mostofi, Y., Sen, P., 2018. Pieapp: Perceptual image-error assessment through pairwise preference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1808–1817.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer. pp. 234–241.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M., 2022. Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 4713–4726.

Shaw, P., Uszkoreit, J., Vaswani, A., 2018. Self-attention with relative position representations. arXiv preprint arXiv:1803.02155 .

Song, Y., Dhariwal, P., Chen, M., Sutskever, I., 2023. Consistency models .

Song, Y., Ermon, S., 2020. Improved techniques for training score-based generative models. Advances in neural information processing systems 33, 12438–12448.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2020. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 .

Stabile, A., Giganti, F., Rosenkrantz, A.B., Taneja, S.S., Villeirs, G., Gill, I.S., Allen, C., Emberton, M., Moore, C.M., Kasivisvanathan, V., 2020. Multiparametric mri for prostate cancer diagnosis: current status and future directions. Nature reviews urology 17, 41–61.

Sun, J., Xu, Z., Shum, H.Y., 2008. Image super-resolution using gradient profile prior, IEEE. pp. 1–8.

Swain, M.J., Ballard, D.H., 1991. Color indexing. International journal of computer vision 7, 11–32.

Villani, C., et al., 2008. Optimal transport: old and new. volume 338. Springer.

Wang, J., Chan, K.C., Loy, C.C., 2023. Exploring clip for assessing the look and feel of images, in: Proceedings of the AAAI conference on artificial intelligence, pp. 2555–2563.

Williams, C., Ahdoot, M., Daneshvar, M.A., Hague, C., Wilbur, A.R., Gomella, P.T., Shih, J., Khondakar, N., Yerram, N., Mehralivand, S., et al., 2022. Why does magnetic resonance imaging-targeted biopsy miss clinically significant cancer? The Journal of urology 207, 95–107.

Wilson, P.F., Gilany, M., Jamzad, A., Fooladgar, F., To, M.N.N., Wodlinger, B., Abolmaesumi, P., Mousavi, P., 2023. Self-supervised learning with limited labeled data for prostate cancer detection in high-frequency ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 70, 1073–1083.

Yang, J., Wright, J., Huang, T., Ma, Y., 2008. Image super-resolution as sparse representation of raw image patches, IEEE. pp. 1–8.

Yang, J., Wright, J., Huang, T.S., Ma, Y., 2010. Image super-resolution via sparse representation. IEEE transactions on image processing 19, 2861–2873.

Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y., 2022. Maniqa: Multi-dimension attention network for no-reference image quality assessment, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1191–1200.

Yue, Z., Wang, J., Loy, C.C., 2023. Resshift: Efficient diffusion model for image super-resolution by residual shifting. Advances in Neural Information Processing Systems 36, 13294–13307.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586–595.

Zhao, K., Pang, K., Hung, A.L.Y., Zheng, H., Yan, R., Sung, K., 2024. Mri super-resolution with partial diffusion models. IEEE transactions on medical imaging 44, 1194–1205.

Zheng, H., Hung, A.L.Y., Miao, Q., Song, W., Scalzo, F., Raman, S.S., Zhao, K., Sung, K., 2024. Atpca-net: anatomical-aware prostate cancer detection network on multi-parametric mri. Scientific Reports 14, 5740.

Zhou, S.R., Zhang, L., Choi, M.H., Vesal, S., Kinnaird, A., Brisbane, W.G., Lughezzani, G., Maffei, D., Fasulo, V., Albers, P., et al., 2025. Promus-net: Artificial intelligence detects more prostate cancer than urologists on micro-ultrasonography. BJU international 136, 1071–1079.