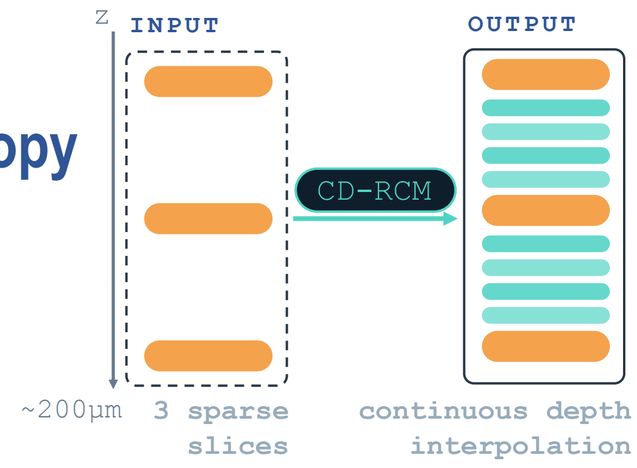
MEDICAL IMAGING × NEURAL RENDERING

CD-RCM Generalizable Continuous-Depth Novel View Synthesis for Reflectance Confocal Microscopy

TOOBA IMTIAZ, MILIND RAJADHYAKSHA, KIVANC KOSE*, JENNIFER DY*

The first feedforward, depth-queryable model that turns a sparse, anisotropic skin z-stack into a dense, isotropic volume — with no per-patient optimization, and sub-second inference.



~6x
resolution anisotropy addressed (lateral 0.5µm vs. axial ~3µm sectioning)

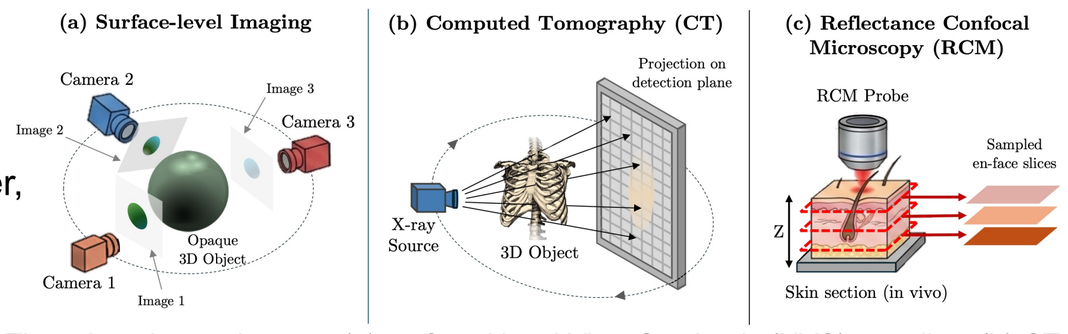
1 pass
feedforward inference — no per-patient / per-stack optimization

4 / 4
metrics beaten vs. classical interpolation (PSNR · SSIM · LPIPS · L_{SPF})

0.8 s
to 10x densify a full stack on a single NVIDIA A6000

1 The problem

RCM can provide *in-vivo* images of skin as depth-stacked en-face optical sections, to obtain noninvasive, cellular-resolution "*optical biopsies*." However, the axial resolution (~3µm sectioning) is roughly 6x coarser than the 0.5µm lateral resolution, so stacks are strongly anisotropic and hard to read across depth, especially near the dermal–epidermal junction. Conventional interpolation lacks anatomical priors, yielding blocky, implausible slices.



Three imaging regimes — (a) surface Novel View Synthesis (NVS): parallax; (b) CT: X-ray attenuation; (c) RCM: axial optical sectioning of internal tissue.

A DISTINCT REGIME

Unlike surface NVS (geometric parallax) or CT (X-ray attenuation), RCM images internal tissue by axial optical sectioning with almost no viewpoint change — and shallow sections optically occlude deeper ones.

2 Goal & contributions

Interpolate intermediate sections to make the volume isotropic — enabling arbitrary-direction, histopathology-like virtual sectioning, without per-patient optimization.

1
First arbitrary-depth NVS formulation for RCM, explicitly handling anisotropic, see-through stacks.

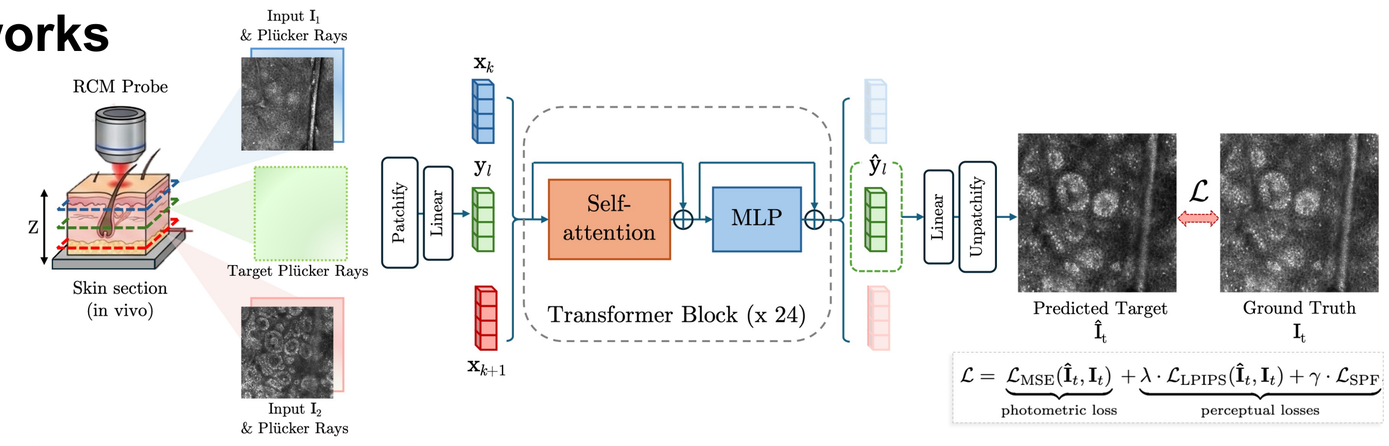
2
A feedforward, depth-queryable model built on RCM's axial imaging physics, which generalizes across skin sites with no per-stack optimization.

3
High-fidelity novel-depth synthesis in one pass, beating classical interpolation on every metric.

DATASET

- 216 in-vivo RCM stacks
- arm & trunk of healthy volunteers (20–50 yrs)
- 50–65 en-face slices, each at 1000×1000 *px* (~0.5µm lateral, ~3µm sectioning, 1.5µm steps)
- stack-level disjoint 156 / 60 train–test split
- Collected with VivaScope 1500

3 How CD-RCM works



Sparse input slices + Plücker rays → shared token space → 24-block decoder-only transformer; target ray tokens condition synthesis, only updated target tokens are decoded to $\hat{I}_t$.

STEP 01
Virtual camera
Model acquisition as a pinhole camera in pure z-translation — no parallax. Encode each slice with Plücker rays.

STEP 02
Unified tokens
Patchify slices + rays (ViT) into one token space; target depths enter as ray tokens that condition synthesis.

STEP 03
Decoder-only transformer
24 bidirectional self-attention + MLP blocks (QK-norm). Decode only updated target tokens.

STEP 04
Skin-specific loss
A DINOv3 ViT adapted to RCM via LoRA (self-supervised, EMA teacher) supervises cellular & textural fidelity.

$$\mathcal{L} = \mathcal{L}_{MSE}(\hat{I}_t, I_t) + \lambda \cdot \mathcal{L}_{LPIPS}(\hat{I}_t, I_t) + \gamma \cdot \mathcal{L}_{SPF}$$

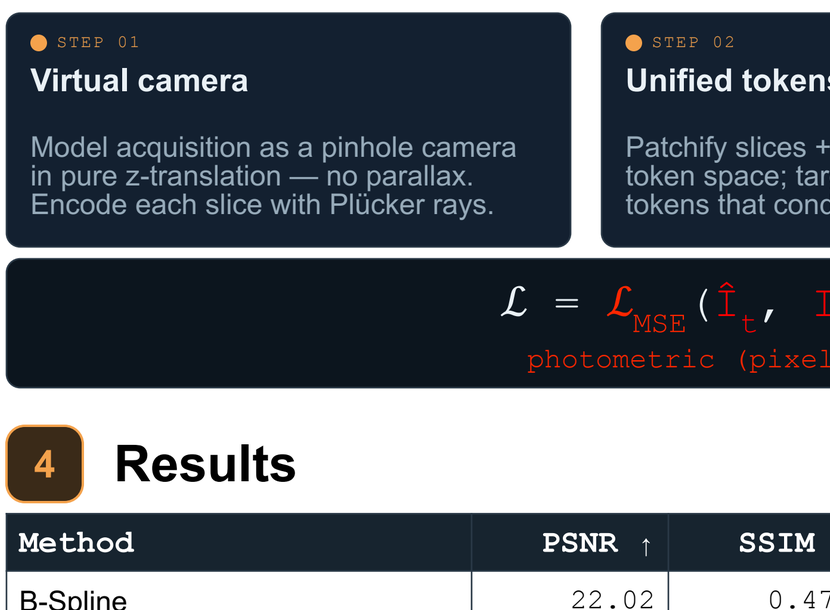
photometric (pixel fidelity) · perceptual = LPIPS · + skin-specific feature loss

4 Results

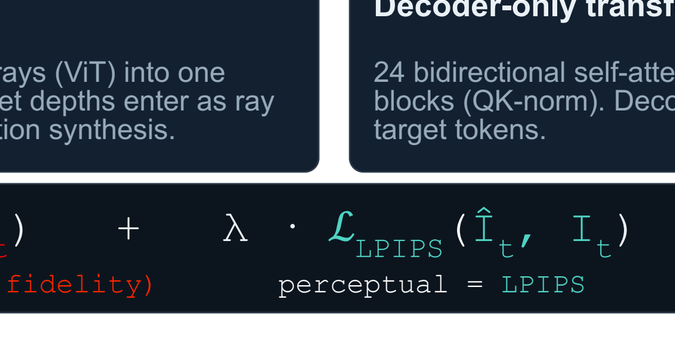
Method	PSNR ↑	SSIM ↑	LPIPS ↓	L_{SPF} ↓
B-Spline	22.02	0.471	0.375	0.161
Cubic Spline	22.09	0.469	0.378	0.163
Gaussian	22.87	0.522	0.477	0.164
CD-RCM (Ours)	**23.36**	**0.581**	**0.314**	**0.145**

NO ACCURACY–SHARPNESS TRADE-OFF

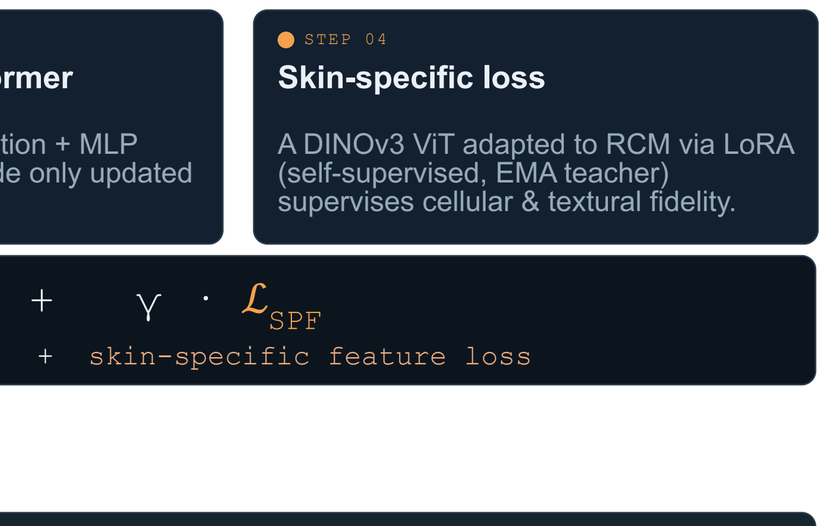
CD-RCM wins on both photometric (PSNR, SSIM) and perceptual (LPIPS, L_{SPF}) metrics. Gaussian interpolation can raise PSNR by blurring — but that erases diagnostic detail. CD-RCM stays photometrically accurate and perceptually sharp.



Novel-depth synthesis vs. interpolation. CD-RCM reconstructs fine cellular structure closest to ground truth; insets show where interpolation distorts.



Isotropic 3D & arbitrary sectioning. Densifying to isotropic sampling yields smooth sagittal, coronal, and oblique cuts — removing staircase artifacts.



Our RCM stack viewer VoxelScope allows viewing the RCM stacks densified using CD-RCM along any viewing direction, including cross-sectional, oblique cuts

TAKEAWAY

CD-RCM is the first **feedforward, continuous-depth** view-synthesis method built for RCM — enabling post-hoc axial densification and arbitrary-orientation virtual sectioning of **in-vivo skin**, a step toward practical continuous 3D dermatologic visualization.

Northeastern University · Memorial Sloan Kettering Cancer Center

Work supported by grants DoD/USAMRAA HT94252410553 and DoD/USAMRAA HT94252410554 and NIH/NCI R01CA199673, U24CA264369