Scale-Agnostic Super-Resolution in MRI using Feature-Based Coordinate Networks

Dave Van Veen, Rogier van der Sluijs, Batu Ozturkler, Arjun Desai, Christian Bluethgen, Robert D. Boutin, Marc H. Willis, Gordon Wetzstein, David Lindell, Shreyas Vasanawala, John Pauly, Akshay S. Chaudhari

{VANVEEN, SLUIJS, OZT, ARJUNDD, BLUCH, BOUTIN, MWILLIS7, GORDONWZ, LINDELL, VASANAWALA, PAULY, AKSHAYSC}@STANFORD.EDU

Stanford University

Editors: Under Review for MIDL 2022

Abstract

We propose using a coordinate network as a decoder for MRI super-resolution. The continuous signal representation of coordinate networks enables this approach to be scale-agnostic, i.e. training over a continuous range of scales and querying at arbitrary resolutions. We evaluate the benefits of denoising for coordinate networks and also compare our method to a convolutional decoder using image quality metrics and a radiologist study.

Keywords: Coordinate networks, super-resolution, MRI.

1. Introduction

High resolution (HR) magnetic resonance imaging (MRI) scans are crucial for many diagnostic imaging tasks. However, tradeoffs with scan time and signal-to-noise ratios (SNR) motivate improved MRI resolution for higher downstream clinical utility. While deep learning (DL) can learn data-driven priors for encoding high-frequency information in super-resolution (SR) tasks, state-of-the-art methods such as EDSR (Lim et al., 2017) are limited to upsampling at fixed, discrete scales due to their convolutional structure. Discrete scales are undesirable for clinical interpretation (Chaudhari et al., 2021); further, training such fixed networks places strict limits on acquiring homogeneous training data.

Here we propose a *scale-agnostic* framework for MRI SR using a coordinate network as a decoder. The continuous nature of this decoder enables (1) querying at arbitrary resolutions (2) decoupling between training and querying scales, e.g. one can train on a continuous range of 1-2× and query at 3×. Throughout the optimization we employ a denoiser to mitigate the network from outputting noise as high-frequency detail. We compare the proposed framework's coordinate decoder against a standard convolutional decoder (EDSR), using image quality metrics and a clinical reader study.

Related Work: Coordinate networks are recent powerful tools for representing signals such as images with fully-connected multi-layer perceptrons (MLPs) by mapping image coordinates to their corresponding pixel values. In contrast to standard pixel-based representations, this representation is continuous w.r.t. network weights, allowing it to model fine detail which is limited by network capacity instead of grid resolution. Coordinate networks are commonly employed for unsupervised representation of a single signal (Lindell et al., 2022; Wu et al., 2021) but cannot incorporate novel high-frequency information for SR. In contrast, supervised methods learn to represent many signals over a shared function space, often with meta-learning or convolutional structure to encode features (Chen et al., 2021).

2. Methods

Given a high-resolution image x_{hr}, we create low-resolution network input via bicubic downsampling, which is mapped to a 2D latent representation with a SR encoder. Subsequently, given a coordinate (c_*, Fig. 1) in the latent space, the coordinate network queries the neighboring four latent codes ($\boldsymbol{c} \coloneqq [c_1, c_2, c_3, c_4]$) such that the decoder's 1D output predicts the grayscale pixel value at those four locations. The pixel value at c_* is estimated as a linear combination of surrounding pixel values at $\boldsymbol{c}$ based on relative distance, hence preventing discontinuities in the output image. Querying over many coordinates produces the predicted image $\hat{x}$. For training we use a consistency loss L_c and denoising loss L_d, i.e. $L_c + \lambda L_d = \|\hat{x} - x_{hr}\|_1 + \lambda\|\hat{x} - D_\sigma(x_{hr})\|_2^2$, where D_σ is a denoiser with strength σ.

Implementation: This framework allows for many choices of encoders or denoisers; for simplicity we choose an EDSR encoder and BM3D denoiser (Dabov et al., 2007). We compare our decoder's continuous representation ("*coord*"), to the same framework with a convolutional decoder, i.e. the original EDSR ("*conv*"), which is not scale-agnostic. *Coord* is similar to *conv* but modifies the decoder to be a five-layer MLP containing 256 hidden units and ReLU activations. We also compare against bicubic interpolation, which can be queried at arbitrary scales but has no prior to incorporate higher frequency information. We evaluate on 2D sagittal slices of the SKM-TEA dataset (Desai et al., 2021).

Reader study details: We perform a reader study with radiologists comparing *coord* and *conv*, both trained on 2× SR. In clinical applications one would want to scale beyond ground-truth; hence at inference we bypass downsampling and perform 2× SR on the ground-truth itself. Both readers used a five-point Likert scale to evaluate randomized side-by-side image pairs on both sharpness and noise.

3. Results and Discussion

Per Table 1, *coord* performs comparably when trained on a range of scales (1-2×, row 2) vs. a fixed scale (2×, row 5). Because *coord* is scale-agnostic, it can be queried at a resolution which is both arbitrary and independent of its training scales. Conversely, *conv*—without additional interpolation—queries only at fixed integer upsampling according to its training

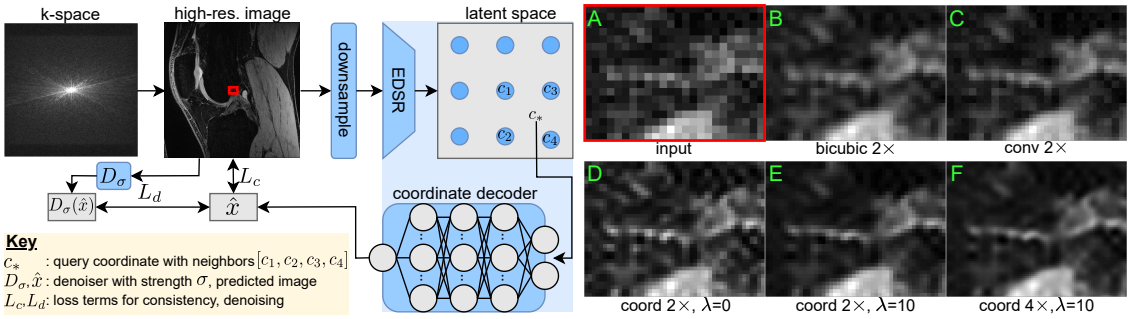
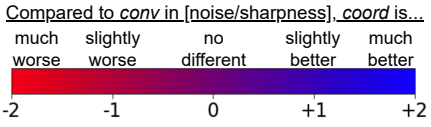


Figure 1: **Left: System overview**. **Right: Image comparison** given input cropped from the red box at left. *Coord* and conv are each trained at 2×. *Coord* can also be queried at 4× without re-training because it is scale-agnostic (F). *Coord* benefits from denoising i.e. $\lambda \neq 0$ (D vs. E), while *conv* does not (see Table 1).

Table 1: **Left: Quantitative scores** (VIF/PSNR). *Coord* obtains similar performance to *conv* (slightly better VIF, slightly worse PSNR) and has the benefit of being agnostic with respect to train and query scales. Below the midline we provide ablations at various training scales and with/without denoising i.e. λ (see discussion, Section 3). **Right: Reader study** scoring criteria (top) and results (bottom) demonstrating a slight overall preference for *coord*.

	Method	λ	Scale-agnostic	Train scale	Query scale (VIF/PSNR) 1.5×	2×	3×
Default	Bicubic	—	✓	—	.87/34.2	.79/31.4	.64/27.6
	Coord	10	✓	1-2×	.92/33.8	.87/31.0	.76/26.9
	Conv	0		2×	—	.82/32.3	—
Ablation	Coord	0	✓	1-2×	.95/33.0	.88/30.2	.74/26.3
	Coord	10	✓	2×	.95/33.2	.89/30.6	.77/26.2
	Conv	10		2×	—	.81/32.2	—



Reader	Noise	Sharpness
1	.70 ± .64	1.4 ± .69
2	.78 ± .46	.02 ± .14
Pooled	.74 ± .56	.70 ± .84

scales. *Conv* was slightly superior to *coord* in PSNR but slightly inferior in VIF, which is more indicative of clinical diagnostic quality (Mason et al., 2019). Unlike *coord* which benefited from denoising regularization (Fig. 1, D vs. E), *conv* did not (row 6), presumably since the convolutional kernel structure inherently acts as a denoiser.

We note the challenge of evaluating clinical potential using image metrics alone. Consider instances where higher metrics scores seemingly do not pertain to better quality: compared to *coord*, $\lambda = 10$ (Fig. 1, E), bicubic interpolation (B) achieves higher PSNR while *coord*, $\lambda = 0$ (D) achieves higher VIF; however, these are perceptually undesirable in terms of sharpness and noise, respectively. Furthermore, quantitative metrics require a ground-truth reference; yet in a clinical setting, the goal is to scale larger than ground-truth resolution. Hence to gain insight beyond these limitations, we present a reader study in Table 1 demonstrating that *coord* is equivalent or slightly preferable to *conv* in terms of perceived sharpness and noise.

In the future, we plan to extensively evaluate across different encoding, decoding, and denoising methods and also assess impact on pixel-level quantitative MRI metrics.

References

Akshay Chaudhari et al. Prospective deployment of deep learning in MRI: a framework for important considerations, challenges, and recommendations for best practices. *JMRI*, 54(2):357–371, 2021.

Yinbo Chen et al. Learning continuous image representation with local implicit image function. In *CVPR*, pages 8628–8638, 2021.

Kostadin Dabov et al. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE - TIP*, 16(8):2080–2095, 2007.

Arjun D Desai et al. SKM-TEA: A dataset for accelerated MRI reconstruction with dense image labels for quantitative clinical evaluation. In *NeurIPS Datasets*, 2021.

Bee Lim et al. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, pages 136–144, 2017.

David B. Lindell et al. BACON: Band-limited coordinate networks for multiscale scene representation. In *CVPR*, 2022.

Allister Mason et al. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE - TMI*, 39(4):1064–1072, 2019.

Qing Wu et al. IREM: High-resolution MRI reconstruction via implicit neural representation. In *MICCAI*, pages 65–74. Springer, 2021.