LEARNING IMAGE-ADAPTIVE GABOR FRAMES FOR SELF-SUPERVISED IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning image representations that adapt to local structures while retaining explicit reconstruction guarantees offers a principled foundation for image restoration. We propose GaborCanvas, a framework for learning image-adaptive Gabor frames. A neural network predicts the scales, orientations, and frequencies of Gabor atoms, aligning them with edges and textures so that relatively few significant coefficients can capture most image structures. We construct corresponding dual frames that ensure perfect reconstruction under suitable conditions. Together, structural adaptation and an explicit reconstruction formula provide an efficient and interpretable image representation that supports coefficient-based processing. We exploit the resulting coefficient sparsity through self-supervised formulations of three tasks: image denoising, inpainting, and super-resolution. The restoration models are trained without access to ground-truth images. GaborCanvas achieves competitive performance across all three tasks.

1 INTRODUCTION

Learning representations that adapt to local image structures while retaining reconstruction guarantees offers a principled foundation for image restoration, especially self-supervised image restoration. Edges, contours, and textures vary across images and locations, calling for representations that adapt to their geometry and frequency content. When representation components align with these structures, relatively few significant coefficients may capture much of the image content. An explicit reconstruction identity complements this adaptation by preserving the full image before task-specific processing and providing a direct way to map modified coefficients back to image space.

Biological studies show that Gabor functions provide a good model of the receptive fields of many simple cells in the primary visual cortex (Daugman, 1985; Jones & Palmer, 1987). These functions describe where image patterns occur and their local orientation and frequency (Lee, 1996). A Gabor atom is a Gaussian window modulated by a complex exponential or, in its real-valued form, a cosine. For spatial coordinates $r \in \mathbb{R}^2$, the real-valued form is

$$\psi_k(r) = \exp\left[-\frac{1}{2}(r-p_k)^\top \Sigma_k^{-1}(r-p_k)\right] \cos\big(2\pi\omega_k^\top (r-p_k) + \varphi_k\big)\,, \tag{1}$$

where p_k is the center, $\Sigma_k \succ 0$ controls the spatial scales and orientation, ω_k specifies the modulation frequency and direction, and φ_k is the phase. These parameters give an interpretable description of directional edges and oscillatory textures within the neighborhood of the location p_k.

An image representation should also support stable coefficient computation and exact reconstruction. Orthonormal bases provide these properties, but classical Gabor systems cannot form an orthonormal basis for $L^2(\mathbb{R})$ while retaining good localization in the spatial and frequency domains (Benedetto et al., 1994). A *frame* generalizes a basis by allowing redundancy while retaining stable representation and exact reconstruction (Daubechies et al., 1995; Ron & Shen, 1997). For a frame $\{\psi_k\}_k$ in a Hilbert space $\mathcal{H}$, an associated dual frame $\{\widetilde{\psi}_k\}_k$ provides

$$c_k = \langle f, \psi_k \rangle, \qquad f = \sum_k c_k \widetilde{\psi}_k, \qquad f \in \mathcal{H}. \tag{2}$$

The ability of Gabor atoms to describe local structures at different scales and orientations has motivated using Gabor frames for image restoration (Ji et al., 2017; 2019). These constructions use predefined atoms, limiting their adaptation to an individual image.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Neural representations offer greater adaptivity. Convolutional neural networks (CNNs) and coordinate-based networks can represent an image by fitting their NN weights to the image (Ulyanov et al., 2018; Sitzmann et al., 2020; Tancik et al., 2020). However, the nonlinear and indirect relationship between NN weights and image structures makes it less straightforward to impose explicit image priors for restoration. Reconstruction also relies on network fitting rather than an explicit perfect reconstruction identity. These concerns motivate the following question:

> *Can we combine neural image adaptation with the interpretability of Gabor atoms and the perfect reconstruction property of frames?*

We introduce **GaborCanvas**, a framework for learning image-adaptive Gabor frames. A neural network predicts the scales, orientations, and frequencies of a set of Gabor atoms at each pixel, aligning the representation with local image content. Figure 1 illustrates how these interpretable parameters describe directional edges and textures. We also construct corresponding dual frames for an explicit perfect reconstruction identity. Together, structural adaptation and exact reconstruction support efficient preservation of image content in a flexible and interpretable representation.

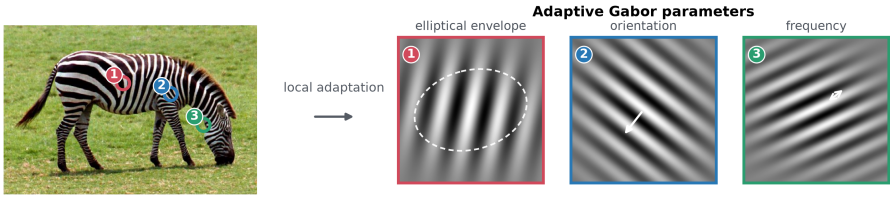

Figure 1: In GaborCanvas, Gabor atoms adapt their elliptical envelopes, orientations, and frequencies to directional local image structures within the marked local regions.

We develop self-supervised formulations for denoising, inpainting, and single-image super-resolution. Denoising learns coefficient thresholding with the recorrupted-to-recorrupted (R2R) loss (Pang et al., 2021); inpainting and super-resolution (SR) combine measurement fidelity with an ℓ_1 penalty on the representation coefficients. The restoration models are trained without clean targets. Experiments demonstrate competitive performance across all three tasks. Our contributions are threefold:

- We propose GaborCanvas, a neural framework that predicts spatially varying Gabor parameters to provide an image-adaptive and interpretable representation of local structures.
- We construct corresponding dual frames and establish sufficient conditions for perfect reconstruction, combining neural adaptation with an explicit reconstruction identity.
- We exploit coefficient sparsity of adaptive Gabor representations through task-specific formulations for self-supervised denoising, inpainting, and SR, achieving competitive performance.

2 Related Work

Analytic Gabor frames: Gabor frame theory provides conditions for stable signal decomposition and reconstruction using localized, modulated windows (Daubechies et al., 1995; Ron & Shen, 1997; Gröchenig, 2001). This explicit spatial-frequency structure is useful for texture analysis and image representation (Manjunath & Ma, 1996; Lee, 1996). Wavelets provide multiscale decompositions, while curvelets, contourlets, and shearlets introduce directional elements to capture image geometry (Daubechies, 1992; Candès & Donoho, 2004; Do & Vetterli, 2005; Kutyniok & Labate, 2012). Multiscale discrete Gabor constructions further combine directional selectivity with frame-based image recovery (Ji et al., 2017; 2019). These analytic designs provide reconstruction guarantees for prescribed systems. GaborCanvas learns spatially varying Gabor parameters and constructs compatible dual frames, extending the use of this structure to image-dependent representations.

Learned dictionaries and frames: Dictionary-learning methods such as K-SVD learn discrete atoms for sparse image-patch representation by alternating between sparse coding and dictionary

updates (Aharon et al., 2006). Data-driven tight frames learn filters from an observed image under a tight-frame constraint, combining image adaptation with an explicit convolution-based reconstruction formula (Cai et al., 2014; Bao et al., 2015). In these works, the learned dictionary or filter bank is shared across spatial locations, while the coefficients vary with local image content. Gabor CNNs introduce Gabor structure into neural feature extraction, using fixed or learned Gabor parameters to capture directional and frequency-selective patterns (Luan et al., 2018; Yuan et al., 2022). GaborCanvas learns spatially varying Gabor atoms by predicting their parameters at each pixel through a neural network and constructs the corresponding dual frame for perfect reconstruction.

Learning neural image representations: Neural image representations encode image content through learned NN weights. Deep Image Prior (DIP) fits a CNN to a noisy image, using the architecture as a prior for denoising (Ulyanov et al., 2018). Implicit neural representations (INR) map spatial coordinates to image values, with Fourier features and sinusoidal activations to capture fine details and high-frequency patterns (Tancik et al., 2020; Sitzmann et al., 2020). Particularly relevant to our work, WIRE introduces complex Gabor activations to exploit spatial-frequency localization for image reconstruction and restoration (Saragadam et al., 2023). GaborCanvas instead learns an explicit image-adaptive Gabor frame and constructs its corresponding dual, providing perfect reconstruction and direct access to representation coefficients for task-specific processing.

Gaussian splatting: Gaussian splatting represents scenes and images using localized Gaussian primitives (Kerbl et al., 2023; Zhang et al., 2024). Recent methods incorporate neural parameter prediction and extend these representations to continuous super-resolution (Zeng et al., 2025; Peng et al., 2026). However, Gaussian primitives have smooth, non-oscillatory profiles and may require many components to represent sharp edges and oscillatory textures. Gabor modulation introduces local frequency and phase control, allowing individual atoms to capture oscillatory intensity variations, as also explored in 3D Gabor Splatting (Watanabe et al., 2025). None of these splatting methods establishes a perfect reconstruction property. GaborCanvas learns a spatially varying Gabor frame and explicitly constructs its corresponding dual to guarantee perfect reconstruction under the stated conditions. This guarantee follows from the frame and dual construction, preserving image information before task-specific processing.

3 Main results

Let $\mathcal{H}$ be a Hilbert space. A collection $\{\phi_k\}_{k\in\mathcal{I}} \subset \mathcal{H}$ is called a *frame* for $\mathcal{H}$ if

$$A\|f\|_{\mathcal{H}}^2 \le \sum_{k\in\mathcal{I}} |\langle f, \phi_k\rangle|^2 \le B\|f\|_{\mathcal{H}}^2, \qquad \forall f \in \mathcal{H},$$

where the constants $0 < A \le B$ are referred to as the lower and upper frame bounds, respectively. Unlike a basis, a frame may be redundant. A system $\{\widetilde{\phi}_k\}_{k\in\mathcal{I}}$ is called a *dual frame* of $\{\phi_k\}_{k\in\mathcal{I}}$ if

$$f = \sum_{k\in\mathcal{I}} \langle f, \widetilde{\phi}_k\rangle \phi_k, \qquad \forall\, f \in \mathcal{H}.$$

3.1 Adaptive Gabor frames in the continuum setting

Let $g(x) = \exp(-x^2/\sigma_0^2)$ with $\sigma_0 > 0$. A classical lattice Gabor system takes the form

$$\{g(x - an)e^{2\pi imbx}\}_{n,m\in\mathbb{Z}}, \tag{3}$$

where $a, b > 0$ are the translation and modulation steps. Consider $\mathcal{H} = L^2([0, L])$ for an integer $L \ge 1$, with $\langle f, h\rangle = \int_0^L f(x)\overline{h(x)}dx$. Setting $a = 1$ and $b = 1/2$ gives a stationary Gabor system

$$\{g(x - n)e^{\pi imx}\}_{0\le n<L,\ m\in\mathbb{Z}}. \tag{4}$$

A finite-interval adaptation of the fiberization argument of Ron & Shen (1997) shows that this system is a frame for $L^2([0, L])$. Proposition 1 in Appendix A.1 gives a self-contained proof.

To align the Gabor system with any input function, we allow the scale to vary with location and the frequency to vary with both indices, which define the system as

$$\mathcal{G} = \left\{ g_{n,m}(x) := \exp\left(-\frac{(x-n)^2}{\sigma_n^2}\right) e^{2\pi i w_{n,m} x} \right\}_{0\le n<L,\ m\in\mathbb{Z}}, \tag{5}$$

where

$$\sigma_n = \sigma_0 + \delta_n, \qquad w_{n,m} = \frac{m}{2} + \epsilon_{n,m}.$$

The real offsets δ_n and $\epsilon_{n,m}$ adapt the reference scales and frequencies to the function.

Theorem 1 (Frame preservation under parameter adaptation)**.** *For $L \geq 1$ and $\sigma_0 > 0$, define*

$$\Delta := \sup_{0 \leq n < L, m \in \mathbb{Z}} |\epsilon_{n,m}|, \qquad \rho := \sup_{0 \leq n < L} |\delta_n|. \tag{6}$$

Let A^*, B^* denote the frame bounds of the Gabor frame defined by equation 4. Then, if

$$\rho \leq \frac{\sigma_0}{16}\sqrt{\frac{A^*}{B^*}}, \qquad \Delta \leq \frac{1}{16\pi L}\sqrt{\frac{A^*}{B^*}}, \tag{7}$$

the nonstationary and non-uniform system defined by equation 5 forms a frame for $L^2([0, L])$:

$$\frac{A^*}{4}\|f\|_2^2 \leq \sum_{n=0}^{L-1}\sum_{m \in \mathbb{Z}} |\langle f, g_{n,m}\rangle|^2 \leq 4B^*\|f\|_2^2, \qquad \forall\, f \in L^2([0, L]). \tag{8}$$

Consequently, this frame admits a dual frame $\widetilde{\mathcal{G}} := \{\widetilde{g}_{n,m}\}_{n,m}$ satisfying:

$$f = \sum_{0 \leq n < L, m \in \mathbb{Z}} \langle f, \widetilde{g}_{n,m}\rangle g_{n,m} = \sum_{0 \leq n < L, m \in \mathbb{Z}} \langle f, g_{n,m}\rangle \widetilde{g}_{n,m}, \quad \forall\, f \in L^2([0, L]).$$

Proof. The complete proof is provided in Appendix A.1. □

The theorem permits input-dependent scales and frequencies within a sufficient admissible range, while retaining perfect reconstruction.

3.2 Adaptive Gabor frames in the discrete setting

Let $L \geq 1$ and $K > 1$ be integers. For $f \in \mathbb{C}^L$, let $f_0 \in \ell^2(\mathbb{Z})$ be its zero extension:

$$f_0(j) = \begin{cases} f(j), & 0 \leq j < L, \\ 0, & \text{otherwise.} \end{cases}$$

Let $I_K = \{-K, \ldots, K-1\}$. For each location n, define

$$R_n : \mathbb{C}^L \to \mathbb{C}^{2K}, \quad (R_n f)(r) = f_0(n + r), \quad r \in I_K,$$

and its adjoint embeds a local vector into the full signal:

$$R_n^* : \mathbb{C}^{2K} \to \mathbb{C}^L, \quad (R_n^* u)(j) = u(j - n), \text{if } (j - n) \in I_K, \text{ and 0 otherwise.}$$

Define the local Gabor atoms by

$$g_0(r) = \exp(-\frac{r^2}{\sigma_0^2}), \quad \psi_{n,m}(r) = \frac{1}{\sqrt{2K}} g_0(r) e^{2\pi i m r/(2K)}, \quad r \in I_K, \quad 0 \leq m < 2K. \tag{9}$$

The corresponding atoms for signal space $\mathbb{C}^L$ are $\overline{\phi}_{n,m} = R_n^* \psi_{n,m}$. Explicitly, the stationary discrete Gabor system is

$$\overline{\mathcal{G}} = \{\overline{\phi}_{n,m} : -K + 1 \leq n \leq L + K - 1,\ 0 \leq m < 2K\}, \tag{10}$$

where

$$\overline{\phi}_{n,m}(j) = \frac{\mathbf{1}_{I_K}(j - n)}{\sqrt{2K}} \exp(-\frac{(j-n)^2}{\sigma_0^2}) e^{2\pi i m(j-n)/(2K)}, \qquad 0 \leq j < L. \tag{11}$$

Similarly, for each index n, we adapt the scale and frequencies of its associated discrete Gabor atom. This leads to a non-stationary discrete Gabor system with non-uniform frequencies defined by

$$\mathcal{G} = \{\phi_{n,m} : -K + 1 \leq n \leq L + K - 1,\ 0 \leq m < 2K\} \tag{12}$$

where

$$\phi_{n,m}(j) = \frac{\mathbf{1}_{I_K}(j - n)}{\sqrt{2K}} \exp\left(-\frac{(j-n)^2}{\sigma_n^2}\right) e^{2\pi i w_{n,m}(j-n)}, \qquad 0 \leq j < L, \tag{13}$$

and

$$\sigma_n = \sigma_0 + \delta_n, \quad w_{n,m} = \frac{m}{2K} + \epsilon_{n,m}.$$

Theorem 2 (Discrete adaptive Gabor frame and its dual)**.** *Consider the system defined in equation 13. If*

$$\Delta := \max_{\substack{n\in\mathcal{N}\\ 0\le m<2K}} |\epsilon_{n,m}| < \frac{1}{4K}, \tag{14}$$

then $\mathcal{G}$ is a frame for $\mathbb{C}^L$. Define the local analysis matrix by

$$G_n = [\psi_{n,0}, \dots, \psi_{n,2K-1}] \in \mathbb{C}^{(2K)\times(2K)},$$

where $\psi_{n,m}(r) = \frac{1}{\sqrt{2K}} \exp(-\frac{r^2}{\sigma_n^2}) e^{2\pi i w_{n,m} r}$, $r \in I_K$. Each G_n is invertible, and we define its associated local dual atoms by

$$[\widetilde{\psi}_{n,0}, \dots, \widetilde{\psi}_{n,2K-1}] = (G_n^{-1})^*, \quad -K+1 \le n \le L+K-1.$$

A dual frame of $\mathcal{G}$ for $\mathbb{C}^L$ is then given by $\widetilde{\phi}_{n,m} = \frac{1}{2K} R_n^* \widetilde{\psi}_{n,m}$. In particular,

$$f = \sum_{n=-K+1}^{L+K-1} \sum_{m=0}^{2K-1} \langle f, \widetilde{\phi}_{n,m}\rangle \phi_{n,m} = \sum_{n=-K+1}^{L+K-1} \sum_{m=0}^{2K-1} \langle f, \phi_{n,m}\rangle \widetilde{\phi}_{n,m}, \quad \forall\, f \in \mathbb{C}^L. \tag{15}$$

Proof. The complete proof is provided in Appendix A.2. □

Unlike the continuum condition, the discrete result requires only positive scales and sufficiently separated local frequencies. The dual is obtained by local matrix inversion followed by constant overlap averaging; it need not be the canonical dual of the Gabor frame. The identity holds for every admissible parameter choice, including one selected by an NN.

Extension to two-dimensional images: The construction in Theorem 2 can be extended to two-dimensional images by vectorizing each finite local region. The local analysis matrix is

$$G_p = [\overrightarrow{\psi_{p,1}}, \dots, \overrightarrow{\psi_{p,M}}] \in \mathbb{C}^{d\times M},$$

where the columns are the vectorized local Gabor atoms and d is the number of pixels in each region.

Extension to real-valued cosine atoms with local redundancy: The construction in Theorem 2 also applies to any local system spanning $\mathbb{R}^{2K}$. For example, replacing complex exponential atoms by local cosine atoms, $\left\{ a_m \cos(\frac{\pi m(r+K'+\frac{1}{2})}{2K'}), r \in I_K \right\}_{0\le m<2K'}$. It forms a DCT-II orthonormal basis when $K' = K$ and a redundant frame when $K' > K$. Multiplication by a positive Gaussian window preserves full row rank of its associated local analysis matrix G_n. It therefore has the local dual formed by the columns of $\widetilde{G}_n = (G_n G_n^\top)^{-1} G_n$. The same embedding and overlap normalization yield a real-valued Gabor frame for $\mathbb{R}^L$ with perfect reconstruction. Its parameters can adapt to image content, provided that the local matrices retain full row rank.

3.3 Learning image-adaptive Gabor frames

Real-valued local Gabor atoms with orientation selectivity: Let Λ denote the sampled image locations. At each $p \in \Lambda$, we define M Gabor atoms on a $q \times q$ local region, with coordinates $u \in [-1,1]^2$ measured relative to its center. Each atom combines an oriented elliptical Gaussian window with a cosine modulation:

$$\psi_{p,m}(u) = \exp\left[-\frac{1}{2} \left(\frac{(e_{p,1}^\top u)^2}{s_{p,1}^2} + \frac{(e_{p,2}^\top u)^2}{s_{p,2}^2} \right) \right] \cos\big(2\pi \omega_{p,m}^\top u + \varphi_{p,m}\big), \quad m = 1, \dots, M, \tag{16}$$

where $e_{p,1} = (\cos\beta_p, \sin\beta_p)^\top$ and $e_{p,2} = (-\sin\beta_p, \cos\beta_p)^\top$ are the principal directions of the window, and $s_{p,1}, s_{p,2} > 0$ control its extent along these directions. The elliptical shape is useful for image edges, which often extend along one direction while varying sharply across the other. Adapting the orientation and scales allows the window to extend along an edge while remaining localized across it. The frequency vector $\omega_{p,m}$ specifies the direction and rate of oscillation, while $\varphi_{p,m}$ specifies its phase. All atoms at a location share the same Gaussian window but have individual frequencies and phases, allowing them to describe different oscillatory patterns within a local region.

Dual-frame construction: Let $R_p : \mathbb{R}^{N_1 \times N_2} \to \mathbb{R}^{q^2}$ extract and vectorize the local region at p, and let $R_p^* : \mathbb{R}^{q^2} \to \mathbb{R}^{N_1 \times N_2}$ embed the vector back into the image with zero padding. For each location p, arrange the associated Gabor atoms into the analysis matrix

$$G_p = \left[\overrightarrow{\psi_{p,1}}, \ldots, \overrightarrow{\psi_{p,M}}\right] \in \mathbb{R}^{q^2 \times M},$$

where the arrow denotes vectorization. We choose $M > q^2$ for redundancy. Following the local-dual construction in Theorem 2, we require G_p to have full row rank, which ensures that $G_p G_p^\top$ is invertible. The local dual synthesis matrix is

$$\left[\overrightarrow{\widetilde{\psi}_{p,1}}, \ldots, \overrightarrow{\widetilde{\psi}_{p,M}}\right] = \widetilde{G}_p = (G_p G_p^\top)^{-1} G_p.$$

Its columns $\{\overrightarrow{\widetilde{\psi}_{p,m}}\}$ are then the vectorized local dual atoms, and

$$G_p \widetilde{G}_p^\top = \widetilde{G}_p G_p^\top = I_{q^2}.$$

Once the parameters of Gabor frames are determined, for every image f, we have

$$\left\{c_{p,m} = \langle R_p f, \overrightarrow{\psi_{p,m}} \rangle\right\}_{p \in \Lambda, 1 \le m \le M}, \quad f = \frac{1}{q^2} \sum_{p \in \Lambda} \sum_{m=1}^{M} c_{p,m}(f)\, R_p^* \overrightarrow{\widetilde{\psi}_{p,m}}. \tag{17}$$

Thus, the predicted Gabor atoms and their constructed duals preserve the image.

Predicting and learning Gabor parameters: A single NN f_θ predicts the Gabor parameters at all spatial locations. For each location p, it takes the local image region v_p as input and outputs

$$\Theta_p = f_\theta(v_p) = \left(\beta_p, s_{p,1}, s_{p,2}, \{\omega_{p,m}, \varphi_{p,m}\}_{m=1}^M\right). \tag{18}$$

The same network weights θ are shared across all locations; the predicted parameters vary because the local inputs differ. Training across sampled regions therefore learns a common prediction rule that produces Gabor atoms adapted to the image content at each location.

Because the adaptive Gabor frame and its dual provide perfect reconstruction, minimizing image reconstruction error alone cannot guide the learning of θ. We instead seek frames that align with image structures and yield sparse coefficients, providing a learned sparsity prior for image restoration. The predictor is trained through task-specific self-supervised objectives, as detailed in Section 4.

4 Experiments on unsupervised image restoration

We evaluate adaptive Gabor sparsity for denoising, SR, and inpainting, with all restoration models trained without clean targets. Architecture and optimization details are given in Appendix D.

4.1 Image denoising

We denoise adaptive Gabor coefficients using learnable soft thresholding (Donoho, 1995):

$$\mathcal{T}_\lambda(a) = \operatorname{sign}(a) \max(|a| - \lambda, 0), \qquad \lambda \ge 0.$$

For a noisy image v, the predicted Gabor frame gives coefficients $c_{p,m} = \langle R_p v, \overrightarrow{\psi_{p,m}} \rangle$. For denoising, θ denotes the trainable parameters of both the Gabor parameter predictor and the thresholding module. We threshold the coefficients and define the denoising map using the corresponding dual atoms:

$$d_\theta(v) = \frac{1}{q^2} \sum_{p \in \Lambda} \sum_{m=1}^{M} \mathcal{T}_{\sigma_{p,m}\tau_{p,m}}(c_{p,m})\, R_p^* \overrightarrow{\widetilde{\psi}_{p,m}}, \tag{19}$$

where $\sigma_{p,m}$ is the estimated coefficient noise standard deviation, and $\tau_{p,m} \ge 0$ is predicted by a thresholding module trained jointly with the frame predictor. Training uses the recorrupted-to-recorrupted (R2R) loss (Pang et al., 2021), defined in Eq. 45 in Appendix D.

Table 1: Mean PSNR (dB)/SSIM for color AWGN denoising. Bold/underlined: best/second-best.

Category	Method	Kodak24		McMaster18	
		$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$	$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$
Traditional	TV (Rudin et al., 1992)	32.24/.878	27.64/.745	33.18/.895	28.68/.800
	Framelet+LDCT (Cai et al., 2009)	32.75/.892	27.77/.720	33.75/.909	28.54/.752
	DT-CWT (Selesnick et al., 2005)	32.31/.878	27.15/.717	33.26/.899	27.97/.766
	DST (Lim, 2010)	32.95/.897	28.13/.758	33.89/.909	29.02/.770
	Gabor (Ji et al., 2017)	31.96/.871	26.14/.676	32.42/.880	26.80/.727
Self(un)-supervised	PD-Denoising (Zhou et al., 2020)	32.10/.893	28.99/.802	30.64/.865	28.20/.789
	ScoreDVI (Cheng et al., 2023)	31.75/.854	28.86/.771	32.38/.885	28.87/.780
	MASH (Chihaoui & Favaro, 2024)	24.25/.627	25.09/.646	26.28/.740	27.01/.758
	R2R (Pang et al., 2021)	32.34/.861	29.23/.793	32.79/.865	29.69/.797
	NLR (Guo et al., 2025)	28.87/.840	29.12/.810	30.43/.883	**30.12/.853**
	Pixel2Pixel (Ma et al., 2025)	32.71/.906	29.31/.818	33.88/.916	29.50/.782
	ZS-HPD (Zhao et al., 2026)	33.41/.920	29.78/.835	34.49/.921	29.89/.822
	Ours	**35.31/.937**	**30.03/.837**	**34.95/.929**	29.93/.835

Table 2: Real-noise denoising: PSNR (dB)/SSIM. Bold/underlined: best/second-best within each category. Params: trainable parameters (M).

Category	Method	SIDD Val		SIDD Ben		FMD		Params (M)
Traditional	BM3D (Dabov et al., 2007)	30.83	/.657	34.07	/.662	31.80	**/.878**	–
	TV (Rudin et al., 1992)	**33.52**	**/.842**	**34.60**	**/.84**	**33.23**	/.863	–
	Gabor (Ji et al., 2017)	32.67	/.794	33.88	/.80	33.23	/.857	–
Supervised	DnCNN (Zhang et al., 2017)	37.73	/.943	37.61	/.941	27.36	/.554	0.668
	MPRNet (Zamir et al., 2021)	39.71	/.958	–		31.06	/.785	20.100
	Restormer (Zamir et al., 2022)	**39.93**	**/.960**	**40.02**	**/.960**	**31.88**	**/.803**	26.112
Self(un)-supervised	Noise2Void (Krull et al., 2019)	27.06	/.651	26.99	/.652	32.05	/.794	0.926
	NBR2NBR (Huang et al., 2021)	27.94	/.604	27.90	/.679	33.46	/.869	1.259
	CVF-SID (Neshatavar et al., 2022)	34.81	**/.944**	34.71	/.917	32.73	/.843	1.186
	R2R (Pang et al., 2021)	35.04	/.844	35.88	/.863	–		0.669
	APBSN (Lee et al., 2022)	36.73	/.878	36.91	**/.931**	31.99	/.836	3.656
	DIP (Ulyanov et al., 2018)	32.11	/.740	–		32.90	/.854	13.400
	PD-Denoising (Zhou et al., 2020)	33.97	/.820	–		33.01	/.856	0.715
	Self2Self (Quan et al., 2020)	29.46	/.595	29.51	/.651	30.76	/.695	1.000
	ZS-N2N (Mansour & Heckel, 2023)	25.59	/.565	30.19	/.428	31.65	/.767	0.020
	ScoreDVI (Cheng et al., 2023)	34.75	/.856	35.39	/.859	33.10	/.865	13.500
	TTT-MIM (Mansour et al., 2024)	33.22	/.638	34.91	/.805	31.41	/.819	7.800
	MASH (Chihaoui & Favaro, 2024)	35.06	/.851	34.80	/.814	33.71	**/.882**	1.000
	Pixel2Pixel (Ma et al., 2025)	30.29	/–	31.87	/.631	33.26	/.830	0.200
	NLR (Guo et al., 2025)	36.32	/.874	36.81	/.874	33.47	**/.882**	0.500
	ZS-HPD (Zhao et al., 2026)	35.57	/.870	–		32.54	/.830	2.200
	Ours	**37.05**	/.888	**37.43**	/.889	**33.87**	/.880	0.592

Gaussian white noise: Table 1 compares denoising on 256×256 center crops from Kodak24 and McMaster18 with $\sigma_{\mathrm{n}} \in \{10, 25\}$. Among GT-free methods, GaborCanvas leads in three settings, with gains up to 1.90 dB on Kodak24 at $\sigma_{\mathrm{n}} = 10$. These margins exceed its 0.19 dB deficit to NLR in the remaining setting, McMaster18 at $\sigma_{\mathrm{n}} = 25$.

Real-world noise: As shown in Table 2, GaborCanvas achieves the highest GT-free PSNR across all three evaluations on SIDD smartphone images (Abdelhamed et al., 2018) and FMD microscopy images (Zhang et al., 2019), although some competitors obtain higher SSIM.

Visual comparisons in Appendix E show fewer artifacts and better detail preservation.

Table 3: SR on Set14. Bold/underlined: best/2nd-best GT-free results. PSNR/SSIM use luminance with s-pixel border crop; LPIPS-VGG uses predicted luminance and bicubic chroma

Method	×2			×4			Params (M)
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
Bicubic (Keys, 1981)	30.315	0.8809	0.1899	26.082	0.7212	0.3540	–
Supervised external training							
SwinIR (Liang et al., 2021)	34.430	0.9249	0.1302	29.110	0.7951	0.2903	11.9
LIIF-EDSR (Chen et al., 2021)	33.662	0.9267	0.1412	28.618	0.7995	0.3069	1.6
LTE-EDSR (Lee & Jin, 2022)	33.721	0.9271	0.1414	28.648	0.8004	0.3062	1.7
ContinuousSR (Peng et al., 2026)	–	–	–	28.410	0.7955	0.2927	44.9
Pretrained diffusion priors							
DDNM (Wang et al., 2023)	31.838	0.9236	0.1155	26.198	0.7589	0.2985	–
DiffPIR (Zhu et al., 2023)	25.685	0.6525	0.3869	22.738	0.5057	0.5058	–
DAPS (Zhang et al., 2025)	30.742	0.8807	0.1763	25.668	0.7055	0.3577	–
GT-free single-image methods							
DIP (Ulyanov et al., 2018)	31.133	0.8808	0.1943	<u>26.643</u>	0.7230	0.3616	13.4
DD (Heckel & Hand, 2019)	29.437	0.8523	0.2642	26.466	<u>0.7387</u>	<u>0.3545</u>	0.100
PIP/Fourier INR (Tancik et al., 2020)	<u>31.177</u>	<u>0.8962</u>	<u>0.1723</u>	26.443	0.7235	0.3570	0.231
INCODE (Kazerouni et al., 2024)	26.404	0.7147	0.3251	21.062	0.3623	0.5811	0.437
I-INR (Haider et al., 2025)	25.395	0.6317	0.3798	17.778	0.2349	0.6244	0.274
SL^2A-INR (Rezaeian et al., 2025)	25.971	0.7183	0.3442	16.740	0.2606	0.6283	0.215
SuperF-single (Jyhne et al., 2026)	28.441	0.8187	0.1876	23.413	0.5586	0.4309	0.198
MSA (Han et al., 2026)	28.005	0.8087	0.2766	23.365	0.5732	0.4966	0.215
GaborCanvas (Ours)	**31.368**	**0.9064**	**0.1665**	**26.653**	**0.7509**	**0.3412**	0.118

4.2 Self-supervised super-resolution

For SR, we extract training regions $\{u_i\}$ from the input image y and downsample them to form self-supervised training pairs $\left\{\left((u_i)\downarrow_s,\, u_i\right)\right\}_i$, where $\downarrow_s$ denotes downsampling by factor s. The reconstruction map uses a smoothed ℓ_1 penalty on coefficients of the frame predicted by f_θ:

$$g_\theta(v) \in \arg\min_{\widehat{u}} \|\widehat{u}\downarrow_s - v\|_2^2 + \lambda \| \{\langle \widehat{u},\, \phi_{n,m}(f_\theta(v))\rangle\}_{n,m} \|_{1,\epsilon}. \tag{20}$$

Then, the Gabor-parameter NN f_θ is trained by

$$\min_\theta \sum_i \|g_\theta((u_i)\downarrow_s) - u_i\|_2^2 .$$

At inference, we apply equation 20 to y. Further details are given in Appendix D.

We evaluate on Set14 (Bevilacqua et al., 2012) at $\times 2$ and $\times 4$, using only the low-resolution luminance image. GaborCanvas is the best performer among all GT-free methods. Visual comparisons are provided in Appendix E.

4.3 Single-image inpainting

We evaluate inpainting with 30% or 50% of the pixels randomly removed from each Set14 image. Following the same self-supervised strategy as in SR, we construct training pairs by further removing a subset of the available pixel values from the input image. The reconstructor g_θ is also implemented by regularizing the estimate with a sparsity prior on the coefficients of the predicted Gabor frame. Then, we train the Gabor-parameter predictor by minimizing the mean squared reconstruction error only at those further masked-out pixels. GaborCanvas achieves the best results among GT-free methods at both mask ratios. Visual comparisons are provided in Appendix E.

Table 4: Inpainting on Set14. Bold/underlined: best/2nd-best GT-free results.

Method	30% missing			50% missing			Params (M)
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
Pretrained diffusion priors							
DDNM (Wang et al., 2023)	35.853	0.9674	0.0325	32.112	0.9329	0.0741	–
DiffPIR (Zhu et al., 2023)	33.517	0.9523	0.0604	29.606	0.8994	0.1263	–
DAPS (Zhang et al., 2025)	30.384	0.8960	0.1306	26.406	0.7898	0.2412	–
GT-free single-image methods							
DIP (Ulyanov et al., 2018)	<u>34.046</u>	<u>0.9543</u>	<u>0.0670</u>	30.168	0.9038	0.1423	13.400
DD (Heckel & Hand, 2019)	32.689	0.9498	0.0760	29.787	<u>0.9061</u>	<u>0.1389</u>	0.215
Fourier INR (Tancik et al., 2020)	31.743	0.9174	0.1315	28.324	0.8373	0.2323	0.248
INCODE (Kazerouni et al., 2024)	33.737	0.9466	0.0840	<u>30.366</u>	0.8941	0.1596	0.207
I-INR (Haider et al., 2025)	33.667	0.9475	0.0795	29.993	0.8910	0.1626	0.274
SL^2A-INR (Rezaeian et al., 2025)	33.141	0.9376	0.0998	29.699	0.8782	0.1814	0.330
MSA (Han et al., 2026)	30.037	0.9004	0.1637	26.570	0.8016	0.2891	0.133
Ours	**34.622**	**0.9620**	**0.0586**	**30.798**	**0.9178**	**0.1256**	0.041

4.4 Ablation studies

Stationary versus non-stationary Gabor frames. We compare three variants. The predefined Gabor frame uses the same fixed local atoms at every pixel. The learned stationary frame optimizes one set of local atoms shared by all pixels. Our non-stationary frame instead learns different local atoms for different pixels, adapting them to local image structure. The predefined parameters also initialize both learned variants.

To evaluate the resulting sparse representation, we use image denoising with coefficient thresholding (Donoho, 1995; Elad, 2006). All variants use 96 atoms, identical noise realizations, and a common normalized threshold tuned for each case. Keeping the threshold fixed across variants isolates the contribution of adapting the Gabor frame parameters. Table 5 shows that learning a stationary frame improves PSNR over the predefined frame by around 0.15 dB. Our non-stationary frame provides a further gain of around 0.5 dB across all four settings, consistent with SSIM. These results suggest that stationary frames require a compromise among diverse image structures, whereas location-dependent atoms of non-stationary frames provide a more effective sparse representation.

Additional evaluation of the efficiency of sparse representation, in terms of k-term approximation accuracy, is provided in Appendix B.1. Some learned local atoms are visualized in Appendix C.2.

Other ablation studies. Four further ablation studies are presented in Appendix B.

Table 5: Spatial-adaptation ablation (mean PSNR/SSIM). All variants use $\tau = 1.4$ at $\sigma_{\mathrm{n}} = 10$ and $\tau = 1.6$ at $\sigma_{\mathrm{n}} = 25$. The learned variants use 800 updates.

	Kodak24		McMaster18	
Variant	$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$	$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$
Predefined Gabor frame	33.686/.9094	28.692/.7685	33.635/.9081	28.841/.7835
Learned stationary Gabor frame	33.848/.9124	28.833/.7753	33.811/.9112	28.999/.7902
Learned nonstationary Gabor frame	**34.370/.9202**	**29.329/.7954**	**34.338/.9178**	**29.551/.8078**

5 Conclusion and future work

We presented GaborCanvas, an image-adaptive Gabor-frame model with interpretable spatial-frequency structure and exact reconstruction under the stated conditions. Results on self-supervised denoising, super-resolution, and inpainting support the value of local adaptation and coefficient sparsity. Future work will extend the framework to other imaging problems.

AI USE STATEMENT

We used generative AI tools to improve the clarity and conciseness of the manuscript, particularly to fit the main text within the nine-page limit. For mathematical content, AI assistance focused on polishing the presentation of author-developed proofs to make them more accessible to readers unfamiliar with the relevant mathematical background. The authors verified the mathematical arguments, reviewed and revised all AI-assisted text, and take full responsibility for the final manuscript.

ETHICS STATEMENT

This work studies mathematical tools for image restoration using existing benchmark datasets. We do not identify specific ethical concerns arising from the methodology or experiments presented in this paper.

REPRODUCIBILITY STATEMENT

The main paper describes the proposed representation and task-specific restoration objectives. Complete proofs are provided in Appendix A, and network architecture, training procedures, and optimization settings are detailed in Appendix D. We will release the implementation on GitHub upon acceptance of the paper.

REFERENCES

Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1692–1700, 2018.

Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11): 4311–4322, 2006.

Chenglong Bao, Hui Ji, and Zuowei Shen. Convergence analysis for iterative data-driven tight frame construction scheme. *Applied and Computational Harmonic Analysis*, 38(3):510–523, 2015.

John J Benedetto, Christopher Heil, and David F Walnut. Differentiation and the Balian-Low theorem. *Journal of Fourier Analysis and Applications*, 1(4):355–402, 1994.

Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *Proceedings of the British Machine Vision Conference*, pp. 135.1–135.10. BMVA Press, 2012. doi: 10.5244/C.26.135.

Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Split Bregman methods and frame based image restoration. *Multiscale Modeling & Simulation*, 8(2):337–369, 2009. doi: 10.1137/090753504.

Jian-Feng Cai, Hui Ji, Zuowei Shen, and Gui-Bo Ye. Data-driven tight frame construction and image denoising. *Applied and Computational Harmonic Analysis*, 37(1):89–105, 2014.

Emmanuel J. Candès and David L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise-C^2 singularities. *Communications on Pure and Applied Mathematics*, 57(2):219–266, 2004. doi: 10.1002/cpa.10116.

Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8628–8638, 2021.

Jun Cheng, Tao Liu, and Shan Tan. Score priors guided deep variational inference for unsupervised real-world single image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12891–12902, 2023.

Hamadi Chihaoui and Paolo Favaro. Masked and shuffled blind spot denoising for real-world images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3025–3034, 2024.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16 (8):2080–2095, 2007.

Ingrid Daubechies. *Ten Lectures on Wavelets*, volume 61. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

Ingrid Daubechies, Henry J. Landau, and Zeph Landau. Gabor time-frequency lattices and the Wexler–Raz identity. *Journal of Fourier Analysis and Applications*, 1(4):437–478, 1995.

John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, 1985.

Minh N. Do and Martin Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106, 2005.

David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41 (3):613–627, 1995.

Michael Elad. Why simple shrinkage is still relevant for redundant representations? *IEEE Transactions on Information Theory*, 52(12):5559–5569, 2006. doi: 10.1109/TIT.2006.885522.

Karlheinz Gröchenig. *Foundations of Time-Frequency Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, 2001. doi: 10.1007/978-1-4612-0003-1.

Qilong Guo, Tianjing Zhang, Zhiyuan Ma, and Hui Ji. Zero-shot blind-spot image denoising via cross-scale non-local pixel refilling. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Ali Haider, Muhammad Salman Ali, Maryam Qamar, Tahir Khalil, Soo Ye Kim, Jihyong Oh, Enzo Tartaglione, and Sung-Ho Bae. I-INR: Iterative implicit neural representations. *arXiv preprint arXiv:2504.17364*, 2025.

Jufeng Han, Shu Wei, Min Wu, Lina Yu, Weijun Li, Linjun Sun, Hong Qin, and Yan Pang. Implicit neural representation with multi-scale sine activation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 21567–21575, 2026. doi: 10.1609/aaai.v40i26.39305.

Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *Proceedings of the International Conference on Learning Representations*, 2019.

Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2Neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14781–14790, 2021.

Hui Ji, Zuowei Shen, and Yufei Zhao. Directional frames for image recovery: Multi-scale discrete Gabor frames. *Journal of Fourier Analysis and Applications*, 23(4):729–757, 2017.

Hui Ji, Zuowei Shen, and Yufei Zhao. Digital Gabor filters do generate MRA-based wavelet tight frames. *Applied and Computational Harmonic Analysis*, 47(1):87–108, 2019.

J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

Sander Riisøen Jyhne, Christian Igel, Morten Goodwin, Per-Arne Andersen, Serge Belongie, and Nico Lang. SuperF: Neural implicit fields for multi-image super-resolution. In *Proceedings of the International Conference on Learning Representations*, 2026.

Amirhossein Kazerouni, Reza Azad, Alireza Hosseini, Dorit Merhof, and Ulas Bagci. INCODE: Implicit neural conditioning with prior knowledge embeddings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1298–1307, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023. doi: 10.1145/3592433.

Robert G. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981. doi: 10.1109/TASSP.1981.1163711.

Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2Void: Learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137, 2019.

Gitta Kutyniok and Demetrio Labate (eds.). *Shearlets: Multiscale Analysis for Multivariate Data*. Birkhäuser, Boston, 2012.

Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1929–1938, 2022.

Tai Sing Lee. Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):959–971, 1996. doi: 10.1109/34.541406.

Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. AP-BSN: Self-supervised denoising for real-world images via asymmetric PD and blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17725–17734, 2022.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using Swin Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021.

Wang-Q Lim. The discrete shearlet transform: A new directional transform and compactly supported shearlet frames. *IEEE Transactions on Image Processing*, 19(5):1166–1180, 2010. doi: 10.1109/TIP.2010.2041410.

Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han, and Jianzhuang Liu. Gabor convolutional networks. *IEEE Transactions on Image Processing*, 27(9):4357–4366, 2018. doi: 10.1109/TIP.2018.2835143.

Qing Ma, Junjun Jiang, Xiong Zhou, Pengwei Liang, Xianming Liu, and Jiayi Ma. Pixel2Pixel: A pixelwise approach for zero-shot single image denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4614–4629, 2025.

B. S. Manjunath and Wei-Ying Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996. doi: 10.1109/34.531803.

Youssef Mansour and Reinhard Heckel. Zero-shot Noise2Noise: Efficient image denoising without any data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14018–14027, 2023.

Youssef Mansour, Xuyang Zhong, Serdar Caglar, and Reinhard Heckel. TTT-MIM: Test-time training with masked image modeling for denoising distribution shifts. In *Proceedings of the European Conference on Computer Vision*, 2024.

Reyhaneh Neshatavar, Mohsen Yavartanoo, Sanghyun Son, and Kyoung Mu Lee. CVF-SID: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2043–2052, 2021.

Long Peng, Anran Wu, Wenbo Li, Peizhe Xia, Xinjie Zhang, Xueyuan Dai, Xin Di, Haoze Sun, Renjing Pei, Yang Wang, Yang Cao, and Zheng-Jun Zha. Pixel to Gaussian: Ultra-fast continuous super-resolution with 2D Gaussian modeling. In *Proceedings of the International Conference on Learning Representations*, 2026.

Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2Self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1890–1898, 2020.

Reza Rezaeian, Moein Heidari, Reza Azad, Dorit Merhof, Hamid Soltanian-Zadeh, and Ilker Hacihaliloglu. SL^2A-INR: Single-layer learnable activation for implicit neural representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 26065–26074, 2025.

Amos Ron and Zuowei Shen. Weyl–Heisenberg frames and Riesz bases in $L_2(\mathbb{R}^d)$. *Duke Mathematical Journal*, 89(2):237–282, 1997. doi: 10.1215/S0012-7094-97-08913-4.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1–4):259–268, 1992.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. WIRE: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18507–18516, 2023.

Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005. doi: 10.1109/MSP.2005.1550194.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473, 2020.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547, 2020.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023.

Haato Watanabe, Kenji Tojo, and Nobuyuki Umetani. 3D Gabor splatting: Reconstruction of high-frequency surface texture using Gabor noise. *arXiv preprint arXiv:2504.11003*, 2025.

Robert M. Young. *An Introduction to Nonharmonic Fourier Series*, volume 93 of *Pure and Applied Mathematics*. Academic Press, revised first edition, 2001.

Ye Yuan, Li-Na Wang, Guoqiang Zhong, Wei Gao, Wencong Jiao, Junyu Dong, Biao Shen, Dongdong Xia, and Wei Xiang. Adaptive Gabor convolutional networks. *Pattern Recognition*, 124: 108495, 2022. doi: 10.1016/j.patcog.2021.108495.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5728–5739, 2022.

Zhaojie Zeng, Yuesong Wang, Tao Guan, Chao Yang, and Lili Ju. Instant GaussianImage: A generalizable and self-adaptive image representation via 2D Gaussian splatting. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27896–27905. IEEE, 2025.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20895–20905, 2025.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. GaussianImage: 1000 FPS image representation and compression by 2D Gaussian splatting. In *Proceedings of the European Conference on Computer Vision*, pp. 327–345, 2024.

Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A Poisson–Gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11710–11718, 2019.

Xiaole Zhao, Qingsong Pang, Xiaobo Zhang, Xun Xu, Xun Gong, Yan Yang, and Tianrui Li. Zero-shot image denoising via hybrid prior-guided pseudo sample generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Yuqian Zhou, Jianbo Jiao, Haibin Huang, Yang Wang, Jue Wang, Honghui Shi, and Thomas Huang. When AWGN-based denoiser meets real noises. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 13074–13081, 2020.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1219–1229, 2023.

A PROOFS

A.1 PROOF OF THEOREM 1

We first establish the frame properties of stationary Gabor systems for $L^2([0, L])$ by adapting the argument for Gabor frames in $L^2(\mathbb{R})$ (Ron & Shen, 1997). Set $q = \lceil L/2 \rceil$.

Proposition 1 (Stationary Gaussian Gabor frame)**.** *For every $\sigma_0 > 0$, the system*

$$\phi_{n,m}(x) = \exp\left(-\frac{(x-n)^2}{\sigma_0^2}\right) e^{\pi i m x}, \qquad x \in [0, L], \tag{21}$$

is a frame for $\mathcal{H}$. In particular, there is a constant $A > 0$ such that

$$A\|f\|_2^2 \le \sum_{n=0}^{L-1} \sum_{m\in\mathbb{Z}} \left|\langle f, \phi_{n,m}\rangle\right|^2 \le B\|f\|_2^2, \qquad B := L(L+1). \tag{22}$$

For each fixed n, the corresponding frequency family also satisfies

$$\sum_{m\in\mathbb{Z}} \left|\langle h, \phi_{n,m}\rangle\right|^2 \le (L+1)\,\|h\|_2^2, \qquad \forall\, h \in \mathcal{H}. \tag{23}$$

Proof. Given $f \in \mathcal{H}$, let $\widetilde{f}$ be its extension by zero outside $[0, L]$. Let $q = \lceil \frac{L}{2} \rceil$. For $t \in [0, 2]$, define

$$F(t) = \left(\widetilde{f}(t), \widetilde{f}(t+2), \ldots, \widetilde{f}(t+2(q-1))\right)^\top \in \mathbb{C}^q.$$

Since $L \le 2q$,

$$\int_0^2 \|F(t)\|_{\mathbb{C}^q}^2 \, dt = \|f\|_2^2. \tag{24}$$

Define the $L \times q$ matrix

$$M(t) = \left[\exp\left(-\frac{(t+2k-n)^2}{\sigma_0^2} \right) \right]_{\substack{0 \le n < L \\ 0 \le k < q}}. \tag{25}$$

Then, we have

$$(M(t)F(t))_n = \sum_{k=0}^{q-1} \widetilde{f}(t+2k) \exp\left(-\frac{(t+2k-n)^2}{\sigma_0^2} \right).$$

Splitting the coefficient integral into intervals of length 2, and using $e^{-2\pi imk} = 1$, we have

$$\begin{aligned} \langle f, \phi_{n,m} \rangle &= \sum_{k=0}^{q-1} \int_0^2 \widetilde{f}(t+2k) \exp\left(-\frac{(t+2k-n)^2}{\sigma_0^2} \right) e^{-\pi im(t+2k)} \, dt \\ &= \int_0^2 (M(t)F(t))_n e^{-\pi imt} \, dt. \end{aligned}$$

The functions $\{2^{-1/2} e^{\pi imt}\}_{m \in \mathbb{Z}}$ form an orthonormal basis for $L^2([0,2])$. By Parseval's identity, we have

$$\sum_{m \in \mathbb{Z}} \left| \langle f, \phi_{n,m} \rangle \right|^2 = 2 \int_0^2 |(M(t)F(t))_n|^2 \, dt. \tag{26}$$

Then,

$$\sum_{n=0}^{L-1} \sum_{m \in \mathbb{Z}} \left| \langle f, \phi_{n,m} \rangle \right|^2 = 2 \int_0^2 \|M(t)F(t)\|_{\mathbb{C}^L}^2 \, dt. \tag{27}$$

We next prove a uniform lower bound for $M(t)$. Let

$$M_0(t) = \left[\exp\left(-\frac{(t+2k-n)^2}{\sigma_0^2} \right) \right]_{n,k=0}^{q-1}$$

be the square submatrix formed by its first q rows; these rows exist because $q \le L$. The identity

$$-\frac{(t+2k-n)^2}{\sigma_0^2} = \frac{2nt - n^2}{\sigma_0^2} + \frac{4nk}{\sigma_0^2} - \frac{(t+2k)^2}{\sigma_0^2}$$

gives the factorization

$$M_0(t) = D_r(t) \, V \, D_c(t),$$

where

$$D_r(t) = \mathrm{diag}_{0 \le n < q} \left(e^{(2nt - n^2)/\sigma_0^2} \right), \quad D_c(t) = \mathrm{diag}_{0 \le k < q} \left(e^{-(t+2k)^2/\sigma_0^2} \right),$$

and $V = [z_k^n]_{n,k=0}^{q-1}$ with $z_k = e^{4k/\sigma_0^2}$ is a Vandermonde matrix with distinct nodes. Hence,

$$\det V = \prod_{0 \le j < k < q} (z_k - z_j) > 0.$$

Both diagonal matrices are invertible too. Thus $M_0(t)$ is invertible for every $t \in [0,2]$. Its entries are continuous in t, so compactness gives

$$\alpha := \min_{t \in [0,2]} \lambda_{\min}(M_0(t)^* M_0(t)) > 0. \tag{28}$$

Consequently,

$$\|M(t)v\|^2 \ge \|M_0(t)v\|^2 \ge \alpha \|v\|^2 \qquad (v \in \mathbb{C}^q).$$

Every entry of $M(t)$ is at most 1, so

$$\|M(t)v\|^2 \le \|M(t)\|_{\mathrm{F}}^2 \|v\|^2 \le Lq \, \|v\|^2.$$

Combining these inequalities with equation 24 and equation 27 proves equation 22 with

$$A := 2\alpha, \qquad B := L(L+1) \geq 2Lq.$$

Finally, each row of $M(t)$ has squared Euclidean norm at most q. Applying the Cauchy–Schwarz inequality in equation 26 gives

$$\sum_{m\in\mathbb{Z}} \left|\langle f, \phi_{n,m}\rangle\right|^2 \leq 2q \int_0^2 \|F(t)\|_{\mathbb{C}^q}^2 \, dt = 2q \, \|f\|_2^2.$$

Since f was arbitrary, this also proves equation 23. □

Our discussion of frequency perturbations uses the following quantitative form of Kadec's theorem, as presented by Young (2001, Chapter 1).

Lemma 1 (Nonharmonic Fourier bounds on an interval of length two)**.** *For the system* $\{2^{-1/2}e_m\}_{m\in\mathbb{Z}}$ *with* $e_m(t) = e^{2\pi i(m/2+\varepsilon_m)t}$*, suppose*

$$d = \sup_{m\in\mathbb{Z}} |\varepsilon_m| < \frac{1}{8}.$$

Then, for $\kappa(d) = 1 - \cos(2\pi d) + \sin(2\pi d)$*, we have* $0 \leq \kappa(d) < 1$*, and* $\{2^{-1/2}e_m\}_{m\in\mathbb{Z}}$ *is a Riesz basis for* $L^2([0,2])$*. In particular, every* $h \in L^2([0,2])$ *satisfies*

$$2(1-\kappa(d))^2\|h\|_2^2 \leq \sum_{m\in\mathbb{Z}} |\langle h, e_m\rangle|^2 \leq 2(1+\kappa(d))^2\|h\|_2^2. \tag{29}$$

We now prove the main result. Recall from the proof of Proposition 1 that the square reference matrix

$$V_0(t) = \left[\exp\left(-\frac{(t+2k-n)^2}{\sigma_0^2}\right)\right]_{n,k=0}^{q-1}, \qquad t \in [0,2], \tag{30}$$

defines the constant

$$\alpha = \min_{t\in[0,2]} \lambda_{\min}(V_0(t)^* V_0(t)) > 0. \tag{31}$$

Positivity of α follows from the Vandermonde factorization given below. This is the same reference constant that gives stationary frame bounds $A = 2\alpha$ and $B = L(L+1)$.

Theorem 3 (Position-dependent scales and perturbed frequencies)**.** *For* $0 \leq n < L$ *and* $m \in \mathbb{Z}$*, let*

$$\sigma_n = \sigma_0 + \delta_n, \qquad w_{n,m} = \frac{m}{2} + \epsilon_{n,m}.$$

Define

$$\rho = \max_{0\leq n<L} |\delta_n|, \qquad \Delta = \sup_{\substack{0\leq n<L \\ m\in\mathbb{Z}}} |\epsilon_{n,m}|.$$

If

$$\rho \leq \frac{\sigma_0}{16}\sqrt{\frac{A}{B}}, \qquad \Delta \leq \frac{1}{16\pi L}\sqrt{\frac{A}{B}}, \tag{32}$$

then the Gabor system

$$\psi_{n,m}(x) = \exp\left(-\frac{(x-n)^2}{\sigma_n^2}\right) e^{2\pi i w_{n,m} x}, \qquad x \in [0,L], \tag{33}$$

is a frame for $\mathcal{H}$*, with the uniform bounds*

$$\frac{A}{4}\|f\|_2^2 \leq \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}} |\langle f, \psi_{n,m}\rangle|^2 \leq 4B\|f\|_2^2, \qquad f \in \mathcal{H}. \tag{34}$$

Moreover, define the frame operator

$$Sf = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}} \langle f, \psi_{n,m}\rangle \psi_{n,m}.$$

Then S is bounded and invertible, and the canonical dual frame *is given by*

$$\widetilde{\psi}_{n,m} = S^{-1}\psi_{n,m}.$$

Consequently, every $f \in \mathcal{H}$ admits the perfect reconstruction formulas

$$f = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}} \langle f, \psi_{n,m}\rangle \widetilde{\psi}_{n,m} = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}} \langle f, \widetilde{\psi}_{n,m}\rangle \psi_{n,m}. \tag{35}$$

Proof. Set

$$s := \sqrt{\frac{A}{B}} = \sqrt{\frac{2\alpha}{L(L+1)}} \le \sqrt{\frac{\alpha}{Lq}}, \qquad 0 < s \le 1.$$

The assumptions imply $\rho \le \sigma_0/16$ and $\Delta \le 1/(16\pi L) < 1/8$. In particular, every σ_n is positive. Moreover,

$$\kappa(\Delta) = 1 - \cos(2\pi\Delta) + \sin(2\pi\Delta) \le 4\pi\Delta \le \frac{1}{4}. \tag{36}$$

Thus the lower and upper norm factors in Lemma 1 may be replaced by $3/4$ and $5/4$.

Step 1: a uniform lower bound for the Gaussian fiber matrix. First, the entries of the reference matrix factor as

$$[V_0(t)]_{n,k} = e^{(2nt-n^2)/\sigma_0^2}\left(e^{4k/\sigma_0^2}\right)^n e^{-(t+2k)^2/\sigma_0^2}.$$

The middle factors form a Vandermonde matrix with distinct nodes e^{4k/σ_0^2}, and the remaining row and column factors are positive. Hence $V_0(t)$ is invertible for every $t \in [0,2]$. Continuity and compactness establish $\alpha > 0$ in equation 31.

Write

$$g_n(x) = e^{-(x-n)^2/\sigma_n^2}, \qquad g_n^0(x) = e^{-(x-n)^2/\sigma_0^2}.$$

Since $\sigma \ge \sigma_0/2$ along the interval between σ_n and σ_0, the derivative bound

$$\sup_{u\in\mathbb{R}}\left|\frac{\partial}{\partial\sigma}e^{-u^2/\sigma^2}\right| = \frac{2}{e\sigma} \le \frac{4}{e\sigma_0} < \frac{2}{\sigma_0}$$

and the mean value theorem give

$$\sup_{x\in\mathbb{R}} |g_n(x) - g_n^0(x)| \le \frac{2\rho}{\sigma_0}. \tag{37}$$

Now define the $L \times q$ matrix

$$M_\sigma(t) = [g_n(t+2k)]_{\substack{0\le n<L,\\ 0\le k<q}}$$

and let $V_\sigma(t)$ be its first q rows. By equation 37,

$$\|V_\sigma(t) - V_0(t)\|_{\mathrm{op}} \le \|V_\sigma(t) - V_0(t)\|_{\mathrm{F}} \le \frac{2q\rho}{\sigma_0} \le \frac{qs}{8} \le \frac{\sqrt{\alpha}}{8},$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm and we used $q \le \sqrt{Lq}$. The triangle inequality and $|g_n| \le 1$ therefore give

$$\frac{7}{8}\sqrt{\alpha}\,\|v\|_2 \le \|M_\sigma(t)v\|_2 \le \sqrt{Lq}\,\|v\|_2, \qquad v \in \mathbb{C}^q, \quad t \in [0,2]. \tag{38}$$

Step 2: apply Young's bounds to the fiber functions. For $f \in \mathcal{H}$, let f_0 denote its zero extension to $\mathbb{R}$. Set

$$F_k(t) = f_0(t+2k), \qquad F(t) = (F_0(t), \dots, F_{q-1}(t))^\top, \qquad t \in [0,2].$$

The intervals of length two partition the zero-extended signal, so

$$\int_0^2 \|F(t)\|_2^2\, dt = \|f\|_2^2. \tag{39}$$

Define

$$h_{n,k}(t) = g_n(t+2k)F_k(t), \qquad H_n(t) = \sum_{k=0}^{q-1} h_{n,k}(t) = (M_\sigma(t)F(t))_n,$$

and the nonharmonic exponentials

$$e_{n,m}(t) = e^{2\pi i(m/2+\epsilon_{n,m})t}.$$

For comparison with the actual Gabor coefficients, introduce

$$d_{n,m} = \langle H_n, e_{n,m}\rangle_{L^2([0,2])}.$$

Apply Lemma 1 separately for each n, with $d = \Delta$, and sum over n. Using equation 36, equation 38, and equation 39 yields

$$\frac{21}{32}\sqrt{A}\,\|f\|_2 \le \|d\|_{\ell^2} \le \frac{5}{4}\sqrt{B}\,\|f\|_2. \tag{40}$$

Here and below, ℓ^2 is over $\{0,\dots,L-1\}\times\mathbb{Z}$.

Step 3: control the phase differences and conclude. Splitting the coefficient integral over intervals of length two gives

$$c_{n,m} := \langle f, \psi_{n,m}\rangle = \sum_{k=0}^{q-1} e^{-4\pi i k\epsilon_{n,m}}\langle h_{n,k}, e_{n,m}\rangle_{L^2([0,2])}, \qquad d_{n,m} = \sum_{k=0}^{q-1}\langle h_{n,k}, e_{n,m}\rangle_{L^2([0,2])}. \tag{41}$$

The factor $e^{-2\pi imk}$ equals one, whereas the remaining phase satisfies

$$|e^{-4\pi i k\epsilon_{n,m}} - 1| \le 4\pi(q-1)\Delta \le 2\pi L\Delta =: r \le \frac{s}{8}.$$

Here $2(q-1) \le L$. By the triangle inequality, Young's upper bound, and Cauchy–Schwarz,

$$\begin{aligned}\|c-d\|_{\ell^2} &\le r\sum_{k=0}^{q-1}\left(\sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}}|\langle h_{n,k}, e_{n,m}\rangle|^2\right)^{1/2} \le \frac{5\sqrt{2}}{4}\,r\sum_{k=0}^{q-1}\left(\sum_{n=0}^{L-1}\|h_{n,k}\|_2^2\right)^{1/2}\\ &\le \frac{5\sqrt{2q}}{4}\,r\left(\sum_{k=0}^{q-1}\sum_{n=0}^{L-1}\|h_{n,k}\|_2^2\right)^{1/2} \le \frac{5}{4}r\sqrt{B}\,\|f\|_2 \le \frac{5}{32}\sqrt{A}\,\|f\|_2.\end{aligned}$$

We used $|g_n| \le 1$ and equation 39 in the last line. Combining it with equation 40 now gives

$$\begin{aligned}\|c\|_{\ell^2} &\ge \left(\frac{21}{32} - \frac{5}{32}\right)\sqrt{A}\,\|f\|_2 = \frac{1}{2}\sqrt{A}\,\|f\|_2,\\ \|c\|_{\ell^2} &\le \left(\frac{5}{4}\sqrt{B} + \frac{5}{32}\sqrt{A}\right)\|f\|_2 \le 2\sqrt{B}\,\|f\|_2.\end{aligned}$$

Therefore, we proved equation 34.

It remains to establish the canonical dual construction. The frame operator S is bounded, positive, and self-adjoint, with

$$\langle Sf, f\rangle = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}}|\langle f, \psi_{n,m}\rangle|^2.$$

Hence the frame bounds imply the operator inequalities

$$\frac{A}{4}I_{\mathcal{H}} \le S \le 4BI_{\mathcal{H}}.$$

Therefore, S is invertible and $\|S^{-1}\|_{\mathrm{op}} \le 4/A$. Defining $\widetilde{\psi}_{n,m} = S^{-1}\psi_{n,m}$, the identities $S^{-1}Sf = f$ and $SS^{-1}f = f$ give

$$f = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}}\langle f, \psi_{n,m}\rangle S^{-1}\psi_{n,m} = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}}\langle S^{-1}f, \psi_{n,m}\rangle\psi_{n,m} = \sum_{n=0}^{L-1}\sum_{m\in\mathbb{Z}}\langle f, S^{-1}\psi_{n,m}\rangle\psi_{n,m},$$

where the last equality uses the self-adjointness of S^{-1}. The upper frame bound ensures unconditional convergence of the synthesis series, and applying the bounded operator S^{-1} preserves this convergence. This proves equation 35. □

A.2 Proof of Theorem 2

Proof. For positive scales $\sigma_n > 0$, we first establish the invertibility of each local analysis matrix G_n. Fix n and set

$$z_{n,m} := e^{2\pi i w_{n,m}}, \qquad m = 0, \dots, 2K-1.$$

For $0 \le \ell < m < 2K$, the frequency perturbation bound gives

$$0 < \frac{1}{2K} - 2\Delta \le w_{n,m} - w_{n,\ell} \le \frac{2K-1}{2K} + 2\Delta < 1.$$

Thus $w_{n,m} - w_{n,\ell}$ is not an integer, so the nodes $z_{n,0}, \dots, z_{n,2K-1}$ are pairwise distinct.

Indexing the rows by $s = r + K = 0, \dots, 2K-1$, we factor G_n as

$$G_n = \frac{1}{\sqrt{2K}} D_n V_n Z_n,$$

where

$$D_n = \mathrm{diag}_{r \in I_K}\left(e^{-r^2/\sigma_n^2}\right), \qquad (V_n)_{s,m} = z_{n,m}^s, \qquad Z_n = \mathrm{diag}_{0 \le m < 2K}\left(z_{n,m}^{-K}\right).$$

Since $\sigma_n > 0$, all diagonal entries of D_n are strictly positive. The matrix Z_n is also invertible. Moreover, V_n is a Vandermonde matrix with

$$\det V_n = \prod_{0 \le \ell < m < 2K} (z_{n,m} - z_{n,\ell}) \neq 0.$$

Therefore, G_n is invertible.

By the definition

$$\widetilde{G}_n := [\widetilde{\psi}_{n,0}, \dots, \widetilde{\psi}_{n,2K-1}] = (G_n^{-1})^*,$$

we have

$$G_n \widetilde{G}_n^* = \widetilde{G}_n G_n^* = I_{2K}. \tag{42}$$

Hence, for every $u \in \mathbb{C}^{2K}$,

$$u = \sum_{m=0}^{2K-1} \langle u, \widetilde{\psi}_{n,m} \rangle \psi_{n,m} = \sum_{m=0}^{2K-1} \langle u, \psi_{n,m} \rangle \widetilde{\psi}_{n,m}.$$

We next establish constant coverage of the signal domain. For each $j = 0, \dots, L-1$,

$$j - n \in I_K \quad \Longleftrightarrow \quad j - K + 1 \le n \le j + K.$$

These are exactly $2K$ integer centers, all contained in $\mathcal{N} = \{-K+1, \dots, L+K-1\}$. Since

$$(R_n^* R_n f)(j) = \mathbf{1}_{I_K}(j-n) f(j),$$

it follows that

$$\sum_{n \in \mathcal{N}} R_n^* R_n = 2K I_L. \tag{43}$$

Finally, using $\phi_{n,m} = R_n^* \psi_{n,m}$ and $\widetilde{\phi}_{n,m} = (2K)^{-1} R_n^* \widetilde{\psi}_{n,m}$, we obtain, for every $f \in \mathbb{C}^L$,

$$\sum_{n \in \mathcal{N}} \sum_{m=0}^{2K-1} \langle f, \widetilde{\phi}_{n,m} \rangle \phi_{n,m} = \frac{1}{2K} \sum_{n \in \mathcal{N}} R_n^* G_n \widetilde{G}_n^* R_n f = \frac{1}{2K} \sum_{n \in \mathcal{N}} R_n^* R_n f = f.$$

Similarly, using $\widetilde{G}_n G_n^* = I_{2K}$ gives

$$\begin{aligned} \sum_{n \in \mathcal{N}} \sum_{m=0}^{2K-1} \langle f, \phi_{n,m} \rangle \widetilde{\phi}_{n,m} &= \frac{1}{2K} \sum_{n \in \mathcal{N}} R_n^* \widetilde{G}_n G_n^* R_n f \\ &= f. \end{aligned}$$

These identities show that both finite families span $\mathbb{C}^L$. They are therefore frames and are dual to each other, proving equation 15. □

B Ablation studies

B.1 $(k+1)$-term approximation of clean images

We examine whether non-stationary Gabor frames represent local image structures more accurately with fewer retained coefficients. Unlike the restoration experiments, this diagnostic trains and evaluates the representations on clean image patches.

For a clean image patch $x_p = R_p f$ at location p, let G_p contain all local atoms, including a fixed low-frequency atom and many high-frequency atoms (learned or fixed). Let $\widetilde{G}_p$ be its dual. We reconstruct the image region using the low-pass coefficient from the low-frequency atom and the k largest high-pass coefficients in magnitude from high-frequency atoms, where k denotes the number of retained high-pass coefficients.

We compare predefined atoms, learned stationary atoms shared across all locations, and learned non-stationary atoms predicted from each location. All variants use the same protocol. The non-stationary predictor uses channels $(32, 48, 64)$. Both learned variants are trained for 800 Adam updates with a learning rate of 10^{-3}, minimizing the average reconstruction MSE at $k \in \{8, 16, 32\}$. Dual-consistency and parameter-offset penalties have weights 0.01 and 10^{-4}, respectively. The predefined bank remains fixed. Training uses 1,024 nonoverlapping luminance patches from each of Kodak24 images 01–12; evaluation uses 512 patches from each of images 13–24. Test images are excluded from fitting and checkpoint selection.

Figure 2 shows that non-stationary frames consistently come with the best approximation with the same number of atoms. At $k = 16$, the predefined, learned stationary, and learned non-stationary frames achieve 16.96, 29.84, and 33.43 dB, respectively. At $k = 32$, they achieve 21.53, 33.72, and 38.38 dB. These results support the value of local adaptation for representing image content with limited coefficients. When all coefficients are retained, every variant achieves a mean patch MSE below 4×10^{-10}.

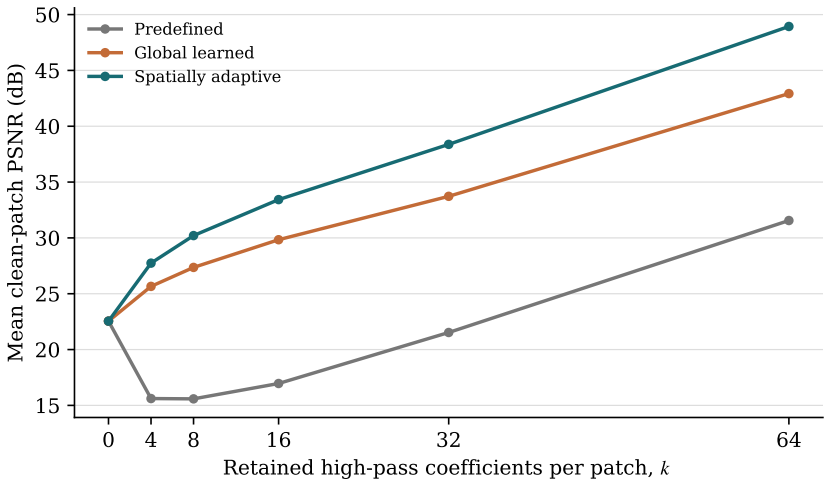

Figure 2: $(k+1)$-term approximation accuracy on held-out Kodak24 images, retaining the k largest high-pass coefficients in magnitude and the low-pass coefficient.

B.2 Image denoising: Learning Gabor versus Gaussian atoms.

We examine whether learning Gabor atoms provides a more effective image representation than learning Gaussian primitives, through the experiments of image denoising. Gaussian atoms describe localized spatial structures, while Gabor atoms additionally encode local frequency and phase to capture edges and oscillatory textures. We replace the Gabor atoms in equation 16 with learned Gaussian atoms, whose centers and anisotropic scales are optimized.

Both variants use the same denoising pipeline, including learnable coefficient thresholding, with identical atom counts, training schedules, and random seeds. Table 6 shows that replacing Gabor atoms with Gaussian atoms reduces PSNR by 2.77 and 2.99 dB at $\sigma_{\mathrm{n}} = 10$ and 25, respectively. These results support the value of frequency modulation in learning an effective sparse representation for images with edges and textures.

Table 6: Effects of atom type of the learned frame on Kodak24 (PSNR/SSIM).

Variant	$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$
Full model	**35.31/.937**	**30.03/.837**
Gaussian atoms	32.54/.888	27.04/.720

B.3 Image denoising: Location-dependent versus uniform thresholding.

We compare learned location-dependent thresholds with a single predefined normalized threshold shared across all locations and atoms. Both variants use spatially adaptive Gabor atoms. Table 7 shows that location-adaptive thresholding achieves higher PSNR by approximately 0.94 and 0.70 dB at $\sigma_{\mathrm{n}} = 10$ and 25, respectively, with higher SSIM in both settings.

Table 7: Effect of location-adaptive versus uniform thresholding on Kodak24 (PSNR/SSIM).

Variant	$\sigma_{\mathrm{n}} = 10$	$\sigma_{\mathrm{n}} = 25$
Location-adaptive thresholding	**35.31/.937**	**30.03/.837**
Uniform thresholding	34.37/.920	29.33/.795

B.4 Image inpainting: Effectiveness of the sparsity-based regularization from adaptive Gabor frames

We evaluate the effectiveness of the sparsity-based regularization from adaptive Gabor frames in image inpainting, comparing reconstruction with and without the sparsity-promoting penalty on frame coefficients. Setting $\lambda = 0$ removes this prior and returns the biharmonic initialization. Table 8 shows that the coefficient prior improves PSNR by 0.84 and 0.49 dB in the two settings. These gains support the effectiveness of adaptive Gabor coefficient sparsity in recovering missing image content beyond the initial completion.

Table 8: Inpainting effectiveness of the regularization with adaptive Gabor frames on Set14 (PSNR/SSIM).

Variant	30% missing	50% missing
with Gabor-based regularization	**34.62/.9620**	**30.80/.9178**
w/o Gabor-based regularization	33.78/.9562	30.31/.9136

B.5 Impact of local atom count on denoising accuracy and model size

We examine how the number M of local Gabor atoms affects denoising accuracy and model size. All variants use Kodak24 at $\sigma_{\mathrm{n}} = 25$, with identical noise realizations, training seeds, thresholding rules, and 200 training updates. Table 9 shows that increasing M from 80 to 96 improves PSNR from 29.80 to 30.03 dB. Increasing M further to 128 provides no additional PSNR gain while adding approximately 16,000 parameters. We therefore use $M = 96$. At $M = 64$, the learned local matrices become numerically rank-deficient and dual reconstruction is unstable; this configuration is omitted from the table.

Table 9: Denoising accuracy and model size for different numbers of local Gabor atoms (Kodak24, $\sigma_{\mathrm{n}} = 25$).

M	Params (M)	PSNR
80	0.545	29.80
96	0.553	**30.03**
128	0.569	29.98

C Additional study

C.1 Numerical stability of dual reconstruction

Our local dual construction requires the analysis matrix G_p to have full row rank, so that $G_p G_p^\top$ is invertible. In practice, reliable reconstruction also requires adequate numerical conditioning: a large condition number makes dual computation sensitive to numerical errors. We therefore examine the rank and numerical stability of the learned local frames.

We sample 512 locations from each Kodak24 image, giving 12,288 local frames per noise level. For each frame, we compute the lower frame bound and condition number:

$$A_p = \sigma_{\min}^2(G_p), \qquad \kappa(G_p) = \frac{\sigma_{\max}(G_p)}{\sigma_{\min}(G_p)}.$$

A positive A_p indicates full row rank, while a smaller $\kappa(G_p)$ indicates better conditioning. We also measure the dual identity error $\|\widetilde{G}_p G_p^\top - I\|_{\max}$, where $\|\cdot\|_{\max}$ denotes the largest absolute matrix entry, and the relative reconstruction error $\|\widehat{x}_p - x_p\|_2 / \|x_p\|_2$.

Table 10 reports statistics over the sampled locations. The 95th percentile is the value at or below which 95% of the measurements lie; the maximum identity error is the largest error across all sampled locations. All sampled frames have positive lower frame bounds. The computed duals also satisfy the reconstruction identity accurately: the maximum identity error is at most 1.2×10^{-4}, and the 95th-percentile relative reconstruction error is below 4.2×10^{-5}. These results support the numerical reliability of the learned frames and their duals in practice.

Table 10: Conditioning and numerical reconstruction accuracy of learned local Gabor frames on Kodak24. "95th pct." denotes the 95th percentile.

σ_{n}	$\min A_p$	Median $\kappa(G_p)$	95th pct. $\kappa(G_p)$	Max. identity error	95th pct. reconstruction error
10	1.48×10^{-3}	69.97	77.68	1.06×10^{-4}	4.04×10^{-5}
25	1.28×10^{-3}	67.32	82.63	1.20×10^{-4}	4.18×10^{-5}

C.2 Visualization of local Gabor atoms

Figures 3 and 4 show predicted atoms at three locations in each of two noisy images. The dominant carriers align with roof and window boundaries in `kodim08`, and with horizontal shutters, petal texture, and a thin branch in `kodim07`.

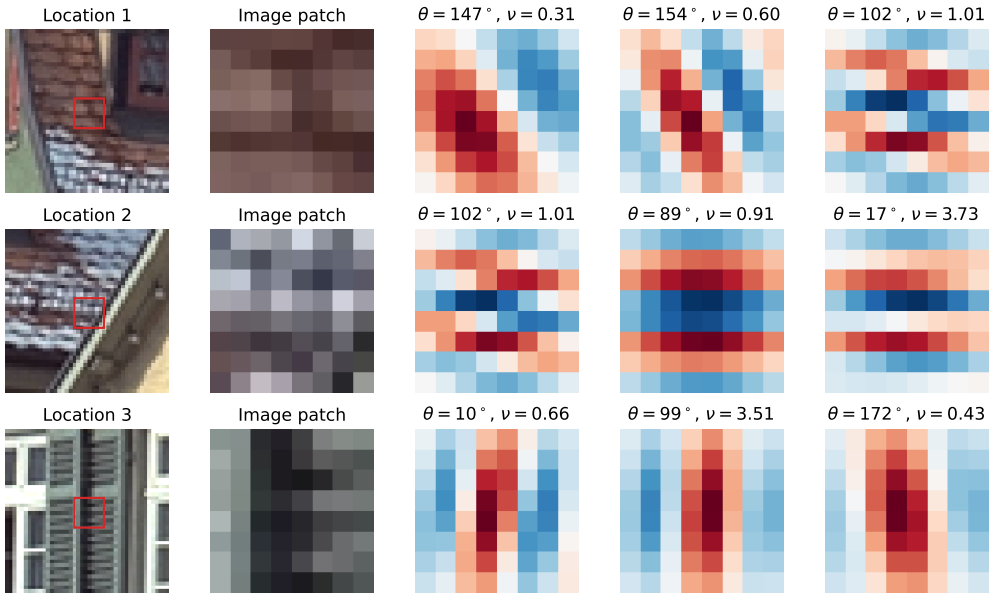

Figure 3: Predicted atoms for three patches from Kodak24 `kodim08` at $\sigma_{\mathrm{n}} = 10$. Each row shows the local image region, selected 8×8 patch, and three atoms with the largest absolute responses. Titles report orientation θ and frequency ν.

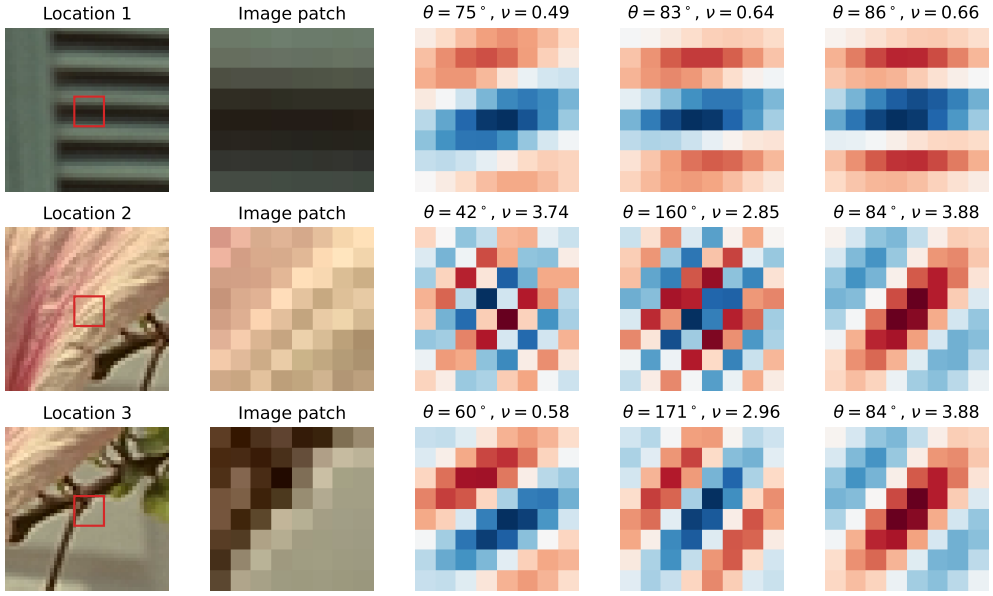

Figure 4: The same visualization for Kodak24 `kodim07`. The selected patches contain horizontal shutters, petal texture, and an oblique branch.

D IMPLEMENTATION DETAILS AND TRAINING

The predictor is a lightweight convolutional encoder shared across spatial locations. Let $\mathsf{C}_{c,s}$ denote a 3×3 convolution with c output channels and stride s, followed by a SiLU activation. The same architecture is used for all tasks:

$$h_p = \mathrm{GAP}(\mathsf{C}_{c_3,2} \circ \mathsf{C}_{c_2,2} \circ \mathsf{C}_{c_1,1}(v_p)), \qquad \Theta_p = \Theta^0 + r \odot \tanh(W h_p + b), \tag{44}$$

where Θ^0 contains the initial Gabor parameters and r sets the allowed parameter ranges. The linear heads (W, b) are initialized to zero, and all predicted atoms are normalized to unit ℓ_2 norm. For all tasks, we use reflection padding before patch extraction and average overlapping patches during folding. The local analysis matrix is formed from the atoms predicted at each location, and its corresponding dual is used for synthesis. The zero-frequency Gabor atom is an unmodulated Gaussian that supplies the low-pass component. We keep it fixed during training and inference, focusing adaptation on the modulated atoms. Reported atom counts exclude this fixed low-pass atom.

D.1 IMAGE DENOISING

We use $(c_1, c_2, c_3) = (32, 48, 64)$ and 8×8 patches with stride 1. The frame has 96 atoms for AWGN and SIDD and 128 atoms for FMD. A lightweight convolutional module maps the noise-normalized coefficient field to nonnegative thresholds using a softplus output. It is trained jointly with the frame predictor.

We use the recorrupted-to-recorrupted (R2R) loss (Pang et al., 2021) to train the denoiser from noisy observations. For an AWGN observation v, we sample an independent perturbation $n \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2 I)$, where σ_{n} is the observation noise standard deviation. Using the denoising map d_θ in Eq. 19, the loss for each sampled perturbation is

$$\mathcal{L}_{\mathrm{R2R}}(\theta) = \mathrm{MSE}(d_\theta(v + \alpha n),\, v - n/\alpha)\,. \tag{45}$$

Here $\alpha > 0$ is the R2R perturbation factor, and MSE is averaged over pixels and channels. Training samples random crops of v and a fresh perturbation at each update. At inference, we apply Eq. 19 to the original observation v.

For synthetic AWGN, we optimize each image for 200 Adam steps using random 48×48 crops. The learning rates are 2×10^{-5} for the frame predictor and 2×10^{-3} for thresholding. We use an R2R perturbation factor $\alpha = 0.5$ and a ridge parameter of 10^{-4} during optimization; the ridge is removed for the final reconstruction. For SIDD, we train for four epochs on 128×128 crops. The corresponding learning rates are 10^{-5} and 2×10^{-4}. For FMD, a separate frame is optimized for each noisy image for 500 steps using random 48×48 crops.

D.2 SELF-SUPERVISED SINGLE-IMAGE SUPER-RESOLUTION

We use $(c_1, c_2, c_3) = (32, 48, 64)$. At $\times 2/\times 4/\times 8$, the local Gabor support sizes are 12×12, 5×5, and 4×4, the strides are $8/4/2$, and the numbers of modulated atoms per location are $384/96/96$, respectively.

We extract I training regions u_i from the observed low-resolution image y and downsample each region once more to form the self-supervised pairs

$$(u_i \downarrow_s, u_i), \qquad i = 1, \ldots, I.$$

These training regions are distinct from the local Gabor supports. In our implementation, each network update uses $I = 1$ region from the upper three quarters of the image, while the lower quarter is reserved for validation. The region dimensions are aligned with the internal downsampling factor s, which is $2/4/4$ for $\times 2/\times 4/\times 8$, respectively.

In equation 20, the smoothed ℓ_1 penalty is defined by

$$\|v\|_{1,\epsilon} := \sum_n \sqrt{|v[n]|^2 + \epsilon}, \qquad \epsilon = 10^{-6}. \tag{46}$$

For each training region, we initialize the reconstruction by bicubic upsampling. We then perform 3 reconstruction rounds. At the beginning of each round, the current Gabor frame is predicted from

the current image estimate and held fixed while equation 20 is approximately solved with 40 Adam updates of the image variable using learning rate 3×10^{-3}. The analysis coefficients are recomputed at every image update, while the Gabor atoms remain fixed throughout that solve. After the 40 image updates, the resulting reconstruction is used to predict a new Gabor frame for the next reconstruction round.

The original training region provides the self-supervised target:

$$\mathcal{L}_{\mathrm{SR}}^{\mathrm{self}}(\theta) = \frac{1}{I} \sum_{i=1}^{I} \mathrm{MSE}(g_\theta(u_i \downarrow_s), u_i)\,. \tag{47}$$

We train the Gabor-parameter predictor by minimizing $\mathcal{L}_{\mathrm{SR}}^{\mathrm{self}}$ with Adam for up to 100 updates per image, with learning rate 3×10^{-4}, differentiating through the reconstruction steps.

At inference, the predictor parameters are fixed. We initialize the estimate by bicubic upsampling of the full observed image y and apply the same reconstruction procedure. The final image estimate is returned as the SR output.

D.3 Self-supervised image inpainting

In image inpainting, we use the following constrained variational form. Let Λ denote the indices of the pixels with available values.

$$g_\theta(v) \in \arg\min_{\widehat{u}} \ \|\widehat{u} - x^0\|_2^2 + \lambda \, \|\{\langle \widehat{u}, \phi_{n,m}(f_\theta(v))\rangle\}_{n,m}\|_{1,\epsilon} \quad \text{s.t.} \quad \widehat{u}|_\Lambda = v|_\Lambda,$$

where $\|\cdot\|_{1,\epsilon}$ is the smoothed ℓ_1 penalty defined in equation 46. The predicted Gabor frame is computed once from the masked input and kept fixed during reconstruction. The constrained problem is solved by alternating between an unconstrained optimization step for the image variable and the observation-restoration projection:

$$\widetilde{v} = \arg\min_{\widehat{u}} \ \|\widehat{u} - x^0\|_2^2 + \lambda \, \|\{\langle \widehat{u}, \phi_{n,m}(f_\theta(v))\rangle\}_{n,m}\|_{1,\epsilon}\,,$$

and

$$\widehat{v}[r] = \begin{cases} v[r], & r \in \Lambda, \\ \widetilde{v}[r], & \text{otherwise.} \end{cases}$$

In practice, the optimization over $\widehat{u}$ is approximated by 8 projected gradient steps. The predicted Gabor frame is kept fixed during these steps, and after each gradient update the retained observed pixels are restored according to the projection above.

We use $(c_1, c_2, c_3) = (24, 32, 48)$ and 8×8 patches with stride 4 and 128 atoms. Here v denotes the masked input supplied to the reconstruction procedure, while x^0 is its data-consistent biharmonic initialization. The predictor constructs the Gabor frame from x^0 and the corresponding observation mask.

For self-supervision, we generate eight masking variants by independently hiding each observed pixel with probability 0.08, with $\mathcal{V}$ denoting the hidden locations in each variant. Each network update samples 64 overlapping crops from these variants, each containing at least one hidden pixel, and reconstructs them from the retained observations. The loss compares the reconstruction $\widehat{x}$ with the original observation y only at the hidden locations:

$$\mathcal{L}_{\mathrm{inp}}^{\mathrm{self}}(\theta, \lambda) = \mathrm{MSE}_{\mathcal{V}}(\widehat{x}, y). \tag{48}$$

This loss jointly trains the frame predictor and the regularization parameter. We optimize θ and $\log \lambda$ using Adam for 1200 network updates per image, with learning rates 8×10^{-5} and 2×10^{-3}, respectively, and initialize $\lambda = 0.1$. We differentiate through the reconstruction steps and use a frame-conditioning penalty with weight 10^{-4}.

At inference, we use all observed pixels without additional masking and apply the same reconstruction procedure, keeping the predictor parameters and λ fixed. The final estimate is synthesized with the corresponding dual frames, and overlapping local reconstructions are averaged to form the full-image output.

E VISUAL COMPARISONS

E.1 IMAGE DENOISING

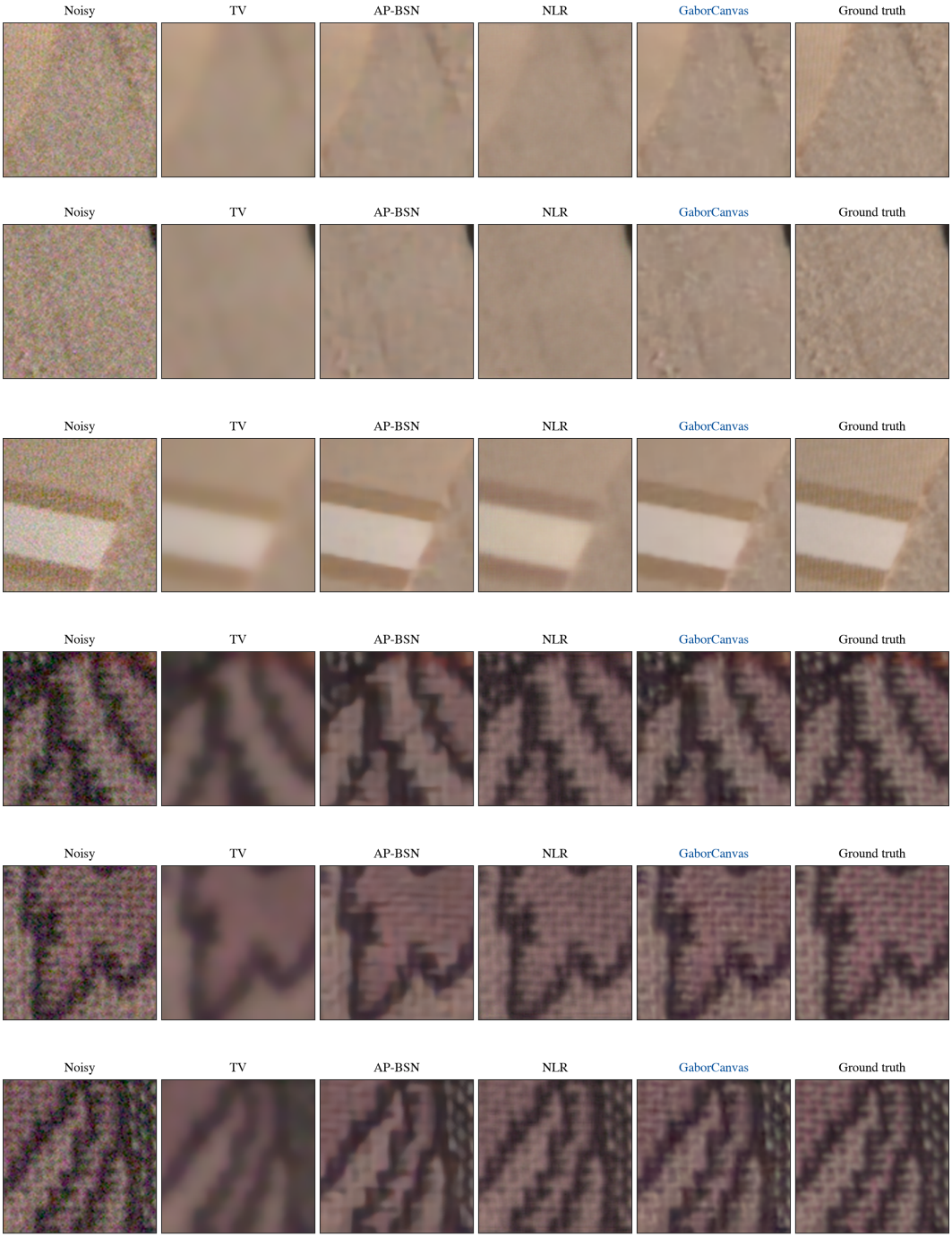

Figure 5: Qualitative denoising comparisons. The first three rows show regions from SIDD validation block 963 (scene 30, block 3). The last three rows show regions from SIDD validation block 1180.

E.2 SINGLE-IMAGE SUPER-RESOLUTION

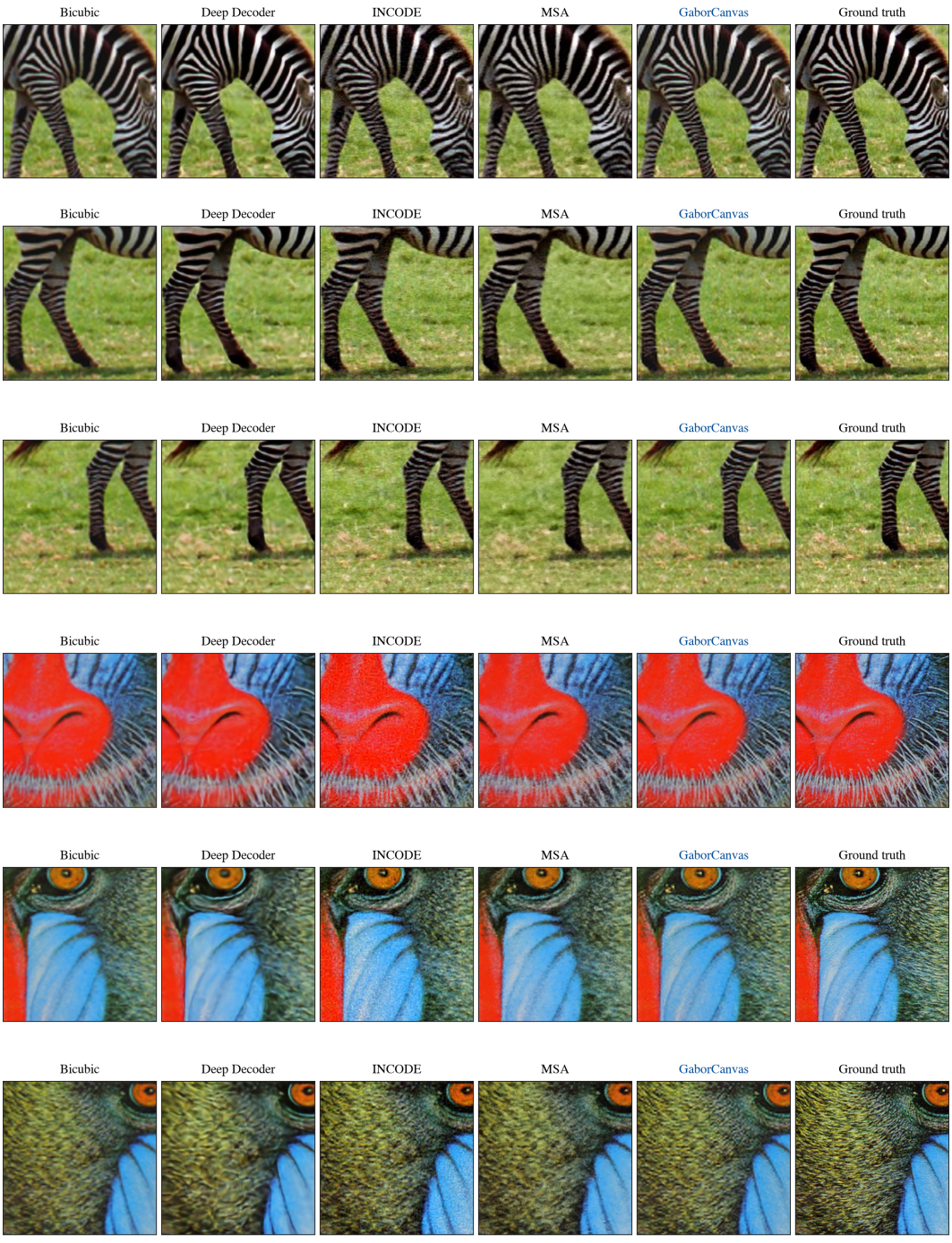

Figure 6: Qualitative super-resolution comparisons. The first three rows show regions from Set14 *zebra* at $\times 2$. The last three rows show regions from Set14 *baboon* at $\times 2$.

E.3 SINGLE-IMAGE INPAINTING

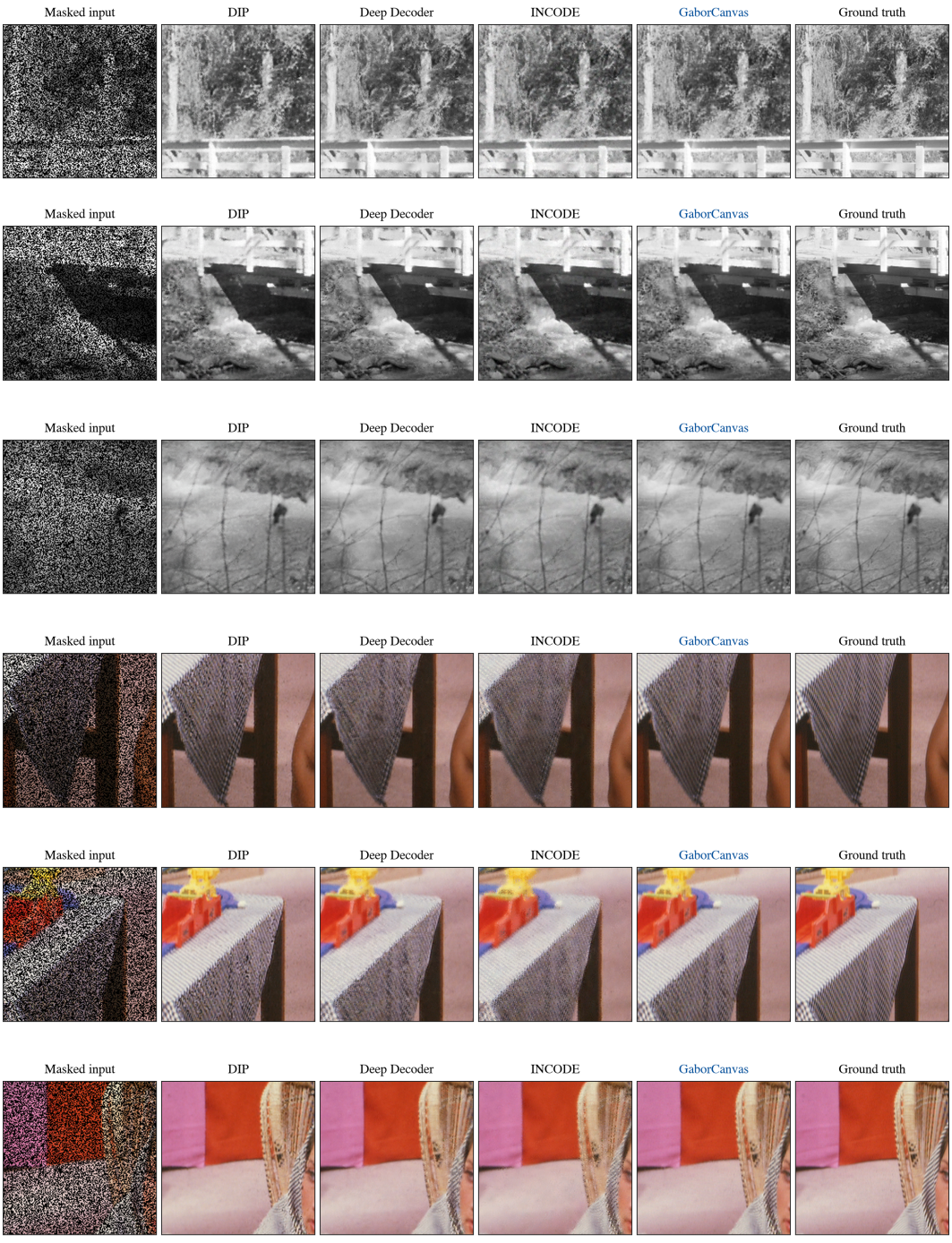

Figure 7: Qualitative inpainting comparisons with 50% randomly missing pixels. The first three rows show regions from Set14 *bridge*. The last three rows show regions from Set14 *barbara*.