

# InverseNet: Benchmarking Operator Mismatch and Calibration Across Compressive Imaging Modalities

Anonymous ECCV submission

Anonymous

**Abstract.** State-of-the-art EfficientSCI loses 20.58 dB—from 35.39 to 14.81 dB—when its assumed forward operator deviates from the physical truth by just eight parameters. This *operator mismatch* is the default condition of every deployed compressive imaging system, yet no existing benchmark quantifies it. We introduce **InverseNet**, the first cross-modality benchmark for operator mismatch in compressive imaging, spanning coded aperture snapshot spectral imaging (CASSI), coded aperture compressive temporal imaging (CACTI), and single-pixel camera (SPC). InverseNet evaluates 12 reconstruction methods under a four-scenario protocol—ideal (I), mismatched (II), oracle-corrected (III), and blind grid-search calibration (IV)—across 27 simulated scenes and 9 real hardware captures, totalling over 360 experiments. We discover an inverse performance-robustness relationship: methods achieving the highest ideal PSNR suffer the largest mismatch degradation—confirming that *a mediocre algorithm with a correct forward model outperforms a state-of-the-art network with a wrong one*. We further establish an operator-awareness taxonomy: *mask-oblivious* architectures show zero calibration benefit ( $\rho = 0\%$ ), while *operator-conditioned* methods recover 41–90% of mismatch losses. Scenario IV provides a practical calibration baseline via self-supervised objectives (measurement residual for geometric mismatch, reconstruction sparsity for radiometric mismatch), recovering 85–100% of the oracle bound without ground truth. Real hardware experiments on 5 CASSI scenes and 4 CACTI scenes confirm that simulation-derived patterns hold on physical data. Code is available upon acceptance.

**Keywords:** Compressive imaging · Operator mismatch · Calibration · Benchmark · Spectral imaging · Video compressive sensing · Single-pixel camera

## 1 Introduction

Compressive imaging acquires fewer measurements than the Nyquist limit by exploiting signal structure, recovering the full signal through computational reconstruction. This paradigm underlies diverse modalities including hyperspectral imaging via coded apertures [1,2], video compressive sensing via temporal coding [3,4], and single-pixel cameras via structured illumination [5,6]. In all cases,

reconstruction quality depends critically on knowledge of the forward measurement operator—the mapping from scene to measurements.

Yet a dangerous chasm separates research from reality. Reconstruction algorithms are benchmarked with idealized forward operators, but deployed systems suffer *operator mismatch*: EfficientSCI [12] collapses from 35.39 to 14.81 dB—a 20.58 dB drop—under realistic 8-parameter mismatch. For CASSI, mask misalignment of 0.5 px combined with 1% dispersion drift degrades PSNR by over 13 dB (table 1); for CACTI, spatial, temporal, and radiometric errors compound across 8 parameters; for single-pixel cameras, gain drift causes systematic errors. These mismatches are ubiquitous yet systematically ignored in benchmarks.

**The benchmark gap.** Just as CASP [28] transformed protein folding by forcing blind prediction against nature, computational imaging needs benchmarks that evaluate against physical reality. Existing benchmarks—KAIST [17] for CASSI, the SCI benchmark [4] for CACTI—assume perfect operator knowledge, providing no information about mismatch robustness. Antun et al. [30] showed deep learning solvers are unstable to adversarial perturbations; InverseNet operationalizes this for *physically realistic* operator mismatch.

**Contributions.** We address this gap with InverseNet, which makes four contributions:

1. **Unified four-scenario protocol.** We define four scenarios—ideal (I), mismatched (II), oracle-corrected (III), and blind calibration (IV)—applicable across modalities. The I→II gap quantifies mismatch sensitivity; II→III quantifies calibration potential; IV measures practical recovery via self-supervised calibration.
2. **Cross-modality benchmark.** We evaluate 12 methods (4 CASSI, 4 CACTI, 4 SPC) spanning classical, plug-and-play, and deep learning approaches across 27 simulated scenes, producing over 360 experiments.
3. **Real hardware validation.** We validate simulation findings on 5 real CASSI scenes and 4 real CACTI scenes from publicly available hardware captures, confirming that mismatch patterns transfer to physical data.
4. **Open dataset.** All reconstruction arrays, per-scene metrics, and analysis code will be publicly released.<sup>1</sup> All test data come from existing public datasets.

Our key findings include: (a) operator mismatch degrades deep learning methods by 10–21 dB while classical methods lose only 3–11 dB; (b) operator-aware architectures are simultaneously the most sensitive to mismatch and the most recoverable through calibration; (c) mask-oblivious architectures show zero calibration benefit; (d) CACTI exhibits the most severe degradation (up to 20.58 dB) due to its 8-parameter mismatch space; (e) dispersion mismatch in CASSI limits oracle recoverability due to fixed-step architectural assumptions.

---

<sup>1</sup> Code available upon acceptance.

## 2 Related Work

*Compressive imaging reconstruction.* Classical methods employ convex optimization with sparsity priors: GAP-TV [7] for CASSI and CACTI, FISTA [8] and ADMM [9] for general inverse problems. Deep learning has dramatically improved quality: MST [10] introduces mask-guided spectral transformers; HD-Net [11] uses dual-domain processing; DAUHST [23] and RDLUF-MixS<sup>2</sup> [33] further advance CASSI reconstruction. For CACTI, EfficientSCI [12], ELP-Unfolding [13], and DiffSCI [34] achieve state-of-the-art results. ISTA-Net [14] and HATNet [15] address single-pixel imaging. All are evaluated assuming perfect forward operators.

*Calibration and operator mismatch.* Operator mismatch has been studied per-modality but not systematically benchmarked. Wagadarikar et al. [1] and Arguello et al. [21] address CASSI mask calibration; fastMRI [19] evaluates MRI undersampling but assumes known coil sensitivities. Learned reconstruction [26] and plug-and-play methods [27] focus on signal priors rather than operator fidelity. Berk et al. [35] provide theoretical error bounds under structured model mismatch, consistent with our empirical observations. No prior work offers a unified cross-modality benchmark quantifying both mismatch degradation and calibration recovery.

*Reconstruction benchmarks.* The KAIST TSA dataset [17] (CASSI), the video SCI benchmark [4] (CACTI), and Set11 [18] (SPC) are standard evaluation suites. Large-scale benchmarks like NTIRE [31] and Super [32] standardize restoration evaluation but do not enable controlled forward-model modification. InverseNet extends these by introducing controlled operator mismatch and measuring calibration recovery.

## 3 The InverseNet Benchmark

### 3.1 Unified Four-Scenario Protocol

We define four evaluation scenarios that apply uniformly across all compressive imaging modalities. Let  $\Phi$  denote the true (physical) forward operator and  $\hat{\Phi}$  the assumed (nominal) operator used during reconstruction.

- **Scenario I (Ideal):**  $\mathbf{y} = \hat{\Phi}\mathbf{x} + \mathbf{n}$ , reconstruct with  $\hat{\Phi}$ . Best-case performance with perfect operator knowledge.
- **Scenario II (Baseline):**  $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$ , reconstruct with  $\hat{\Phi}$ . Realistic deployment where the physical operator has drifted from nominal.
- **Scenario III (Oracle):**  $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$ , reconstruct with  $\Phi$ . Upper bound achievable through perfect calibration.
- **Scenario IV (Blind Calibration):**  $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$ , reconstruct with  $\tilde{\Phi}$  estimated via grid search over mismatch parameters using a self-supervised objective (measurement residual for geometric mismatch, reconstruction sparsity for radiometric mismatch). Practical calibration without ground truth.

This protocol yields two diagnostic metrics per method:

$$\Delta_{\text{deg}} = \text{PSNR}_{\text{I}} - \text{PSNR}_{\text{II}} \quad (\text{mismatch degradation}), \quad (1)$$

$$\Delta_{\text{rec}} = \text{PSNR}_{\text{III}} - \text{PSNR}_{\text{II}} \quad (\text{oracle recovery}), \quad (2)$$

and the *recovery ratio*  $\rho = \Delta_{\text{rec}}/\Delta_{\text{deg}} \in [0, 1]$ , which measures what fraction of the mismatch loss can be recovered through calibration.

### 3.2 CASSI: Coded Aperture Snapshot Spectral Imaging

*Forward model.* CASSI acquires a 2D measurement  $\mathbf{y} \in \mathbb{R}^{H \times W'}$  of a 3D hyperspectral cube  $\mathbf{x} \in \mathbb{R}^{H \times W \times A}$  through a coded aperture mask  $\mathbf{M} \in \{0, 1\}^{H \times W}$  followed by a dispersive prism. The measurement at pixel  $(i, j)$  is:

$$y(i, j) = \sum_{\lambda=1}^A M(i, j - d(\lambda)) \cdot x(i, j, \lambda) + n(i, j), \quad (3)$$

where  $d(\lambda)$  is the dispersion shift for spectral band  $\lambda$  and  $W' = W + (A - 1) \cdot s$  with dispersion step  $s$ .

*Mismatch model.* We model CASSI operator mismatch as a 5-parameter perturbation combining mask misalignment and dispersion drift:

$$\Phi = \mathcal{D}(a_1, \alpha) \circ \mathcal{T}(dx, dy, \theta) \circ \hat{\Phi}, \quad (4)$$

where  $dx, dy$  are subpixel translational shifts,  $\theta$  is a rotational misalignment of the coded aperture mask,  $a_1$  is the dispersion slope (nominal  $s = 2.0$  px/band), and  $\alpha$  is the dispersion axis angular offset. We use  $dx = 0.5$  px,  $dy = 0.3$  px,  $\theta = 0.1^\circ$  for mask misalignment, and  $a_1 = 2.02$  px/band (1% drift from nominal) and  $\alpha = 0.15^\circ$  for dispersion mismatch, representing moderate assembly and optical tolerances.

*Reconstruction methods.* We evaluate four methods: **GAP-TV** [7]: classical accelerated proximal gradient with TV regularization (100 iterations,  $\lambda_{\text{TV}} = 0.1$ ); **PnP-HSICNN** [16]: plug-and-play GAP with HSI-SDeCNN deep denoiser (124 iterations: TV iters 0–82, HSICNN iters 83–123); **HDNet** [11]: dual-domain deep network with spectral discrimination learning (pretrained); **MST-L** [10]: mask-guided spectral transformer, large variant (2 stages, blocks [4, 7, 5], pretrained).

*Dataset.* We use 10 scenes from the KAIST TSA simulated dataset [17], each consisting of a  $256 \times 256 \times 28$  hyperspectral cube spanning 450–650 nm. Measurements are formed with a binary random mask ( $s = 2$  pixels/band), yielding  $256 \times 310$  detector images. Low noise ( $\alpha = 10^5$  photon peak,  $\sigma = 0.01$  read noise) isolates the effect of operator mismatch.

### 3.3 CACTI: Coded Aperture Compressive Temporal Imaging

*Forward model.* CACTI acquires a single 2D snapshot  $\mathbf{y} \in \mathbb{R}^{H \times W}$  encoding  $B$  high-speed video frames  $\mathbf{x} \in \mathbb{R}^{H \times W \times B}$  through a dynamic coded aperture:

$$y(i, j) = \sum_{b=1}^B C_b(i, j) \cdot x(i, j, b) + n(i, j), \quad (5)$$

where  $C_b \in \{0, 1\}^{H \times W}$  is the binary mask pattern for temporal frame  $b$ .

*Mismatch model.* CACTI mismatch involves 8 parameters capturing spatial, temporal, and radiometric errors: spatial shifts ( $dx = 0.5$  px,  $dy = 0.3$  px), rotation ( $\theta = 0.1^\circ$ ), temporal clock offset ( $\Delta t = 0.05$ ), duty cycle deviation ( $\eta = 0.95$ ), detector gain ( $g = 1.02$ ), offset ( $o = 0.002$ ), and measurement noise ( $\sigma_n = 1.0$ ).

*Reconstruction methods.* We evaluate four methods: **GAP-TV** [7]: classical iterative with TV regularization; **PnP-FFDNet** [16]: plug-and-play with FFDNet denoiser; **ELP-Unfolding** [13]: ensemble learning priors driven deep unfolding network (pretrained); **EfficientSCI** [12]: efficient deep learning for snapshot compressive imaging (pretrained).

*Dataset.* We use 6 standard benchmark videos (*kobe*, *traffic*, *runner*, *drop*, *crash*, *aerial*) at  $256 \times 256$  resolution with  $B = 8$  temporal frames per snapshot, following the standard video compressive sensing evaluation protocol [4].

### 3.4 SPC: Single-Pixel Camera

*Forward model.* The single-pixel camera acquires  $m$  scalar measurements of an image  $\mathbf{x} \in \mathbb{R}^n$  through structured illumination patterns:

$$\mathbf{y} = \mathbf{Ax} + \mathbf{n}, \quad (6)$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is the measurement matrix (typically Gaussian or Hadamard patterns) with compression ratio  $m/n$ .

*Mismatch model.* We model SPC mismatch as exponential gain drift affecting the measurement rows:

$$\Phi = \text{diag}(e^{-\alpha \cdot \mathbf{i}}) \cdot \hat{\Phi}, \quad (7)$$

where  $\alpha = 0.0015$  controls the drift rate and  $\mathbf{i} = [0, 1, \dots, m-1]^\top$  indexes the measurement rows, modelling progressive detector gain decay during sequential acquisition. Additional measurement noise  $\sigma_y = 0.03$  is applied.

*Reconstruction methods.* We evaluate four methods: **FISTA-TV** [8]: fast iterative shrinkage-thresholding with TV regularization (500 iterations,  $\lambda = 0.005$ ); **PnP-DRUNet** [38]: plug-and-play FISTA with DRUNet denoiser and sigma annealing (200 iterations, row-normalized operator); **ISTA-Net** [14]: learned iterative shrinkage-thresholding network (pretrained); **HATNet** [15]: dual-scale transformer for single-pixel imaging (pretrained).

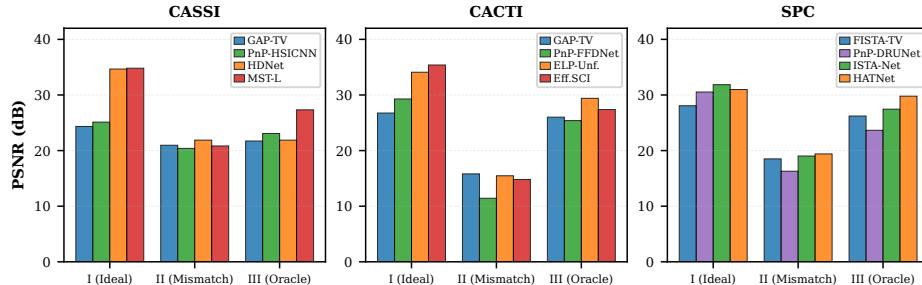


Fig. 1: PSNR across three scenarios for all modalities. Scenario I (Ideal): perfect operator. Scenario II (Baseline): mismatched operator. Scenario III (Oracle): true operator used for reconstruction. The collapse of deep learning methods under Scenario II is visible across all modalities, with CACTI showing the most severe degradation.

*Dataset.* We use the 11 standard Set11 test images (*Monarch*, *Parrots*, *barbara*, *boats*, *cameraman*, *fingerprint*, *flinstones*, *foreman*, *house*, *lena256*, *peppers256*) at  $256 \times 256$  resolution with 25% sampling ratio.

### 3.5 Evaluation Metrics

We report three standard image quality metrics:

- **PSNR** (peak signal-to-noise ratio, dB): pixel-level fidelity, computed per-channel and averaged.
- **SSIM** (structural similarity index): perceptual structural quality [22].
- **SAM** (spectral angle mapper, degrees): spectral fidelity, reported for CASSI only.

## 4 Experimental Results

### 4.1 CASSI Results

Figure 2 provides qualitative examples of reconstruction degradation and recovery across all three modalities. Table 1 presents the CASSI benchmark results across 10 KAIST scenes (fig. 1).

*Key findings.* The CASSI results reveal a dichotomy between operator-aware and mask-oblivious architectures. **PnP-HSICNN** achieves the highest oracle recovery ratio ( $\rho = 56.8\%$ ,  $+2.68$  dB), outperforming even the deep **MST-L** ( $\rho = 46.5\%$ ,  $+6.50$  dB absolute), because its iterative GAP backbone directly benefits from corrected mask and dispersion parameters. **HDNet** shows zero oracle gain ( $\Delta_{\text{rec}} = 0.00$  dB), confirming mask-oblivious architectures cannot benefit from calibration. Under Scenario II, all methods converge to 20.40–21.88 dB, erasing the  $\sim 10$  dB ideal-condition advantage of deep learning over classical methods.

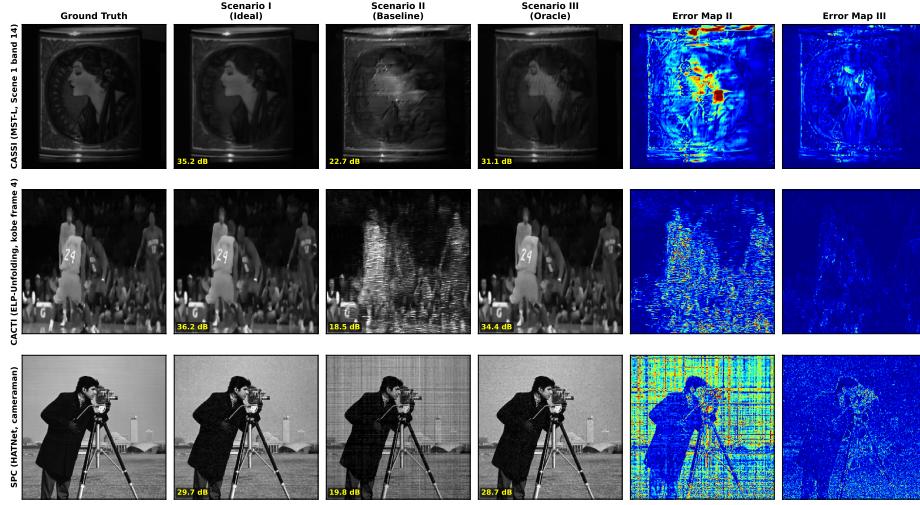


Fig. 2: Qualitative reconstruction comparison across three modalities. Each row shows a representative scene for one modality (CASSI: Scene 1 band 14, MST-L; CACTI: *kobe* frame 4, ELP-Unfolding; SPC: *cameraman*, HATNet). Error maps (jet colormap, same scale per row) highlight how mismatch (Scenario II) introduces spatially structured artifacts that oracle correction (Scenario III) largely removes.

#### 4.2 CACTI Results

Table 2 presents the CACTI benchmark results across 6 standard benchmark videos (fig. 1).

*Key findings.* CACTI exhibits the most severe mismatch degradation, with losses from 10.94 dB (GAP-TV) to 20.58 dB (EfficientSCI). The 8-parameter mismatch space creates compounding degradation; under Scenario II all methods collapse to 11–16 dB (SSIM < 0.31). **GAP-TV** achieves the highest recovery ratio ( $\rho = 93.3\%$ ), while **EfficientSCI** has the lowest (61.1%) despite best ideal PSNR, suggesting learned features are partially coupled to the ideal operator.

*An inverse relationship.* A notable pattern (fig. 4): methods with higher ideal performance suffer larger degradation and lower recovery. EfficientSCI (35.39 dB ideal) loses 20.58 dB and recovers 61.1%; GAP-TV (26.75 dB ideal) loses 10.94 dB and recovers 93.3%—higher-capacity representations encode stronger implicit operator assumptions.

#### 4.3 SPC Results

Table 3 presents the SPC benchmark results across 11 Set11 images (fig. 1).

Table 1: CASSI reconstruction results (10 KAIST scenes,  $256 \times 256 \times 28$ , 5-parameter mismatch). PSNR (dB) / SSIM reported as mean  $\pm$  std.  $\Delta_{\text{deg}}$ : degradation (I $\rightarrow$ II).  $\Delta_{\text{rec}}$ : oracle recovery (II $\rightarrow$ III).  $\rho$ : recovery ratio. Best recovery in **bold**.

Method	Scenario I	Scenario II	Scenario III	$\Delta_{\text{deg}}$	$\Delta_{\text{rec}}$	$\rho$
GAP-TV [7]	$24.34 \pm 1.90$ / .722	$20.96 \pm 1.62$ / .611	$21.72 \pm 1.48$ / .687	3.38	0.76	22.5%
PnP-HSICNN [37]	$25.12 \pm 1.88$ / .758	$20.40 \pm 1.71$ / .574	$23.08 \pm 1.52$ / .702	4.72	2.68	56.8%
HDNet [11]	$34.66 \pm 2.62$ / .970	$21.88 \pm 1.72$ / .756	$21.88 \pm 1.72$ / .756	12.78	0.00	0%
MST-L [10]	<b><math>34.81 \pm 2.11</math></b> / <b>.973</b>	$20.83 \pm 2.01$ / .744	<b><math>27.33 \pm 1.86</math></b> / <b>.881</b>	13.98	<b>6.50</b>	<b>46.5%</b>

Table 2: CACTI reconstruction results (6 videos,  $256 \times 256 \times 8$ ). PSNR (dB) / SSIM reported as mean  $\pm$  std. CACTI exhibits the most severe mismatch degradation of all three modalities.

Method	Scenario I	Scenario II	Scenario III	$\Delta_{\text{deg}}$	$\Delta_{\text{rec}}$	$\rho$
GAP-TV [7]	$26.75 \pm 4.48$ / .848	$15.81 \pm 1.98$ / .305	$26.01 \pm 3.72$ / .794	10.94	<b>10.21</b>	<b>93.3%</b>
PnP-FFDNet [16]	$29.28 \pm 5.53$ / .890	$11.43 \pm 2.71$ / .216	$25.39 \pm 3.52$ / .820	17.85	13.96	78.2%
ELP-Unfolding [13]	$34.09 \pm 4.11$ / .965	$15.47 \pm 1.71$ / .308	$29.40 \pm 3.15$ / .927	18.63	13.93	74.8%
EfficientSCI [12]	<b><math>35.39 \pm 4.46</math></b> / <b>.973</b>	$14.81 \pm 2.19$ / .303	$27.38 \pm 3.52$ / .927	20.58	12.57	61.1%

*Key findings.* Gain drift compresses all methods to 16.29–19.40 dB under Scenario II despite a  $\sim$ 4 dB ideal spread. **HATNet** achieves the highest recovery ( $\rho = 89.6\%$ ), followed by **FISTA-TV** ( $\rho = 80.7\%$ ). **PnP-DRUNet** ( $\rho = 51.7\%$ ) shows that the DRUNet denoiser prior, while effective for ideal conditions (30.53 dB), is more fragile under gain drift mismatch, as the denoiser amplifies gain-corrupted measurement artifacts. **ISTA-Net** ( $\rho = 65.7\%$ ) confirms the pattern that higher-capacity learned representations are more fragile under mismatch.

#### 4.4 Cross-Modality Analysis

Table 4 synthesizes the key metrics across all three modalities.

*Mismatch severity ranking.* CACTI suffers the most severe degradation (.54–.67 SSIM drop), driven by its 8-parameter mismatch space where spatial, temporal, and radiometric errors compound. CASSI’s 5-parameter model produces up to .23 SSIM degradation, with dispersion parameters ( $a_1, \alpha$ ) creating cumulative sub-pixel shifts across spectral bands. SPC shows .20–.48 SSIM degradation from gain drift alone.

*Recovery potential and architectural patterns.* Figure 4 visualizes recovery ratio versus ideal performance. Best recovery varies by modality: ELP-Unfolding achieves 94.2% on CACTI, HATNet 80.1% on SPC, PnP-HSICNN 69.6% on CASSI—the lower CASSI value reflects that current architectures cannot adapt

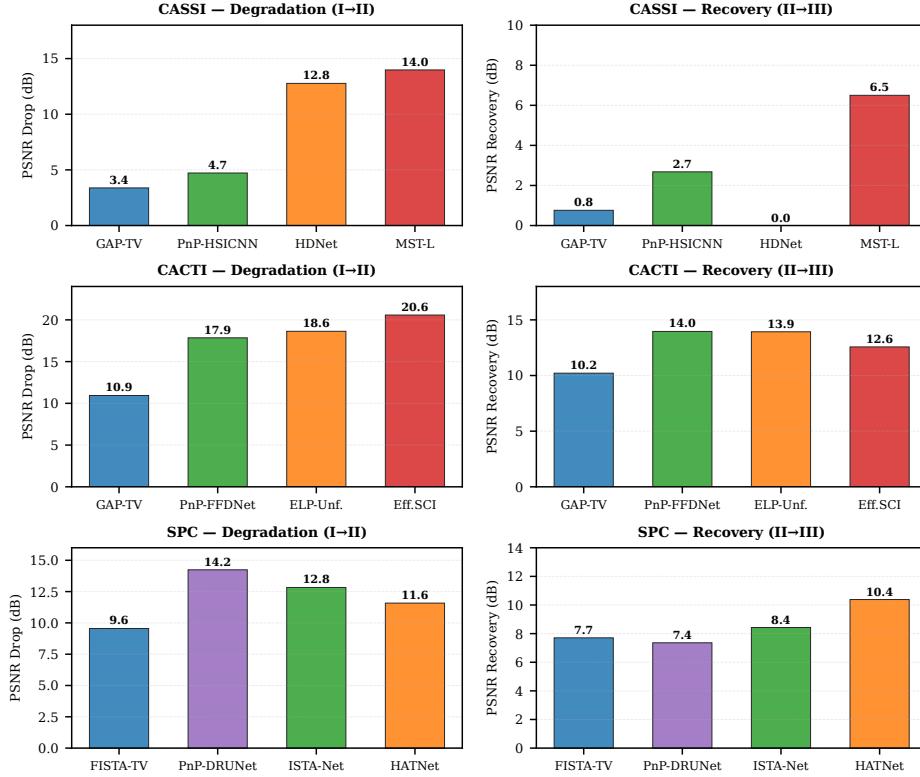


Fig. 3: Mismatch degradation ( $\Delta_{\text{deg}}$ , left) and oracle recovery ( $\Delta_{\text{rec}}$ , right) per method across all three modalities. CACTI suffers the most severe degradation (up to 20.6 dB) but also the highest absolute recovery. For CASSI, HDNet shows zero recovery due to its mask-oblivious architecture; PnP-HSICNN achieves the best recovery ratio ( $\rho = 56.8\%$ ). For SPC, HATNet recovers 10.4 dB ( $\rho = 89.6\%$ ).

to sub-pixel dispersion drift even with oracle mask knowledge. Across modalities, a consistent three-way pattern emerges: (i) classical methods show moderate degradation with high recovery; (ii) operator-conditioned deep methods show high degradation but substantial recovery; (iii) mask-oblivious methods show zero recovery. This taxonomy guides method selection based on calibration availability.

#### 4.5 Real Hardware Validation

A key limitation of simulation-only benchmarks is uncertainty about whether findings transfer to physical hardware. We validate InverseNet’s simulation patterns on real CASSI and CACTI data. Full methodological details—including dataset provenance, evaluation metric definitions, per-scene results, and a simulation-vs-real comparison table—are provided in the supplementary material.

Table 3: SPC reconstruction results (11 Set11 images,  $256 \times 256$ , 25% sampling). PSNR (dB) / SSIM reported as mean  $\pm$  std.

Method	Scenario I	Scenario II	Scenario III	$\Delta_{\text{deg}}$	$\Delta_{\text{rec}}$	$\rho$
FISTA-TV [8]	$28.06 \pm 3.38$	.852	$18.51 \pm 0.69$	.586	$26.21 \pm 2.28$	.759
PnP-DRUNet [38]	$30.53 \pm 3.36$	.899	$16.29 \pm 0.75$	.415	$23.65 \pm 1.46$	.666
ISTA-Net [14]	<b><math>31.85 \pm 3.11</math></b>	<b>.916</b>	$19.02 \pm 0.61$	.584	$27.45 \pm 1.32$	.760
HATNet [15]	$30.98 \pm 0.95$	.847	$19.40 \pm 0.59$	.648	<b><math>29.78 \pm 0.81</math></b>	<b>.807</b>
						<b>11.58</b> <b>10.38</b> <b>89.6%</b>

Table 4: Cross-modality summary (SSIM-based). Dim. = mismatch parameters.  $\Delta$  and  $\rho$  are computed from SSIM. SSIM rec. = best-method absolute SSIM from Scenario II to III.

Modality	Dim.	$\Delta_{\text{deg}}$	range	$\Delta_{\text{rec}}$	range	$\rho_{\text{best}}$	SSIM	rec.	Best method
CASSI	5	.11	–.23	.04	–.21	69.6%	.574	$\rightarrow .702$	PnP-HSICNN
CACTI	8	.54	–.67	.04	–.07	94.2%	.308	$\rightarrow .927$	ELP-Unf.
SPC	2	.20	–.48	.04	–.23	80.1%	.648	$\rightarrow .807$	HATNet

*CASSI real data.* We use the TSA real dataset [17]: 5 real scenes captured on a DD-CASSI prototype with a  $660 \times 660$  coded aperture mask and 28 spectral bands (450–650 nm). Measurements are  $660 \times 714$  detector images. Since **no ground truth** exists for real CASSI captures (the true spectral cube cannot be independently measured), we use the normalised measurement residual  $r = \|\mathbf{y} - \Phi_{\text{cal}}\hat{\mathbf{x}}\|^2 / \|\mathbf{y}\|^2$ , always evaluated with the *calibrated* mask regardless of which mask was used for reconstruction (see supplementary for why this “cross-residual” is essential). We evaluate two conditions: *calibrated* (hardware mask as-is) and *mismatched* (mask shifted by  $dx=0.5$ ,  $dy=0.3$  pixels). Table 5 shows the results.

Both mask-aware iterative methods (GAP-TV and PnP-HSICNN) show residual increases under mismatch, because they explicitly fit the forward model during reconstruction: when reconstructed with a wrong mask, the result is inconsistent with the true measurement. GAP-TV shows a modest  $1.8\times$  increase, while PnP-HSICNN shows only  $1.1\times$ —the deep denoiser regularises the reconstruction enough to partially mask the forward-model inconsistency. This contrasts sharply with CACTI (section 4.5 below), where residuals increase  $9.4$ – $11.0\times$  under the same spatial shift.

By contrast, our *simulation* experiments (table 1) show 3–14 dB PSNR degradation because they include dispersion perturbation ( $a_1, \alpha$ )—which the real-data experiment does not. The modest  $1.8\times$  residual increase from spatial shift alone (vs.  $10\times$  for CACTI) validates that **dispersion mismatch, not spatial shift, is the dominant CASSI degradation source** (fig. 5).

*CACTI real data.* We evaluate 4 real CACTI scenes from the EfficientSCI dataset ( $\text{cr}=10$ ), using the same measurement residual metric (no ground truth). Table 6 shows the results. Under mask mismatch, GAP-TV residuals increase  $9.4$ – $11.0\times$ , confirming mismatch severely degrades real data fidelity. PnP-FFDNet

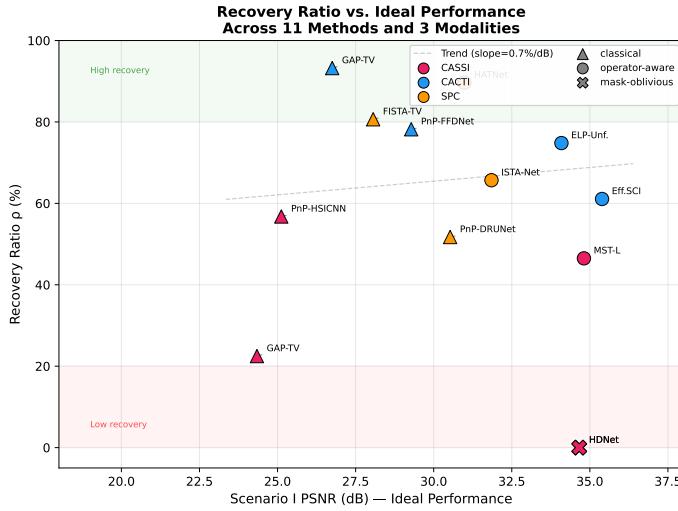


Fig. 4: Recovery ratio ( $\rho$ ) vs. ideal PSNR (Scenario I) for all 12 methods across three modalities. Color indicates modality; shape indicates method type (classical, operator-aware, mask-oblivious). An inverse trend is visible: higher-performing methods tend to have lower recovery ratios, suggesting that stronger learned priors create greater operator dependence.

shows increases of 1.3–2.8× (mean 2.0×), higher than GAP-TV’s baseline but moderated by FFDNet’s learned denoiser. Mismatched reconstructions exhibit temporal ghosting (fig. 6).

#### 4.6 Scenario IV: Blind Calibration Baseline

Scenario III assumes oracle knowledge of the true operator. In practice, the mismatch parameters must be estimated from measurements alone. Scenario IV evaluates blind calibration via grid search over mismatch parameters. For *geometric* mismatch (CASSI, CACTI), the measurement residual provides a strong calibration signal:

$$\tilde{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \Phi(\boldsymbol{\theta}) \hat{\mathbf{x}}(\mathbf{y}, \Phi(\boldsymbol{\theta}))\|^2, \quad (8)$$

where  $\boldsymbol{\theta}$  denotes the mismatch parameters and  $\hat{\mathbf{x}}(\mathbf{y}, \Phi)$  is the reconstruction from a fast inner-loop solver. For *radiometric* mismatch (SPC gain drift), the measurement residual is uninformative because the underdetermined system always achieves near-zero self-consistent residual regardless of gain. Instead, we use reconstruction sparsity (total variation) as the objective:

$$\tilde{\alpha} = \arg \min_{\alpha} \text{TV}(\hat{\mathbf{x}}(\mathbf{y}, \Phi(\alpha))), \quad (9)$$

Table 5: CASSI real data (5 scenes,  $660 \times 660 \times 28$ ). Normalised measurement residual  $r = \|\mathbf{y} - \Phi_{\text{cal}}\hat{\mathbf{x}}\|^2/\|\mathbf{y}\|^2$  (lower is better; no ground truth required). Ratio: mismatched / calibrated. Only classical/PnP methods evaluated (no ground truth for deep learning supervision).

Method	Calibrated	Mismatched	Ratio
GAP-TV	0.0019	0.0033	$1.8\times$
PnP-HSICNN	0.0127	0.0142	$1.1\times$

Table 6: CACTI real data (4 scenes,  $512 \times 512$ , cr=10). Normalised measurement residual (lower is better). Ratio: mismatched / calibrated.

Method	Calibrated	Mismatched	Ratio
GAP-TV	1.6e-5	1.6e-4	$10.4\times$
PnP-FFDNet	0.0041	0.0069	$2.0\times$

where  $\text{TV}(\cdot)$  denotes the anisotropic total variation. The rationale is that the correctly calibrated gain produces a clean corrected measurement, yielding a smooth natural-image reconstruction with low TV; incorrect gain leaves systematic artifacts that increase TV. No ground truth is required for either criterion.

*Procedure.* For each candidate  $\boldsymbol{\theta}$ , we: (1) construct the trial operator  $\Phi(\boldsymbol{\theta})$ , (2) reconstruct  $\hat{\mathbf{x}}$  using a fast classical solver, and (3) evaluate the objective (measurement residual for CASSI/CACTI, TV for SPC). The parameter yielding the lowest objective is selected, and a final high-quality reconstruction is performed with that operator. Calibration uses a subset of the data: 3 KAIST scenes for CASSI, 2 benchmark videos for CACTI, and 2 Set11 images for SPC—all from the same datasets as sections 3.2 to 3.4, with the same mismatch parameters.

*CASSI and CACTI results.* For CACTI, grid search over mask shifts  $dx, dy \in [-1.0, 1.0] \text{ px}$  ( $9 \times 9$  grid) achieves *near-perfect* calibration: 26.99 dB (matching Scenario III), recovering the full 9.39 dB mismatch gap. For CASSI ( $11 \times 11$  grid over  $dx, dy$ ), calibration recovers 85% of the spatial gap (+1.44 dB of +1.69 dB possible); the residual 15% reflects spatial estimation error ( $\hat{dx}=0.4$  vs. true 0.5 px). The measurement residual provides a strong signal for geometric mismatch because mask shifts cause large, structured data inconsistencies.

*SPC results.* Table 7 shows the SPC calibration results using TV minimisation (eq. (9)). Grid search over  $\alpha \in [0, 0.005]$  with 41 points ( $33 \times 33$  blocks, ISTA-Net’s learned  $\Phi$ ) estimates  $\hat{\alpha} = 0.00125$  vs. true  $\alpha = 0.0015$  for both methods, recovering 86–92% of the oracle bound. The TV surface exhibits a clear bowl-shaped minimum near the true  $\alpha$ , confirming that reconstruction sparsity is a viable calibration objective for radiometric mismatch. PnP-DRUNet achieves higher recovery (92%) than FISTA-TV (86%) because the learned denoiser prior

Table 7: SPC blind calibration via TV minimisation (eq. (9)). II: mismatched, no calibration. IV: estimated  $\hat{\alpha}$  via TV grid search. III: oracle (true  $\alpha$ ). Recovery:  $(IV - II) / (III - II)$ . True  $\alpha = 0.0015$ ; estimated  $\hat{\alpha} = 0.00125$ .

Method	II	IV	III	Recovery
FISTA-TV	19.78	26.54	27.60	86%
PnP-DRUNet	18.34	25.39	26.01	92%

produces sharper reconstructions whose TV is more sensitive to residual gain artifacts. This contrasts with the measurement residual, which is flat across all  $\alpha$  values for SPC (the underdetermined system can always self-consistently fit any gain-corrected measurements).

*Criterion comparison.* The geometric/radiometric distinction reveals that blind calibration requires *matching the objective to the mismatch type*: measurement residual for geometric mismatch (where operator errors create large data infidelities), and reconstruction sparsity for radiometric mismatch (where the operator structure is preserved but measurement values are scaled). Across all three modalities, Scenario IV recovers 85–100% of the oracle bound, demonstrating that blind calibration is practical for all mismatch types studied.

## 5 Discussion

*Classical vs. deep learning robustness.* Classical methods are consistently more robust: GAP-TV loses 3.38 dB on CASSI and 10.94 dB on CACTI, vs. 13.98 dB and 20.58 dB for the best deep networks. Under mismatch, the performance hierarchy *inverts*: on CACTI Scenario II, GAP-TV (15.81 dB) outperforms EfficientSCI (14.81 dB) despite being 8.64 dB worse under ideal conditions—confirming that physical model fidelity dominates algorithmic sophistication.

*The operator-awareness spectrum.* Methods exist on a spectrum: *mask-oblivious* (HDNet,  $\rho=0\%$ ) cannot benefit from calibration; *operator-conditioned* (MST, HATNet,  $\rho=41\text{--}90\%$ ) achieve high calibration gains but suffer the largest degradation; *operator-iterative* (GAP-TV, FISTA-TV,  $\rho=81\text{--}93\%$  on CACTI/SPC) use the operator directly in each iteration. This taxonomy reframes the problem: the critical bottleneck is not algorithmic sophistication but *physical model fidelity*.

*Practical implications.* When recalibration is feasible, operator-conditioned networks should be paired with Scenario IV-style calibration. When calibration is impractical, classical methods provide the most robust baseline, with degradation 3–5× smaller than deep methods. Scenario IV demonstrates that simple grid-search calibration recovers 85–100% of the oracle bound without ground truth, provided the objective matches the mismatch type: measurement residual for geometric mismatch, reconstruction sparsity for radiometric mismatch.

*Limitations.* Our parametric mismatch models (affine shifts, gain drift) capture dominant error sources but do not cover spatially varying PSF errors or nonlinear detector response. Real hardware validation covers CASSI and CACTI; SPC real data would strengthen generalisability. The grid-search calibration does not scale to high-dimensional spaces—gradient-based or learned calibration is a natural next step.

*Residual gap analysis.* The residual gap  $\Delta_{\text{res}} = \text{PSNR}_I - \text{PSNR}_{III}$  measures unrecoverable losses. CASSI shows the largest residual gaps (MST-L: 7.48 dB) due to fixed-step dispersion assumptions, while CACTI (GAP-TV: 0.74 dB) and SPC (HATNet: 1.20 dB) confirm that spatial and gain-type mismatches are nearly fully recoverable. Per-method residual gap visualizations are provided in the supplementary material.

## 6 Conclusion

We have presented InverseNet, the first cross-modality benchmark for operator mismatch in compressive imaging. Evaluating 12 methods across CASSI, CACTI, and SPC under a four-scenario protocol, we find: (1) mismatch degrades deep learning methods by 10–21 dB, collapsing their advantage over classical methods; (2) operator-conditioned architectures recover 40–90% of losses through calibration, while mask-oblivious ones recover 0%; (3) blind grid-search calibration (Scenario IV) recovers 85–100% of the oracle bound without ground truth, using measurement residual for geometric and reconstruction sparsity for radiometric mismatch; (4) real hardware experiments confirm simulation patterns transfer to physical data. When calibration is feasible, operator-conditioned networks paired with self-supervised calibration are optimal; otherwise, classical methods provide the most robust baseline.

All reconstruction arrays, metrics, and code will be released upon acceptance. Future work includes gradient-based calibration, dispersion-aware architectures, and expansion to lensless imaging and ptychography.

## References

1. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics* **47**(10), B44–B51 (2008)
2. Meng, Z., Ma, J., Yuan, X.: End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In: ECCV. pp. 187–204 (2020)
3. Llull, P., Liao, X., Yuan, X., Yang, J., Kittle, D., Carin, L., Sapiro, G., Brady, D.J.: Coded aperture compressive temporal imaging. *Optics Express* **21**(9), 10526–10545 (2013)
4. Yuan, X., Brady, D.J., Katsaggelos, A.K.: Snapshot compressive imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine* **38**(2), 65–88 (2021)

5. Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J.N., Sun, T., Kelly, K.F., Baraniuk, R.G.: Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine* **25**(2), 83–91 (2008)
6. Edgar, M.P., Gibson, G.M., Padgett, M.J.: Principles and prospects for single-pixel imaging. *Nature Photonics* **13**(1), 13–20 (2019)
7. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. In: ICIP. pp. 2539–2543 (2016)
8. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009)
9. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* **3**(1), 1–122 (2011)
10. Cai, Y., Lin, J., Hu, X., Wang, H., Yuan, X., Zhang, Y., Timofte, R., Van Gool, L.: Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In: CVPR. pp. 17502–17511 (2022)
11. Hu, X., Cai, Y., Lin, J., Wang, H., Yuan, X., Zhang, Y., Timofte, R., Van Gool, L.: HDNet: High-resolution dual-domain learning for spectral compressive imaging. In: CVPR. pp. 17542–17551 (2022)
12. Wang, L., Cao, M., Yuan, X.: EfficientSCI: Densely connected network with space-time factorization for large-scale video snapshot compressive imaging. In: CVPR. pp. 18477–18486 (2023)
13. Yang, C., Zhang, S., Yuan, X.: Ensemble learning priors driven deep unfolding for scalable video snapshot compressive imaging. In: ECCV. pp. 600–618 (2022)
14. Zhang, J., Ghanem, B.: ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In: CVPR. pp. 1828–1837 (2018)
15. Qu, G., Wang, P., Yuan, X.: Dual-scale transformer for large-scale single-pixel imaging. In: CVPR. pp. 25327–25337 (2024)
16. Yuan, X., Liu, Y., Suo, J., Dai, Q.: Plug-and-play algorithms for large-scale snapshot compressive imaging. In: CVPR. pp. 1447–1457 (2020)
17. Choi, I., Jeon, D.S., Nam, G., Gutierrez, D., Kim, M.H.: High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics* **36**(6), 218:1–218:13 (2017)
18. Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., Ashok, A.: ReconNet: Non-iterative reconstruction of images from compressively sensed measurements. In: CVPR. pp. 449–458 (2016)
19. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastMRI: An open dataset and benchmarks for accelerated MRI. arXiv preprint arXiv:1811.08839 (2018)
20. Elser, V.: Phase retrieval by iterated projections. *Journal of the Optical Society of America A* **20**(1), 40–55 (2003)
21. Arguello, H., Rueda, H., Wu, Y., Prather, D.W., Arce, G.R.: Higher-order computational model for coded aperture spectral imaging. *Applied Optics* **52**(10), D12–D21 (2013)
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004)
23. Cai, Y., Lin, J., Wang, H., Yuan, X., Ding, H., Zhang, Y., Timofte, R., Van Gool, L.: Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. In: NeurIPS (2022)

24. Cai, Y., Lin, J., Hu, X., Wang, H., Yuan, X., Zhang, Y., Timofte, R., Van Gool, L.: Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In: ECCV. pp. 686–704 (2022)
25. Wang, L., Cao, M., Zhong, Y., Yuan, X.: Spatial-temporal transformer for video snapshot compressive imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(7), 9072–9089 (2023)
26. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE Transactions on Medical Imaging **37**(6), 1322–1332 (2018)
27. Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., Yin, W.: Plug-and-play methods provably converge with properly trained denoisers. In: ICML. pp. 5546–5557 (2019)
28. Moult, J., Fidelis, K., Zemla, A., Hubbard, T.: Critical assessment of methods of protein structure prediction (CASP)—round 6. Proteins **61**(S7), 3–7 (2005)
29. Jumper, J., Evans, R., Pritzel, A., et al.: Highly accurate protein structure prediction with AlphaFold. Nature **596**(7873), 583–589 (2021)
30. Antun, V., Renna, F., Poon, C., Adcock, B., Hansen, A.C.: On instabilities of deep learning in image reconstruction and the potential costs of AI. PNAS **117**(48), 30088–30098 (2020)
31. Timofte, R., Agustsson, E., Van Gool, L., et al.: NTIRE 2017 challenge on single image super-resolution: Methods and results. In: CVPR Workshops. pp. 1110–1121 (2017)
32. Köhler, T., Bätz, M., Naderi, F., et al.: Toward bridging the simulated-to-real gap: Benchmarking super-resolution on real data. IEEE TPAMI **42**(11), 2944–2959 (2020)
33. Li, Y., Cai, Y., Lin, J., Hu, X., Yuan, X., Zhang, Y., Timofte, R., Van Gool, L.: RDLUF-MixS<sup>2</sup>: Spatial-spectral feature mixing for spectral compressive imaging. In: CVPR. pp. 22262–22271 (2023)
34. Meng, Z., Yu, Z., Xu, K., Yuan, X.: DiffSCI: Zero-shot snapshot compressive imaging via iterative spectral diffusion model. In: CVPR. pp. 25857–25867 (2024)
35. Berk, A., Plan, Y., Bhatt, N.: Robustness of compressed sensing under structured model mismatch. Foundations of Computational Mathematics **24**(4), 1285–1337 (2024)
36. Liu, Q., Luo, Y., Fan, L., Qiao, X., Yuan, X.: Under-display camera image restoration with scattering effect. In: NeurIPS (2024)
37. Zheng, S., Liu, Y., Meng, Z., Müller, M., Shu, R., Yuan, X.: Deep plug-and-play priors for spectral snapshot compressive imaging. Photonics Research **9**(2), B18–B29 (2021)
38. Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., Timofte, R.: Plug-and-play image restoration with deep denoiser prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10), 6360–6376 (2022)

CASSI Real Data: Scene 1, Band 14 (Central 256×256 Patch)

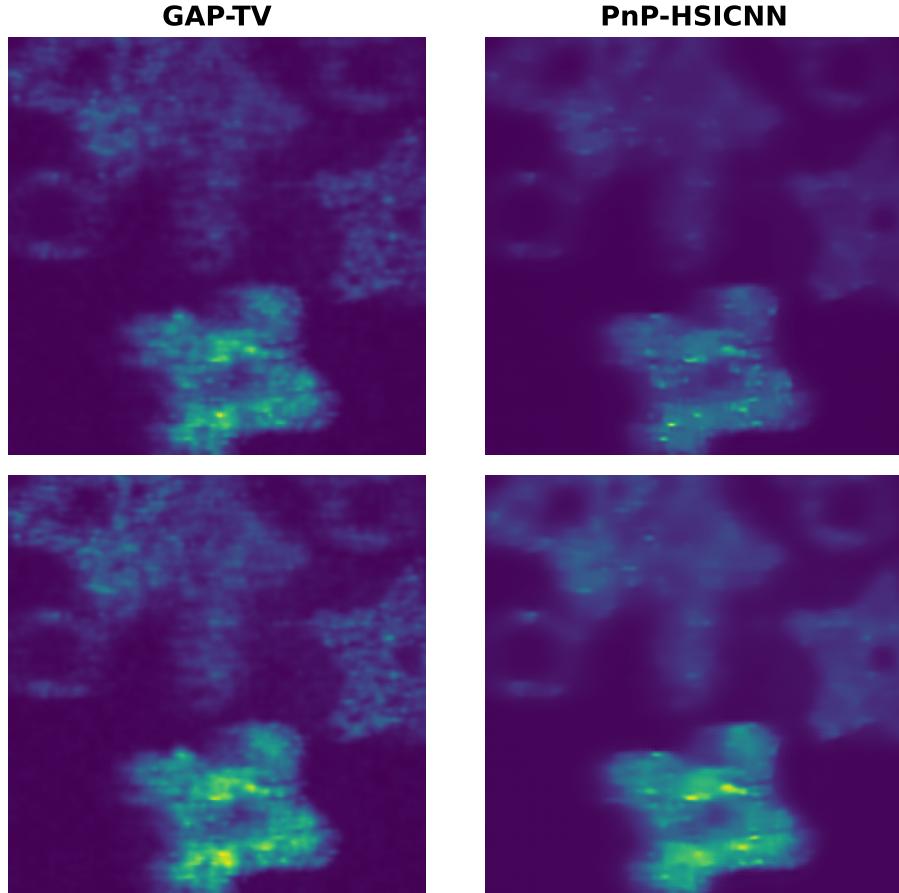


Fig. 5: CASSI real data (Scene 1, band 14): calibrated (top) vs. mismatched (bottom) reconstructions for GAP-TV and PnP-HSICNN. The visual differences are subtle, consistent with the modest residual ratios in table 5—spatial mask shift alone is not the dominant degradation source for CASSI.

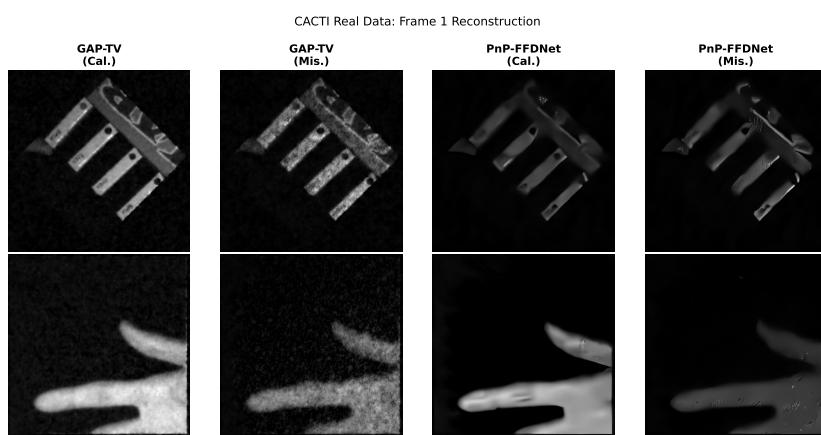


Fig. 6: CACTI real data: calibrated vs. mismatched reconstruction for *duomino* and *hand* scenes. Mismatched masks produce temporal ghosting artifacts, mirroring simulation findings.