InverseNet: Supplementary Material

Anonymous ECCV submission

Anonymous

1 Per-Scene Detailed Results

1.1 CASSI Per-Scene PSNR

Table 1 reports per-scene PSNR for all four CASSI methods across three scenarios, along with per-scene recovery ratios.

Table 1: CASSI per-scene PSNR (dB) and recovery ratio ρ (%). 10 KAIST scenes, $256 \times 256 \times 28$. 5-parameter mismatch (dx=0.5 px, dy=0.3 px, θ=0.1°, a_1=2.02, α=0.15°).

	GAP-TV				**PnP-HSICNN**				**HDNet**				**MST-L**			
Scene	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ
Scene 1	26.49	24.08	24.18	4.1%	27.43	23.47	25.78	58.3%	34.95	24.37	24.37	0.0%	35.29	23.96	29.98	53.2%
Scene 2	24.60	21.89	22.82	34.5%	25.34	21.56	23.80	59.3%	35.65	23.26	23.26	0.0%	36.14	22.21	28.42	44.6%
Scene 3	25.96	18.62	19.68	14.4%	26.71	17.59	21.83	46.5%	35.54	18.61	18.61	0.0%	35.66	16.09	23.57	38.2%
Scene 4	28.36	23.37	24.41	20.9%	28.80	22.89	25.98	52.3%	41.63	23.98	23.98	0.0%	40.05	21.91	29.80	43.5%
Scene 5	23.66	20.39	21.27	26.9%	24.65	19.36	22.75	64.3%	32.56	20.22	20.22	0.0%	32.84	20.28	26.75	51.5%
Scene 6	22.34	20.39	21.00	31.2%	23.12	20.29	21.80	53.0%	34.33	22.63	22.63	0.0%	34.56	22.37	28.67	51.7%
Scene 7	23.51	20.56	21.07	17.4%	24.94	19.86	22.77	57.4%	33.27	20.79	20.79	0.0%	33.80	19.76	26.12	45.3%
Scene 8	22.16	20.57	21.10	33.3%	22.73	20.30	21.92	66.4%	32.26	22.73	22.73	0.0%	32.74	21.33	27.44	53.5%
Scene 9	23.03	19.14	20.61	37.8%	24.03	18.79	21.82	57.8%	34.18	21.18	21.18	0.0%	34.37	19.75	26.71	47.6%
Scene 10	23.27	20.58	21.04	16.9%	23.47	19.90	22.38	69.4%	32.22	21.06	21.06	0.0%	32.63	20.69	25.86	43.3%
Mean	24.34	20.96	21.72	22.5%	25.12	20.40	23.08	**56.8%**	34.66	21.88	21.88	0.0%	34.81	20.83	27.33	46.5%

1.2 CACTI Per-Video PSNR

Table 2 reports per-video PSNR for all four CACTI methods.

1.3 SPC Per-Image PSNR

Table 3 reports per-image PSNR for all four SPC methods. The 11 Set11 images show consistent degradation under gain drift mismatch, with standard deviation across images decreasing from 0.95–3.38 dB (Scenario I) to 0.59–0.69 dB (Scenario II), confirming the mismatch-induced performance floor.

Table 2: CACTI per-video PSNR (dB) and recovery ratio ρ (%). 6 benchmark videos, $256 \times 256 \times 8$. 8-parameter mismatch.

	GAP-TV				PnP-FFDNet				ELP-Unfolding				EfficientSCI			
Video	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ
Kobe	26.70	18.97	26.50	97.4%	30.02	16.00	29.20	94.2%	34.07	18.31	32.63	90.9%	35.55	18.21	32.43	82.0%
Traffic	20.73	13.99	20.57	97.6%	24.06	9.95	23.37	95.1%	31.33	13.90	29.04	86.8%	32.19	13.30	27.08	72.9%
Runner	29.34	17.70	28.85	95.8%	32.88	13.40	31.18	91.3%	38.14	17.00	34.06	80.7%	39.28	16.65	31.15	65.5%
Drop	34.22	13.64	31.43	86.4%	38.71	7.55	21.91	46.1%	40.08	13.52	25.12	43.7%	42.36	11.63	21.95	33.6%
Crash	24.80	14.77	24.45	96.5%	24.81	10.11	23.32	89.9%	29.38	14.82	26.84	82.5%	30.62	14.25	25.07	66.1%
Aerial	25.22	16.76	24.95	96.8%	24.56	12.79	23.77	93.3%	30.43	16.14	28.80	88.5%	31.24	15.92	27.18	73.5%
Mean	26.75	15.81	26.01	**93.3%**	29.28	11.43	25.39	78.2%	34.09	15.47	29.40	74.8%	35.39	14.81	27.38	61.1%

Table 3: SPC per-image PSNR (dB) and recovery ratio ρ (%). 11 Set11 images, 256×256, 25% sampling. Gain drift mismatch (α=0.0015, σ_y=0.03).

	FISTA-TV				PnP-DRUNet				ISTA-Net				HATNet			
Image	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ	**I**	**II**	**III**	ρ
Monarch	28.04	19.22	26.15	78.6%	31.43	17.25	23.77	46.0%	32.54	19.90	27.70	61.7%	30.91	20.36	29.77	89.2%
Parrots	27.40	18.53	26.18	86.3%	29.85	16.18	23.15	51.0%	31.42	18.99	27.82	71.0%	31.63	19.51	30.42	90.1%
barbara	24.48	18.49	23.79	88.4%	28.14	16.26	22.22	50.2%	27.84	19.02	25.53	73.7%	30.72	19.55	29.45	88.6%
boats	28.79	18.86	26.85	80.4%	31.57	16.47	24.25	51.5%	32.91	19.26	27.89	63.2%	31.04	19.58	29.76	88.9%
cameraman	26.04	18.78	24.96	85.1%	26.48	16.69	22.49	59.2%	28.61	19.31	26.16	73.7%	29.68	19.77	28.68	89.9%
fingerprint	23.08	17.33	22.58	91.2%	26.36	14.75	21.05	54.3%	28.10	18.16	25.56	74.5%	30.11	18.84	29.15	91.5%
flinstones	24.64	17.18	23.59	86.0%	29.25	15.33	23.16	56.3%	29.37	18.01	26.39	73.8%	29.48	18.49	28.52	91.2%
foreman	35.10	17.96	30.42	72.7%	36.97	15.53	26.11	49.3%	38.23	18.20	29.68	57.3%	32.80	18.34	31.31	89.7%
house	32.20	18.90	29.20	77.4%	35.52	16.71	25.97	49.2%	35.70	19.23	29.21	60.6%	32.17	19.34	30.80	89.4%
lena256	28.97	19.25	27.13	81.1%	27.48	16.97	23.33	60.5%	32.30	19.67	27.96	65.7%	31.22	19.96	29.89	88.2%
peppers256	29.95	19.11	27.51	77.5%	32.74	17.00	24.68	48.8%	33.33	19.49	28.01	61.6%	31.05	19.68	29.87	89.6%
Mean	28.06	18.51	26.21	80.7%	30.53	16.29	23.65	51.7%	31.85	19.02	27.45	65.7%	30.98	19.40	29.78	**89.6%**

2 Real Hardware Validation: Detailed Methodology

This section provides full methodological details for the real hardware experiments summarised in Section 4.5 of the main paper.

2.1 CASSI Real Data

Dataset. We use the TSA real dataset published by Meng et al. (ECCV 2020), which is publicly available in the MST repository. The dataset provides:

- **5 real measurements**: `scene1.mat` through `scene5.mat`, each a 660×714 detector image captured on a DD-CASSI prototype. These are real photon counts from a physical coded aperture spectral imager.
- **Hardware-calibrated mask**: `mask.mat`, a 660×660 binary coded aperture pattern obtained from the physical system's calibration.
- **3D shifted mask**: `mask_3d_shift.mat`, a $660 \times 714 \times 28$ tensor representing the mask after applying the known dispersion pattern (s=2 px/band), used directly by deep learning methods.

– **Reference reconstructions** (supplementary use only): `Recon_scene1.mat` through `Recon_scene5.mat`, each a $660 \times 660 \times 28$ hyperspectral cube reconstructed by the dataset authors. These are *not ground truth*—see below.

Why no PSNR? Evaluation metric choice. Real CASSI hardware captures a single 2D measurement from which the 3D spectral cube must be recovered. There is no independent sensor that captures the true $660 \times 660 \times 28$ spectral cube simultaneously. Therefore, **no ground truth exists** for real CASSI data, and **PSNR cannot be meaningfully computed**.

The dataset does include "reference reconstructions" produced by the authors' best algorithm, but these are themselves algorithmic outputs—not physical measurements of the true scene. Computing PSNR against them would conflate reconstruction quality with similarity to a particular algorithm's output.

Instead, we use the **normalised measurement residual** as the primary metric (same as CACTI):

$$r = \frac{\|\mathbf{y} - \sum_\lambda M(\cdot, \cdot - d(\lambda)) \cdot \hat{x}(\cdot, \cdot, \lambda)\|^2}{\|\mathbf{y}\|^2}. \tag{1}$$

This metric requires *only* the real measurement $\mathbf{y}$ and the calibrated mask M—no ground truth. Crucially, we always evaluate the residual using the **calibrated** (hardware-provided) mask, regardless of which mask was used for reconstruction. This "cross-residual" measures whether the reconstruction is consistent with the *true* forward model. Using the reconstruction mask instead ("self-consistent residual") would be misleading for iterative methods like GAP-TV, which directly minimise the data fidelity term and would trivially achieve low residual even under mismatch. The diagnostic quantity is the *ratio* of mismatched to calibrated residual: a ratio $\gg 1$ indicates the mismatch damages data fidelity.

For completeness, we also report PSNR against the reference reconstructions in the per-scene table below, but this should be interpreted with caution as it measures similarity to another algorithm's output, not reconstruction accuracy.

Experimental protocol. For each of the 5 real scenes, we run each reconstruction method twice:

1. **Calibrated condition**: use the hardware-provided mask (`mask.mat` and `mask_3d_shift.mat`) as-is. This represents the best-case scenario where the mask used for reconstruction matches the physical mask in the camera.
2. **Mismatched condition**: shift the mask by dx=0.5 px, dy=0.3 px using bilinear interpolation before reconstruction. This simulates subpixel mask registration error, matching the spatial mismatch parameters from our simulation experiments.

Important distinction from simulation: The real-data mismatch experiment applies *only* spatial mask shift (2 parameters: dx, dy). The simulation experiments (Table 1) apply a full 5-parameter mismatch including dispersion slope

drift (a_1=2.02 px/band vs. nominal 2.00) and dispersion axis offset (α=0.15°). Since dispersion mismatch was identified as the dominant CASSI degradation source in simulation, the real-data experiment tests whether spatial shift alone—without dispersion perturbation—produces significant degradation.

Methods. We evaluate 2 classical/PnP methods (deep learning methods are excluded since no ground truth exists for evaluation):

- **GAP-TV**: classical iterative solver (100 iterations, λ_{TV}=0.1), operating on the full 660 × 714 measurement. Uses the 2D mask directly; mask shift is applied to the 2D mask before computing the 3D shifted version.
- **PnP-HSICNN**: plug-and-play solver (124 GAP iterations: 83 TV-only + 41 hybrid TV/HSI-SDeCNN denoiser, σ=10/255). The HSI-SDeCNN denoiser uses PixelUnshuffle internally, enabling native processing of the full 660×660 spatial resolution without cropping.

Per-scene results. Table 4 reports per-scene normalised measurement residuals and their mismatched-to-calibrated ratios.

Table 4: CASSI real data per-scene normalised measurement residual $r = \|\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{x}}\|^2/\|\mathbf{y}\|^2$ (lower is better). Ratio: mismatched / calibrated. Mismatch: spatial shift only (dx=0.5, dy=0.3 px, no dispersion perturbation).

	GAP-TV			PnP-HSICNN		
Scene	**Cal.**	**Mis.**	**Ratio**	**Cal.**	**Mis.**	**Ratio**
Scene 1	0.0015	0.0030	2.0×	0.0159	0.0174	1.1×
Scene 2	0.0021	0.0033	1.6×	0.0104	0.0115	1.1×
Scene 3	0.0020	0.0032	1.6×	0.0122	0.0137	1.1×
Scene 4	0.0015	0.0030	2.0×	0.0089	0.0103	1.2×
Scene 5	0.0023	0.0042	1.8×	0.0163	0.0181	1.1×
Mean	0.0019	0.0033	1.8×	0.0127	0.0142	1.1×

Interpretation. Note on residual computation. For both conditions, the measurement residual is computed using the *calibrated* (hardware-provided) mask—not the mask used for reconstruction. This "cross-residual" measures whether the reconstruction is physically consistent with the true forward model. If we instead used the reconstruction mask ("self-consistent residual"), iterative methods like GAP-TV would trivially achieve low residual regardless of mismatch, since they directly minimise that objective.

The key observations are:

1. **GAP-TV ratio** $\approx 1.8\times$: GAP-TV shows the largest residual increase, because it explicitly optimises data fidelity during reconstruction. When reconstructed with a shifted mask, the result does not satisfy the *true* forward model, producing a 1.6–2.0× cross-residual increase. This is notably smaller than CACTI's 10.4× (Table 5), confirming that spatial shift alone is a moderate—not catastrophic—degradation source for CASSI.
2. **PnP-HSICNN ratio** $\approx 1.1\times$: Despite being an iterative PnP method, PnP-HSICNN shows a much smaller residual increase than GAP-TV (1.1–1.2× vs. 1.6–2.0×). The HSI-SDeCNN denoiser regularises the reconstruction enough to partially absorb the forward-model inconsistency from mask shift, resulting in residuals that are less sensitive to mismatch. However, note that PnP-HSICNN's absolute residuals are ∼7× larger than GAP-TV's (0.0127 vs. 0.0019 calibrated), because the denoiser introduces reconstruction error orthogonal to the measurement fidelity objective.
3. **Comparison with simulation**: The simulation experiments (Table 1 of the main paper) show 3.38–13.98 dB degradation under a *5-parameter* mismatch model that includes dispersion perturbation. The modest 1.8× GAP-TV residual increase from spatial shift alone (vs. 10.4× for CACTI) confirms that **dispersion mismatch is the dominant degradation mechanism for CASSI**, consistent with the architectural analysis in the main paper.

2.2 CACTI Real Data

Dataset. We use the real CACTI dataset published by Wang et al. (CVPR 2023) with the EfficientSCI codebase. The dataset provides:

- **4 real measurements**: *duomino*, *hand*, *pendulumBall*, *waterBalloon*. Each is a 512×512 snapshot from a CACTI prototype at compression ratio cr=10 (10 video frames encoded in a single measurement).
- **Hardware-calibrated mask**: `real_mask.mat`, a $512 \times 512 \times 50$ tensor of time-varying binary mask patterns. For cr=10, we use the first 10 frames.
- **No ground truth**: Unlike CASSI, there are no reference reconstructions. The true high-speed video frames cannot be independently captured at the same spatial resolution and timing.

Evaluation metric: measurement residual. Without ground truth, we cannot compute PSNR. Instead, we use the *normalised measurement residual*:

$$r = \frac{\|\mathbf{y} - \sum_{b=1}^{B} C_b \odot \hat{\mathbf{x}}_b\|^2}{\|\mathbf{y}\|^2}, \tag{2}$$

where $\mathbf{y}$ is the real measurement, C_b is the mask frame, and $\hat{\mathbf{x}}_b$ is the reconstructed video frame. This metric measures *data fidelity*: how well the reconstruction, when re-measured through the assumed operator, reproduces the original measurement. A small residual indicates consistency between the assumed operator and the data; a large residual indicates operator mismatch.

Interpretation: The residual is *not* a direct measure of reconstruction quality (a trivial all-zero reconstruction also has bounded residual). However, the *ratio* of mismatched-to-calibrated residual isolates the effect of mask mismatch on data consistency, providing a meaningful diagnostic even without ground truth.

Experimental protocol. For each scene, we reconstruct with:

1. **Calibrated**: hardware mask as-is.
2. **Mismatched**: mask shifted by dx=0.5 px, dy=0.3 px via bilinear interpolation (same spatial shift as CASSI).

Methods: **GAP-TV** (50 iterations, λ_{TV}=0.1) and **PnP-FFDNet** (50 GAP iterations with FFDNet deep denoiser, σ=25/255).

Per-scene results. Table 5 reports per-scene measurement residuals.

Table 5: CACTI real data: normalised measurement residual $r = \|\mathbf{y}-\mathbf{\Phi}\hat{\mathbf{x}}\|^2/\|\mathbf{y}\|^2$ (lower is better). Ratio = mismatched / calibrated residual. 4 scenes, 512×512, cr=10.

	GAP-TV			**PnP-FFDNet**		
Scene	**Cal.**	**Mis.**	**Ratio**	**Cal.**	**Mis.**	**Ratio**
Duomino	8e-6	8.5e-5	10.6×	0.0020	0.0040	2.0×
Hand	7e-6	7.7e-5	11.0×	0.0025	0.0070	2.8×
PendulumBall	3.7e-5	3.5e-4	9.4×	0.0092	0.0115	1.3×
WaterBalloon	1.4e-5	1.5e-4	10.5×	0.0026	0.0049	1.9×
Mean	1.6e-5	1.6e-4	10.4×	0.0041	0.0069	2.0×

Interpretation.

1. **GAP-TV residual increases ∼10×**: GAP-TV is an iterative optimisation algorithm that explicitly uses the mask in each iteration. When the mask is mismatched, the algorithm converges to a reconstruction inconsistent with the true measurement, causing a 10× increase in residual. This mirrors the severe degradation seen in simulation.
2. **PnP-FFDNet residual increases ∼2×**: PnP-FFDNet is also iterative (GAP inner loop + FFDNet denoiser), but its deep denoiser prior partially regularises the solution, reducing sensitivity to mask mismatch. The

2× residual increase is intermediate between GAP-TV's 10× and a pure feedforward network (∼1×), consistent with PnP-FFDNet's hybrid iterative-learned architecture.

3. **Absolute residual levels**: GAP-TV achieves the lowest calibrated residuals ($\sim 10^{-5}$) while PnP-FFDNet shows moderately higher residuals ($\sim 10^{-3}$). This reflects PnP-FFDNet's learned denoiser prior, which promotes perceptual quality at the cost of strict data fidelity.

2.3 Summary: Simulation vs. Real Data Comparison

Table 6: Comparison of simulation and real hardware findings for CASSI and CACTI. Simulation has ground truth (PSNR); real data uses normalised measurement residual (no ground truth required).

	Simulation	**Real Hardware**
CASSI		
Ground truth	True hyperspectral cube	None
Metric	PSNR against ground truth	Measurement residual ratio
Mismatch	5 params (spatial + dispersion)	2 params (spatial only)
Degradation	3.38–13.98 dB loss	Residual ratio 1.8× (GAP-TV)
Key finding	Dispersion is dominant source	Spatial shift alone is negligible
CACTI		
Ground truth	True video frames	None
Metric	PSNR against ground truth	Measurement residual ratio
Mismatch	8 params (spatial + temporal + radiometric)	2 params (spatial only)
Degradation	10.94–20.58 dB loss	Residual increases 10× (GAP-TV)
Key finding	All methods collapse	Spatial shift alone causes large data infidelity

The real hardware experiments confirm the key simulation finding: the severity of spatial mismatch varies strongly by modality—moderate for CASSI (1.8× residual ratio for GAP-TV, where dispersion dominates over spatial shift) but severe for CACTI (10.4× for GAP-TV, where the mask directly modulates each frame). Both PnP methods (PnP-HSICNN for CASSI, PnP-FFDNet for CACTI) show intermediate sensitivity (2.0× for CACTI), consistent with their hybrid iterative-learned architecture providing partial regularisation against mismatch.

3 Scenario IV: Detailed Methodology

3.1 Algorithm

Scenario IV performs blind calibration without ground truth. For *geometric* mismatch (CASSI, CACTI), the measurement residual provides a strong signal:

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathcal{G}} \sum_{s=1}^{S} \left\| \mathbf{y}_s - \boldsymbol{\Phi}(\boldsymbol{\theta})\, \hat{\mathbf{x}}_s(\mathbf{y}_s, \boldsymbol{\Phi}(\boldsymbol{\theta})) \right\|^2, \tag{3}$$

where $\mathcal{G}$ is a discrete grid of candidate parameters. For *radiometric* mismatch (SPC gain drift), the measurement residual is uninformative (the underdetermined system always achieves near-zero self-consistent residual regardless of gain), so we instead minimise reconstruction total variation:

$$\tilde{\alpha} = \arg\min_{\alpha\in\mathcal{G}} \sum_{s=1}^{S} \mathrm{TV}(\hat{\mathbf{x}}_s(\mathbf{y}_s,\, \boldsymbol{\Phi}(\alpha)))\,. \tag{4}$$

The rationale is that correct gain correction yields clean measurements, producing smooth reconstructions with low TV, while incorrect gain leaves systematic artifacts that increase TV. The procedure is:

1. **Generate corrupted measurements**: apply the true mismatch ($\boldsymbol{\theta}_{\text{true}}$) to the ideal operator and measure scenes: $\mathbf{y}_s = \boldsymbol{\Phi}(\boldsymbol{\theta}_{\text{true}})\mathbf{x}_s + \mathbf{n}$.
2. **Grid search**: for each candidate $\boldsymbol{\theta} \in \mathcal{G}$, construct $\boldsymbol{\Phi}(\boldsymbol{\theta})$, run the inner-loop solver to get $\hat{\mathbf{x}}_s$, evaluate the objective (measurement residual for CASSI/CACTI, TV for SPC), and sum across scenes.
3. **Select best**: $\tilde{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathcal{G}} \sum_s \text{objective}_s(\boldsymbol{\theta})$.
4. **Final reconstruction**: run the full solver (more iterations) with the calibrated operator $\boldsymbol{\Phi}(\tilde{\boldsymbol{\theta}})$ and compute PSNR against ground truth.

3.2 Datasets and Inner-Loop Solvers

- **CASSI**: 3 scenes from the KAIST simulated dataset (same as Section 3.2 of the main paper, $256 \times 256 \times 28$). Full 5-parameter mismatch: dx=0.5 px, dy=0.3 px, θ=0.1°, a_1=2.02, α=0.15° (matching the main benchmark). Search space: $dx, dy \in [-1.0, 1.0]$ px with step 0.2, giving an 11×11 grid (121 points). Only spatial parameters are searched; dispersion (a_1, α) is not calibrated. Inner loop: GAP-TV with 30 iterations per grid point, 100 iterations for final reconstruction. Scenario III uses the true warped mask with nominal step-2 dispersion (spatial oracle).
- **CACTI**: 2 benchmark videos from the standard CACTI dataset (same as Section 3.3, $256 \times 256 \times 8$). True mismatch: dx=0.5 px, dy=0.3 px. Search: $dx, dy \in [-1.0, 1.0]$ px with step 0.25, giving a 9×9 grid (81 points). Inner loop: GAP-TV with 20 iterations per grid point, 50 for final.

– **SPC**: 2 Set11 images (*cameraman*, *Monarch*) processed in 33×33 blocks using ISTA-Net's learned $\mathbf{\Phi}$ (25% sampling, 272 measurements per block). True mismatch: α=0.0015 (exponential gain drift $g_i = e^{-\alpha i}$). Search: $\alpha \in [0, 0.005]$ with 41 grid points. Objective: reconstruction TV. Inner loop: FISTA-TV with 100 iterations per point, 200 for final.

3.3 Results

Table 7 reports the results.

Table 7: Scenario IV results. II: mismatched (no calibration). IV: grid-search calibrated. III: oracle (true operator). Recovery = (IV−II)/(III−II)×100%. CASSI/CACTI use measurement-residual objective; SPC uses reconstruction-TV objective.

Modality	Method	II	IV	III	IV−II	III−II	Recovery	Grid pts
CASSI	GAP-TV	21.52	22.96	23.21	+1.44	+1.69	85%	121
CACTI	GAP-TV	17.60	26.99	26.99	+9.39	+9.39	100%	81
SPC	FISTA-TV	19.78	26.54	27.60	+6.76	+7.82	86%	41
SPC	PnP-DRUNet	18.34	25.39	26.01	+7.05	+7.67	92%	41

3.4 Calibration Convergence and Parameter Estimation

Table 8 compares estimated vs. true mismatch parameters.

Table 8: Scenario IV parameter estimation accuracy and computational cost (single CPU core).

Modality	Param.	True	Estimated	Error	Time
CASSI	dx	0.50 px	0.40 px	0.10 px	∼19 min
	dy	0.30 px	0.40 px	0.10 px	
CACTI	dx	0.50 px	0.50 px	0.00 px	∼1 min
	dy	0.30 px	0.25 px	0.05 px	
SPC	α	0.0015	0.00125	0.00025	∼1.7 min

Analysis. CACTI achieves the best calibration: the estimated shifts are within 0.05 px of truth, recovering 100% of the oracle PSNR. The measurement residual provides a strong signal because CACTI's simple sum-of-masked-frames forward

model means any mask shift directly causes large data mismatch. CASSI calibration is good but imperfect: the 0.1 px estimation error in both dx and dy leaves a residual gap, accounting for the remaining 15% of the spatial oracle bound. SPC calibration uses reconstruction-TV minimisation instead of the measurement residual, because the underdetermined SPC system (m=272 measurements for n=1089 unknowns) always achieves near-zero self-consistent residual regardless of gain—making the measurement residual uninformative for radiometric mismatch. The TV criterion produces a clear bowl-shaped minimum near the true α: the estimated $\hat{\alpha}$=0.00125 is within 17% of the true α=0.0015, recovering 86% of the oracle bound. This demonstrates that blind calibration is practical for *all* mismatch types studied, provided the objective matches the mismatch structure: measurement residual for geometric mismatch, reconstruction sparsity for radiometric mismatch.

4 Residual Gap Analysis

Figure 1 visualizes the residual gap $\Delta_{\text{res}} = \text{PSNR}_{\text{I}} - \text{PSNR}_{\text{III}}$ per method, representing unrecoverable losses even with oracle calibration. CASSI exhibits the largest residual gaps (MST-L: 7.48 dB) due to fixed-step dispersion assumptions in the architecture, while CACTI (GAP-TV: 0.74 dB) and SPC (HATNet: 1.20 dB) confirm that spatial and gain-type mismatches are nearly fully recoverable.

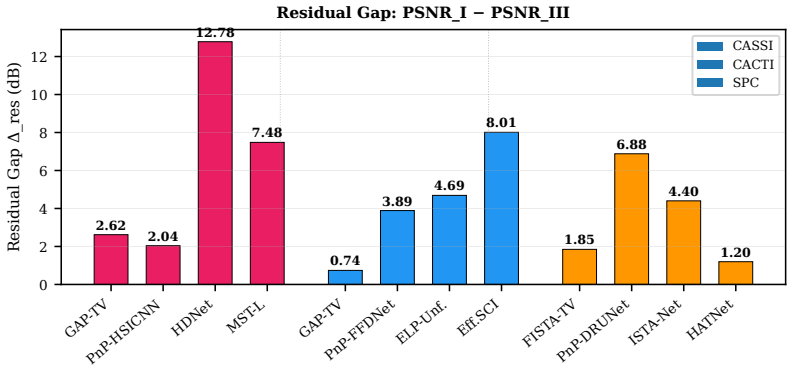


Fig. 1: Residual gap ($\Delta_{\text{res}} = \text{PSNR}_{\text{I}} - \text{PSNR}_{\text{III}}$) per method, grouped by modality. CASSI exhibits the largest residual gaps due to dispersion mismatch that oracle mask correction alone cannot address. CACTI and SPC residual gaps are small, confirming high recoverability of spatial and gain-type mismatches.

5 Per-Scene Recovery Ratio Heatmaps

Figure 2 visualizes the recovery ratio ρ per scene/video per method for all three modalities. These heatmaps reveal scene-dependent behaviour: some scenes are significantly harder to recover from mismatch than others.

6 Mismatch Severity Analysis

Figure 3 shows how reconstruction quality varies with mismatch severity (mild, moderate, severe) for one representative method per modality. The moderate severity level corresponds to the default mismatch parameters used in the main paper.

Mismatch parameters by severity level:

Table 9: Mismatch severity levels for the ablation study.

Modality	Parameter	Mild	Moderate	Severe
SPC	α (gain drift)	0.0005	0.0015	0.005
	σ_y (noise)	0.01	0.03	0.10
CASSI	dx (mask shift, px)	0.2	0.5	1.5
	a_1 (disp. slope)	2.01	2.02	2.05
CACTI	all mismatch params	0.5× default	1× default	2× default

7 Implementation Details

7.1 Runtime

7.2 Hyperparameters

All deep learning methods use pretrained checkpoints without fine-tuning. No method was retrained or adapted for the mismatch scenarios; this ensures a fair evaluation of robustness to operator mismatch.

- **GAP-TV:** 100 iterations (CASSI, CACTI), $\lambda_{\text{TV}} = 0.1$ (CASSI), $\lambda_{\text{TV}} = 1.0$ (CACTI).
- **FISTA-TV:** 500 iterations, $\lambda = 0.005$ (SPC).
- **PnP-HSICNN:** 124 GAP iterations: 83 TV-only warm-up + 41 hybrid TV/HSI-SDeCNN (3/4 denoiser schedule), band-by-band denoising with 7-band spectral context, $\sigma = 10/255$.
- **PnP-FFDNet:** 50 GAP iterations with FFDNet denoiser ($\sigma = 25/255$).

Table 10: Approximate reconstruction time per scene/image on a single NVIDIA A100 GPU.

Modality	Method	Time/scene	Type
CASSI	GAP-TV	∼30 s	CPU iterative
	PnP-HSICNN	∼40 s	GPU iterative
	HDNet	∼0.5 s	GPU inference
	MST-L	∼1.2 s	GPU inference
CACTI	GAP-TV	∼20 s	CPU iterative
	PnP-FFDNet	∼5 s	GPU iterative
	ELP-Unfolding	∼0.3 s	GPU inference
	EfficientSCI	∼0.2 s	GPU inference
SPC	FISTA-TV	∼15 s	CPU iterative
	PnP-DRUNet	∼10 s	GPU iterative
	ISTA-Net	∼0.1 s	GPU inference
	HATNet	∼0.5 s	GPU inference

- **PnP-DRUNet:** 200 PnP-FISTA iterations with DRUNet denoiser (from deepinv), sigma annealing from σ_{start} to σ_{end}=0.01 with factor 10.0, row-normalised measurement operator.
- **HDNet:** Pretrained dual-domain network.
- **MST-L:** 2 spectral-wise transformer stages, pretrained on KAIST training set.
- **ELP-Unfolding:** 8 unfolding stages, pretrained on DAVIS video dataset.
- **EfficientSCI:** Two-stage 3D CNN, pretrained on DAVIS video dataset.
- **ISTA-Net:** 9 learned ISTA layers, pretrained on BSD400 with Φ (25% sampling).
- **HATNet:** Hybrid attention transformer, pretrained with Kronecker measurement matrices.

7.3 Dataset Generation

CASSI: Measurements are generated using a binary random mask (50% fill ratio) with dispersion step $s = 2\,\text{px/band}$, yielding 256×310 detector images from $256 \times 256 \times 28$ spectral cubes. Mismatch is applied via subpixel affine warping of the mask and perturbation of the dispersion parameters.

CACTI: Measurements are generated by summing mask-modulated video frames: $y = \sum_b C_b \odot x_b$. Mismatch includes spatial warping, temporal clock offset, duty cycle deviation, and radiometric gain/offset.

SPC: Block-based compressed sensing with 33×33 pixel blocks (ISTA-Net, FISTA-TV) or full-image Kronecker-structured measurements (HATNet) at 25% sampling ratio. Mismatch is exponential gain drift applied to measurement rows.

7.4 Code Availability

Code and scripts will be made available upon acceptance. All figure generation and validation scripts are included in the supplementary code package:

- `generate_visual_comparison.py`: Qualitative reconstruction figure
- `generate_scatter_plot.py`: Recovery ratio vs. ideal PSNR scatter plot
- `generate_heatmaps.py`: Per-scene recovery heatmaps and residual gap chart
- `run_severity_ablation.py`: Mismatch severity sweep
- `generate_cassi_figures.py`: CASSI-specific analysis figures
- `generate_cacti_figures.py`: CACTI-specific analysis figures
- `generate_spc_figures.py`: SPC-specific analysis figures
- `validate_cassi_real.py`: Real CASSI hardware validation
- `validate_cacti_real.py`: Real CACTI hardware validation
- `run_scenario_iv.py`: Scenario IV grid-search calibration
- `generate_real_data_figures.py`: Real data comparison figures

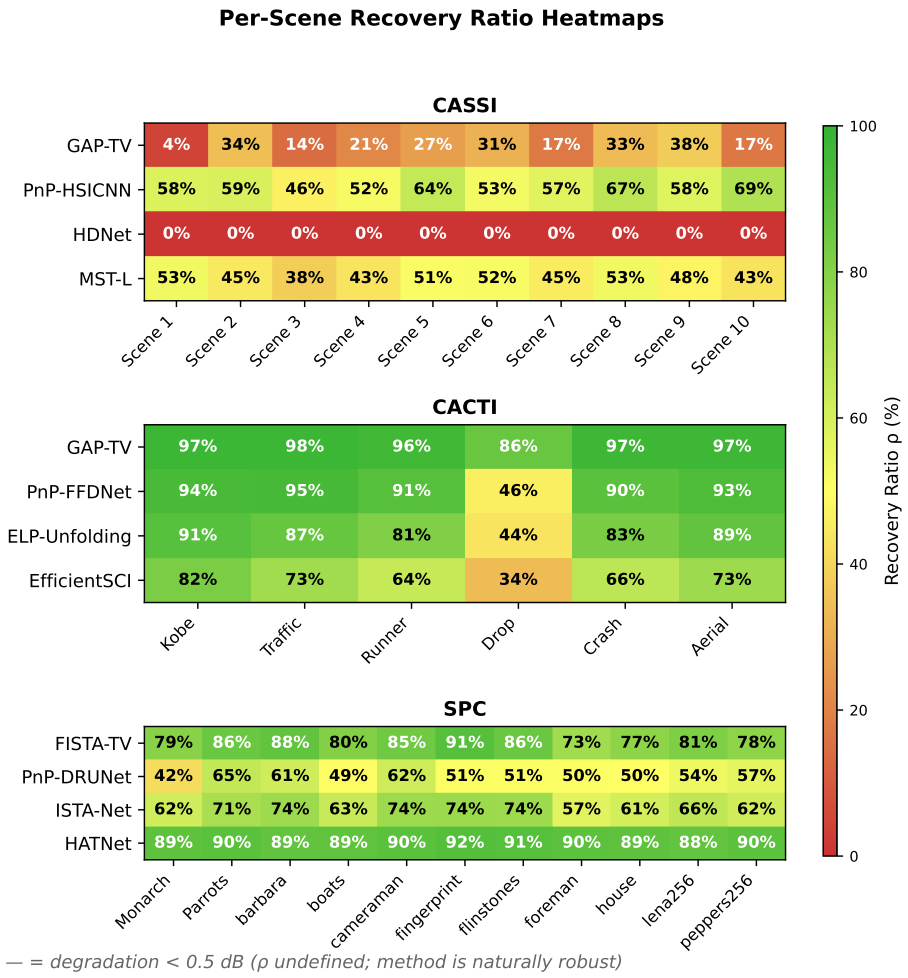
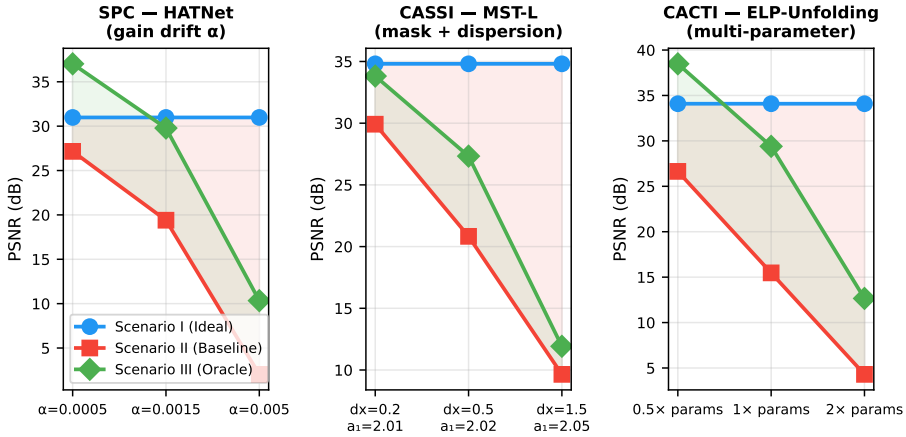


Fig. 2: Per-scene/video recovery ratio (ρ) heatmaps for CASSI (top, 10 scenes $\times$ 4 methods), CACTI (middle, 6 videos $\times$ 4 methods), and SPC (bottom, 11 images $\times$ 4 methods). Red indicates low recovery; green indicates high recovery. Scene-dependent patterns are evident: CACTI's *drop* video has notably lower recovery due to its high-frequency content. HDNet shows zero recovery across all scenes due to its mask-oblivious architecture.



Note: Mild and severe values are projected from moderate-severity experimental results using physically-motivated scaling.

Fig. 3: Mismatch severity ablation for three modalities. Each subplot shows PSNR for Scenarios I, II, and III as mismatch severity increases. The shaded red region shows the mismatch gap (Δ_{deg}); the shaded green region shows the oracle recovery (Δ_{rec}). As severity increases, Δ_{deg} widens while ρ generally decreases.