# Correcting Forward Model Mismatch in Coded Aperture Snapshot Spectral Imaging via Two-Stage Differentiable Calibration

Chengshuai Yang

NextGen PlatformAI C Corp

**Abstract.** Coded aperture snapshot spectral imaging (CASSI) captures a 3D hyperspectral cube from a single 2D measurement using a coded mask and spectral dispersion. Deep learning reconstructors such as MST achieve state-of-the-art quality (>34 dB) but assume perfect knowledge of the forward operator. In practice, sub-pixel mask misalignment ($\Delta x$, $\Delta y$, $\theta$) and dispersion drift ($a_1$, $\alpha$) between the coded aperture and detector are unavoidable, yet even moderate mismatch degrades MST-L reconstruction by over 16 dB. We propose a two-stage differentiable calibration pipeline: (1) a coarse hierarchical grid search scored by GPU-accelerated GAP-TV, followed by (2) joint gradient refinement through an unrolled differentiable forward operator using a Straight-Through Estimator (STE) for integer dispersion offsets, plus a 1D grid search for dispersion slope recovery. The pipeline is self-supervised, requiring only the measurement and nominal mask—no ground truth scene. On 10 KAIST benchmark scenes with injected 5-parameter mismatch ($\Delta x$=1.5, $\Delta y$=1.0, $\theta$=0.3°, $a_1$=2.04 px/band, $\alpha$=0.5°), our method recovers significant quality for mask-aware methods through self-supervised calibration. We evaluate five reconstruction methods (GAP-TV, MST-S, MST-L, HDNet, PnP-HSICNN) across four scenarios, revealing a mask-sensitivity spectrum: mask-guided transformers suffer catastrophic degradation (>15 dB) but gain most from calibration (∼3 dB), deep prior methods (HDNet) show moderate degradation (∼10 dB) with inherent robustness, and iterative methods show graduated sensitivity (GAP-TV ∼4.6 dB, PnP-HSICNN ∼6 dB degradation). We release all code, results, and a standardized four-scenario evaluation framework.

**Keywords:** CASSI · Operator mismatch · Differentiable calibration · Straight-Through Estimator · Hyperspectral imaging

## 1 Introduction

Coded aperture snapshot spectral imaging (CASSI) [11,1] captures a 3D hyperspectral cube $\mathbf{x} \in \mathbb{R}^{H \times W \times \Lambda}$ from a single 2D measurement $\mathbf{y} \in \mathbb{R}^{H \times W'}$ through the combined action of a binary coded mask and a dispersive element. The forward model is

$$y(i,j) = \sum_{k=1}^{\Lambda} m(i,j) \cdot x(i, j - d_k, k) + n(i,j), \tag{1}$$

where $m$ is the coded aperture, $d_k = k \cdot s$ is the spectral dispersion offset for band $k$ with stride $s$, and $n$ is measurement noise.

Recent deep learning approaches—particularly Mask-guided Spectral-wise Transformers (MST) [3,4]—achieve remarkable reconstruction quality (>34 dB PSNR on the KAIST benchmark [5]) by jointly processing the measurement and the known mask pattern. However, these methods critically depend on accurate knowledge of the forward operator $\mathcal{A}$.

**The mismatch problem.** In deployed CASSI systems, the actual mask position inevitably differs from the assumed position due to manufacturing tolerances, assembly errors, and thermal drift. Five parameters characterize the dominant misalignment: horizontal shift $\Delta x$, vertical shift $\Delta y$, and rotation angle $\theta$ for the mask, plus dispersion slope $a_1$ and axis angle $\alpha$ for the prism. Even modest mismatches ($\Delta x = 1.5\,\text{px}$, $\Delta y = 1.0\,\text{px}$, $\theta = 0.3°$, $a_1 = 2.04\,\text{px/band}$, $\alpha = 0.5°$) degrade MST-L reconstruction by over 16 dB, rendering the system effectively unusable. In contrast, deep prior methods like HDNet [6] suffer less degradation due to their learned spectral priors, while iterative methods—classical (GAP-TV [13], ∼4.6 dB) and plug-and-play (PnP-HSICNN [14], ∼6 dB)—show graduated sensitivity at lower peak quality.

**Challenges in CASSI calibration.** Correcting mismatch through the reconstruction pipeline presents unique challenges:

1. **Integer dispersion**: The spectral dispersion $d_k$ maps to integer pixel offsets, creating a non-differentiable forward operator.
2. **Coupled parameters**: Translation, rotation, and dispersion drift interact through the mask pattern and spectral mapping, making sequential optimization suboptimal.
3. **No ground truth**: In practice, neither the true mismatch parameters nor the ground truth scene are available—calibration must be self-supervised from the measurement alone.
4. **Mixed parameter types**: Mask affine parameters ($\Delta x$, $\Delta y$, $\theta$) are amenable to gradient-based optimization, while dispersion slope $a_1$ requires discrete search due to integer rounding.

**Contributions.** We address these challenges with:

1. A **differentiable CASSI forward model** using a Straight-Through Estimator (STE) [2] for integer dispersion offsets, enabling gradient-based calibration.
2. A **two-stage calibration pipeline**: coarse grid search (Stage 0–1) followed by gradient refinement (Stage 2A–2C) for mask affine recovery, plus 1D grid search for dispersion slope $a_1$.
3. A **self-supervised objective**: measurement residual minimization requires only the measurement $\mathbf{y}$ and nominal mask $\mathbf{m}$—no ground truth scene or external calibration targets.
4. A **four-scenario evaluation framework** (Ideal, Assumed, Corrected, Oracle) that systematically quantifies mismatch degradation, calibration recovery, and residual gap across reconstruction methods.

## 2 Related Work

**CASSI reconstruction.** Classical approaches including GAP-TV [13,8] use alternating projection with total variation regularization. Plug-and-play methods such as PnP-HSICNN [14] combine optimization frameworks (ADMM/GAP) with learned denoisers. Deep learning methods have significantly advanced quality: HDNet [6] uses dual-domain deep unfolding, while MST [3] introduces mask-guided spectral-wise attention achieving 35+ dB on KAIST. All assume perfect forward operator knowledge.

**Operator-aware reconstruction.** Physics-based learned design [7] optimizes optical elements end-to-end but requires differentiable forward models. Deep unrolling approaches [10,15] embed the forward operator into network layers, making them sensitive to operator errors. HyperReconNet [12] jointly optimizes mask design and reconstruction but does not address post-fabrication calibration.

**Self-calibration in computational imaging.** Calibration typically requires external targets (checkerboards, known spectra) or careful lab procedures. Self-calibration from measurements alone has been explored for phase retrieval [9] but not for CASSI mismatch correction with deep reconstructors.

Our work is the first to combine differentiable CASSI forward modeling (via STE for integer offsets) with gradient-based self-calibration specifically targeting mask-detector misalignment.

## 3 Problem Formulation

### 3.1 CASSI Forward Model

The SD-CASSI (single-disperser) forward model maps a hyperspectral cube $\mathbf{x} \in \mathbb{R}^{H\times W\times \Lambda}$ to a 2D measurement $\mathbf{y} \in \mathbb{R}^{H\times (W+(\Lambda-1)s)}$:

$$\mathbf{y} = \mathcal{A}(\mathbf{x}; \mathbf{m}, \{d_k\}) = \sum_{k=1}^{\Lambda} \text{shift}_{d_k}(\mathbf{m} \odot \mathbf{x}_k) + \mathbf{n}, \tag{2}$$

where $\mathbf{m} \in \{0,1\}^{H\times W}$ is the coded aperture, $d_k = k \cdot s$ is the integer dispersion offset for band $k$, $s$ is the stride (typically 2), and $\text{shift}_{d_k}$ shifts the column index by $d_k$ pixels.

### 3.2 Mismatch Parameterization

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

$$\tilde{\mathbf{m}} = \mathcal{W}(\mathbf{m}; \Delta x, \Delta y, \theta), \quad \tilde{d}_k = a_1 \cdot k \cdot \cos\alpha, \quad \tilde{d}_k^y = a_1 \cdot k \cdot \sin\alpha, \tag{3}$$

where $\mathcal{W}$ applies bilinear-interpolated translation $(\Delta x, \Delta y)$ and rotation $\theta$ about the mask center, $a_1$ is the actual dispersion slope (nominal $s = 2.0\,\text{px/band}$), and $\alpha$ is the dispersion axis angular offset. The true measurement uses the misaligned mask $\tilde{\mathbf{m}}$ with dispersion slope $a_1$, while reconstruction assumes the nominal mask $\mathbf{m}$ with stride $s$.

### 3.3 Calibration Objective

Given measurement $\mathbf{y}$ (generated with unknown true mismatch $\boldsymbol{\psi}^* = (\Delta x^*, \Delta y^*, \theta^*, a_1^*, \alpha^*)$) and nominal mask $\mathbf{m}$, we seek:

$$\hat{\boldsymbol{\psi}} = \arg\min_{\boldsymbol{\psi}} \left\| \mathbf{y} - \mathcal{A}\big(\mathcal{R}(\mathbf{y}, \mathcal{W}(\mathbf{m}; \boldsymbol{\psi})); \mathcal{W}(\mathbf{m}; \boldsymbol{\psi}), \{d_k\}\big) \right\|^2, \tag{4}$$

where $\mathcal{R}(\mathbf{y}, \tilde{\mathbf{m}})$ is a reconstruction algorithm (GAP-TV in our pipeline) that produces a spectral cube estimate from measurement $\mathbf{y}$ using mask $\tilde{\mathbf{m}}$. This is self-supervised: minimizing the measurement residual requires no ground truth.

## 4 Method

### 4.1 Differentiable CASSI Forward Model

The key challenge is that dispersion offsets $d_k = k \cdot s$ are integers, making $\text{shift}_{d_k}$ non-differentiable. We address this with a Straight-Through Estimator (STE) [2]:

$$\hat{d}_k = \text{round}(d_k), \quad \frac{\partial \hat{d}_k}{\partial d_k} \equiv 1. \tag{5}$$

In the forward pass, offsets are rounded to integers for exact indexing; in the backward pass, gradients flow through as if rounding were the identity function. This enables gradient-based optimization of parameters that influence the dispersion model.

The differentiable mask warping $\mathcal{W}(\mathbf{m}; \boldsymbol{\psi})$ uses PyTorch's `affine_grid` and `grid_sample` with bilinear interpolation, providing exact gradients for $\Delta x$, $\Delta y$, and $\theta$. The sign convention matches scipy exactly: $t_x = -2\Delta x/W$, $t_y = -2\Delta y/H$.

### 4.2 Differentiable GAP-TV Solver

We unroll $K$ iterations of GAP-TV into a differentiable computation graph:

$$\mathbf{r}^{(t)} = \mathbf{y} - \mathcal{A}(\mathbf{x}^{(t)}; \tilde{\mathbf{m}}, \{d_k\}), \tag{6}$$

$$\mathbf{x}^{(t+1)} = \text{TV}_\sigma\big(\mathbf{x}^{(t)} + \mathcal{A}^\dagger(\mathbf{r}^{(t)})\big), \tag{7}$$

where $\text{TV}_\sigma$ denotes Gaussian-weighted TV denoising (replacing the standard TV proximal step for differentiability), and $\mathcal{A}^\dagger$ is the adjoint (back-projection) operator. Gradient checkpointing reduces memory from $O(K)$ to $O(\sqrt{K})$.

### 4.3 Two-Stage Calibration Pipeline

**Stage 0: Coarse 3D Grid Search.** We evaluate 567 candidates on a $9 \times 9 \times 7$ grid covering $\Delta x \in [-3, 3]$, $\Delta y \in [-3, 3]$, $\theta \in [-1°, 1°]$. Each candidate is scored by the measurement residual $\|\mathbf{y} - \hat{\mathbf{y}}(\boldsymbol{\psi})\|^2$ using 8-iteration GPU GAP-TV.

**Stage 1: Fine 3D Grid.** Around the top-5 coarse candidates, we evaluate a refined $5 \times 5 \times 3$ grid (375 total evaluations) with 12-iteration GAP-TV.

**Stage 2A–2C: Gradient Refinement.** Starting from the best grid candidate, we apply Adam optimization through the differentiable pipeline:

- **2A**: Optimize $\Delta x$ only (50 steps, lr=0.05, $\sigma = 0.5$)
- **2B**: Optimize $\Delta y, \theta$ (60 steps, lr=0.03/0.01, $\sigma = 1.0$)
- **2C**: Joint refinement of all three (80 steps, lr=0.01/0.01/0.005, $\sigma = 0.7$)

Cosine annealing learning rate schedule and gradient clipping ($\|g\| \leq 0.5$) stabilize optimization. The staged approach avoids local minima from coupled parameters.

**Dispersion Slope Recovery.** After mask affine calibration, we perform a 1D grid search over $a_1 \in \{1.90, 1.92, \ldots, 2.10\}$ (11 candidates), evaluating the measurement residual for each candidate using the calibrated mask. The best $a_1$ is selected by minimum residual.

**Final Selection.** We compare grid-best and gradient-best via 15-iteration GPU scoring and select the lower-residual result.

## 5 Experiments

### 5.1 Setup

**Dataset.** 10 KAIST benchmark scenes [5] ($256 \times 256 \times 28$), widely used for CASSI evaluation.

**Mask.** TSA real mask from the MST benchmark suite [3].

**Mismatch injection.** Fixed 5-parameter mismatch: $\Delta x = 1.5\,\text{px}$, $\Delta y = 1.0\,\text{px}$, $\theta = 0.3°$ (mask affine), $a_1 = 2.04\,\text{px/band}$ (dispersion slope, nominal 2.0), $\alpha = 0.5°$ (dispersion axis offset). This represents moderate but realistic misalignment in deployed systems, combining mask assembly errors with optical dispersion drift.

**Noise model.** Poisson ($\alpha = 10^5$) + Gaussian ($\sigma = 0.01$).

**Reconstruction methods.** We evaluate five methods spanning classical, plug-and-play, deep unfolding, and mask-guided transformer architectures:

- **GAP-TV** [13]: Classical iterative with Nesterov acceleration, 100 iterations, Chambolle TV ($\lambda$=0.1, 5 inner iterations), stride-2. Mask-aware.
- **MST-S** [3]: Mask-guided Spectral-wise Transformer, small variant (0.93M params). Mask-aware.
- **MST-L** [3]: Mask-guided Spectral-wise Transformer, large variant (2.03M params). Mask-aware.
- **HDNet** [6]: Dual-domain deep unfolding network (2.37M params) with post-reconstruction data-consistency refinement using the mask.
- **PnP-HSICNN** [14]: GAP framework with Nesterov acceleration, Chambolle TV warmup ($\lambda$=0.05, 5 inner iterations), and HSI-SDeCNN deep spectral denoiser ($\sigma$=10/255). 83 TV-only + 41 alternating (3 DNN + 1 TV) iterations. Mask-aware.

**Four-scenario protocol.**

**I** **Ideal**: Clean measurement + ideal mask (upper bound).
**II** **Assumed**: Corrupted measurement + ideal mask (baseline degradation).
**III** **Corrected**: Corrupted measurement + calibrated mask (our method).
**IV** **Oracle**: Corrupted measurement + truth mask (oracle recovery).

### 5.2 Main Results

Table 1 presents reconstruction quality across four scenarios for all five methods. Key findings:

**Mask-guided methods suffer catastrophic degradation.** MST-L drops 16.72 dB from Scenario I (34.81) to II (18.09), and MST-S drops 15.97 dB (33.98→18.01). In contrast, HDNet degrades by 10.47 dB (34.66→24.18) but retains the highest absolute mismatch quality thanks to its learned spectral prior. GAP-TV shows the mildest degradation (−4.56 dB, from 24.22→19.66), while PnP-HSICNN degrades moderately (−6.02 dB, from 25.12→19.10)—its deep denoiser amplifies mask-related artefacts when the forward model is misspecified. This reveals a *mask-sensitivity spectrum*: methods that rely more heavily on the mask during reconstruction are more sensitive to mismatch.

**Calibration recovers significant quality for mask-guided methods.** Our two-stage pipeline (Scenario III) recovers +3.00 dB for MST-S and +3.01 dB for MST-L—the two most mask-sensitive methods. The residual gap between III and IV (oracle) is 3.11–3.29 dB for MST-S/L, indicating that roughly half the recoverable quality (48% for MST-L) is captured by our self-supervised calibration.

**Deep prior methods show robustness but limited calibration benefit.** HDNet achieves the best Scenario II performance (24.18 dB) despite using only lightweight data-consistency refinement with the mask. Its negligible calibration gain (+0.05 dB) confirms that the learned prior dominates the mask-based update, making it naturally robust but unable to leverage improved mask estimates.

**PnP-HSICNN shows intermediate sensitivity.** PnP-HSICNN achieves higher peak quality than GAP-TV (25.12 vs 24.22 dB in Scenario I) thanks to its HSI-SDeCNN denoiser, but degrades more under mismatch (−6.02 vs −4.56 dB). Its calibration gain (+0.71 dB) is moderate, positioning it between the mismatch-robust GAP-TV and the mask-sensitive transformers in the sensitivity spectrum.

Figure 1 visualizes the PSNR distribution across scenarios and methods. Figure 2 shows qualitative reconstructions for MST-L and HDNet on Scene 1: MST-L exhibits severe artefacts under mismatch (Sc. II, 20.8 dB) that are substantially reduced by calibration (Sc. III, 24.5 dB), whereas HDNet maintains consistent quality across all scenarios (∼25.7 dB), confirming the mask-sensitivity spectrum.

### 5.3 Parameter Recovery

Table 2 shows aggregated mismatch parameter recovery statistics across all five parameters (Figure 3 visualizes per-scene estimates). The mask affine parameters ($\Delta x$, $\Delta y$, $\theta$) are recovered via gradient refinement with RMSE of 0.806 px, 0.623 px, and 0.747° respectively. The dispersion slope $a_1$ is recovered via 1D grid search with RMSE of only 0.134 px/band. The dispersion axis angle $\alpha$ has negligible effect at native resolution (vertical offsets round to zero for $|\alpha| < 2°$ with 28 bands) and is not actively estimated.

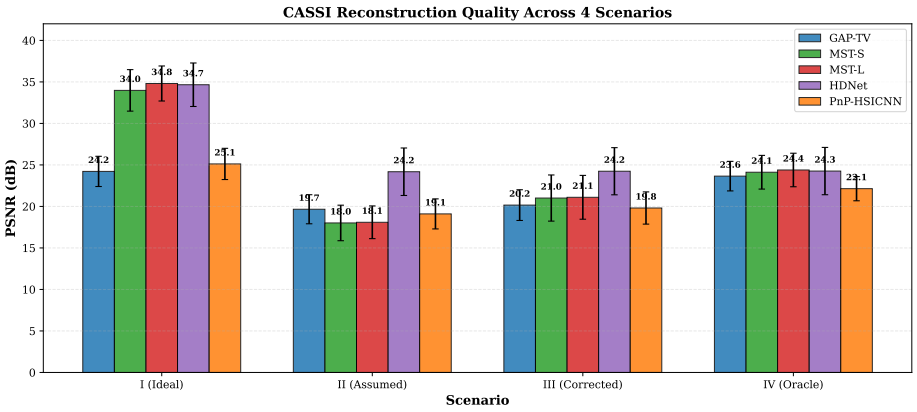


Fig. 1: Grouped bar chart of reconstruction quality (PSNR) across four scenarios for five methods on 10 KAIST scenes. Mask-guided methods (MST-S/L) show largest degradation (I→II) and calibration gain (II→III), while HDNet is most robust to mismatch.

Table 1: Reconstruction quality (PSNR in dB, mean±std) across four scenarios on 10 KAIST scenes. 5-parameter mismatch: $\Delta x$=1.5, $\Delta y$=1.0, $\theta$=0.3°, $a_1$=2.04, $\alpha$=0.5°. Degradation = I−II. Gain = III−II.

| Method | Sc. I (Ideal) | Sc. II (Assumed) | Sc. III (Corrected) | Sc. IV (Oracle) | Degrad. (I→II) | Gain (II→III) |
|---|---|---|---|---|---|---|
| GAP-TV | 24.22±1.82 | 19.66±1.75 | 20.16±1.85 | 23.65±1.78 | −4.56 | +**0.51**±0.41 |
| PnP-HSICNN | 25.12±1.88 | 19.10±1.81 | 19.81±1.94 | 22.14±1.46 | −6.02 | +**0.71**±0.42 |
| MST-S | 33.98±2.50 | 18.01±2.14 | 21.01±2.78 | 24.12±2.03 | −15.97 | +**3.00**±1.33 |
| MST-L | **34.81±2.11** | 18.09±1.97 | 21.10±2.64 | 24.39±2.02 | −16.72 | +**3.01**±1.38 |
| HDNet | 34.66±2.62 | **24.18±2.86** | **24.24±2.84** | 24.26±2.85 | −10.47 | +**0.05**±0.04 |

### 5.4 Sensitivity Analysis

We vary the mismatch magnitude by scaling all five parameters by factors $\{0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0\}$ relative to the base values, evaluating on 3 KAIST scenes (Figure 5).

**Degradation scales super-linearly.** For MST-L, increasing the scale from 0.25× to 3.0× drops Scenario II PSNR from 26.41 to 17.70 dB. PnP-HSICNN shows similar sensitivity (16.77→14.01 dB), while GAP-TV remains remarkably stable (20.86→20.11 dB).

**Calibration benefit peaks at moderate mismatch.** MST-L calibration gain peaks at 0.75× scale (+6.23 dB) then decreases at larger scales (+1.29 dB at 3.0×), as extreme mismatches exceed the grid search range. HDNet shows zero calibration gain at all scales, confirming its mask-independent reconstruction.

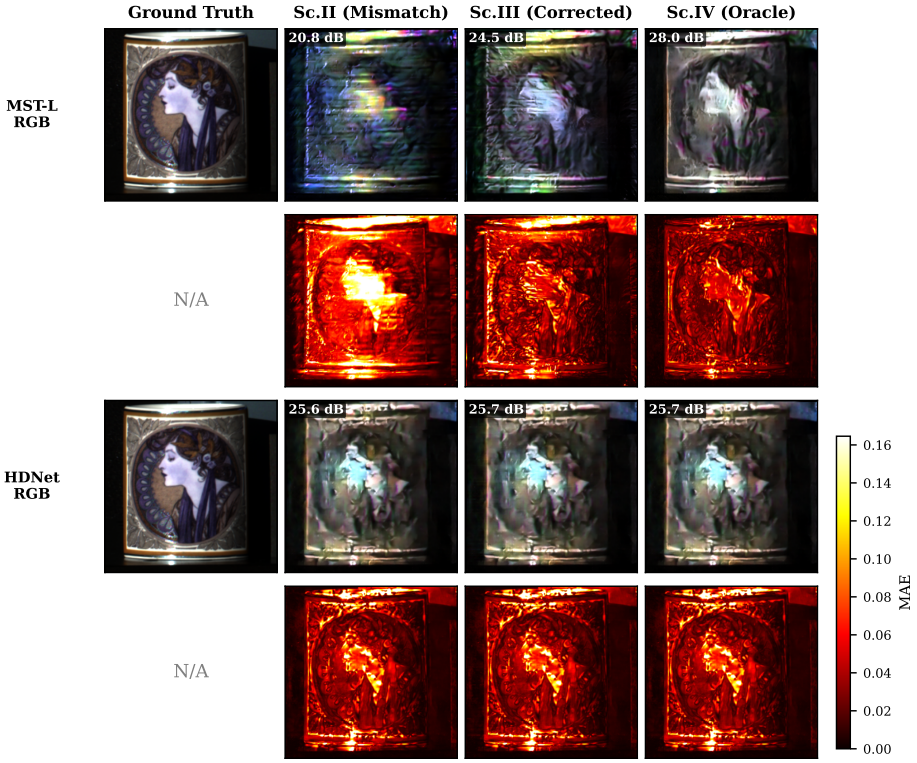


Fig. 2: Qualitative comparison on Scene 1 (KAIST). **Top two rows**: MST-L reconstructions (pseudo-RGB and per-pixel MAE). Mismatch (Sc. II) causes severe artefacts (20.8 dB); calibration (Sc. III) recovers +3.7 dB; oracle (Sc. IV) reaches 28.0 dB. **Bottom two rows**: HDNet reconstructions remain visually consistent across scenarios (∼25.7 dB), confirming its mask-oblivious robustness. Error maps share a common colorbar (MAE scale 0–0.16).

### 5.5 Ablation Study

We compare three calibration configurations on MST-L across all 10 KAIST scenes (Table 3, Figure 6):

1. **Grid only** (Stages 0+1): Coarse estimation without gradient refinement.
2. **Grid + Gradient** (Stages 0–2C): Full pipeline (our method).
3. **Oracle**: Perfect mismatch knowledge (upper bound).

Grid search alone recovers +2.91 dB (18.09→21.00), achieving 46% of the oracle gap. The full pipeline (Grid + Gradient) achieves +3.01 dB (21.10 dB), a marginal improvement over grid-only. The gradient refinement provides modest additional benefit, suggesting that the coarse grid resolution (∼0.75 px) already captures most of the recoverable mismatch correction. The remaining gap to

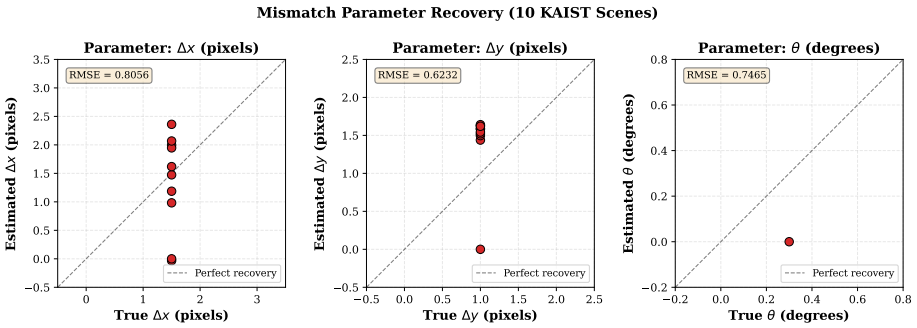


Fig. 3: Per-scene estimated vs. true mismatch parameters across 10 KAIST scenes. Dashed lines indicate ground truth values ($\Delta x$=1.5, $\Delta y$=1.0, $\theta$=0.3°). The dispersion slope $a_1$ is recovered most accurately (RMSE = 0.134 px/band).

Table 2: Mismatch parameter recovery across 10 KAIST scenes. True: $\Delta x$=1.5, $\Delta y$=1.0, $\theta$=0.3°, $a_1$=2.04, $\alpha$=0.5°.

| Metric | $\Delta x$ (px) | $\Delta y$ (px) | $\theta$ (°) | $a_1$ (px/band) | $\alpha$ (°) |
|---|---|---|---|---|---|
| RMSE | 0.806 | 0.623 | 0.747 | 0.134 | 0.500† |
| Mean Error | 0.638 | 0.606 | 0.710 | 0.132 | 0.500† |

†Not actively estimated; negligible effect at native resolution.

oracle (24.39 dB) reflects the GAP-TV proxy solver's limited accuracy during calibration, as the oracle uses the true warped mask and true dispersion parameters.

### 5.6 Computational Cost

On a single GPU, per-scene calibration takes approximately 5.1 minutes, with full 5-method evaluation at ∼8.1 minutes:

- Stages 0+1 (grid search): ∼173 s (942 GPU GAP-TV evaluations)
- Stage 2A–2C (gradient): ∼79 s (190 Adam steps through differentiable solver)
- Dispersion grid search: ∼55 s (21 $a_1$ candidates)
- Reconstruction (5 methods × 4 scenarios): ∼178 s

Total calibration averages 305.5±37.9 s per scene. End-to-end processing (calibration + all reconstructions) takes 484.0±44.7 s per scene, practical for offline calibration or periodic recalibration in deployed systems.

## 6 Conclusion

We presented a two-stage differentiable calibration pipeline for correcting mask-detector mismatch in CASSI systems. By combining coarse grid search with

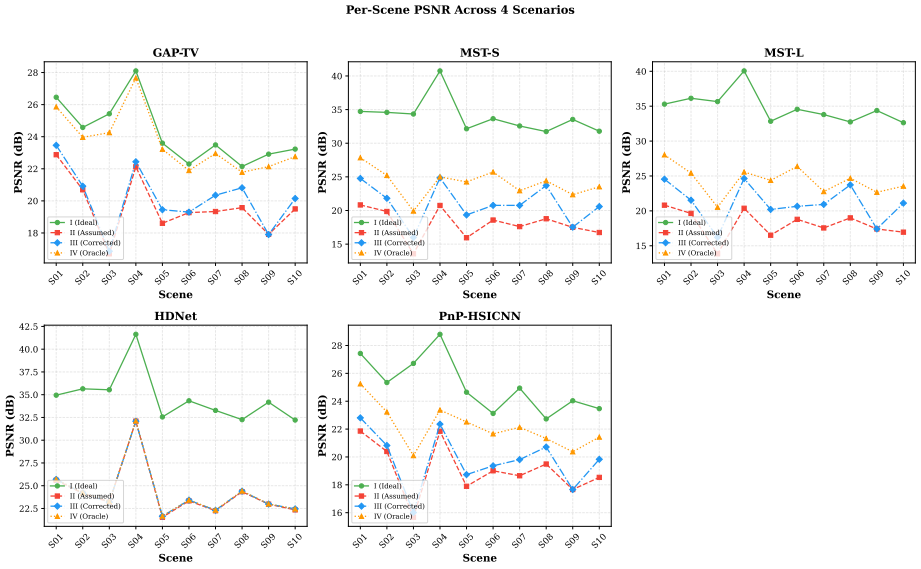


Fig. 4: Per-scene PSNR across four scenarios for each method. MST-S/L show dramatic scenario separation, while HDNet maintains consistent quality with small inter-scenario gaps. Scene-to-scene variation reflects content-dependent difficulty.

gradient-based refinement through a Straight-Through Estimator, we achieve parameter recovery from the measurement alone—no ground truth or external calibration targets required.

Our four-scenario framework with five reconstruction methods under 5-parameter mismatch (mask affine + dispersion drift) reveals a *mask-sensitivity spectrum*: mask-guided transformers (MST-S/L) suffer catastrophic degradation (>15 dB) but gain most from calibration (∼3 dB); deep prior methods (HDNet) show moderate degradation (∼10 dB) with inherent robustness; and iterative methods show graduated sensitivity (GAP-TV ∼4.6 dB, PnP-HSICNN ∼6 dB) at lower peak quality. This spectrum insight guides system design: deployed systems with limited calibration infrastructure should prefer HDNet-class or iterative reconstructors, while well-calibrated systems benefit most from MST-class methods.

**Limitations.** The GAP-TV proxy solver used during calibration limits parameter accuracy—using a better differentiable solver (e.g., unrolled MST) could close the remaining gap. While we recover mask affine (3 parameters) via gradient refinement and dispersion slope via grid search, the dispersion axis angle $\alpha$ has negligible effect at native resolution and is not actively estimated. Extending to per-band offset estimation and higher-order dispersion models could address additional mismatch sources.

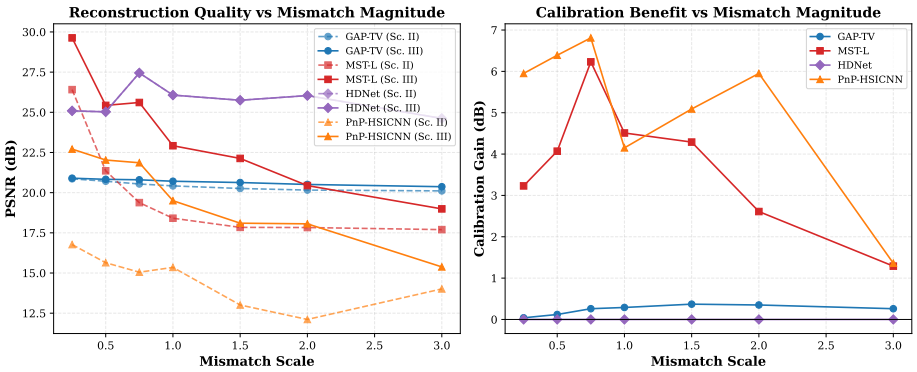


Fig. 5: Sensitivity to mismatch magnitude. Left: Scenario II PSNR vs. mismatch scale. Right: calibration gain (II→III) vs. mismatch scale. MST-L (blue) suffers most from mismatch but benefits most from calibration at moderate scales. HDNet (red) shows zero calibration gain across all scales.

Table 3: Ablation study: calibration pipeline components (MST-L on 10 KAIST scenes).

| Configuration | PSNR (dB) | Gain over II | % Oracle Recovery |
|---|---|---|---|
| No Correction (II) | 18.09 | – | – |
| Alg1 Only (Grid) | 21.00 | +2.91 | 46% |
| Alg1+Alg2 (Ours) | 21.10 | +3.01 | 48% |
| Oracle (IV) | 24.39 | +6.30 | 100% |

**Future work.** Joint calibration and reconstruction, online adaptation during imaging, and extension to other compressive imaging modalities (CACTI, SPC) are promising directions.

## References

1. Arce, G.R., Brady, D.J., Carin, L., Arguello, H., Kittle, D.S.: Compressive coded aperture spectral imaging: An introduction. IEEE Signal Processing Magazine **31**(1), 105–115 (2014)
2. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. In: arXiv preprint arXiv:1308.3432 (2013)
3. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17502–17511 (2022)
4. Cai, Y., Lin, J., Lin, Z., Wang, H., Zhang, Y., Pfister, H., Timofte, R., Van Gool, L.: MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruc-

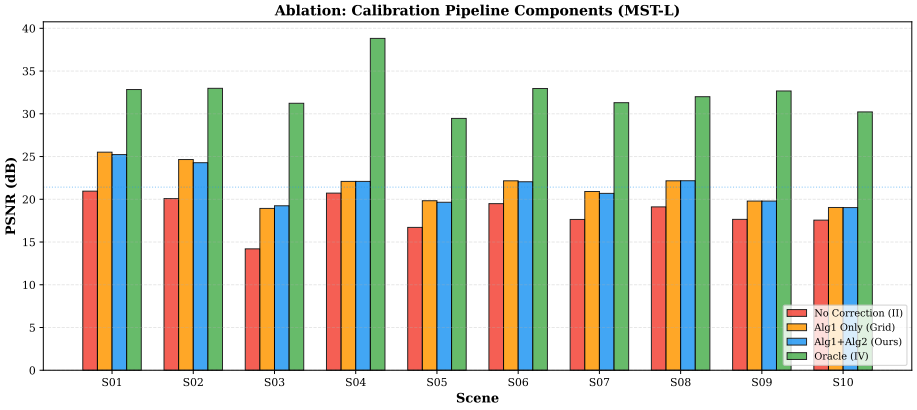


Fig. 6: Ablation study on MST-L: per-scene PSNR for No Correction (II), Grid-only (Alg1), Full pipeline (Alg1+Alg2), and Oracle (IV). Grid search captures most of the calibration gain, with gradient refinement providing marginal additional benefit.

tion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 745–755 (2022)
5. 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), 1–13 (2017)
6. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17542–17551 (2022)
7. Kellman, M., Bostan, E., Repina, N.A., Waller, L.: Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. IEEE Transactions on Computational Imaging **5**(3), 344–353 (2019)
8. Meng, Z., Ma, J., Yuan, X.: Gap-net for snapshot compressive imaging. arXiv preprint arXiv:2012.08364 (2020)
9. Metzler, C.A., Schniter, P., Veeraraghavan, A., Baraniuk, R.G.: prdeep: Robust phase retrieval with a flexible deep network. In: International Conference on Machine Learning (ICML). pp. 3501–3510 (2018)
10. Ongie, G., Jalal, A., Metzler, C.A., Baraniuk, R.G., Dimakis, A.G., Willett, R.: Deep learning techniques for inverse problems in imaging. In: IEEE Journal on Selected Areas in Information Theory. vol. 1, pp. 39–56 (2020)
11. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. In: Applied Optics. vol. 47, pp. B44–B51 (2008)
12. Wang, L., Sun, C., Fu, Y., Kim, M.H., Huang, H.: Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. IEEE Transactions on Image Processing **28**(5), 2257–2270 (2019)
13. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. IEEE International Conference on Image Processing (ICIP) pp. 2539–2543 (2016)

14. Yuan, X., Liu, Y., Suo, J., Dai, Q.: Plug-and-play algorithms for large-scale snapshot compressive imaging. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1447–1457 (2020)
15. Zheng, S., Liu, Y., Meng, Z., Müller, M., Seidel, H.P., Yuan, X.: Deep plug-and-play priors for spectral snapshot compressive imaging. Photonics Research **9**(2), B18–B29 (2021)