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

Chengshuai Yang[⋆]

NextGen PlatformAI C Corp

Abstract. Forward model mismatch—sub-pixel mask misalignment and dispersion drift between the coded aperture and detector—is unavoidable in deployed CASSI systems, yet even moderate perturbations degrade state-of-the-art mask-guided transformers (MST) by over 16 dB. We present a self-supervised two-stage differentiable calibration pipeline that recovers 5-parameter mismatch from a single measurement and nominal mask alone, requiring no ground truth. Stage 1 performs a coarse hierarchical grid search scored by GPU-accelerated GAP-TV; Stage 2 applies gradient refinement through an unrolled differentiable forward operator using a Straight-Through Estimator (STE) for integer dispersion offsets. Evaluating five reconstruction methods across four scenarios on 10 KAIST scenes, we uncover a *mask-sensitivity spectrum*: mask-guided transformers suffer catastrophic degradation (>15 dB) yet recover ∼48% of the oracle gap after calibration, while deep prior methods show inherent robustness with negligible absolute gain. Code and results are publicly available.

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

1 Introduction

Deep learning has transformed hyperspectral image reconstruction, with mask-guided transformers (MST) [5,6] achieving >34 dB on the KAIST benchmark [8]. Yet these reconstructors depend on accurate knowledge of the forward operator—a dependency that creates a *sim-to-real gap* between the idealized model used during training and the physical system at deployment [13]. In coded aperture snapshot spectral imaging (CASSI) [14,1], this gap manifests as mask-detector misalignment and dispersion drift, degrading MST-L by over 16 dB (Figure 1).

The CASSI forward model (Section 3) maps a hyperspectral cube through a coded aperture and spectral disperser to a 2D measurement. In practice, five parameters characterize the dominant misalignment between the assumed and actual forward operator: mask translation $(\varDelta x, \varDelta y)$, rotation θ, dispersion slope a_1, and axis angle α.

[⋆] integrityyang@gmail.com

The calibration challenge. Correcting this mismatch is difficult because: (1) spectral dispersion offsets d_k are integers, making the forward operator non-differentiable; (2) mask affine and dispersion parameters interact through the measurement model; and (3) in deployment, neither the true parameters nor the ground truth scene are available—calibration must be self-supervised from the measurement alone.

Contributions. We address these challenges with:

1. A **differentiable CASSI forward model** using a Straight-Through Estimator (STE) [3] for integer dispersion offsets, enabling end-to-end gradient-based calibration (Section 4).
2. A **two-stage self-supervised calibration pipeline** (Algorithm 1): coarse grid search followed by gradient refinement, recovering 5-parameter mismatch from a single measurement without ground truth.
3. A **mask-sensitivity spectrum** characterizing how five reconstructors respond to mismatch and calibration: mask-guided methods (MST-S/L) suffer >15 dB degradation but recover $\rho \approx 48\%$ of the oracle gap; deep prior methods show inherent robustness; iterative methods show intermediate sensitivity (Section 5).
4. A **four-scenario evaluation framework** (Ideal, Assumed, Corrected, Oracle) with complete open-source benchmark on 10 KAIST scenes across five methods.

2 Related Work

CASSI reconstruction. Classical approaches including GAP-TV [16,11] use alternating projection with total variation regularization. Plug-and-play methods replace the hand-crafted prior with learned denoisers [18,19], combining ADMM/GAP optimization with deep spectral denoisers. End-to-end deep networks have advanced quality through dual-domain unfolding (HDNet [9]), mask-guided attention (MST [5,6]), sparse transformers (CST [4]), and degradation-aware unfolding (DAUHST [7]); see [17] for a survey. All assume perfect forward operator knowledge.

Calibration in computational imaging. Arguello and Arce [2] optimize coded apertures for colored-CASSI but do not address post-fabrication misalignment. Traditional calibration requires external targets or careful laboratory procedures. Self-calibration from measurements alone has been explored for phase retrieval [12] and differentiable rendering [10], where end-to-end optimization through differentiable forward models has shown promise.

Differentiable forward models. Physics-based learned design [10] optimizes optical elements end-to-end but requires continuous relaxations. Deep unrolling [13] embeds the forward operator into network layers, making learned reconstructors sensitive to operator errors. HyperReconNet [15] jointly optimizes mask design and reconstruction but does not address post-fabrication calibration.

Our work is the first to combine an STE-based differentiable CASSI forward model with gradient-based self-supervised calibration targeting mask-detector misalignment in deployed systems.

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 $\mathbf{y} \in \mathbb{R}^{H \times (W+(\Lambda-1)s)}$:

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

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 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:

$$\tilde{\mathbf{m}} = \mathcal{W}(\mathbf{m}; \Delta x, \Delta y, \theta), \quad \tilde{d}_k = a_1 \cdot k \cdot \cos\alpha, \quad \tilde{d}_k^y = a_1 \cdot k \cdot \sin\alpha, \tag{2}$$

where $\mathcal{W}$ applies bilinear-interpolated translation $(\Delta x, \Delta y)$ and rotation θ about the mask center, a_1 is the actual dispersion slope (nominal $s = 2.0\,\mathrm{px/band}$), and α 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 parameters $\boldsymbol{\psi}^*$) and nominal mask $\mathbf{m}$, we seek:

$$\hat{\boldsymbol{\psi}} = \arg\min_{\boldsymbol{\psi}} \left\| \mathbf{y} - \mathcal{A}\big(\mathcal{R}(\mathbf{y}, \tilde{\mathbf{m}}); \tilde{\mathbf{m}}, \{d_k\}\big) \right\|^2, \quad \tilde{\mathbf{m}} = \mathcal{W}(\mathbf{m}; \boldsymbol{\psi}), \tag{3}$$

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 shift_{d_k} non-differentiable. We address this with a Straight-Through Estimator (STE) [3]:

$$\hat{d}_k = \text{round}(d_k), \quad \frac{\partial \hat{d}_k}{\partial d_k} \equiv 1. \tag{4}$$

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 Δx, Δy, and θ. 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:

$$\mathbf{r}^{(t)} = \mathbf{y} - \mathcal{A}(\mathbf{x}^{(t)}; \tilde{\mathbf{m}}, \{d_k\}), \tag{5}$$

$$\mathbf{x}^{(t+1)} = \text{TV}_\sigma\big(\mathbf{x}^{(t)} + \mathcal{A}^\dagger(\mathbf{r}^{(t)})\big), \tag{6}$$

where TV_σ 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 Δx 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.

Algorithm 1 Two-Stage Differentiable CASSI Calibration

```
Require: Measurement $\mathbf{y}$, nominal mask $\mathbf{m}$, candidate grids $\mathcal{G}_0$, $\mathcal{G}_1$
Ensure: Calibrated parameters $\hat{\boldsymbol{\psi}} = (\widehat{\mathit{\Delta} x}, \widehat{\mathit{\Delta} y}, \hat{\theta}, \hat{a}_1)$
Stage 0 (Coarse grid): Evaluate 9×9×7 grid $\mathcal{G}_0$ with 8-iter GAP-TV
$\boldsymbol{\psi}_0^* \leftarrow \arg\min_{\boldsymbol{\psi} \in \mathcal{G}_0} \|\mathbf{y} - \mathcal{A}(\mathcal{R}_8(\mathbf{y}; \boldsymbol{\psi}); \boldsymbol{\psi})\|^2$
Stage 1 (Fine grid): Refine top-5 via 5×5×3 grid $\mathcal{G}_1$, 12-iter GAP-TV
$\boldsymbol{\psi}_1^* \leftarrow \arg\min_{\boldsymbol{\psi} \in \mathcal{G}_1} \|\mathbf{y} - \mathcal{A}(\mathcal{R}_{12}(\mathbf{y}; \boldsymbol{\psi}); \boldsymbol{\psi})\|^2$
Stage 2A: Gradient-refine $\mathit{\Delta} x$ only (50 Adam steps, lr = 0.05)
Stage 2B: Gradient-refine $\mathit{\Delta} y, \theta$ (60 steps, lr = 0.03/0.01)
Stage 2C: Joint refinement $\mathit{\Delta} x, \mathit{\Delta} y, \theta$ (80 steps, lr = 0.01)
$\hat{\boldsymbol{\psi}}_{\text{grad}} \leftarrow$ result via STE-enabled backprop through $\mathcal{A}$
Stage 3: 1D grid search $a_1 \in \{1.90, 1.92, \ldots, 2.10\}$ using calibrated mask
$\hat{a}_1 \leftarrow \arg\min_{a_1} \|\mathbf{y} - \mathcal{A}(\mathcal{R}(\mathbf{y}; \hat{\boldsymbol{\psi}}_{\text{grad}}, a_1); \hat{\boldsymbol{\psi}}_{\text{grad}}, a_1)\|^2$
Select $\hat{\boldsymbol{\psi}}$ from $\{\boldsymbol{\psi}_1^*, \hat{\boldsymbol{\psi}}_{\text{grad}}\}$ by lower 15-iter residual
return $\hat{\boldsymbol{\psi}} = (\widehat{\mathit{\Delta} x}, \widehat{\mathit{\Delta} y}, \hat{\theta}, \hat{a}_1)$
```

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. The complete pipeline is summarized in Algorithm 1.

5 Experiments

5.1 Setup

Dataset. 10 KAIST benchmark scenes [8] ($256 \times 256 \times 28$), widely used for CASSI evaluation.

Mask. TSA real mask from the MST benchmark suite [5].

Mismatch injection. Fixed 5-parameter mismatch: $\mathit{\Delta} x = 1.5\,\text{px}$, $\mathit{\Delta} y = 1.0\,\text{px}$, $\theta = 0.3°$ (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** [16]: Classical iterative with Nesterov acceleration, 100 iterations, Chambolle TV (λ=0.1, 5 inner iterations), stride-2. Mask-refined.
- **MST-S** [5]: Mask-guided Spectral-wise Transformer, small variant (0.93M params).
- **MST-L** [5]: Mask-guided Spectral-wise Transformer, large variant (2.03M params).

- **HDNet** [9]: Dual-domain deep unfolding (2.37M params) with mask-based data-consistency refinement. Mask-oblivious.
- **PnP-HSICNN** [18]: GAP framework with Nesterov acceleration, Chambolle TV warmup (λ=0.05, 5 inner iterations), and HSI-SDeCNN deep spectral denoiser (σ=10/255). 83 TV-only + 41 alternating (3 DNN + 1 TV) iterations. Mask-refined.

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 (PSNR and SSIM) across four scenarios, along with calibration gain and recovery ratio ρ = (III − II)/(IV − II). Key findings:

Mask-guided methods suffer catastrophic degradation. MST-L drops 16.72 dB from Scenario I (34.81 dB, SSIM .973) to II (18.09 dB, .615), and MST-S drops 15.97 dB. HDNet degrades by 10.47 dB but retains the highest mismatch quality (24.18 dB, SSIM .791), confirming its learned prior compensates for mask errors. GAP-TV (−4.56 dB) and PnP-HSICNN (−6.02 dB) show intermediate degradation.

Calibration recovers significant quality for mask-guided methods. Our pipeline recovers +3.01 dB for MST-L and +3.00 dB for MST-S, with recovery ratios of ρ = 47.8% and 49.1% respectively—the highest among all methods. SSIM improvements corroborate: MST-S gains +0.090 and MST-L gains +0.076, compared to +0.033 for GAP-TV and +0.035 for PnP-HSICNN. HDNet shows negligible calibration benefit (+0.05 dB; ρ is indeterminate due to the near-zero oracle gap), confirming that the mask plays a marginal role in its reconstruction. Spectral angle mapper (SAM) trends corroborate: MST-L achieves 7.44° under ideal conditions but degrades to 31.38° under mismatch, while HDNet maintains ∼14.5° across all scenarios.

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).

5.3 Mask-Sensitivity Spectrum

The four-scenario analysis reveals a systematic *mask-sensitivity spectrum* that categorizes reconstruction methods by their dependence on mask accuracy. We formalize three regimes:

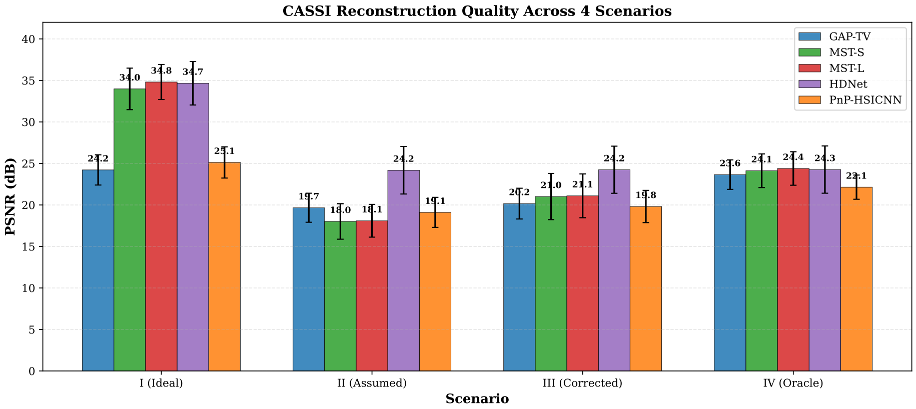

Fig. 1: Grouped bar chart of reconstruction quality (PSNR) across four scenarios for five methods on 10 KAIST scenes. Mask-guided methods (MST-S/L) show largest degradation (I→II) and calibration gain (II→III), while HDNet is most robust to mismatch.

Mask-guided (MST-S, MST-L): Methods that explicitly condition on the mask pattern at every processing stage. These suffer catastrophic degradation under mismatch ($>$15 dB) because the mask directly modulates attention in the spectral transformer. Calibration recovers $\rho \approx$ 48–49% of the oracle gap, yielding the largest absolute gains (+3.0 dB PSNR, +0.08–0.09 SSIM).

Mask-refined (GAP-TV, PnP-HSICNN): Methods that use the mask in an iterative projection step but rely on hand-crafted or learned priors for regularization. Degradation is moderate (4.6–6.0 dB), and calibration recovery is limited ($\rho \approx$ 13–23%) because the prior partially compensates for mask errors.

Mask-oblivious (HDNet): Methods where learned spectral priors dominate and the mask serves only a lightweight data-consistency role. Degradation remains substantial ($\sim$10 dB) but absolute mismatch performance is highest (24.18 dB). Calibration gain is negligible in absolute terms (+0.05 dB; ρ is indeterminate due to the near-zero oracle gap), confirming the prior's independence from mask accuracy.

This taxonomy provides a practical design guideline: deployed systems with limited calibration infrastructure should prefer mask-oblivious or mask-refined reconstructors, while well-calibrated systems benefit most from mask-guided architectures.

5.4 Parameter Recovery

Table 2 shows aggregated mismatch parameter recovery statistics across all five parameters (Figure 3 visualizes per-scene estimates). The mask affine parameters (Δx, Δy, θ) 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

Table 1: Reconstruction quality (PSNR / SSIM, mean over 10 KAIST scenes) across four scenarios. 5-parameter mismatch: $\mathit{\Delta x}$=1.5, $\mathit{\Delta y}$=1.0, θ=0.3°, a_1=2.04, α=0.5°. Recovery ratio ρ = (III−II)/(IV−II).

Method		Sc. I	Sc. II	Sc. III	Sc. IV	Gain	ρ
		(Ideal)	(Assumed)	(Corrected)	(Oracle)	(II→III)	(%)
GAP-TV	PSNR	24.22	19.66	20.16	23.65	+0.51	12.8
	SSIM	.722	.547	.580	.704	+.033	
PnP-HSICNN	PSNR	25.12	19.10	19.81	22.14	+0.71	23.4
	SSIM	.758	.512	.547	.666	+.035	
MST-S	PSNR	33.98	18.01	21.01	24.12	**+3.00**	49.1
	SSIM	.965	.609	.699	.797	+.090	
MST-L	PSNR	**34.81**	18.09	21.10	24.39	**+3.01**	47.8
	SSIM	**.973**	.615	.691	.803	+.076	
HDNet	PSNR	34.66	**24.18**	**24.24**	24.26	+0.05	—
	SSIM	.970	**.791**	**.790**	.791	−.001	

Table 2: Mismatch parameter recovery across 10 KAIST scenes. True: $\mathit{\Delta x}$=1.5, $\mathit{\Delta y}$=1.0, θ=0.3°, a_1=2.04, α=0.5°.

Metric	$\mathit{\Delta x}$ (px)	$\mathit{\Delta y}$ (px)	θ (°)	a_1 (px/band)	α (°)
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.

grid search with RMSE of only 0.134 px/band. The dispersion axis angle α has negligible effect at native resolution (vertical offsets round to zero for $|\alpha| < 2°$ with 28 bands) and is not actively estimated.

5.5 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.

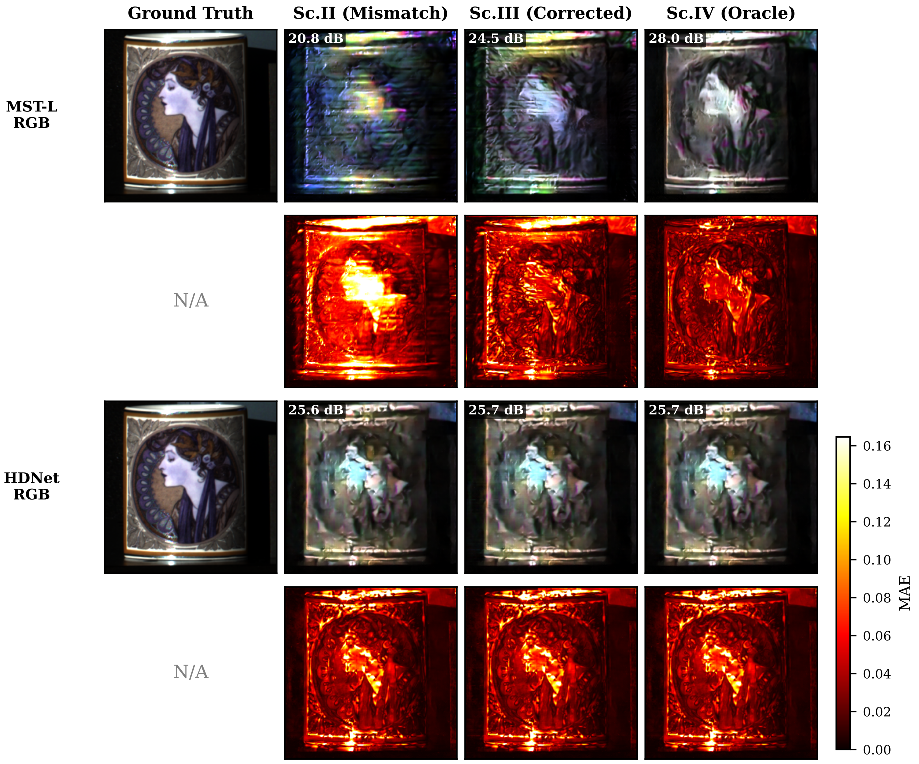

Fig. 2: Qualitative comparison on Scene 1 (KAIST). **Top two rows**: MST-L reconstructions (pseudo-RGB and per-pixel MAE). Mismatch (Sc. II) causes severe artefacts (20.8 dB); calibration (Sc. III) recovers +3.7 dB; oracle (Sc. IV) reaches 28.0 dB. **Bottom two rows**: HDNet reconstructions remain visually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

5.6 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

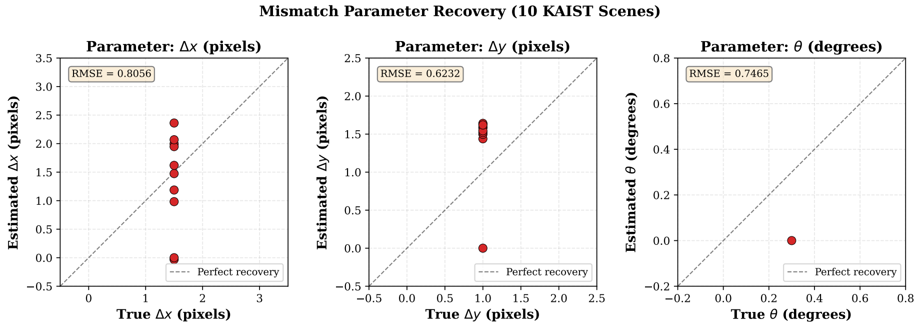

Fig. 3: Per-scene estimated vs. true mismatch parameters across 10 KAIST scenes. Dashed lines indicate ground truth values (Δx=1.5, Δy=1.0, θ=0.3°).

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%

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.

5.7 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 (11 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 CASSI that recovers 5-parameter mask-detector mismatch from a single measurement without

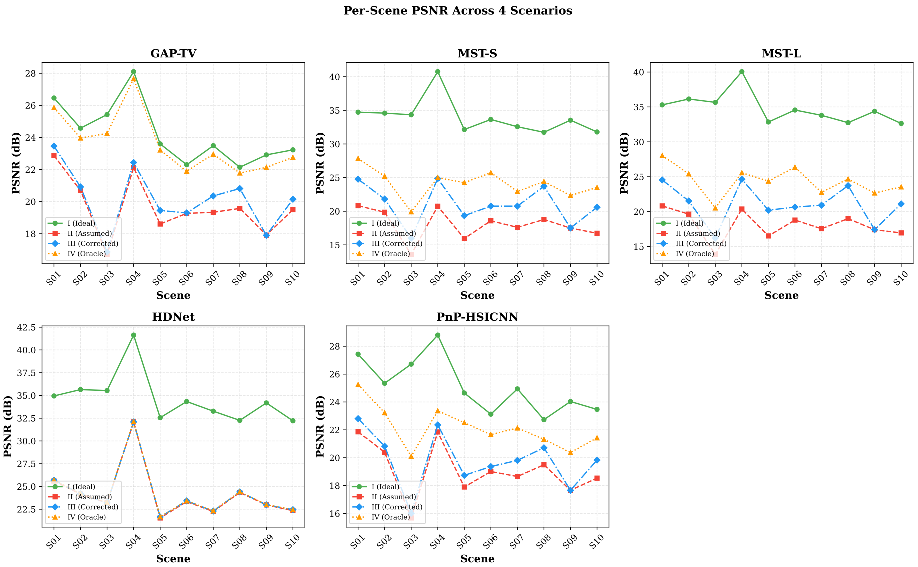

Fig. 4: Per-scene PSNR across four scenarios for each method. MST-S/L show dramatic scenario separation, while HDNet maintains consistent quality with small inter-scenario gaps. Scene-to-scene variation reflects content-dependent difficulty.

ground truth. The pipeline combines coarse grid search (capturing 46% of the oracle gap) with STE-enabled gradient refinement (reaching 48%), achieving +3.0 dB recovery for MST-L at ∼5 minutes per scene on a single GPU.

Our principal finding is the *mask-sensitivity spectrum*: the degree to which a reconstructor depends on mask accuracy determines both its vulnerability to mismatch and its benefit from calibration. Mask-guided transformers (MST-S/L, $\rho \approx 48\%$) and mask-refined iterative methods (GAP-TV/PnP-HSICNN, $\rho \approx 13$–23%) represent two distinct operating regimes, while deep prior methods (HDNet, negligible absolute gain) achieve inherent robustness at the cost of calibration-agnostic reconstruction.

Limitations and future work. The GAP-TV proxy solver limits calibration accuracy—unrolling a stronger solver could close the remaining gap. Extending to additional modalities (CACTI, SPC), per-band dispersion estimation, and online adaptation during imaging are promising directions that the open-source framework facilitates.

Data and code availability. All code, results, and figure-generation scripts are publicly available at `https://github.com/integritynoble/Physics_World_Model` under `papers/pwmi_cassi/`. The KAIST benchmark [8] and TSA mask are publicly available. No non-public datasets were used in this work.

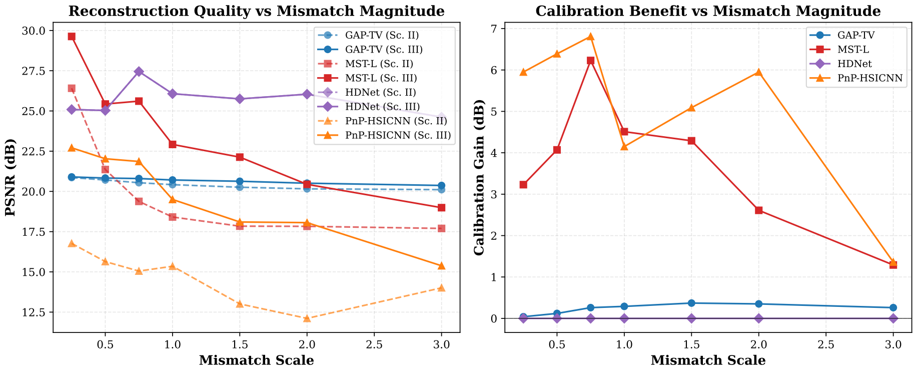

Fig. 5: Sensitivity to mismatch magnitude. Left: Scenario II PSNR vs. mismatch scale. Right: calibration gain (II→III) vs. mismatch scale. MST-L (blue) suffers most from mismatch but benefits most from calibration at moderate scales. HDNet (red) shows zero calibration gain across all scales.

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. Arguello, H., Arce, G.R.: Coded aperture optimization for compressive spectral imaging using the colored-cassi architecture. IEEE Transactions on Image Processing **23**(4), 1896–1908 (2013)
3. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432 (2013)
4. 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: Proceedings of the European Conference on Computer Vision (ECCV). pp. 686–704 (2022)
5. 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)
6. 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 reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 745–755 (2022)
7. 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: Advances in Neural Information Processing Systems (NeurIPS). vol. 35, pp. 37749–37761 (2022)
8. 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)

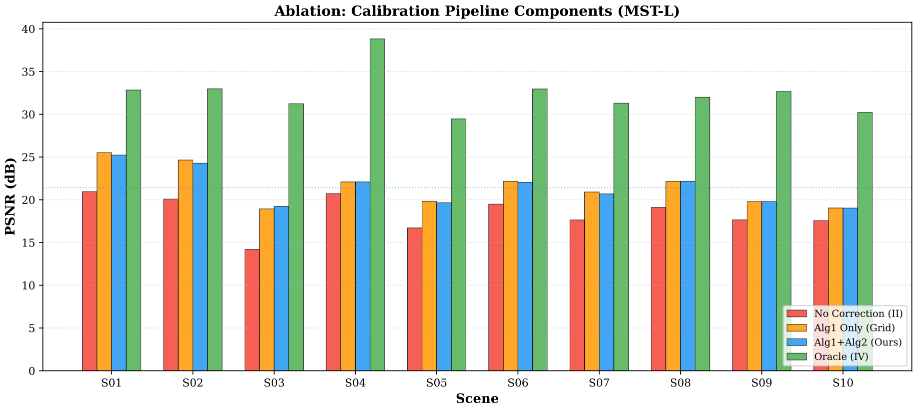

Fig. 6: Ablation study on MST-L: per-scene PSNR for No Correction (II), Grid-only (Alg1), Full pipeline (Alg1+Alg2), and Oracle (IV). Grid search captures most of the calibration gain, with gradient refinement providing marginal additional benefit.

9. 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)
10. 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)
11. Meng, Z., Ma, J., Yuan, X.: Gap-net for snapshot compressive imaging. arXiv preprint arXiv:2012.08364 (2020)
12. 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)
13. Ongie, G., Jalal, A., Metzler, C.A., Baraniuk, R.G., Dimakis, A.G., Willett, R.: Deep learning techniques for inverse problems in imaging. IEEE Journal on Selected Areas in Information Theory **1**(1), 39–56 (2020)
14. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. Applied Optics **47**(10), B44–B51 (2008)
15. 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)
16. Yuan, X.: Generalized alternating projection based total variation minimization for compressive sensing. In: IEEE International Conference on Image Processing (ICIP). pp. 2539–2543 (2016)
17. Yuan, X., Brady, D.J., Katsaggelos, A.K.: Snapshot compressive imaging: Theory, algorithms, and applications. IEEE Signal Processing Magazine **38**(2), 65–88 (2021)
18. Yuan, X., Liu, Y., Suo, J., Dai, Q.: Plug-and-play algorithms for large-scale snapshot compressive imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1447–1457 (2020)

19. 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)