InverseNet: A CASP-Inspired Benchmark for Operator Mismatch in Compressive Imaging

Chengshuai Yang*

NextGen PlatformAI C Corp

Abstract. Compressive imaging faces a critical *sim-to-real crisis*: models trained on idealized forward operators fail catastrophically when deployed on real hardware. Operator mismatch—the gap between assumed and true forward operators—degrades deep learning reconstruction by 10–21 dB, yet no existing benchmark measures this effect. 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 11 reconstruction methods under a standardized three-scenario protocol—ideal (I), mismatched (II), and oracle-corrected (III)—across 27 test scenes and over 240 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*. On CACTI, state-of-the-art EfficientSCI loses 20.58 dB under mismatch, while classical GAP-TV recovers 93% of its own mismatch loss through oracle calibration. We further establish a mask-awareness taxonomy—mask-oblivious architectures show zero calibration benefit ($\rho = 0\%$), while mask-conditioned methods recover 41–90% of mismatch losses depending on mismatch type. All reconstruction arrays, per-scene metrics, and analysis code are publicly released. By providing a standardized, reproducible evaluation of operator mismatch across modalities, InverseNet aims to catalyze an "AlphaFold moment" for computational imaging—shifting the field's focus from ideal-condition leaderboards to real-world deployment robustness.

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

* integrityyang@gmail.com

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 has formed between research and reality—a *sim-to-real valley of death*. Reconstruction algorithms are developed and benchmarked using idealized forward operators, but when these models are deployed on real optical bench setups, performance collapses. The scale of this collapse is staggering: EfficientSCI [12] reconstructs video at 35.39 dB from ideal measurements but collapses to 14.81 dB—a 20.58 dB drop—under realistic 8-parameter mismatch. This failure mode is not an edge case; it is the default condition of every compressive imaging system in the field.

In practice, the assumed forward operator inevitably differs from the true physical operator. For CASSI systems, mask misalignment of even 0.5 pixels and 0.1° rotation, combined with dispersion slope drift of 1% and axis offset of 0.15°, can degrade peak signal-to-noise ratio (PSNR) by over 13 dB [1]. For CACTI, clock drift, duty cycle variations, and detector gain offsets collectively introduce multi-parameter mismatch. For single-pixel cameras, gain drift in the digital micromirror device or photodetector causes systematic measurement errors. These *operator mismatches* are ubiquitous in deployed systems yet are systematically ignored in reconstruction benchmarks.

The CASP analogy. In biology, the Critical Assessment of protein Structure Prediction (CASP) challenge [28] transformed protein folding by forcing blind prediction against nature's ground truth, ultimately driving the AlphaFold breakthrough [29]. Computational imaging needs its own CASP moment: a benchmark that evaluates algorithms not against idealized simulations but against the messy reality of physical measurement systems. InverseNet is designed to fill this role.

The benchmark gap. Existing benchmarks for compressive imaging reconstruction—such as the KAIST hyperspectral dataset for CASSI [17] and the video compressive sensing benchmark for CACTI [4]—assume perfect operator knowledge. Methods are evaluated only under Scenario I (ideal operator), providing no information about robustness to operator mismatch or the potential benefit of calibration. This creates a critical blind spot: a method that achieves state-of-the-art PSNR under ideal conditions may catastrophically fail in practice if it is highly sensitive to operator errors. Antun et al. [30] demonstrated that deep learning solvers for inverse problems are unstable to adversarial perturbations; InverseNet operationalizes this finding into a systematic cross-modality benchmark that quantifies instability under *physically realistic* operator mismatch.

Contributions. Our central hypothesis is that *a mediocre algorithm with a correct physical model outperforms a state-of-the-art algorithm with a wrong one*—and our results confirm this across all three modalities. We address this gap with InverseNet, which makes three contributions:

1. **Unified three-scenario protocol.** We define a standardized evaluation framework with three scenarios—ideal (I), mismatched (II), and oracle-corrected

(III)—applicable across compressive imaging modalities. The gap between Scenarios I and II quantifies mismatch sensitivity; the recovery from II to III quantifies the upper bound of calibration benefit.
2. **Cross-modality benchmark.** We evaluate 11 reconstruction methods (4 CASSI, 4 CACTI, 3 SPC) spanning classical optimization (GAP-TV, FISTA-TV) and deep learning (MST, HDNet, EfficientSCI, ELP-Unfolding, ISTA-Net, HATNet) across 27 test scenes, producing over 240 reconstruction experiments.
3. **Open dataset.** We publicly release all reconstruction arrays, per-scene metrics (PSNR, SSIM, SAM), and analysis code at `https://github.com/integritynoble/Physics_World_Model` to enable reproducible operator-mismatch research. All underlying 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) mask-aware architectures (MST, ELP-Unfolding) are simultaneously the most sensitive to mismatch and the most recoverable through calibration; (c) mask-oblivious architectures (HDNet) show zero calibration benefit; (d) CACTI exhibits the most severe mismatch degradation (up to 20.58 dB) due to its multi-parameter mismatch space; (e) dispersion mismatch in CASSI creates larger degradation than mask spatial mismatch alone, with limited oracle recoverability due to fixed-step architectural assumptions.

2 Related Work

Compressive imaging reconstruction. Classical reconstruction methods for compressive imaging employ convex optimization with sparsity-promoting regularization. GAP-TV [7] uses the generalized alternating projection framework with total variation (TV) regularization, applicable to both CASSI and CACTI. FISTA [8] and ADMM [9] provide efficient solvers for ℓ_1-regularized inverse problems. Deep learning methods have dramatically improved reconstruction quality: MST [10] introduces mask-guided spectral transformers for CASSI; HDNet [11] uses dual-domain processing; DAUHST [23] combines deep unfolding with hierarchical spectral transformers; CST [24] leverages cross-stage spectral attention. For video compressive sensing, EfficientSCI [12] and ELP-Unfolding [13] achieve state-of-the-art results, while STFormer [25] introduces spatial-temporal transformers. ISTA-Net [14] and HATNet [15] address single-pixel imaging. All these methods are developed and evaluated assuming perfect forward operators.

Calibration and operator mismatch. Operator mismatch has been studied in specific modalities but not systematically benchmarked. For CASSI, Wagadarikar et al. [1] identified mask misalignment as a key error source, and subsequent work [21] proposed calibration procedures. For MRI, the fastMRI benchmark [19] evaluates undersampling patterns but assumes known coil sensitivities. Phase retrieval literature has examined model mismatch in coherent imaging [20]. Recent work on robust reconstruction has explored distributional robustness and uncertainty

quantification in inverse problems [26], and plug-and-play methods with convergence guarantees [27] provide frameworks for handling model uncertainty. However, no prior work provides a unified cross-modality benchmark quantifying both the degradation from mismatch and the recovery potential from calibration.

Reconstruction benchmarks. The KAIST TSA dataset [17] provides 10 simulated hyperspectral scenes for CASSI evaluation. The video compressive sensing benchmark [4] provides grayscale video sequences for CACTI. The Set11 dataset [18] is standard for single-pixel camera evaluation. Large-scale benchmarks like NTIRE [31] have standardized image restoration challenges, and the SupER dataset [32] provides real optical data for super-resolution, but neither enables controlled modification of the forward model estimate. All these benchmarks evaluate reconstruction quality under ideal conditions only. InverseNet extends them by introducing controlled operator mismatch and measuring calibration recovery, creating the first cross-modality operator-mismatch benchmark.

3 The InverseNet Benchmark

3.1 Unified Three-Scenario Protocol

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

- **Scenario I (Ideal):** The measurement is formed with the ideal operator $\hat{\mathbf{\Phi}}$, and reconstruction uses the same ideal operator. This represents the best-case performance with perfect operator knowledge: $\mathbf{y} = \hat{\mathbf{\Phi}}\mathbf{x} + \mathbf{n}$, reconstruct with $\hat{\mathbf{\Phi}}$.
- **Scenario II (Baseline):** The measurement is formed with the true (mismatched) operator $\mathbf{\Phi}$, but reconstruction still uses the assumed operator $\hat{\mathbf{\Phi}}$. This represents the realistic deployment scenario where the physical operator has drifted from its assumed value: $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$, reconstruct with $\hat{\mathbf{\Phi}}$.
- **Scenario III (Oracle):** The measurement is formed with the true operator $\mathbf{\Phi}$ (same as Scenario II), but reconstruction uses the true operator as oracle knowledge: $\mathbf{y} = \mathbf{\Phi}\mathbf{x}+\mathbf{n}$, reconstruct with $\mathbf{\Phi}$. This represents the upper bound achievable through perfect calibration.

This protocol yields two diagnostic metrics per method:

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

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

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.

The Ladder of Pain. The three-scenario protocol can be understood as a snapshot from a graduated "Ladder of Pain"—a curriculum of increasing difficulty from nominal (ideal) through geometric, temporal, radiometric, and combined mismatch. Our current benchmark evaluates the combined (hardest) tier; future rounds will report per-tier results to identify which mismatch types are most damaging for each modality.

Toward forward-model estimation. While InverseNet currently evaluates reconstruction under known mismatch (Scenario III provides oracle operator knowledge), the benchmark design supports a more ambitious task: *forward-model estimation*, where participants must reverse-engineer the physical state of the measurement system from calibration data alone. This transforms the challenge from "reconstruct given the operator" to "identify the operator, then reconstruct"—a strictly harder problem that better reflects real-world deployment.

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 $\mathbf{x} \in \mathbb{R}^{H \times W \times \Lambda}$ 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:

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

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

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

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

where dx, dy are subpixel translational shifts, θ is a rotational misalignment of the coded aperture mask, a_1 is the dispersion slope (nominal $s = 2.0\,\text{px/band}$), and α is the dispersion axis angular offset. We use $dx = 0.5\,\text{px}$, $dy = 0.3\,\text{px}$, $\theta = 0.1^\circ$ for mask misalignment, and $a_1 = 2.02\,\text{px/band}$ (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, $\lambda_{\text{TV}} = 0.1$); **HDNet** [11]: dual-domain deep network with spectral discrimination learning (pretrained); **MST-S** [10]: mask-guided spectral transformer, small variant (2 stages, blocks $[2,2,2]$, 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 256×310 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:

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

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 ($dx = 0.5\,\text{px}$, $dy = 0.3\,\text{px}$), rotation ($\theta = 0.1°$), 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 256×256 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:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{6}$$

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 multiplicative gain drift affecting the measurement matrix:

$$\mathbf{\Phi} = \text{diag}(1 + \alpha \cdot \mathbf{g}) \cdot \hat{\mathbf{\Phi}}, \tag{7}$$

where $\alpha = 0.0015$ controls the drift magnitude and $\mathbf{g}$ is a per-row gain perturbation vector. Additional measurement noise $\sigma_y = 0.03$ is applied.

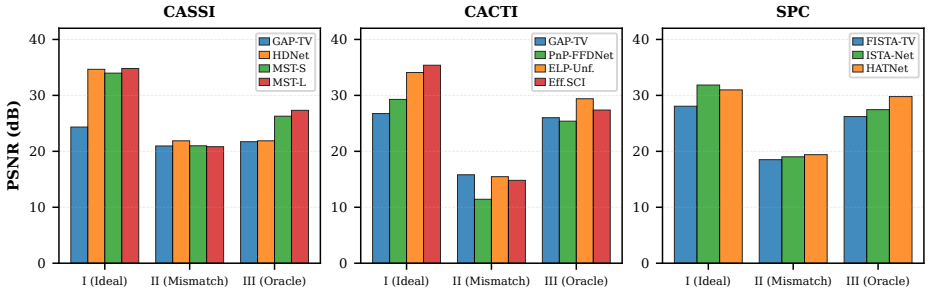

Fig. 1: PSNR across three scenarios for all modalities. Scenario I (Ideal): perfect operator. Scenario II (Baseline): mismatched operator. Scenario III (Oracle): true operator used for reconstruction. The collapse of deep learning methods under Scenario II is visible across all modalities, with CACTI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

Reconstruction methods. We evaluate three methods: **FISTA-TV** [8]: fast iterative shrinkage-thresholding with TV regularization (500 iterations, $\lambda = 0.005$); **ISTA-Net** [14]: learned iterative shrinkage-thresholding network (pretrained); **HATNet** [15]: dual-scale transformer for single-pixel imaging (pretrained).

Dataset. We use the 11 standard Set11 test images (*Monarch*, *Parrots*, *barbara*, *boats*, *cameraman*, *fingerprint*, *flinstones*, *foreman*, *house*, *lena256*, *peppers256*) at 256×256 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).

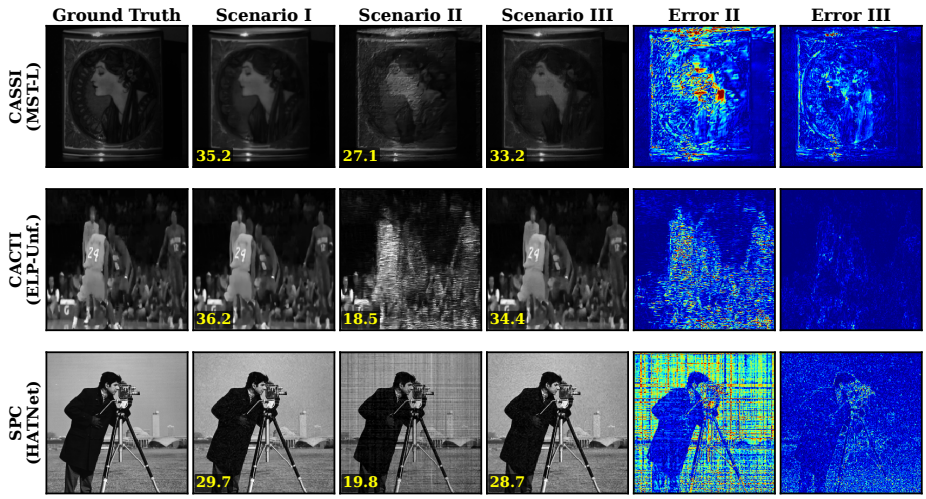

Fig. 2: Qualitative reconstruction comparison across three modalities. Each row shows a representative scene for one modality (CASSI: Scene 1 band 14, MST-L; CACTI: *kobe* frame 4, ELP-Unfolding; SPC: *cameraman*, HATNet). Error maps (jet colormap, same scale per row) highlight how mismatch (Scenario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

Key findings. The CASSI results under the 5-parameter mismatch model reveal a striking dichotomy between mask-aware and mask-oblivious architectures, with dispersion mismatch creating significantly larger degradation than mask spatial mismatch alone. **MST-L** achieves the best ideal performance (34.81 dB) and the highest oracle recovery (+6.50 dB, $\rho = 46.5\%$), making it the optimal choice when calibration is available. However, the moderate recovery ratio reflects the limitation that current architectures assume fixed integer dispersion steps and cannot fully correct sub-pixel dispersion drift even with oracle mask knowledge. **HDNet**, which processes only the initial spectral estimate without mask input, shows *zero* oracle gain ($\Delta_{\text{rec}} = 0.00$ dB), confirming that mask-oblivious architectures cannot benefit from operator calibration. **GAP-TV** shows moderate mismatch degradation ($\Delta_{\text{deg}} = 3.38$ dB) with a 22.5% recovery ratio, demonstrating that even classical iterative solvers benefit from oracle calibration when tuned to competitive strength (24.34 dB ideal). Under Scenario II (uncorrected mismatch), all deep learning methods converge to a narrow performance range (20.83–21.88 dB), erasing the ∼10 dB advantage they hold over classical methods under ideal conditions.

SSIM and SAM analysis. SSIM trends mirror PSNR findings. MST-L recovers from 0.744 to 0.881 SSIM under oracle correction, while HDNet remains fixed at 0.756. For SAM, MST-L improves from 23.92° (Scenario II) to 11.74° (Sce-

Table 1: CASSI reconstruction results (10 KAIST scenes, 256 × 256 × 28, 5-parameter mismatch). PSNR (dB) / SSIM reported as mean ± std. Δ_{deg}: degradation (I→II). Δ_{rec}: oracle recovery (II→III). ρ: recovery ratio. Best recovery in **green**.

Method	Scenario I	Scenario II	Scenario III	Δ_{deg}	Δ_{rec}	ρ
GAP-TV [7]	24.34±1.90 / .722	20.96±1.62 / .611	21.72±1.48 / .687	3.38	0.76	22.5%
HDNet [11]	34.66±2.62 / .970	21.88±1.72 / .756	21.88±1.72 / .756	12.78	0.00	0%
MST-S [10]	33.98±2.50 / .965	20.99±2.08 / .771	26.28±1.88 / .870	12.99	5.29	40.7%
MST-L [10]	**34.81**±2.11 / **.973**	20.83±2.01 / .744	**27.33**±1.86 / **.881**	13.98	**6.50**	**46.5%**

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

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

nario III), compared to the ideal value of 7.44°. The residual SAM gap (11.74° vs. 7.44°) reflects the spectral distortion from uncorrected dispersion mismatch, which mask-only oracle correction cannot fully address.

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 of all three modalities, with losses ranging from 10.94 dB (GAP-TV) to 20.58 dB (EfficientSCI). The 8-parameter mismatch space—encompassing spatial, temporal, and radiometric errors—creates compounding degradation that is far more destructive than the 5-parameter mismatch in CASSI. Under Scenario II, all methods collapse to 11–16 dB, with SSIM dropping below 0.31, indicating near-complete reconstruction failure.

GAP-TV achieves the highest recovery ratio ($\rho = 93.3\%$), recovering 10.21 dB of its 10.94 dB loss. This demonstrates that classical iterative methods, which directly incorporate the forward operator in each iteration, are highly responsive to operator correction. **EfficientSCI**, despite achieving the best ideal performance (35.39 dB), has the lowest recovery ratio (61.1%), suggesting that its learned features are partially coupled to the ideal operator in ways that oracle mask knowledge cannot fully address.

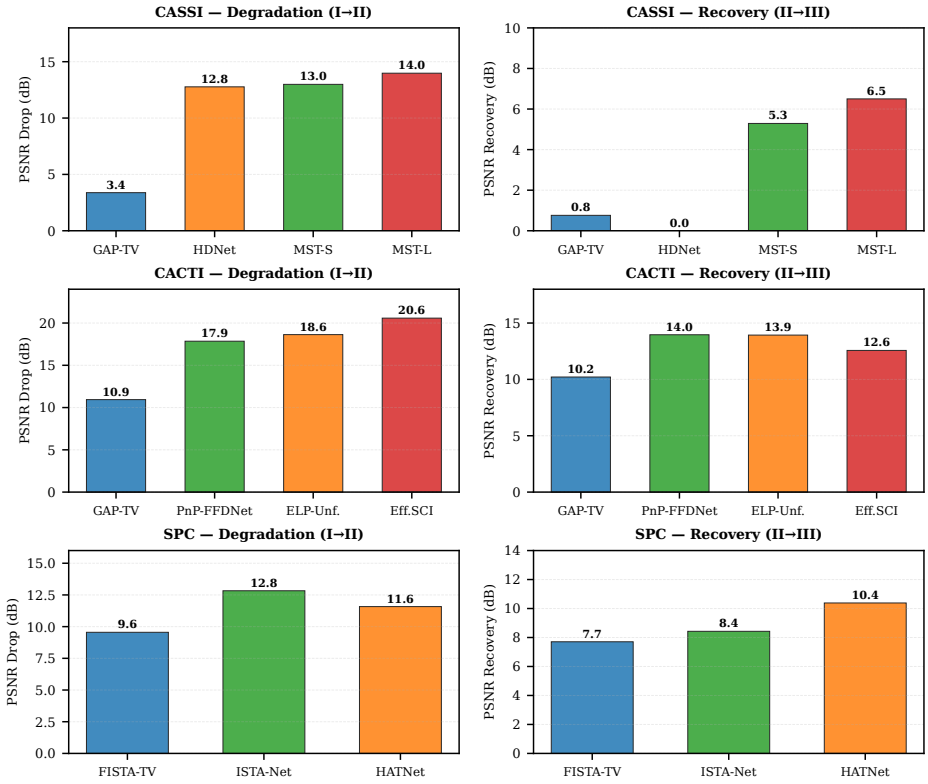

Fig. 3: Mismatch degradation (Δ_{deg}, left) and oracle recovery (Δ_{rec}, right) per method across all three modalities. CACTI suffers the most severe degradation (up to 20.6 dB) but also the highest absolute recovery. For CASSI, HDNet shows zero recovery due to its mask-oblivious architecture; MST-L achieves the best recovery (+6.50 dB, $\rho = 46.5\%$). For SPC, HATNet recovers 10.4 dB ($\rho = 89.6\%$).

An inverse relationship. A notable pattern emerges, also visible in the cross-modality scatter plot (fig. 4): methods with higher ideal performance suffer larger mismatch degradation and achieve lower recovery ratios. EfficientSCI (best ideal: 35.39 dB) loses 20.58 dB and recovers 61.1%; GAP-TV (worst ideal: 26.75 dB) loses 10.94 dB and recovers 93.3%. This suggests that higher-capacity learned representations encode stronger implicit assumptions about the operator, making them more fragile when those assumptions are violated.

4.3 SPC Results

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

Key findings. The SPC results show that gain drift uniformly degrades all methods to a narrow 18.51–19.40 dB range under Scenario II, despite a 3.79 dB

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

Method	Scenario I	Scenario II	Scenario III	Δ_{deg}	Δ_{rec}	ρ
FISTA-TV [8]	28.06±3.38 / .852	18.51±0.69 / .586	26.21±2.28 / .759	9.55	7.71	80.7%
ISTA-Net [14]	**31.85**±3.11 / **.916**	19.02±0.61 / .584	27.45±1.32 / .760	12.83	8.43	65.7%
HATNet [15]	30.98±0.95 / .847	19.40±0.59 / .648	**29.78**±0.81 / **.807**	11.58	**10.38**	**89.6%**

Table 4: Cross-modality summary. For each modality, we report the range of mismatch degradation (Δ_{deg}), oracle recovery (Δ_{rec}), and the best recovery ratio (ρ_{best}) across methods. Mismatch dim. is the number of mismatch parameters.

Modality	Mismatch dim.	Δ_{deg} range	Δ_{rec} range	ρ_{best}	Best method
CASSI	5	3.38–13.98	0.00–6.50	46.5%	MST-L
CACTI	8	10.94–20.58	10.21–13.96	93.3%	GAP-TV
SPC	2	9.55–12.83	7.71–10.38	89.6%	HATNet

spread under ideal conditions. **HATNet** achieves the highest oracle recovery (+10.38 dB, $\rho = 89.6\%$), recovering nearly to its ideal performance (29.78 vs. 30.98 dB). **ISTA-Net**, despite the best ideal performance (31.85 dB), achieves only 65.7% recovery ratio, consistent with the CACTI observation that higher-capacity learned representations are more fragile under mismatch. **FISTA-TV** achieves a balanced 80.7% recovery ratio with the lowest ideal performance.

Mismatch uniformity. Under Scenario II, the standard deviation across images decreases dramatically (from 0.95–3.38 dB to 0.59–0.69 dB), indicating that gain drift mismatch creates a performance floor independent of image content. This is consistent with the multiplicative nature of gain drift, which affects all measurement rows similarly.

4.4 Cross-Modality Analysis

Table 4 synthesizes the key metrics across all three modalities.

Mismatch severity ranking. CACTI suffers the most severe degradation (10.94–20.58 dB), followed by CASSI (3.38–13.98 dB) and SPC (9.55–12.83 dB). The CACTI severity is driven by its high-dimensional mismatch space (8 parameters vs. 5 for CASSI and 2 for SPC), where spatial, temporal, and radiometric errors compound. With the 5-parameter mismatch model, CASSI now surpasses SPC in maximum degradation, driven by the dispersion parameters (a_1, α) which create cumulative sub-pixel shifts across spectral bands. For CASSI, the range reflects the architectural divide: GAP-TV's iterative optimization adapts partially to the corrupted measurement (3.38 dB loss), while mask-aware deep networks lose over 13 dB.

Table 5: Mismatch parameters per modality. CASSI: mask spatial misalignment + dispersion drift (5 parameters). CACTI: spatial + temporal + radiometric mismatch (8 parameters). SPC: multiplicative gain drift.

Modality	**Parameter**	**Value**	**Physical interpretation**
CASSI	dx	0.5 px	Horizontal mask shift
	dy	0.3 px	Vertical mask shift
	θ	0.1°	Mask rotation
	a_1	2.02 px/band	Dispersion slope (nominal 2.0)
	α	0.15°	Dispersion axis offset
CACTI	dx	0.5 px	Horizontal mask shift
	dy	0.3 px	Vertical mask shift
	θ	0.1°	Mask rotation
	Δt	0.05	Clock offset
	η	0.95	Duty cycle deviation
	g	1.02	Detector gain
	o	0.002	Detector offset
	σ_n	1.0	Measurement noise
SPC	α	0.0015	Gain drift magnitude
	σ_y	0.03	Measurement noise

Recovery potential. Figure 4 visualizes the recovery ratio versus ideal performance for all methods. The best recovery ratio varies by modality: GAP-TV achieves 93.3% on CACTI, HATNet achieves 89.6% on SPC, and MST-L achieves 46.5% on CASSI. The lower CASSI recovery ratio reflects the fundamental challenge of dispersion mismatch: current architectures assume fixed integer dispersion steps ($s = 2$ px/band) and cannot adapt to the true dispersion slope ($a_1 = 2.02$ px/band) even when oracle mask knowledge is provided. This suggests that the recovery potential depends on both the method architecture and the mismatch structure, with dispersion-type mismatches being significantly harder to correct than spatial mismatches alone. Notably, the best-recovering method is not always the one with the highest ideal performance—a finding with practical implications for system design.

Architectural patterns. Across modalities, we observe consistent patterns: (i) classical methods (GAP-TV, FISTA-TV) show moderate degradation and high recovery ratios; (ii) mask-aware deep methods (MST, ELP-Unfolding, HATNet) show high degradation but substantial recovery; (iii) mask-oblivious deep methods (HDNet) show moderate degradation but zero recovery. This three-way classification provides a practical taxonomy for selecting reconstruction methods based on whether calibration is available.

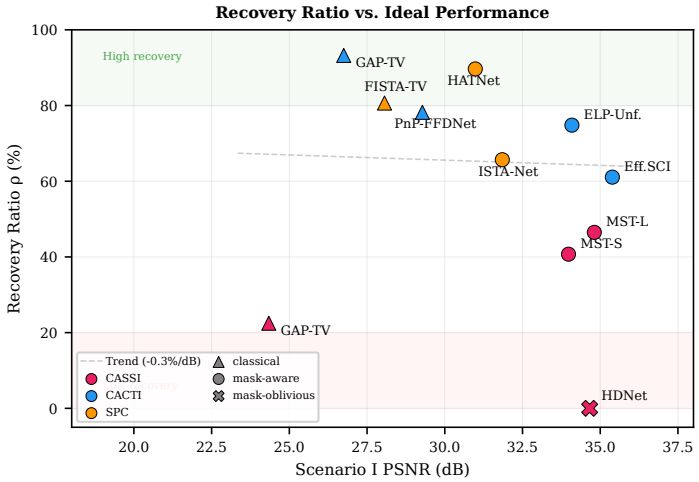

Fig. 4: Recovery ratio (ρ) vs. ideal PSNR (Scenario I) for all 11 methods across three modalities. Color indicates modality; shape indicates method type (classical, mask-aware, mask-oblivious). An inverse trend is visible: higher-performing methods tend to have lower recovery ratios, suggesting that stronger learned priors create greater operator dependence.

5 Discussion

Classical vs. deep learning robustness. A consistent finding across all modalities is that classical optimization methods are more robust to operator mismatch than deep learning methods. 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). However, this robustness comes at the cost of lower ideal performance (24.34 dB vs. 34.81 dB on CASSI). The practical implication is that when calibration is unavailable or imperfect, classical methods may outperform deep learning: on CACTI Scenario II, GAP-TV (15.81 dB) outperforms PnP-FFDNet (11.43 dB) by 4.38 dB despite being 2.53 dB worse under ideal conditions.

The sim-to-real collapse. Under mismatch, the performance hierarchy inverts. On CACTI Scenario II, the best deep network (EfficientSCI, 35.39 dB ideal) scores 14.81 dB—barely above GAP-TV's 15.81 dB. The 8.64 dB advantage that EfficientSCI holds under ideal conditions is not merely erased but *inverted*: GAP-TV outperforms it by 1.0 dB under realistic deployment conditions. This confirms our central hypothesis: a mediocre algorithm with a correct forward model beats a state-of-the-art algorithm with a wrong one.

The mask-awareness spectrum. Our results reveal that reconstruction methods exist on a spectrum of mask-awareness:

- *Mask-oblivious* (HDNet): The mask is not an input; reconstruction quality depends only on the measurement. No calibration benefit is possible ($\rho = 0\%$).
- *Mask-conditioned* (MST-S, MST-L, HATNet): The mask explicitly conditions the reconstruction network. These methods achieve moderate-to-high calibration gains ($\rho = 41$–90%) but suffer the largest mismatch degradation. The recovery ratio depends on mismatch type: spatial mask mismatch is highly recoverable (SPC HATNet: 90%), while dispersion mismatch limits recovery (CASSI MST-L: 47%).
- *Operator-iterative* (GAP-TV, FISTA-TV): The forward operator is used in each optimization iteration. These methods achieve high recovery ratios on spatial and gain-type mismatches ($\rho = 81$–93% on CACTI and SPC), though dispersion mismatch limits CASSI recovery ($\rho = 23\%$).

This taxonomy reframes the reconstruction problem: the critical bottleneck is not algorithmic sophistication but *physical model fidelity*. Mask-conditioned architectures are optimal when calibration is feasible, while mask-oblivious architectures provide a stable (but suboptimal) fallback.

Oracle recovery as calibration upper bound. Scenario III provides the upper bound for calibration benefit, achievable only with perfect knowledge of the true operator. In practice, calibration algorithms estimate the mismatch parameters with finite precision, so the actual calibration gain will be less than Δ_{rec}. The large recovery values we observe (up to 13.96 dB for CACTI, 10.38 dB for SPC, 6.50 dB for CASSI) strongly motivate the development of practical calibration methods, even if they recover only a fraction of the oracle bound. The CASSI results further suggest that dispersion-aware reconstruction architectures—which can adapt their spectral shifting to the true dispersion parameters—could significantly improve recovery beyond the 46.5% achieved by current fixed-step methods.

Residual gap analysis. Figure 5 visualizes the residual gap $\Delta_{\text{res}} = \text{PSNR}_{\text{I}} - \text{PSNR}_{\text{III}}$, which represents unrecoverable losses due to measurement corruption and architectural limitations in the oracle correction. For CASSI MST-L, $\Delta_{\text{res}} = 7.48$ dB; for CACTI GAP-TV, $\Delta_{\text{res}} = 0.74$ dB; for SPC HATNet, $\Delta_{\text{res}} = 1.20$ dB. The larger CASSI residual gap reflects the dispersion mismatch that oracle mask correction alone cannot address: current architectures (MST, HDNet) use fixed integer dispersion steps, so the 1% slope drift ($a_1 = 2.02$ vs. nominal 2.0 px/band) creates cumulative sub-pixel errors across 28 spectral bands that persist even with perfect mask knowledge. The small residual gaps for CACTI and SPC confirm that for spatial and gain-type mismatches, oracle correction nearly recovers ideal performance.

Method-specific insights. **HDNet** presents an instructive case: its mask-oblivious architecture processes only the initial spectral estimate (28 channels from the shift-back operation), making it inherently insensitive to the reconstruction mask.

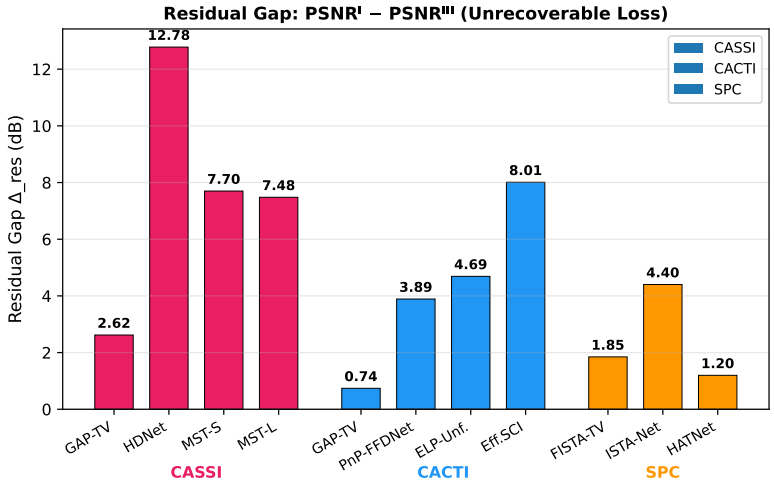

Fig. 5: Residual gap ($\Delta_{\text{res}} = \text{PSNR}_{\text{I}} - \text{PSNR}_{\text{III}}$) per method, grouped by modality. CASSI exhibits the largest residual gaps due to dispersion mismatch that oracle mask correction alone cannot address. CACTI and SPC residual gaps are small, confirming high recoverability of spatial and gain-type mismatches.

While this provides stability under mismatch (21.88 dB in both Scenarios II and III), it permanently forfeits any calibration benefit from the 12.78 dB degradation. **EfficientSCI** on CACTI reveals that state-of-the-art performance under ideal conditions can be a liability: its 35.39 dB ideal performance drops to 14.81 dB under mismatch, a 20.58 dB loss that is only 61.1% recoverable. This suggests that EfficientSCI encodes strong implicit assumptions about the CACTI operator that extend beyond what oracle mask knowledge can correct.

6 Conclusion

We have presented InverseNet, the first cross-modality benchmark for operator mismatch in compressive imaging. By evaluating 11 reconstruction methods across CASSI, CACTI, and SPC under a standardized three-scenario protocol, we establish several key findings: (1) operator mismatch degrades deep learning methods by 10–21 dB, collapsing the performance advantage over classical methods; (2) mask-aware architectures can recover 40–90% of mismatch losses through oracle calibration, with recovery depending on mismatch type; (3) mask-oblivious architectures provide mismatch stability at the cost of zero calibration benefit; (4) CACTI's high-dimensional mismatch space creates the most severe degradation across modalities.

These findings have direct implications for the design of compressive imaging systems. When calibration is feasible, mask-conditioned deep networks (MST-L for CASSI, HATNet for SPC) should be preferred for their high recovery

potential. When calibration is unavailable, classical methods (GAP-TV) provide the most robust baseline.

Dataset release. We release all 240+ reconstruction arrays, per-scene metrics, and analysis code. The benchmark is extensible: new modalities, methods, and mismatch models can be added by implementing the three-scenario protocol.

Future directions. A natural evolution is to replace oracle operator knowledge (Scenario III) with a *calibration packet*—a standardized set of diagnostic measurements (flat-field, dark-field, sparse-field, strobe-field) from which participants must estimate the true operator. This transforms InverseNet from a reconstruction benchmark into a system-identification challenge, analogous to how CASP evolved from homology modeling to end-to-end structure prediction [28]. Our CASSI results highlight a specific opportunity: dispersion-aware architectures that parameterize the spectral shift function (rather than assuming fixed integer steps) could significantly improve recovery from dispersion mismatch. Additionally, incorporating learned robustness through mismatch-aware training could bridge the gap between mask-conditioned and mask-oblivious architectures. Future InverseNet rounds will adopt CASP-style blind evaluation: participants submit containerized solvers evaluated on sealed test sets with hidden mismatch parameters, preventing overfitting to known test scenes and enabling fair longitudinal comparison.

Broader context. InverseNet serves as the targeting system rail for the Physics World Model (PWM), providing a durable evaluation infrastructure for computational imaging. While reconstruction solvers are continuously replaced with improved methods, the three-scenario protocol, scoring formulas, and benchmark datasets remain fixed, ensuring fair longitudinal comparison.

Data and code availability. All InverseNet benchmark data—including 240+ reconstruction arrays (NPZ format), per-scene metrics (PSNR, SSIM, SAM), mismatch parameter configurations, and figure-generation scripts—are publicly available at `https://github.com/integritynoble/Physics_World_Model` under the `papers/inversenet/` directory. The underlying test images come from publicly available sources: the KAIST TSA hyperspectral dataset [17] (CASSI), the standard video compressive sensing benchmark [4] (CACTI), and the Set11 dataset [18] (SPC). No non-public or restricted-access datasets were used in this work. All pretrained model weights were obtained from the respective authors' public repositories as cited.

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)