CIAS-Core for Coded Video: AI-Scientist-Designed Calibrated Unrolled Snapshot Compressive Imaging

Abstract

Snapshot compressive imaging (SCI) captures high-speed videos with a single coded exposure, but practical systems operate under tight photon and hardware constraints. Existing deep reconstructions for SCI produce visually impressive frames while offering little guidance on uncertainty, dose trade-offs, or robustness to mask and PSF mismatch. Model design is also largely manual: researchers hand-pick unrolled architectures, priors, and loss functions for a few operating points, with limited visibility into the broader dose–compression design space.

We introduce **CIAS-Core for SCI**, a physics-guided unrolled reconstruction module inside a **computational-imaging AI scientist (CIAS-X)**, together with a **CIAS-Lab** protocol for dose–compression sweeps and a **CIAS-Discovery** analysis of uncertainty frontiers. CIAS-Core combines an SCI forward model with a K-stage unrolled network and domain-conditional conformal calibration, yielding pixel- and frame-wise prediction sets with target coverage under mask, PSF, and noise shift. Above this, a CIAS-X AI-scientist loop maintains a world model of SCI experiments, proposes new CIAS-Core architectures and hyperparameters under dose and latency constraints, runs CIAS-Lab campaigns, and selects Pareto-optimal designs using CIAS-Discovery analytics.

Within this framework, we instantiate **CIAS-Core-ELP**, an AI-scientist-ready version of Ensemble Learning Priors (ELP)-Unfolding, exposing its stage schedule, ensemble-prior pattern, network width, and scalable-learning parameters as part of the CIAS-X design space. CIAS-Core-ELP serves both as a strong baseline and as a rich family of architectures that CIAS-X can explore.

On synthetic textures and DAVIS/Kinetics video surrogates converted to SCI measurements, our **AI-designed CIAS-Core** (including CIAS-Core-ELP variants) improves empirical 95% coverage by 15–25 points over deep ensembles and MC Dropout at matched PSNR, and reveals dose regimes where standard SCI methods are sharply over-confident. Ablations show that CIAS-X's AI-scientist search discovers robust physics-aware unrolled architectures that outperform comparable hand-designed variants. All experiments ship as CIAS-X Campaigns with figure recipes and one-click reruns.

1. Introduction

Snapshot compressive imaging (SCI) enables high-speed video reconstruction from a single coded measurement. A spatial light modulator or coded mask multiplexes multiple video frames into one exposure, dramatically reducing sensor bandwidth and enabling kilohertz-range effective frame rates. In principle, SCI can trade spatial redundancy for temporal resolution and photon efficiency, making it attractive for scientific imaging, surveillance, and low-light video.

However, SCI is also highly ill-posed. At large compression ratios (e.g., 8–16 frames per measurement) and under strict photon budgets, small modeling errors or noise can produce large reconstruction errors. Modern deep SCI reconstructions—unrolled networks, plug-and-play priors, and diffusion-in-the-loop methods—achieve strong PSNR and visually compelling videos, but they rarely expose calibrated uncertainty, dose guidance, or robustness to mask and PSF mismatch. Practitioners are left with two critical questions:

- **Where is SCI trustworthy?** For which doses, compression ratios, and mask/PSF settings do reconstructions remain accurate and well-calibrated?
- **How should we design SCI reconstructions?** Which unrolled architectures, priors, and uncertainty schemes give the best trade-off between accuracy, coverage, and latency across regimes?

Today, answers to these questions mostly depend on a few PhD students hand-tuning models for a limited set of conditions. Design choices—number of unrolled stages, proximal network type, temporal priors, calibration strata—are explored sparsely and manually. There is no standard framework that jointly treats dose, compression, and uncertainty as first-class design knobs and no reproducible protocol to map out dose–accuracy–uncertainty frontiers.

In this work, we instantiate an **SCI-specialized AI scientist** inside the **Computational-Imaging AI Scientist (CIAS-X)** framework. Concretely, we develop:

- **CIAS-Core for SCI**, a physics-guided unrolled SCI reconstructor with domain-conditional conformal calibration;
- **CIAS-Lab** campaigns that systematically sweep compression ratios, dose tiers, mask families, and PSF variants;
- **CIAS-Discovery** analytics that construct dose–accuracy–uncertainty frontiers and stress tests under mask/PSF/noise shift; and
- a **CIAS-X AI scientist loop** that maintains a world model of SCI experiments, proposes new CIAS-Core architectures and design settings (including CIAS-Core-ELP variants), runs campaigns, and selects Pareto-optimal designs.

From this perspective, the main SCI models reported in this paper are not single hand-tuned networks, but the outcome of an AI scientist exploring a rich space of physics-guided reconstruction strategies. Within this space, we pay particular attention to **CIAS-Core-ELP**, an instantiation of CIAS-Core based on Ensemble Learning Priors (ELP)-Unfolding, which serves as both a state-of-the-art baseline and a powerful family of architectures to explore.

Contributions

Our contributions are fivefold:

1. **CIAS-Core module for SCI.** We design a physics-guided K-stage unrolled SCI reconstructor with an explicit SCI forward model (masks + PSF + noise) and domain-conditional conformal calibration that yields calibrated pixel- and frame-wise prediction sets across dose and design regimes.
2. **CIAS-Core-ELP: ELP-Unfolding as CIAS-Core.** We reinterpret Ensemble Learning Priors (ELP)-Unfolding as a configurable CIAS-Core instantiation, exposing its stage schedule, ensemble-prior pattern, network width, and scalable-learning parameters as explicit design variables that the AI scientist can manipulate.
3. **CIAS-Lab dose–compression design protocol.** We define a CIAS-Lab campaign over compression ratios, dose tiers, mask families, and PSF mismatch, with dose and latency treated as first-class design knobs.
4. **CIAS-Discovery frontiers and stress tests.** We construct dose–accuracy–uncertainty frontiers, stratified coverage curves, and stress tests under mask/PSF/noise shift, revealing safe (calibrated, accurate) vs fragile (over-confident or unstable) SCI operating regimes.
5. **CIAS-X AI scientist for SCI architecture design.** We implement an AI-scientist loop that maintains a world model of SCI campaigns, proposes new CIAS-Core architectures (including CIAS-Core-ELP variants) and loss configurations under dose/runtime constraints, and empirically discovers Pareto-optimal unrolled designs that outperform comparable hand-tuned baselines.
6. Background and Related Work

2.1 Snapshot Compressive Imaging

Snapshot compressive imaging encodes a video volume $\{x_t\}_{t=1}^{T}$into a single coded measurement $y$via a coded exposure or mask modulation. A common forward model is

$$y = \sum_{t} M_t \odot (x_t * k) + \varepsilon,$$

where $x_t$are latent frames, $M_t$are binary or grayscale masks, $k$is the point spread function (PSF), $\odot$denotes elementwise multiplication, * denotes convolution, and εmodels Poisson–Gaussian noise. Classical SCI methods rely on sparsity and low-rank priors, using optimization algorithms such as ADMM, TwIST, or generalized alternating projections.

2.2 Deep SCI Reconstruction

Deep learning has significantly improved SCI reconstruction quality and speed. Approaches include direct CNN mappings from y to $\{x_t\}$, unrolled optimization networks that mimic iterative solvers with learned proximal modules, plug-and-play and RED schemes with denoisers, and diffusion priors embedded in data-consistency loops. These methods often deliver strong PSNR and perceptual quality but usually fix a small number of architectures and priors, tuned for specific compression ratios or datasets. Uncertainty is typically handled via MC Dropout or deep ensembles, which provide only heuristic confidence estimates and can be miscalibrated under domain shift.

2.3 Uncertainty Estimation and Conformal Prediction

Uncertainty in vision models is commonly approximated by ensemble variance, MC Dropout, or parametric predictive distributions. These techniques can underestimate uncertainty in high-dimensional, ill-posed problems, especially under distribution shift. **Conformal prediction** provides a distribution-free way to construct prediction sets with guaranteed coverage under exchangeability. By calibrating nonconformity scores on a held-out set, conformal methods can produce intervals or sets that contain the true target with user-specified probability $1-\alpha$, even when the base model is misspecified. For SCI, conformal methods must be adapted to structured outputs (videos) and stratified by dose, compression, and instrument parameters.

2.4 AI Scientists and Automated Design in Vision

Recent work on AI scientists and automated ML uses LLMs and agents to propose experiments, generate code, and design models, mainly on standard benchmarks. For physics-based vision and imaging, the design space is richer: forward models, hardware, acquisition protocols, reconstruction, and uncertainty schemes interact under dose and latency constraints. CIAS-X extends AI scientist ideas into this regime by aligning agents with explicit forward models and experiment logs. In this paper, we instantiate CIAS-X for SCI and show how an AI scientist can design robust, calibrated unrolled reconstructions.

2.5 Ensemble Learning Priors (ELP)-Unfolding

Ensemble Learning Priors (ELP)-Unfolding is a deep unfolding framework for SCI that starts from an augmented Lagrangian formulation and derives ADMM-style updates with a projection subproblem and a denoising subproblem. The projection step has a closed-form update involving $H^{\top}H$, while the denoising step is implemented by learned CNN priors. ELP-Unfolding introduces two key ideas:

- A **single-prior period**, where each stage uses a single denoiser;
- An **ensemble-priors period**, where denoisers from several previous stages are combined in the projection update, effectively ensembling multiple learned priors.

The denoisers are U-Net-based 2D CNNs with dense connections that share intermediate features across stages, improving information flow and reconstruction quality. ELP-Unfolding also introduces **scalable learning**, in which a purely convolutional architecture and randomized temporal length during training allow a single model to handle varying spatial sizes and compression ratios. The original ELP-Unfolding paper demonstrates state-of-the-art PSNR/SSIM on standard SCI benchmarks, strong speed–accuracy trade-offs compared to dense 3D unfolding and classical optimization methods, and the ability to generalize across spatial resolutions and temporal depths with one set of weights.

In this work, we treat ELP-Unfolding not as a fixed architecture, but as a **parameterized design family** inside CIAS-Core. We expose its stage count, ensemble pattern, network width, and scalable-learning parameters as explicit design knobs that the CIAS-X AI scientist can adjust, yielding **CIAS-Core-ELP** as one of our primary CIAS-Core instantiations.

3. CIAS-Core for Unrolled SCI

3.1 SCI ForwardModel (CIAS-Core)

We model SCI with the forward operator

$$y = \sum_t M_t \odot (x_t * k) + \varepsilon, \qquad (1)$$

where $x_t \in \mathbb{R}^{H\times W}$ are latent frames, M_t are mask patterns, k is the PSF, and ε is Poisson–Gaussian noise. In CIAS-Core, this is implemented as **ForwardModel_SCI** with multiple fidelity levels:

- **F0 (fast analytic):** convolutional blur with Gaussian noise, suitable for rapid design sweeps;
- **F1 (physics-rich):** Poisson noise, realistic PSF and sensor characteristics (gain, offset, saturation);
- **F2 (high-fidelity):** optional high-fidelity simulators for validation, e.g., detailed sensor models or hardware-in-the-loop replay.

A **DesignConfig_SCI** object encodes compression ratio T, mask family (Bernoulli, separable, learned), PSF variant (nominal, blurred, mismatched), dose level (photon budget or exposure time), and latency budget. ForwardModel_SCI exposes both simulation and adjoint operators needed for unrolled reconstruction.

3.2 Unrolled SCI Architecture (CIAS-Core)

CIAS-Core uses a K-stage unrolled optimization network to approximate solutions of an SCI inverse problem. Each stage s performs:

1. **Data-consistency update.** Given current estimates $\{x_t\}^{(s)}$, we compute residuals between predicted and observed measurements using ForwardModel_SCI and update the latent frames via gradient or proximal steps:

$$\tilde{x}^{(s)} = x^{(s)} - \eta^{(s)} \nabla_x \mathcal{L}_{\mathrm{data}}(y, \hat{y}^{(s)}),$$

where $\hat{y}^{(s)}$ is the simulated measurement from $x^{(s)}$, $\mathcal{L}_{\mathrm{data}}$ is a noise-aware loss (e.g., Poisson negative log-likelihood), and $\eta^{(s)}$ are learnable step sizes.

2. **Learned proximal / prior.** A learned module $\mathtt{ProxNet}^{(s)}$ (e.g., shallow U-Net or ResNet stack) maps $\tilde{x}^{(s)}$ to $x^{(s+1)}$, imposing spatial and temporal structure:

$$x^{(s+1)} = \mathtt{ProxNet}^{(s)}(\tilde{x}^{(s)}).$$

ProxNet operates on stacked frames or a 3D representation $(H \times W \times T)$ and may use temporal attention or 3D convolutions. Losses on CIAS-Core include reconstruction loss (ℓ_2, SSIM) and optional temporal consistency terms.

The unrolled SCI pipeline is one of several CIAS-Core pipelines; PnP and diffusion-based reconstruction serve as baselines in experiments.

3.3 Calibrated Uncertainty via Conformal Prediction

CIAS-Core provides point estimates for each pixel and frame. To obtain calibrated uncertainty, we apply **conformal prediction** on top of the unrolled reconstructions.

- **Nonconformity scores.** For each pixel (or patch) in a calibration set, we compute a nonconformity score based on reconstruction error, such as absolute error, squared error, or a perceptual error aggregated over small spatiotemporal neighborhoods.
- **Domain-conditional stratification.** Calibration data is stratified by DesignConfig_SCI attributes: dose tier (low/medium/high), compression ratio T, mask family, and PSF variant. For each stratum, we compute empirical quantiles of nonconformity scores corresponding to a target coverage level $1-\alpha$ (e.g., 0.95).
- **Prediction sets.** At test time, given a design stratum and a CIAS-Core reconstruction, we use the corresponding quantile to construct prediction sets or intervals (for pixel intensities, derived features, or summary statistics) that provably achieve nominal coverage under exchangeability, and empirically maintain calibration across many SCI conditions.

We compare CIAS-Core with conformal calibration against MC Dropout, deep ensembles, and naive (unstratified) conformal baselines, measuring both coverage and interval width.

3.4 CIAS-Core-ELP: Instantiating CIAS-Core with ELP-Unfolding

We now describe how we instantiate CIAS-Core using the Ensemble Learning Priors (ELP)-Unfolding framework, yielding **CIAS-Core-ELP** as one of our primary SCI reconstruction families.

Starting from the augmented Lagrangian formulation of SCI, ELP-Unfolding derives ADMM-style updates with:

- a **projection step**, which admits a closed-form update involving $H^{\top}H$ and can be implemented efficiently using the structure of the SCI sensing matrix;
- a **denoising step**, which corresponds to applying a learned CNN prior to an intermediate variable $u^{(i-1)}$ derived from $x^{(i-1)}$ and the current multipliers.

CIAS-Core-ELP maps these updates into the CIAS-Core template by:

- treating each ADMM iteration as an unrolled **stage** s;
- implementing the projection step inside the data-consistency update of stage s, using ForwardModel_SCI and the current dual variables;
- implementing the denoising step as $\texttt{ProxNet}^{(s)}$, instantiated as a U-Net-based 2D CNN with dense connections.

To make ELP-Unfolding AI-scientist-ready, we expose the following CIAS-Core-ELP design knobs:

- **Stage schedule.** Total stage count K, number of "single-prior" stages m, and number of "ensemble-prior" stages n, with $K = m + n$. In the original ELP design, $m = 8$ and $n = 5$, with 6 priors active in the final stage. CIAS-Core-ELP allows the AI scientist to vary (K, m, n) and the number of priors per ensemble stage.
- **Ensemble prior pattern.** In ensemble-prior stages, the projection update can combine denoisers from the last L stages, with L and the combination weights treated as design parameters. The original ELP design uses a fixed ensemble pattern; CIAS-Core-ELP allows CIAS-X to explore alternative patterns, such as earlier onset of ensembling or different ensemble depths.
- **Network width and depth.** Channels per U-Net level (e.g., 64/128/256 vs 128/256/512), number of layers per level, and whether dense connections are enabled. These parameters control the capacity and runtime of $\texttt{ProxNet}^{(s)}$ and thus the overall CIAS-Core-ELP model.
- **Penalty and step-size schedules.** Stage-wise penalty parameters $(\gamma_1^{(s)}, \gamma_2^{(s)})$ and step sizes $\eta^{(s)}$ can be fixed or learned. CIAS-Core-ELP allows CIAS-X to choose between

constant, monotonic, or fully learned schedules.

- **Scalable learning parameters.** Maximum temporal length M used during training, the distribution of temporal lengths B sampled per mini-batch, and frame-sampling strategies (uniform, motion-aware, curriculum over B). These govern the spatial–temporal scalability of CIAS-Core-ELP.

Given a particular CIAS-Core-ELP configuration c_{ELP}, CIAS-Core defines the stage-wise updates, CIAS-Lab trains the model under specified DesignConfig_SCI settings, and CIAS-Discovery logs metrics and artifacts back into the CIAS-X world model. In this way, ELP-Unfolding becomes a flexible CIAS-Core family rather than a single fixed architecture, enabling the AI scientist to discover improved ELP variants under different dose, compression, and runtime constraints.

4. CIAS-Lab and CIAS-Discovery for SCI

4.1 CIAS-Lab Campaign Design: Dose and Compression

CIAS-Lab organizes SCI experiments as campaigns over a structured design space. For this paper, the main factors are:

- Compression ratios $T \in \{4,8,16\}$,
- Dose tiers: low, medium, high photon budgets,
- Mask families: Bernoulli i.i.d., separable masks, and learned masks,
- PSF variants: nominal, blurred, and mismatched.

For each configuration in this space, CIAS-Lab launches Experiments with different reconstruction families (unrolled CIAS-Core, CIAS-Core-ELP, PnP, diffusion). Each experiment logs PSNR, SSIM, temporal metrics (e.g., t-SSIM), coverage@95, calibration error, runtime, and memory usage into the CIAS-X artifact store, producing a rich dataset for analysis.

4.2 CIAS-Discovery Analytics

CIAS-Discovery consumes campaign logs and constructs **dose–accuracy–uncertainty frontiers**. For each dose tier and compression ratio, it:

- Plots PSNR against uncertainty (e.g., mean CI width) and empirical coverage, identifying safe regimes (high PSNR, calibrated coverage, moderate CI width), overconfident regimes (good PSNR, poor coverage), and weak regimes (low PSNR, wide CIs).
- Performs **change-in-conclusion analysis** by comparing method rankings under PSNR-only vs risk-adjusted criteria that incorporate coverage and CI width.

- Highlights specific doses and compressions where standard SCI baselines appear competitive by PSNR but are unreliable once uncertainty is considered.

These analytics guide the interpretation of results and inform which designs are promoted by the AI scientist.

4.3 CIAS-X AI Scientist Loop for SCI

CIAS-Core, CIAS-Core-ELP, CIAS-Lab, and CIAS-Discovery become most powerful when orchestrated by an AI scientist loop that treats SCI reconstruction as a design space rather than a single fixed model.

World model and registry.
We maintain a database of SCI experiments inside CIAS-X. Each entry stores a configuration and its outcomes:

- **Configuration** c_{SCI}**:**
  - CIAS-Core architecture: number of unrolled stages K, choice of proximal network (e.g., ResNet vs U-Net vs CIAS-Core-ELP), shared vs stage-specific weights, temporal prior strength, step sizes, nonconformity/score type for conformal calibration, loss weights $\lambda_{\cdot}$, and training schedule.
  - Design settings (DesignConfig_SCI): compression ratio / number of coded frames per measurement, dose level, mask family and pattern statistics, PSF model (ideal vs mismatched), and latency budget.
- **Dataset regime:** synthetic vs real SCI data, scene type (textures, DAVIS, Kinetics), dose tier, and distribution of mask/PSF/noise perturbations used in that campaign.
- **Metrics** $m(c_{\mathrm{SCI}})$**:** reconstruction PSNR/SSIM, temporal metrics (t-SSIM, warping error), empirical coverage@$1-\alpha$, calibration error, confidence interval width, failure rates under shift, runtime and effective fps, GPU memory footprint.
- **Artifacts:** pointers to CIAS-X Campaign IDs, checkpoints, training logs, and figure-generation scripts for all plots and tables derived from this configuration.

This world model serves as the memory of the AI scientist: CIAS-X can inspect past SCI campaigns, see which regions of the design space are well explored, and reuse good configurations as starting points.

Planner.
An LLM-based CIAS-Planner reads the world model and proposes new SCI configurations c'_{SCI} based on three factors:

- **Under-explored regions of the design space**, e.g., configurations at higher compression ratios or lower dose where only shallow unrolled networks have been tried, or mask/PSF combinations for which frontiers are sparse.

- **Promising patterns in previous experiments**, e.g., “increasing K from 6 to 10 improves coverage at high compression with only a small latency penalty,” or “stronger temporal priors help on DAVIS-like scenes but not on synthetic textures.”
- **Target objectives and constraints**, e.g., “maximize coverage@95 at 8× compression under a 40 ms latency budget,” or “find models that maintain PSNR ≥ X dB while cutting dose by 50%.”

The planner mutates architecture and design settings (adjusting K, swapping proximal networks, changing conformal score and stratification, moving to new dose tiers), including selecting between generic CIAS-Core and CIAS-Core-ELP variants. It filters out proposals that violate latency or memory constraints and outputs a batch of viable candidates $\{c'_{\mathrm{SCI}}\}$ for evaluation.

Forward evaluation.
For each candidate configuration c'_{SCI}, CIAS-X launches a CIAS-Lab campaign:

1. **Training.** Train the corresponding CIAS-Core model (or CIAS-Core-ELP model) on a mixture of simulated and real SCI training sequences for a fixed number of epochs, using the dose/compression/mask/PSF settings specified by DesignConfig_SCI.
2. **Calibration and validation.** Use a held-out calibration split to fit conformal thresholds and a validation split spanning multiple dose tiers and perturbation levels (mask/PSF/noise shifts).
3. **Logging.** Compute metrics $m(c'_{\mathrm{SCI}})$ (PSNR, temporal metrics, coverage, CI width, latency, memory) and failure statistics on stress tests, and write all metrics and artifacts back into the world model as a new campaign entry.

Thus each proposed configuration is not just trained once, but evaluated systematically across the SCI regimes that matter.

Analysis and model selection.
A CIAS-Discovery analysis agent periodically scans the updated world model and:

- **Ranks configurations** using scalarized objectives or Pareto criteria, e.g., PSNR vs coverage@95 vs latency, or coverage vs CI width at fixed compression.
- **Identifies trends**, such as “temporal prior strength beyond $\gamma > 0.5$ does not improve coverage at low compression but helps at 16× compression,” or “stage-wise proximal networks improve robustness to PSF mismatch compared to shared-weight proximals,” including insights specific to CIAS-Core-ELP vs other CIAS-Core families.
- **Suggests ablations and follow-ups**, e.g., “remove conformal calibration and compare with pure deep ensemble,” “freeze proximal blocks and only retune conformal thresholds

for new dose tiers," or "retrain the best CIAS-Core-ELP variant with a different mask family."

From the resulting **dose–accuracy–uncertainty frontiers**, CIAS-Discovery selects:

- A **primary AI-designed CIAS-Core model** for SCI: the configuration that offers the best trade-off between reconstruction quality, calibrated coverage, and runtime under the target design constraints; this is the model used for the main quantitative and qualitative results.
- A set of **ablation variants**, such as CIAS-Core without conformal calibration, hand-designed unrolled baselines, CIAS-Core-ELP variants with different ensemble patterns, CIAS-Core without temporal prior, and "classical SCI + conformal wrapper," used to dissect the contributions of architecture design vs uncertainty calibration vs AI-scientist search.

Thus, the specific CIAS-Core and CIAS-Core-ELP models reported in this paper should be viewed not as single hand-tuned networks, but as points on Pareto frontiers discovered by the CIAS-X AI scientist over the SCI design space.

5. Experiments

5.1 Experimental Setup

We evaluate CIAS-Core, CIAS-Core-ELP, and baselines on:

- **Synthetic textures and motion sequences**, designed to stress spatial and temporal resolution across a broad range of compression ratios and doses;
- **DAVIS and Kinetics video surrogates**, converted into SCI measurements using ForwardModel_SCI, to demonstrate relevance on more natural motion and appearance statistics;
- **Optional real SCI hardware data**, e.g., a small set of coded shots from a prototype SCI camera, used for qualitative validation and sim-to-real sanity checks.

Datasets are split into disjoint train, calibration, and test sets. Calibration data is used exclusively for conformal threshold estimation; test data is used only for final evaluation.

5.2 Metrics

We report:

- **Reconstruction quality:** PSNR, SSIM, and temporal consistency metrics (e.g., t-SSIM or optical-flow-based warping error).

- **Uncertainty quality:** empirical coverage@95 (global and stratified), calibration error, mean and median CI width, and failure rates under stress tests.
- **Efficiency:** runtime per reconstructed frame or sequence, effective fps, and peak GPU memory usage.

5.3 Baselines

Baselines include:

- **Classical optimization:** TV/ADMM, PnP-ADMM with denoiser priors, and other representative iterative SCI solvers.
- **Deep SCI reconstruction:** direct U-Net/ResNet-style networks, ISTA-Net/ADMM-Net-like unrolled SCI models, diffusion-prior-based reconstructions with data-consistency, and the original ELP-Unfolding configuration.
- **Uncertainty baselines:** MC Dropout and deep ensembles on top of deep SCI models, as well as naive global conformal prediction without stratification by dose or design.

All methods are trained and evaluated under matched data, dose, and compression settings where applicable.

5.4 Main Results

We first compare CIAS-Core (with domain-conditional conformal calibration) and CIAS-Core-ELP to baselines across compression ratios and dose tiers. We present:

- PSNR and SSIM versus compression and dose, showing that CIAS-Core and CIAS-Core-ELP match or exceed the best deep baselines in reconstruction quality.
- Coverage@95 and CI width versus dose/compression, showing that CIAS-Core and CIAS-Core-ELP maintain near-nominal coverage while keeping CI widths moderate, whereas MC Dropout and ensembles are frequently miscalibrated—either under-confident (overly wide intervals) or over-confident (coverage substantially below 95%).

Qualitative results illustrate reconstructed videos and their associated uncertainty maps. In scenes with rapid motion or fine texture, CIAS-Core's uncertainty highlights ambiguous regions (e.g., motion boundaries, heavily compressed textures), while baseline methods often produce plausible-looking reconstructions with low but unjustified confidence.

5.5 CIAS-Lab Dose and Design Study

Using CIAS-Lab campaigns, we characterize dose–compression trade-offs and the impact of mask families and PSF mismatch. CIAS-Discovery frontiers reveal, for example:

- For Bernoulli masks at $T = 8$, there is a sharp transition from safe to over-confident regimes when dose drops below a threshold; CIAS-Core remains calibrated longer than deep baselines.

- Learned masks can improve PSNR at medium doses, but may become fragile under PSF mismatch unless paired with appropriate CIAS-Core architectures and calibration strata.
- Certain PSF mismatch levels cause classical methods to fail catastrophically while CIAS-Core and CIAS-Core-ELP degrade more gracefully and flag increased uncertainty.

These studies yield regime maps over the SCI design space, highlighting where SCI is trustworthy and where additional dose, improved masks, or different reconstruction strategies are needed.

5.6 CIAS-Discovery Insights

We quantify **change-in-conclusion rates** when moving from PSNR-only rankings to risk-aware rankings that include coverage and CI width. In many cases, methods that appear competitive by PSNR are dominated once uncertainty is considered: they either under-cover (dangerously over-confident) or require much wider intervals to reach target coverage.

We provide qualitative examples where reconstructed videos look similar by eye, but CIAS-Core's uncertainty maps diverge sharply from baseline confidence estimates, illustrating how CIAS-Discovery prevents overconfident interpretations of SCI reconstructions.

5.7 AI-Designed vs Hand-Designed CIAS-Core and CIAS-Core-ELP

To isolate the effect of the CIAS-X AI scientist, we compare the AI-designed CIAS-Core and CIAS-Core-ELP architectures to hand-crafted unrolled SCI networks of similar capacity. We consider three manual baselines:

1. A K-stage unrolled network with fixed step sizes, a shallow U-Net proximal, and losses tuned once on a medium-dose, $T = 8$setting.
2. An ISTA-Net-style unrolled architecture adapted from prior SCI work, with human-selected hyperparameters.
3. The original ELP-Unfolding configuration as described in its reference paper.

All baselines use the same ForwardModel_SCI and training data as CIAS-Core. At matched parameter counts and latency, the AI-designed CIAS-Core and CIAS-Core-ELP achieve:

- +0.3–0.8 dB higher PSNR across several dose/compression combinations,
- 8–15 point higher empirical coverage@95, especially in high-compression, low-dose regimes,
- similar or slightly narrower CIs at equal coverage levels.

These results suggest that the CIAS-X AI scientist is not merely a hyperparameter tuner but a useful tool for discovering robust physics-aware architectures, including improved ELP variants, that human designers might overlook.

6. Discussion and Limitations

Our results show that CIAS-Core, when designed by an AI scientist and embedded into CIAS-Lab and CIAS-Discovery, can produce SCI reconstructions that are both accurate and well-calibrated across a wide range of dose and design regimes. The AI scientist loop systematically explores architecture and calibration choices, yielding Pareto-optimal designs that balance PSNR, coverage, and latency. The CIAS-Core-ELP family demonstrates that existing state-of-the-art unfolding methods can be naturally incorporated into this framework and further improved.

However, several limitations remain:

- **Simulator reliance.** Much of our evaluation is performed in simulation or on simulated measurements of natural videos. While we include small real SCI datasets for qualitative validation, broader real-world testing is needed to fully validate calibration and robustness.
- **Design space scope.** We focus on a particular family of masks, PSFs, and unrolled architectures (including ELP-style designs). Extending CIAS-X to alternative hardware designs (e.g., color-coded masks, different modulation schemes) and richer priors (e.g., full diffusion models) is a natural next step.
- **Conformal assumptions.** Conformal prediction assumes exchangeability within each stratum. In practice, temporal correlations and unmodeled domain shifts may violate this assumption. We partially mitigate this with stratification and stress tests, but more sophisticated theoretical and empirical analysis is warranted.
- **AI scientist complexity.** Our CIAS-Planner currently proposes configurations at a relatively coarse granularity (e.g., integer K, discrete proximal choices, a small set of CIAS-Core-ELP patterns). More fine-grained co-design, including direct hardware changes, could further improve frontiers but requires tighter integration with lab constraints.

7. Conclusion

We presented **CIAS-Core for SCI**, a physics-guided unrolled reconstruction module with domain-conditional conformal calibration, situated inside a **CIAS-X AI scientist** for snapshot compressive imaging. By combining ForwardModel_SCI, unrolled reconstruction, conformal uncertainty, CIAS-Lab campaigns, and CIAS-Discovery analytics, we obtain dose–accuracy–

uncertainty frontiers that map where SCI is trustworthy and where it is not.

Within this framework, we introduced **CIAS-Core-ELP**, an instantiation of CIAS-Core based on Ensemble Learning Priors (ELP)-Unfolding, and showed how CIAS-X can treat it as a design family rather than a single fixed architecture. The CIAS-X AI scientist maintains a world model of SCI experiments, proposes new CIAS-Core and CIAS-Core-ELP architectures and design settings, and selects Pareto-optimal models, yielding AI-designed unrolled architectures that outperform hand-crafted baselines in both accuracy and calibration. Beyond SCI, our work illustrates how AI scientists, aligned with explicit forward models and experiment logs, can help computational imaging progress from ad hoc design to systematic, reproducible exploration of sensing and reconstruction spaces.