PROJECT PROPOSAL

CURE

Calibrated Uncertainty in Restoration via Spectral Estimation

Bridging Classical Signal Processing and Diffusion Models for Trustworthy Image Restoration

Advanced Statistical Image Processing
Semester Project Proposal

Table of Contents

1. Executive Summary

Diffusion models have become the state-of-the-art for image restoration, capable of generating diverse, plausible reconstructions from degraded images. However, a critical question remains unanswered: **are the uncertainty estimates from diffusion posterior sampling actually calibrated?** That is, if the model expresses 90% confidence, is it correct 90% of the time?

This project proposes **CURE (Calibrated Uncertainty in Restoration via Spectral Estimation)**, a framework that uses classical signal processing theory—specifically, the closed-form uncertainty from Wiener filtering—to diagnose, understand, and correct the calibration failures of diffusion-based restoration.

Key Innovation: Unlike prior work that focuses on reconstruction quality (PSNR, LPIPS), CURE focuses on calibration quality (ECE, coverage). We use classical Bayesian restoration as a provably-correct reference to evaluate and improve learned methods.

2. Problem Statement & Motivation

2.1 The Core Problem

Image restoration from degraded observations (blurry, noisy, incomplete) is fundamentally ill-posed. The mathematically correct output is not a single image, but a posterior distribution over plausible reconstructions:

$$p(x\_clean \mid y\_degraded)$$

Diffusion models can sample from this posterior, producing multiple diverse reconstructions. From these samples, we can estimate uncertainty (e.g., pixel-wise variance). But here's the problem:

The uncertainty estimates are not calibrated. A model claiming 90% confidence may be correct only 60% of the time, or 99% of the time. Without calibration, uncertainty estimates are meaningless for decision-making.

2.2 Why Calibration Matters

- **Medical imaging:** Radiologists need to know which regions of a reconstruction are reliable
- **Scientific imaging:** Astronomers need valid error bars on deconvolved telescope images
- **Autonomous systems:** Robots must know when their perception is unreliable
- **Human-AI collaboration:** Users need trustworthy confidence to make informed decisions

2.3 The Opportunity

Classical Bayesian restoration (Wiener filtering) provides **closed-form, provably calibrated uncertainty** under known degradation models. This 50-year-old theory has been largely ignored by the deep learning community, but it offers:

1. **Ground-truth calibration targets** for evaluating learned methods
2. **Frequency-domain insights** about where uncertainty should be high vs. low
3. **Theoretical tools** for understanding why diffusion posteriors fail

3. Research Questions

Primary Research Question

Can classical signal processing theory diagnose, explain, and correct the calibration failures of diffusion posterior sampling for image restoration?

Sub-Questions

RQ1 (Characterization): How are diffusion posteriors miscalibrated as a function of frequency band, degradation type, and severity?

RQ2 (Diagnosis): Does the degradation operator's frequency response $|H(f)|$ predict which frequencies will be miscalibrated?

RQ3 (Correction): Can we use Wiener filter uncertainty as a reference to recalibrate diffusion posteriors?

RQ4 (Trade-offs): What is the relationship between calibration quality and reconstruction quality (PSNR, LPIPS)?

4. Technical Background

4.1 Classical Bayesian Restoration

Consider the linear degradation model:

$$y = Hx + n$$

where y is observed, x is clean, H is degradation (blur, downsampling), and $n \sim N(0, \sigma^2 I)$ is noise.

With Gaussian prior $x \sim N(0, C_x)$, the posterior is also Gaussian with **closed-form mean and covariance**:

$$\mu\_post = C_x H^T (H C_x H^T + \sigma^2 I)^{-1} y$$
$$\Sigma\_post = C_x - C_x H^T (H C_x H^T + \sigma^2 I)^{-1} H C_x$$

Key insight: In the Fourier domain (for circular convolution), the posterior variance per frequency is:

$$\Sigma(f) = \sigma^2 S_x(f) / [\sigma^2 + |H(f)|^2 S_x(f)]$$

This shows that uncertainty is high where $|H(f)| \approx 0$ (information destroyed) and low where $|H(f)|$ is large (information preserved).

4.2 Diffusion Posterior Sampling

Diffusion models learn a score function $s\_\theta(x\_t, t) \approx \nabla \log p\_t(x\_t)$. For posterior sampling:

$$\nabla \log p(x|y) = \nabla \log p(x) + \nabla \log p(y|x)$$

Methods like DPS approximate this by adding likelihood guidance during the reverse diffusion process. Multiple samples give uncertainty estimates.

4.3 The Calibration Gap

Diffusion posteriors are NOT the true posterior because:

4. Score network is imperfect (training error)
5. Guidance approximation introduces bias
6. Discretization of the SDE/ODE introduces error
7. Learned prior may not match true image distribution

These errors compound in unknown ways, leading to miscalibrated uncertainty.

5. Proposed Methodology: The CURE Framework

CURE consists of four interconnected modules:

Module A	Module B	Module C	Module D
Classical Baselines	Diffusion Sampling	Spectral Calibration Analysis	CURE Correction

5.1 Module A: Classical Baselines with Closed-Form Uncertainty

Purpose: Establish provably-correct calibration references.

Implementation:

8. Wiener filter in Fourier domain with exact posterior variance computation
9. Tikhonov regularization with uncertainty via inverse Hessian diagonal
10. Gaussian MRF prior with sparse Cholesky for posterior sampling

Outputs: Restored images + pixel-wise variance maps + frequency-resolved variance

5.2 Module B: Diffusion Posterior Sampling

Purpose: Generate posterior samples from diffusion models for uncertainty estimation.

Implementation:

11. Implement DPS (Chung et al., 2023) using pretrained EDM
12. Generate N=50-100 posterior samples per test image
13. Compute empirical variance per pixel and per frequency band

Outputs: Sample sets + uncertainty maps + frequency-resolved variance estimates

5.3 Module C: Spectral Calibration Analysis (Novel Contribution)

Purpose: Characterize miscalibration in the frequency domain.

This is novel! No prior work computes calibration metrics per frequency band or correlates with degradation frequency response.

Implementation:

14. Partition Fourier space into K radial frequency bands
15. For each band, compute Expected Calibration Error (ECE)
16. Compute interval coverage: Does 90% CI contain truth 90% of the time?
17. Correlate miscalibration with $|H(f)|$ (degradation frequency response)

Hypothesis: Miscalibration will be worst at frequencies where $|H(f)| \approx 0$ (nulls of blur kernel), because the model must 'hallucinate' without measurement support.

5.4 Module D: CURE Calibration Correction (Novel Contribution)

Purpose: Improve calibration using classical uncertainty as reference.

Approach 1 - Post-Hoc Spectral Recalibration:

Learn a mapping $\sigma^2\_cal(f) = g(\sigma^2\_diff(f), |H(f)|, \sigma\_n)$ trained to match classical variance on a calibration set.

Approach 2 - Wiener-Guided Sampling:

Add soft constraint during diffusion: penalize posteriors deviating from Wiener uncertainty in low-SNR bands.

Approach 3 - Hybrid Score Combination:

Frequency-dependent mixing of diffusion score and classical Gaussian posterior score.

6. Implementation Roadmap

Phase 1: Foundations (Weeks 1-4)

Weeks 1-2: Classical Baselines

Task	Details
Day 1-3	Set up environment: PyTorch, SciPy, matplotlib. Download FFHQ 256x256 (subset of 1000 images for development).
Day 4-7	Implement Wiener filter in Fourier domain. Verify correctness on synthetic data where ground truth is known.
Day 8-10	Implement posterior variance computation. Visualize variance maps for different blur kernels.
Day 11-14	Implement degradation operators: Gaussian blur (σ=1,2,3,4), motion blur (length 11,21,31), additive Gaussian noise (σ=0.01,0.05,0.1).

Deliverable: wiener.py with WienerFilter class that returns (restored_image, variance_map, frequency_variance)

Validation: On synthetic Gaussian data, verify posterior variance matches analytical formula exactly.

Weeks 3-4: Diffusion Posterior Sampling

Task	Details
Day 1-3	Download pretrained EDM model (FFHQ 256x256). Set up diffusers or guided-diffusion codebase.
Day 4-7	Implement DPS guidance. Start with reference implementation from DPS paper GitHub.
Day 8-10	Verify DPS works on simple denoising task. Debug any numerical issues.
Day 11-14	Implement multi-sample generation (N=50). Compute empirical mean and variance from samples.

Deliverable: dps_sampler.py that generates N samples and returns (samples, mean, variance_map)

Validation: Qualitative check that samples are diverse but consistent with degraded input.

Phase 2: Calibration Analysis (Weeks 5-8)

Weeks 5-6: Calibration Metrics Implementation

Task	Details
Day 1-4	Implement reliability diagram computation. Bin predictions by confidence, compute accuracy per bin.

Day 5-7	Implement ECE (Expected Calibration Error) for regression: bin by predicted variance, compute actual MSE.
Day 8-10	Implement interval coverage: compute 90% CI from samples, measure how often truth is inside.
Day 11-14	Implement FREQUENCY-RESOLVED versions: partition Fourier space into 8-16 radial bands, compute metrics per band.

Deliverable: calibration.py with compute_ece(), compute_coverage(), compute_reliability_diagram(), all with per-frequency options

Weeks 7-8: Calibration Analysis Experiments

Task	Details
Day 1-4	Run classical methods on test set (100 images). Verify calibration is near-perfect (as theory predicts).
Day 5-8	Run DPS on same test set. Generate 50 samples per image. Compute all calibration metrics.
Day 9-11	Create visualizations: frequency-resolved ECE plots, reliability diagrams, coverage vs frequency.
Day 12-14	Correlate miscalibration with $\|H(f)\|$. Plot ECE(f) vs $\|H(f)\|$ to test hypothesis.

Deliverable: Analysis notebook with all plots + written findings about where/why diffusion fails

Phase 3: CURE Method Development (Weeks 9-12)

Weeks 9-10: Post-Hoc Recalibration

Task	Details
Day 1-4	Split data into train/val/test for recalibration. Design input features: (σ^2_diff(f), |H(f)|, σ_n, f).
Day 5-8	Train small MLP to predict calibrated variance from diffusion variance. Use Wiener variance as target.
Day 9-11	Evaluate on held-out test set. Measure improvement in ECE, coverage.
Day 12-14	Ablate: which input features matter most? How does generalization across degradation types work?

Deliverable: recalibration.py with trained model + evaluation results

Weeks 11-12: Wiener-Guided Sampling (Advanced)

Task	Details
Day 1-4	Design regularizer R(x) that penalizes deviation from Wiener uncertainty in low-SNR bands.
Day 5-8	Integrate regularizer into DPS guidance. Tune hyperparameters (regularization strength, which bands).
Day 9-11	Evaluate: does guided sampling improve calibration? What's the cost in PSNR/LPIPS?
Day 12-14	Compare all methods: Classical, DPS, DPS+Recalibration, DPS+Wiener-Guided.

Deliverable: cure_sampler.py with Wiener-guided DPS + full comparison tables

Phase 4: Final Experiments & Writeup (Weeks 13-16)

Weeks 13-14: Comprehensive Experiments

- Run all methods on full test sets (FFHQ, DIV2K, BSD68)
- Test all degradation types (Gaussian blur, motion blur, downsampling, inpainting)
- Ablation studies: number of samples, guidance strength, frequency band granularity
- Create publication-quality figures and tables

Weeks 15-16: Final Report & Presentation

- Write final report (aim for workshop paper length: 8 pages)
- Prepare slides and demo
- Polish code repository with documentation
- (Optional) Submit to workshop (CVPR/ICCV workshops, NeurIPS workshop)

7. Required Reading List

7.1 Classical Signal Processing (Read First)

These provide the theoretical foundation. Read before Week 3.

#	Paper	What to Learn
1	Kaipio & Somersalo, 'Statistical and Computational Inverse Problems' (2005), Ch. 3-4 https://iucat.iu.edu/catalog/6150292	Bayesian formulation of inverse problems, posterior derivation
2	Andrews & Hunt, 'Digital Image Restoration' (1977), Ch. 6-7	Wiener filter derivation, frequency domain analysis
3	Gonzalez & Woods, 'Digital Image Processing' (2018), Ch. 5	Degradation models, blur PSF, noise models

7.2 Diffusion Models for Inverse Problems (Core)

Essential papers for understanding the diffusion posterior sampling approach. Read before Week 5.

#	Paper	What to Learn
4	Chung et al., 'Diffusion Posterior Sampling for General Noisy Inverse Problems' ICLR 2023	DPS algorithm, likelihood guidance, implementation details
5	Song et al., 'Pseudoinverse-Guided Diffusion Models for Inverse Problems' ICLR 2023	ΠGDM, pseudoinverse approximation, comparison to DPS
6	Kawar et al., 'Denoising Diffusion Restoration Models' NeurIPS 2022	DDRM, SVD-based approach, linear inverse problems
7	Karras et al., 'Elucidating the Design Space of Diffusion-Based Generative Models' NeurIPS 2022	EDM framework, pretrained models we'll use

7.3 Uncertainty and Calibration (Critical)

Papers on calibration metrics and uncertainty quantification. Read before Week 6.

#	Paper	What to Learn
8	Guo et al., 'On Calibration of Modern Neural Networks' ICML 2017	ECE definition, reliability diagrams, temperature scaling

9	Kuleshov et al., 'Accurate Uncertainties for Deep Learning Using Calibrated Regression' ICML 2018	Calibration for regression (not classification), recalibration methods
10	Angelopoulos & Bates, 'A Gentle Introduction to Conformal Prediction' 2022	Conformal prediction, coverage guarantees (alternative approach)

7.4 Related Work to Cite and Differentiate From

CRITICAL: You must cite these and clearly explain how your work differs.

#	Paper	Relationship to CURE
11	Thaker et al., 'Frequency-Guided Posterior Sampling' ICCV 2025	MUST CITE. Similar name but different goal: they optimize reconstruction quality, we optimize calibration.
12	Kou et al., 'BayesDiff: Estimating Pixel-wise Uncertainty' ICLR 2024	They estimate uncertainty via Laplace approx; we analyze calibration and use classical reference.
13	Belhasin et al., 'Principal Uncertainty Quantification' CVPR 2024	They use conformal prediction for coverage; we use frequency analysis + classical calibration.
14	Scopacasa & Villa, 'Can Diffusion Models Provide Rigorous UQ?' 2025	They benchmark on synthetic posteriors; we focus on real images + propose correction method.
15	Xiao et al., 'Frequency-Aware Guidance for Blind IR' ECCV 2024	They use wavelet guidance for blind restoration; we focus on calibration, not blind setting.

8. Plagiarism Avoidance Guide

This section is critical. Follow these guidelines carefully.

8.1 Code Attribution

What You Can Use	What You Must Do
DPS official implementation (github.com/DPS2022/diffusion-posterior-sampling)	Add comment: '# Based on DPS implementation, modified for CURE'
EDM pretrained models from Karras et al.	Cite paper in report; add 'Pretrained model from [Karras 2022]' in code
Standard libraries (scipy.fft, torch, numpy)	No citation needed for standard functions
Calibration metrics code from uncertainty-toolbox	Cite package; note if you modified it

8.2 Writing Attribution

Equations and Derivations:

- Wiener filter formula: Cite Kaipio & Somersalo or Andrews & Hunt
- DPS guidance equation: Cite Chung et al. 2023
- ECE definition: Cite Guo et al. 2017

Claims and Observations:

- If FGPS paper says 'diffusion generates images hierarchically in frequency domain': cite them
- If your experiments show 'diffusion is miscalibrated at blur nulls': this is YOUR finding, no citation needed
- If BIPSDA says 'diffusion posteriors are not rigorous': cite them when making similar observation

8.3 What Makes CURE Novel (Claim These)

These are your original contributions. You do not need to cite others for these ideas:

18. **Using Wiener filter closed-form covariance as a calibration reference.** No prior work uses classical uncertainty as a calibration target for diffusion.
19. **Computing calibration metrics (ECE, coverage) per frequency band.** Existing work computes calibration globally, not spectrally.
20. **Correlating miscalibration with degradation frequency response |H(f)|.** This diagnosis framework is new.
21. **Post-hoc recalibration conditioned on frequency and degradation.** Using classical theory to inform recalibration is novel.

8.4 Differentiation Statements (Use in Paper)

Include statements like these in your Related Work section:

"Thaker et al. [FGPS] propose frequency-guided sampling to improve reconstruction quality. In contrast, CURE focuses on calibration quality, using frequency analysis to diagnose and correct uncertainty estimates rather than to improve PSNR."

"BayesDiff [Kou et al.] estimates pixel-wise uncertainty via Laplace approximation. CURE complements this by analyzing whether such estimates are calibrated, and proposes corrections using classical Wiener filter uncertainty as a reference."

"While prior work on diffusion restoration [DPS, ΠGDM, DDRM] evaluates reconstruction quality (PSNR, LPIPS), we focus on the overlooked problem of calibration, measuring whether uncertainty estimates are statistically meaningful."

9. Expected Contributions

9.1 Intellectual Contributions

Contribution 1 - Spectral Calibration Diagnostic Framework: A methodology for analyzing diffusion posterior calibration in the frequency domain. This could become a standard diagnostic tool for any diffusion-based inverse problem solver.

Contribution 2 - Empirical Characterization of Miscalibration: First systematic study showing where and why diffusion posteriors fail to be calibrated, with concrete evidence linking miscalibration to degradation structure.

Contribution 3 - CURE Algorithm: A calibration correction method using classical uncertainty as a physics-informed constraint.

9.2 Practical Deliverables

- Open-source code repository with all methods implemented
- Pretrained recalibration models for common degradation types
- Comprehensive benchmark comparing classical vs. diffusion calibration
- Tutorial notebook explaining frequency-resolved calibration analysis

9.3 Interview Talking Points

Use these to describe your project:

- "I showed that diffusion models hallucinate in predictable frequency bands—exactly where classical signal processing theory says information is lost."
- "I used 50-year-old Wiener filter theory to diagnose and fix a limitation of 2023-era generative models."
- "I developed a calibration framework that measures whether a model knows what it doesn't know—critical for deploying AI in medicine or science."
- "Most ML research optimizes accuracy; I focused on trustworthiness, which is what matters for real-world deployment."

10. Risk Assessment

Risk	Prob	Impact	Mitigation
DPS implementation issues	Medium	High	Start from official repo; test on simple cases first; have DDRM as backup
CURE doesn't improve calibration	Medium	High	Even negative result is valuable; pivot to analysis-only contribution if needed

Compute constraints	Low	Medium	Use smaller resolution (64x64) for development; scale up only for final experiments
Scope creep	High	Medium	Stick to timeline; cut optional experiments; focus on core story
Similar work published	Low	High	Monitor arXiv weekly; differentiate by emphasizing classical + spectral angle

— End of Proposal —