Estimation of Model Parameters through Machine Learning

Anushka Polapally

June 2025

Abstract

This paper investigates how machine learning can solve inverse problems in the context of image classification, focusing on reconstructing input features from classification outputs. This project explored computational strategies that merge the mathematical foundations of inverse problems with the practical methods of computer vision. By combining partial differential equation (PDE) modeling and optimization-based deep learning, the work demonstrates how model parameters and image features can be estimated efficiently and stably, even in the presence of noise.

1 Notation

Functions can take multiple inputs; for example, if we have a vector $x \in \mathbb{R}^n$, $x = (x_1, x_2, \ldots, x_n)$. For a function $f(x)$, $f : \mathbb{R}^n \to \mathbb{R}$, partial derivatives are written as $\partial_i f(x) = \frac{\partial f(x)}{\partial x_i}$.

$$f(x, y) = 3x + 2y, \quad \partial_x f(x, y) = 3, \quad \partial_y f(x, y) = 2$$

For multivariate functions $f : \mathbb{R}^n \to \mathbb{R}$, the gradient, divergence, and Laplacian are defined as

$$\nabla f(x) = (\partial_1 f(x), \ldots, \partial_n f(x)) \in \mathbb{R}^n,$$

$$\nabla \cdot v(x) = \sum_{i=1}^{n} \partial_i v_i(x), \quad -\Delta f(x) = \sum_{i=1}^{n} \partial_i^2 f(x).$$

The Laplacian represents the sum of second derivatives of f with respect to spatial coordinates.

2 The Forward Problem: Heat Equation

The heat equation describes the temporal evolution of temperature (or concentration) in a medium:

$$\partial_t u - \alpha \Delta u = 0 \quad \text{on } \Omega \times [0, 1], \tag{1}$$

with initial condition $u(x,0) = u_0(x)$ and thermal diffusivity $\alpha > 0$. Solving the *forward problem* means determining $u(x,t)$ given α and u_0.

Discretizing $\partial_t u$ and Δu using finite differences yields a numerical system:

$$\frac{u^{k+1} - u^k}{h_t} - \alpha L u^k = 0,$$

where L is the discrete Laplacian matrix:

$$L = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}.$$

This can be advanced in time using an explicit or implicit scheme such as Euler's or Runge–Kutta (RK2/Heun's) methods.

3 The Inverse Problem

Inverse problems seek unknown parameters or inputs that produce a desired output under a known forward model. For the heat equation, the goal is to recover α or u_0 from observed final states $u(x,1)$:

$$\partial_t u - \alpha \Delta u = 0, \quad u(x,0) = u_0, \quad u(x,1) = u_{\text{obs}}.$$

We define the forward operator

$$g(\alpha) = u_\alpha(x,1),$$

and the measured data as

$$u_{\text{obs}} = g(\alpha) + \eta(x),$$

where η represents additive Gaussian noise. Recovering α from u_{obs} constitutes an ill-posed problem—small perturbations in u_{obs} can lead to large variations in α. Regularization and data-driven models are therefore required to ensure stability.

4 Machine Learning Formulation

Machine learning provides a data-driven way to approximate the inverse mapping g^{-1}. Given training pairs (u_{obs}, α), we can learn a model

$$\hat{\alpha} = \mathcal{N}_\theta(u_{\text{obs}}),$$

where $\mathcal{N}_\theta$ denotes a neural network parameterized by θ. The training objective minimizes

$$\mathcal{L}(\theta) = \|\mathcal{N}_\theta(u_{\text{obs}}) - \alpha\|_2^2.$$

For problems without direct supervision, a hybrid physics-informed loss combines data fidelity with the PDE residual:

$$\mathcal{L}(\theta) = \|u_{\text{obs}} - g(\mathcal{N}_\theta(u_{\text{obs}}))\|_2^2 + \lambda \|\partial_t u - \mathcal{N}_\theta(u_{\text{obs}})\Delta u\|_2^2.$$

This structure constrains the learned inverse mapping to respect the physical model while leveraging data for regularization.

5 Inverse Problems in Image Classification

The inverse problem framework extends naturally to computer vision. Let $F : \mathbb{R}^n \to \Delta^{K-1}$ be a classifier mapping image x to class probabilities $p = F(x)$. Given a target class or feature representation $\hat{p}$, the goal is to reconstruct a plausible image $\tilde{x}$ such that $F(\tilde{x}) \approx \hat{p}$:

$$\tilde{x} = \arg\min_{x \in [0,1]^n} \mathcal{L}_{\text{task}}(F(x), \hat{p}) + \lambda \mathcal{R}(x), \tag{2}$$

where $\mathcal{R}(x)$ serves as a regularization term. Common choices include:

- ℓ_2 priors: penalize large pixel magnitudes,
- Total Variation (TV): promote spatial smoothness,
- Learned priors: restrict reconstructions to a generative model's manifold.

Optimization-based inversion can be solved via gradient descent or through latent-space optimization in a pretrained generator $G(z)$. In the latter case, one seeks

$$z^\star = \arg\min_z \mathcal{L}_{\text{task}}(F(G(z)), \hat{p}) + \beta \|z\|^2, \quad \tilde{x} = G(z^\star).$$

This approach mirrors classical inverse problem regularization, where G implicitly encodes prior knowledge of valid images.

6 Experiments and Findings

Experiments were performed on MNIST and CIFAR-10 using convolutional neural networks. The inverse problem was tested under three conditions:

1. Reconstruction from target class probabilities,
2. Reconstruction from intermediate feature maps,
3. Parameter estimation under noisy measurements.

Results showed that:

- Simple pixel-based optimization reconstructs recognizable patterns but is highly sensitive to noise.

- TV regularization improves stability and edge definition, analogous to diffusion-based smoothing.
- Generative priors, such as VAEs or GANs, yield the most realistic reconstructions, preserving semantic structure.
- A hybrid approach combining PDE regularization with learned priors provides the best balance between fidelity and stability.

7 Discussion

The experiments highlight a fundamental parallel between physical inverse problems and feature reconstruction in neural networks. Both domains rely on regularization to constrain ill-posed mappings. The diffusion term $\alpha \Delta u$ in the heat equation acts analogously to the TV prior in image reconstruction, smoothing solutions while preserving important structure.

Moreover, integrating physics-informed loss terms into neural architectures offers a pathway toward interpretable, stable models that generalize well. This approach bridges classical numerical analysis with modern data-driven learning.

8 Conclusion

This research explored how machine learning can address inverse problems for image classification and PDE-based systems. By uniting forward modeling, optimization, and neural representation learning, we demonstrated that input features or model parameters can be reconstructed from partial or noisy outputs. The work establishes a foundation for using deep learning to solve inverse problems across physics and vision, emphasizing interpretability, stability, and data efficiency.

Acknowledgments

This project was completed under the mentorship of Professor Andreas Mang, whose expertise in computational mathematics and inverse problems provided essential guidance throughout the research.