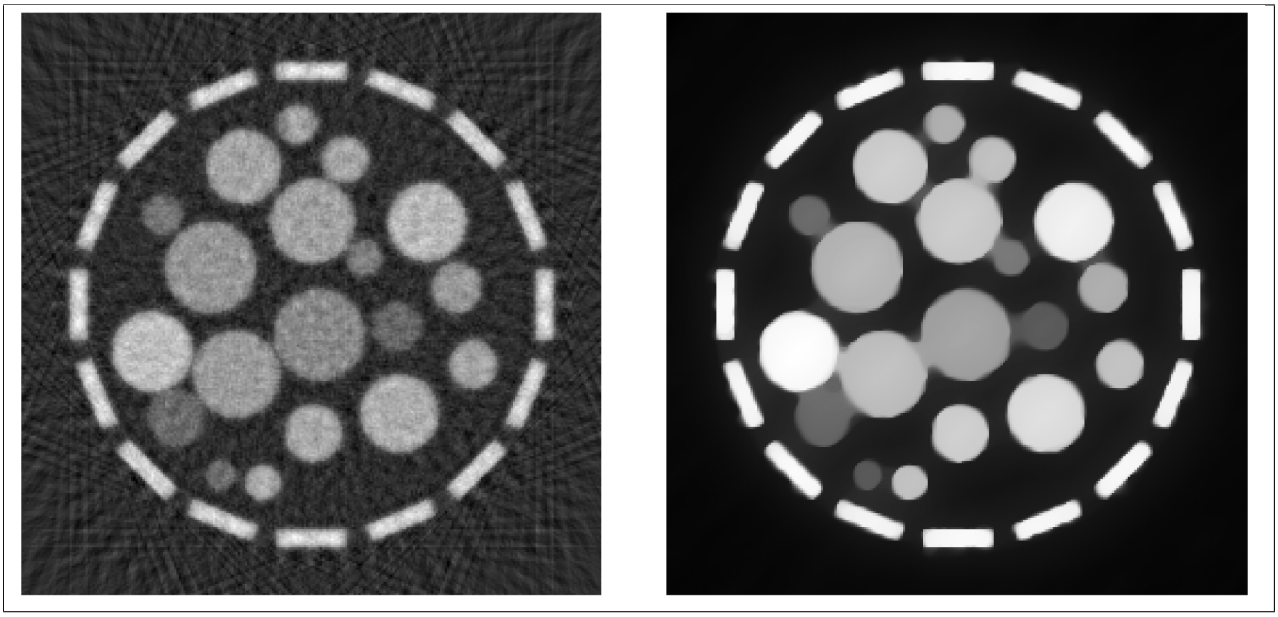
Universiteit Utrecht

Faculteit Bètawetenschappen

Standardizing CT Reconstructions via Learned Mappings Between Iterative Methods

Course: Inverse Problems in Imaging
Master Mathematical Sciences, Utrecht University

Gijs Bartholomeus, Jochem Lange, Pepijn Vermeulen



Example CT slice comparison: Tikhonov vs. TV-regularized reconstruction

Supervisors:

Prof. Dr. Tristan van Leeuwen
Utrecht University

Prof. Dr. Pim de Jong
University Medical Center Utrecht

May 2025

Abstract

In modern clinical practice, medical images are routinely shared across institutions for diagnostic, research, and machine learning purposes. However, differences in CT scanner hardware, acquisition settings, and reconstruction algorithms can lead to significant variations in image appearance, even for identical patients. These inconsistencies complicate visual interpretation and limit the generalizability of automated analysis methods. This is especially problematic for diseases where subtle image features—such as texture, noise levels, or edge sharpness—can influence clinical decisions, as in interstitial lung disease (ILD) or early fibrosis detection [1, 2].

In this work, we construct a controlled toy model that isolates the effect of differing reconstruction algorithms. Using a fixed synthetic sinogram with Poisson noise, we reconstruct two images using distinct variational methods: one with Tikhonov (quadratic) regularization and one with Total Variation (TV) regularization tailored to Poisson noise. We propose a transformation that maps an image from one reconstruction method onto another, effectively harmonizing outputs without access to raw projection data. The transformation is obtained by solving a least squares problem, regularized by a classification CNN that steers the harmonization process. Experiments on a toy dataset show promising results.

Contents

1 Introduction

Medical imaging plays a central role in diagnostics, treatment planning, and longitudinal monitoring. To facilitate collaboration, hospitals and research centers often share CT data across institutions. However, this data is typically shared in its reconstructed form—as 3D intensity maps—rather than as raw projection (sinogram) data. While convenient, this introduces a critical limitation: different scanners, manufacturers, and reconstruction algorithms (e.g., filtered backprojection, statistical iterative methods) can produce notably different image appearances from the same underlying anatomy [1, 3].

These discrepancies are not merely cosmetic. In diseases such as interstitial lung disease (ILD), early-stage fibrosis, or pulmonary nodules, subtle differences in texture, noise, or edge sharpness can influence both human diagnosis and algorithmic performance [2, 4]. As a result, comparing patients across institutions—or training robust machine learning models—requires some form of image harmonization or standardization.

Computed Tomography (CT) is a non-invasive imaging technique that reconstructs cross-sectional images of internal anatomy by measuring the attenuation of X-rays through the body. In a typical 2D model, a fan-beam of X-rays passes through the patient at multiple angles. At each angle, detectors record the cumulative attenuation along straight paths—these measurements are known as line integrals. The full set of measurements, taken across many angles, forms the so-called *sinogram*.

Mathematically, this is described by the Radon transform, which maps an attenuation function $u(x, y)$ to its integrals along lines. Discretizing both the image and measurements, we write:

$$u \in \mathbb{R}^{N_x \times N_y}, \quad f \in \mathbb{R}^{N_\theta \times N_s}, \quad f = Ru,$$

where R is the discrete Radon operator. Each row of R corresponds to one line integral through the object.

The inverse problem of reconstructing u from f is ill-posed: small measurement errors (e.g., due to noise or sparse angular sampling) can lead to large artifacts in the image. Furthermore, physical effects like beam hardening, scattering, and detector blur can distort the data [5]. To obtain stable reconstructions, modern algorithms regularize the inversion by combining a data-fidelity term with prior knowledge about the image, typically via variational formulations.

In this work, we construct a simplified test case to isolate one specific source of variability: the reconstruction algorithm. We generate synthetic sinogram data with fixed acquisition parameters and Poisson noise, and then reconstruct two different images from the same data:

- A basic Tikhonov-regularized reconstruction (least-squares fidelity and ℓ_2 regularization);
- A more advanced Total Variation (TV) reconstruction using a Poisson noise-informed likelihood.

This setup allows us to study reconstruction-induced variation in isolation. We then train a neural network to learn to distinguish the different types of reconstruction methods. We then use this network in a minimalization procedure that transforms from the Tikhonov-based reconstruction to the TV-based one, effectively harmonizing the image appearance. Our goal is to explore whether such learned mappings could eventually serve as a practical solution for post hoc harmonization of CT images from heterogeneous sources—especially in settings where raw sinogram data is unavailable.

2 Poisson Noise in CT Imaging

In practical CT acquisition, the measurements are corrupted by quantum noise due to the discrete nature of photon detection. The expected (noiseless) sinogram is given by $f = Ru$,

where R is the discrete Radon transform of the attenuation map u. However, the actual measured sinogram f^δ is noisy and follows a Poisson distribution:

$$f_i^\delta \sim \text{Poisson}(f_i) = \text{Poisson}([Ru]_i),$$

where each f_i^δ represents the detected number of X-ray photons along ray i. This model captures the physics of photon counting, in which noise arises from the random arrival of photons.

This Poisson assumption is physically well-justified: X-ray photon detection is a discrete, memoryless counting process, where photons arrive independently at the detector. Under these conditions, and assuming a constant X-ray intensity over a short time interval, the number of detected photons per detector ray follows a Poisson distribution. This makes Poisson noise a natural model for CT, PET, and other photon-limited imaging modalities modalities [6, 7].

Moreover, the Poisson model becomes increasingly accurate in low-dose regimes, where the number of incident photons is small. In such settings, Gaussian approximations (which rely on high-count central limit arguments) break down, while the discrete nature of Poisson statistics remains valid.
Unlike Gaussian noise, which is additive and signal-independent, Poisson noise is *signal-dependent*: the variance of the noise is equal to its mean,

$$\text{Var}(f_i^\delta) = \mathbb{E}[f_i^\delta] = f_i.$$

This means that stronger signals have proportionally higher fluctuations, while low-intensity regions exhibit high relative noise. Importantly, Poisson noise is neither purely additive nor multiplicative. As shown in [8], it is better described as a functional dependency on the signal itself:

$$f^\delta = \delta(f),$$

where δ is a random process whose distribution depends on f. This distinction sets Poisson noise apart from typical additive models like Gaussian or Laplace, and from multiplicative models such as Gamma noise.

3 Discrete Inverse Problem

Computed Tomography (CT) aims to reconstruct a two-dimensional attenuation map $u(x, y)$ from line integrals of X-ray absorption through an object, measured at various angles. Mathematically, this is described by the Radon transform $\mathcal{R}u(\theta, s)$, which maps a function u to its integrals over lines oriented at angle θ and offset s from the origin:

$$\mathcal{R}u(\theta, s) = \int_{-\infty}^{\infty} u(x\cos\theta - t\sin\theta, x\sin\theta + t\cos\theta)\, dt.$$

This models the physics of X-ray projection: each value $\mathcal{R}u(\theta, s)$ represents the total attenuation along a line perpendicular to angle θ and displaced by s. The collection of such measurements across angles and detector positions forms the so-called *sinogram*, named after the sinusoidal curves traced by point sources in the sinogram domain.
In the discrete setting, we denote the Radon transform by a linear operator $R : \mathbb{R}^{N_x N_y} \to \mathbb{R}^{N_\theta N_s}$, where $u \in \mathbb{R}^{N_x \times N_y}$ is discretized into a vector and $f = Ru \in \mathbb{R}^{N_\theta \times N_s}$ is the corresponding discretized sinogram. Each row in f contains the measurements for a fixed angle θ, while each column corresponds to a detector location s.

The goal of image reconstruction is to recover u from f, but since the problem is ill-posed and measurements are corrupted by noise, this is typically posed as a regularized optimization problem:

$$\hat{u} = \arg\min_{u \geq 0} \Big\{ \underbrace{\frac{1}{2}\|Ru - f\|_W^2}_{\text{data-fidelity}} + \underbrace{\beta\, \mathcal{R}(u)}_{\text{regularization}} \Big\}, \tag{3.1}$$

where $W = \text{diag}(w_i)$ is a diagonal weighting matrix, often chosen as $w_i = 1/\sigma_i^2$ to reflect measurement variance. The regularization term $\mathcal{R}(u)$ encodes prior assumptions on the image, such as smoothness or sparsity, and serves to stabilize the inversion.
This variational formulation captures the core structure of many modern iterative reconstruction algorithms used in CT, such as iDose and IMR. For a detailed derivation of the Radon geometry and sinogram formation, we refer to the course notes on CT reconstruction [5].

3.1 Filtered Backprojection

FBP solves the unregularized pseudo-inverse by approximating

$$R^+ \approx R^T H,$$

where H is a Fourier-domain filter applying $|\omega|$ and an apodization window. Discrete sampling yields aliasing: H amplifies high frequencies ($|\omega| > \omega_c$) causing streak artifacts. No noise model means minimization of $\|Ru - f\|_2$ is unaddressed. Filtered Backprojection (FBP) alone fails to produce diagnostically reliable images because it treats the inverse Radon transform as a one-shot analytic operation, ignoring both the discrete sampling and stochastic nature of CT data. In practice, FBP amplifies high-frequency noise via the ramp filter $|\omega|$, leading to streak and grain artifacts when projections are limited or noisy[8]. Moreover, finite angular sampling creates gaps in the Fourier domain, causing aliasing that manifests as structured artifacts in the reconstructed slice. Crucially, FBP omits any statistical weighting or regularization, so all measurements—regardless of their noise level—contribute equally, which degrades image quality under low-dose conditions[6]. Without modeling beam hardening, detector blur, or Poisson photon statistics, FBP cannot correct for these physical effects, resulting in cupping and edge-distortion artifacts that hinder clinical interpretation. To generate diagnostically useful reconstructed images, more complicated methods are necessary.

4 Reconstruction Methods Compared

As outlined in the introduction, differences in reconstruction algorithms can significantly affect CT image appearance, even when based on identical sinogram data. In this section, we examine two representative variational reconstruction methods that differ in both their data fidelity terms and regularization strategies. These serve as the endpoints of our toy problem: the "source" and "target" image domains between which we attempt to learn a harmonizing transformation. We begin with Tikhonov regularization, a classical approach that assumes Gaussian noise and penalizes global image energy. Then, we introduce a more sophisticated method that combines a Poisson noise-informed data term with Total Variation (TV) regularization, which is known for preserving edges and suppressing artifacts in low-dose scenarios. By reconstructing both from the same noisy sinogram, we isolate the algorithmic difference as the sole source of image variation.

4.1 Tikhonov Regularization (Least Squares)

This method uses a quadratic penalty to enforce smoothness:

$$\mathcal{R}(u) = \|Lu\|_2^2,$$

where L is often the identity or a discrete gradient operator. This leads to:

$$\hat{u} = \arg\min_u \frac{1}{2}\|Ru - f\|_2^2 + \beta\|u\|_2^2.$$

This objective is convex and has a closed-form solution (or can be solved efficiently using gradient descent), but oversmooths edges.

4.2 Total Variation + Poisson Likelihood

Accurate modeling of Poisson noise is critical, especially for low-dose CT, where photon counts are low. In such cases, classical least-squares data fidelity fails to capture the correct noise behavior. Accurate modeling of Poisson noise is critical, especially for low-dose CT, where photon counts are low. In such cases, classical least-squares data fidelity fails to capture the correct noise behavior. The commonly used term $\|Ru - f^\delta\|_2^2$ is in fact Gaussian-inspired: it arises as the negative log-likelihood under the assumption that the measurement noise is independent and normally distributed with constant variance, i.e., $f_i^\delta \sim \mathcal{N}([Ru]_i, \sigma^2)$. A full derivation of the least-squares likelihood from a Gaussian model is provided in Appendix A.

While this assumption works well in high-photon or high-SNR settings, it becomes inaccurate when the noise is signal-dependent, as in photon-limited imaging. Instead, Poisson-informed reconstruction algorithms replace this with the negative log-likelihood of the Poisson model:

$$D(u; f^\delta) = \sum_i \Big([Ru]_i - f_i^\delta \log([Ru]_i) \Big),$$

leading to the optimization problem:

$$\min_{u \geq 0} \Big\{ D(u; f^\delta) + \beta \mathcal{R}(u) \Big\}.$$

For the regularization term $\mathcal{R}(u)$, we employ Total Variation (TV) regularization:

$$\mathcal{R}(u) = \|\nabla u\|_1.$$

TV penalizes the total amount of gradient in the image, which promotes piecewise-constant reconstructions and preserves sharp edges. This makes it especially suitable for medical imaging, where important structures—such as tissue boundaries or lesions—often correspond to edges. In contrast, Tikhonov regularization enforces global smoothness via an ℓ_2 penalty and tends to oversmooth these edges, leading to loss of critical anatomical detail.
Combining TV with the Poisson data term, we solve:

$$\hat{u} = \arg\min_{u \geq 0} \left\{ \sum_i ([Ru]_i - f_i \log[Ru]_i) + \beta\|\nabla u\|_1 \right\}.$$

This formulation is robust to noise and retains fine structural information, but is computationally more intensive due to the nonsmooth nature of the TV norm and the non-quadratic data term.

5 Learned Transformations Between Reconstructions

Given two reconstructions u_A (e.g., Tikhonov) and u_B (e.g., TV+Poisson), we aim to learn a mapping:

$$\Phi : u_\mathrm{A} \mapsto u_\mathrm{B}.$$

Such a mapping is inherently ill-posed. The forward Tikhonov reconstruction discards high-frequency content and smooths edges, collapsing multiple distinct ground-truth images into

nearly identical outputs. Small perturbations in u_{A}, whether from noise, discretization, or parameter variations, can correspond to large and non-unique differences in u_{B}. In u_{A}, fine structures are either implicitly represented or entirely lost, while u_{B} may sharpen or reintroduce edges based on slight gradient cues. This means there is no continuous, one-to-one inverse mapping from u_{A} back to u_{B}. Such non-uniqueness and instability violate Hadamard's criteria for well-posedness, making Φ sensitive to input perturbations and reliant on strong learned priors. However, if successful, this learned transformation would enable harmonization between reconstruction methods potentially transforming low-quality outputs into high-fidelity equivalents suitable for clinical use.

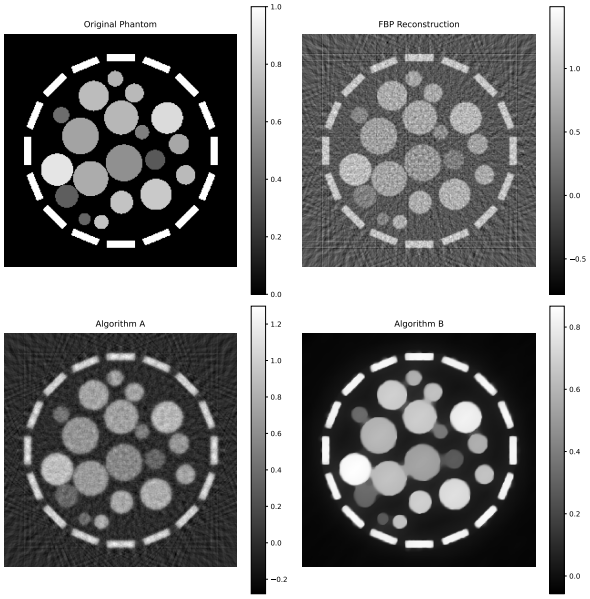


Figure 1: Comparison of FBP, Tikhonov u_{A}, and TV+ Poisson likelihood u_{B} reconstructions on a synthetic phantom with Poisson noise with 1/40 gain. The reconstruction algorithms used 1000 steps of stepsize 10^{-4} and were initialized using the FBP. The discrete Radon transform used 50 angles. Note how both algorithm A and B visually perform better than the FBP, and the clear difference in the handling of noise between A and B.

Figure 1 illustrates the visual differences between FBP, Tikhonov (u_{A}), and TV+Poisson (u_{B}) reconstructions of a synthetic phantom with added Poisson noise. Notably, while both u_{A} and u_{B} improve upon FBP, the difference in noise handling and edge sharpness between A and B is striking.

To achieve this mapping, we first train a convolutional neural network (CNN) to classify images as either u_{A} or u_{B}. This classification capability provides a natural regularizer for a subsequent optimization that transforms one image type into the other. The harmonization procedure minimizes:

$$\min_x \Big\{ \underbrace{\lambda \mathrm{CE}\big(f(x),\, y_{\mathrm{target}}\big)}_{\substack{\text{classification}\\ \text{loss}}} + \underbrace{\|x - x_0\|_2^2}_{\substack{\text{fidelity}\\ \text{loss}}} \Big\},$$

where x denotes the pixel values of the transformed image, x_0 is the original image (e.g., u_{A}), and $f(x)$ is the classification output of the CNN. The cross-entropy loss $\mathrm{CE}(f(x), y_{\mathrm{target}})$ encourages x to be recognized as u_{B}, while the ℓ_2 fidelity term preserves similarity to the input. This approach has the advantage of being invertible: it can harmonize images in both directions between A and B.

The training dynamics of the CNN are shown in Figure 2, where the cross-entropy loss on both training and validation sets converges to near-zero values.

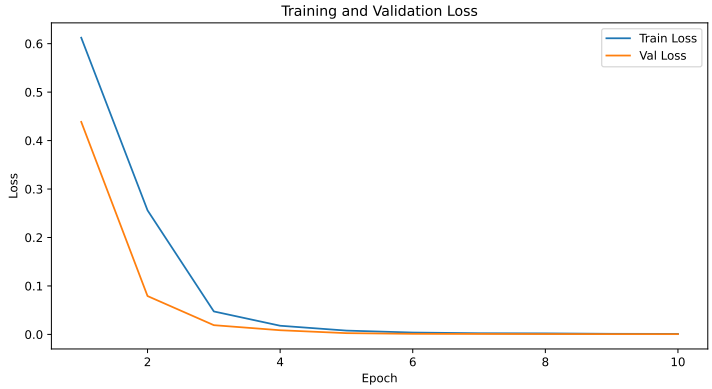


Figure 2: Training and validation loss of the recognizing CNN, given by cross-entropy loss. The algorithm was trained using 10 phantoms, cut and rotated to give 2000 images, of which 1600 were used as training data, and 400 as validation images, which gave a 100.00% accuracy.

6 Results and Discussion

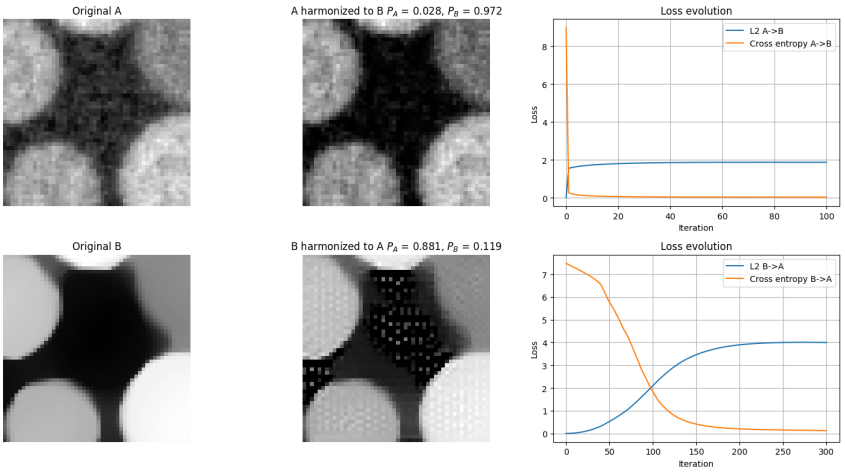


Figure 3: An example of the application of the harmonizing algorithm on a sample patch of our reconstructed phantom.

Implementing the procedure mentioned above starts with reconstructing an image with poisson-noise in three different ways, raw FPB, Tikhonov regularization (algorithm A), and TV + Poisson likelihood (algorithm B). The results are shown in figure 1. It is clear to see that both algorithm A and B show less artifacts and than the FBP, but differ greatly from each other. Algorithm B removes the noise successfully, but blurs some sharp boundaries. The overlapping circles seem to bleed into each other, while algorithm A struggles with the noise, as expected. The rest of the project entails the anylisis and manipulation of these kinds of images.

To train the CNN that recognizes the reconstruction method, we cut the reconstructed images up in many ways to generate a set of training data. Using the cross-entropy loss, the CNN learned to tell the methods apart with a 100.00% succes rate 2

With the recognition CNN in place, we can now implement the minimization procedure previously outlined in Equation **??**. The results are shown in Figure 3. Several key features are immediately noticeable: transforming from A (Tikhonov) to B (TV+Poisson) reduces noise in the background (the dark regions), reflecting a characteristic property of the Poisson+TV

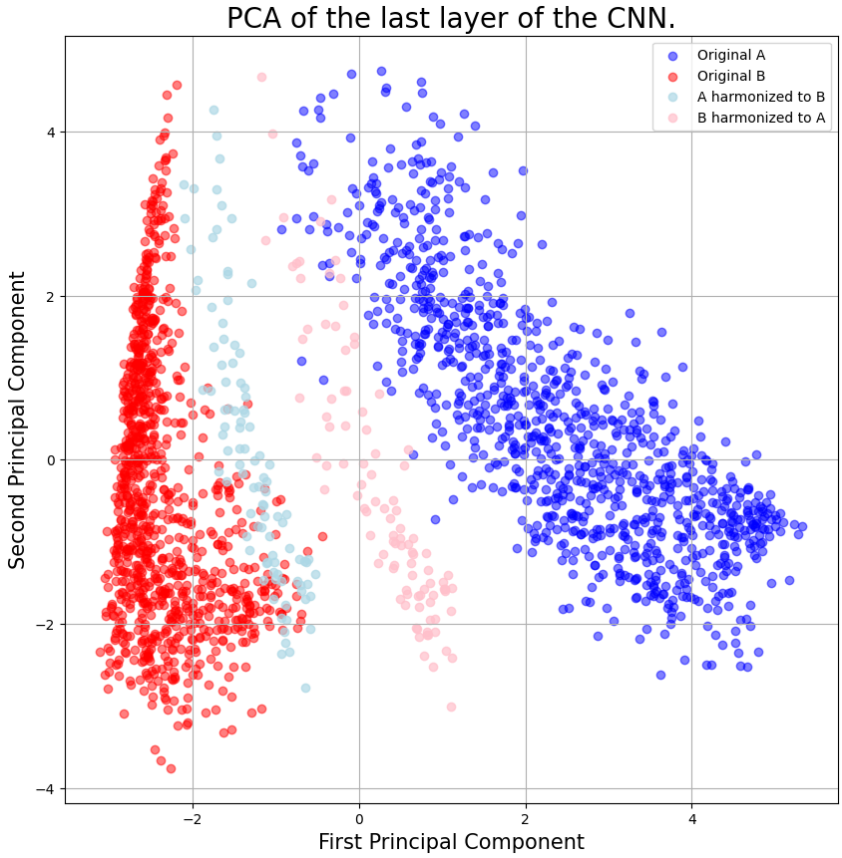


Figure 4: PCA of the last layer of the CNN with explained variance ratio [0.17 0.072].

reconstruction algorithm. However, the noise on the circular structures is not completely removed. Conversely, when transforming a type B image to type A, the algorithm appears to add noise, as both the background and the circular structures acquire additional specks. Notably, this artificial noise is visually quite different from the Poisson noise observed in actual Tikhonov reconstructions, suggesting limitations in the model's capacity to replicate authentic noise patterns.

The graphs on the right of Figure 3 illustrate the interplay between the fidelity and classification losses during optimization. The equilibrium reached is governed by the chosen value of λ, which we fixed at $\lambda = 10$ based on visual assessment of the outputs. Interestingly, the time required to reach a loss minimum varies markedly between directions: transforming $A \to B$ converges roughly an order of magnitude faster than $B \to A$. The harmonization procedure successfully alters the classification of the images to the target type, achieving a confidence of 0.97 for $A \to B$ and 0.81 for $B \to A$. However, the resulting images do not convincingly resemble genuine reconstructions of the target class. This suggests that the space of possible harmonized images is too broad and does not align with the true distribution of images produced by the desired reconstruction method.

A clear illustration of this issue emerges in the principal component analysis (PCA) of the CNN's final layer, shown in Figure 4. Here, we observe a neat separation between the A and B classes, with the harmonized images appearing as outliers. This indicates that while the harmonization process successfully shifts an image just beyond the classification boundary, it fails to refine the image to genuinely resemble the target class. Since the classification confidence of the harmonized images is already high, the regularization term in Equation **??** becomes negligible, leaving no further "push" to adjust the image. The PCA visualization vividly demonstrates that classification alone is an insufficient regularizer, and that a more nuanced approach is needed to

achieve convincing harmonization.
A key challenge we now face is that the harmonized images do not originate from the same underlying data distribution as the true images of either class, as evidenced by Figure 4. One promising way to address this could be to incorporate a variational autoencoder (VAE) framework [9]. By learning a latent space that captures the natural variability of true images from both reconstruction methods, we can constrain the harmonized outputs to resemble genuine images from the target reconstruction method. Specifically, we would expect to see two distinct clusters in the latent space corresponding to A and B, and by steering new images toward the centroid of the target cluster, we could enforce a more realistic and consistent harmonization.

7 Conclusion and Outlook

In this work, we have demonstrated a controlled framework for isolating the impact of reconstruction algorithms on CT image appearance. By generating a single synthetic sinogram with Poisson noise and applying both Tikhonov (ℓ_2) and TV+Poisson variational reconstructions, we quantified the systematic differences between these methods. A simple CNN was then trained to distinguish between the two image domains with perfect accuracy, and subsequently used via a combined cross-entropy and fidelity loss to "translate" Tikhonov outputs into TV-like images. Although our harmonization algorithm succeeded in changing the classifier's label, its outputs often lie outside the distribution of truly TV-reconstructed images, highlighting the intrinsic ill-posedness and non-uniqueness of post-hoc mappings. Possibilities to explore for greater harmonizing succes include the integration of perceptual or texture losses and anatomy-aware constraints in the algorithm for more realistic edge and noise characteristics.[10, 11].
Ultimately, while learned post-processing offers a promising route to scanner-agnostic CT, its success hinges on combining data-driven flexibility with strong priors both statistical and physical to ensure that harmonized images remain diagnostically valid across diverse clinical settings.

References

[1] Jean-Philippe et al. Fortin. "Harmonization of multi-site imaging data: methods and challenges". In: *NeuroImage* 183 (2018), pp. 560–576.

[2] Simon L. F. et al. Walsh. "Imaging biomarkers in interstitial lung disease: a systematic review". In: *European Respiratory Review* 30.160 (2021).

[3] Russell T. et al. Shinohara. "Statistical normalization techniques for magnetic resonance imaging". In: *NeuroImage: Clinical* 6 (2014), pp. 9–19.

[4] Zhiyuan et al. Zhou. "CT image harmonization using deep learning". In: *Medical Physics* 49.6 (2022), pp. 3737–3750.

[5] Tristan Van Leeuwen and Christoph Brune. "Ill-posedness and regularisation". In: *10 Lectures on Inverse Problems and Imaging*. Matheon, Berlin, 2012.

[6] Ge Wang, Hengyong Yu, and Bruno De Man. "Image reconstruction for CT using statistical models and iterative algorithms". In: *Physics in Medicine and Biology* 53.20 (2008), R319–R361.

[7] T. Le, R. Chartrand, and T. J. Asaki. "Poisson intensity reconstruction with total variation regularization". In: *IEEE International Conference on Image Processing*. 2003, pp. 645–648.

[8] Tristan Van Leeuwen and Christoph Brune. "Ill-posedness and regularisation". In: *10 Lectures on Inverse Problems and Imaging*. Matheon, Berlin, 2012.

[9] Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: (2016). arXiv: 1603.08155 [cs.CV]. URL: https://arxiv.org/abs/1603.08155.

[11] Christian Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 105–114. DOI: 10.1109/CVPR.2017.19.

A Derivation of Gaussian Log-Likelihood

Assume each measured sinogram value f_i^δ is related to the true projected value $[Ru]_i$ through additive white Gaussian noise:

$$f_i^\delta = [Ru]_i + \varepsilon_i, \quad \text{with } \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Then, the probability density of observing f_i^δ given u is:

$$p(f_i^\delta \mid u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(f_i^\delta - [Ru]_i)^2\right).$$

Assuming independence across measurements, the total likelihood is:

$$p(f^\delta \mid u) = \prod_i p(f_i^\delta \mid u).$$

Taking the negative log-likelihood (and dropping constants that don't depend on u):

$$-\log p(f^\delta \mid u) = \sum_i \frac{1}{2\sigma^2}(f_i^\delta - [Ru]_i)^2 + \text{const}.$$

Minimizing this expression is equivalent to minimizing:

$$\sum_i (f_i^\delta - [Ru]_i)^2 = \|Ru - f^\delta\|_2^2.$$

This shows that the commonly used least-squares data term corresponds to a maximum likelihood estimator under an i.i.d. Gaussian noise assumption.