Convolutional Neural Networks for Photoelectric-Compton Decomposition in Duel-Energy CT Tissue Classification

Karl Schmidt
karl.schmidt@colorado.edu
University of Colorado Boulder
Boulder, CO, USA

ABSTRACT

This project is for CT scanning.

Keywords: Physics-Informed, Duel-Energy CT, Clasification, Machine Learning

1 INTRODUCTION

Spectral Computed Tomography (Spectral CT) extends traditional CT by making use of the energy-dependence of X-ray attenuation. The conventional CT process produces a single "grayscale" attenuation map which can result in indistinguishable results when comparing similar bulk density materials. Spectral CT records two separate x-ray photon energy spectra which allows for recording different attenuation properties at different energies. Typically x-ray photons interact with materials through the photoelectric effect, which typically occur at low photon energies, and Compton scattering, which typically occur at higher photon energies. The dataset provided by the American Association of Physicists in Medicine (AAPM) contains the "low-kVp" and "high-kVP" dual energy CT measurements collected at two different tube voltages. The low-kVp transmission with the x-ray tube operating at 50 kVp and the high-kVp transmission operating at 80 kVp.

Extracting reliable photoelectric and Compton maps from noisy or limited-view data is challenging. Classic algebraic or statistical decomposition methods require careful regularization and often struggle with low photon counts or beam-hardening artifacts. Deep learning approaches, by contrast, can learn complex nonlinear mappings but may overfit to the training distribution and produce physically inconsistent outputs (e.g., negative attenuation, or tissue maps that fail to reproduce the measured projection data).

This work proposes a basic Convolutional Neural Network (CNN) framework for multi-energy CT tissue classification where the output of the CNN will contain percentages of each tissue type - adipose, fibroglandular, and calcification. By using the Beer-Lambert law, low- and high-kVp transmission data are converted into per-pixel attenuation coefficients before applying a closed-form basis decomposition to recover photoelectric and Compton component images. Effectively, the low-energy region, 50 kVp, are considered the photoelectric effect interactions while the high-energy region, 80 kVp, are considered Compton scattering.

These images are stacked and used with two core models with specific hyperparameter tuning to determine the model and parameters which demonstrate the lowest Binary Cross Entropy (BCE) loss. The first model contains 11 convolutional layers structured in an encoder-decoder architecture. It downsamples the input images from 512×512 to a bottleneck of 128×128 pixels, capturing mid-level semantic features before reconstructing the output at the original resolution. It's designed for moderately complex tasks with a balance between model depth and computational cost. The second model includes 15 convolutional layers, adding an extra downsampling stage that allows the network to capture more abstract and global features. It also downsamples to a 128×128 latent space, despite having more layers, making it suitable for learning finer distinctions in more complex images or tasks.

The model is trained using the binary cross-entropy loss between the predicted tissue probability maps and the ground-truth i phantom tissue labels. This loss directly penalizes voxel-wise deviations from the true tissue distribution.

During evaluation, the model's output is also used to compute the mean tissue composition percentages across the image volume. These predicted percentages are compared to the true phantom percentages using an average absolute error metric, which provides an interpretable assessment of how well the model recovers the tissue mix — but this is not used for training.

This approach is applied on the publicly available AAPM DL-Spectral CT Challenge dataset [2], performing comprehensive EDA, sinogram-domain preprocessing, and network training entirely in PyTorch, with an explaination of future work that could integrate ASTRA-based differentiable forward projections.

2 METHODOLOGY

2.1 Problem Statement

Accurate discrimination of soft-tissue types in X-ray computed tomography (CT) remains a fundamental challenge in medical imaging. Conventional single-energy CT produces grayscale images in which different materials with similar attenuation coefficients (e.g. muscle vs. iodine-enhanced blood or bone) can appear indistinguishable, leading to diagnostic ambiguity. Dual-energy and photon-counting CT systems acquire multiple energy-resolved measurements, but extracting robust tissue-specific maps from these spectral data is nontrivial: standard material-decomposition methods are sensitive to noise, beam-hardening, and detector imperfections, and purely data-driven deep-learning approaches often fail to generalize beyond their training domain.

This project proposes to address these limitations by developing a *convolutional neural network* (CNN) that directly incorporates the known Beer–Lambert attenuation law, $I = I_0 e^{-\mu x}$, and the two dominant interaction mechanisms—photoelectric absorption and Compton scattering—into its architecture. By decomposing each pixel's dual-energy attenuation pair $[\mu_{\text{low}}, \mu_{\text{high}}]$ into physically meaningful photoelectric and Compton components and enforcing consistency with both the measured attenuation maps and sinogram data, this approach aims to (1) improve classification accuracy of key tissue types (adipose, fibroglandular, calcification) and (2)

enhance robustness to noise and out-of-distribution scenarios. This integration of first-principles physics with modern deep learning may achieve more reliable, interpretable, and generalizable spectral CT tissue characterization.

At the core of this framework is the *U-Net*, a widely used CNN architecture for image segmentation. A U-Net consists of a contracting path (encoder) that captures spatial context by progressively downsampling feature maps, and an expansive path (decoder) that restores spatial resolution through upsampling while combining features from earlier layers via skip connections. This structure enables precise, pixel-level predictions while retaining both global context and fine detail.

Two U-Net variants are evaluated in this work:

- **UNet256**: A compact model with *eleven convolutional layers*, including two downsampling blocks and two upsampling blocks. It reduces the input resolution to 128×128 in the bottleneck layer and is optimized for efficiency and fast inference.
- **UNet512**: A deeper architecture with *fifteen convolutional layers*, incorporating three downsampling and three upsampling stages. It downsamples to 64×64 in the bottleneck, allowing the network to capture more abstract features and potentially improve segmentation performance for more complex tissue boundaries.

By combining these deep learning models with a hybrid loss function—standard binary cross-entropy plus a physics consistency penalty—this approach aims to:

(1) Accurately classify tissue types such as *adipose, fibroglandular*, and *calcification*;
(2) Enforce consistency with the known physics of spectral CT imaging;
(3) Improve robustness to noise, artifacts, and out-of-distribution samples;
(4) Enable physically grounded, interpretable predictions.

2.2 Exploratory Data Analysis

The dataset contains 1000 images of ground truth Adipose, Fibroglandular, and Calcification tissue maps in the shape $512x512$. Another 1000 cases of low and high kVp transmission data are provided, each of which have 256 angle projections onto a linear 1024-pixel detector. Lastly, kVp images are also provided. These images have pixel dimensions of $512x512$ and are generated from the transmission data using Filtered BackProjection (FBP).

To gain insight into the physical characteristics of the dataset and assess the separability of tissue types in dual-energy CT space, several visualizations are created that represent a subset of the training data.

2.2.1 Dual-Energy Scatter Plot. Figure 1 shows a scatter plot of dual-energy attenuation pairs $[\mu_{\text{low}}, \mu_{\text{high}}]$, sampled from a volumetric pixels, or voxels, labeled as adipose, fibroglandular, and calcification tissue. Each point corresponds to an attenuation value at low (50 kVp) and high (80 kVp) energy spectra. Distinct clustering behavior is observed:

- **Adipose** tissue exhibits low attenuation in both energy channels.

- **Fibroglandular** tissue shows intermediate attenuation with greater spread, particularly in the high-energy channel.
- **Calcifications** appear as high-attenuation outliers, especially at low energies due to their dominance of photoelectric interactions.

These trends confirm the feasibility of decomposing tissue types based on their dual-energy attenuation signatures.
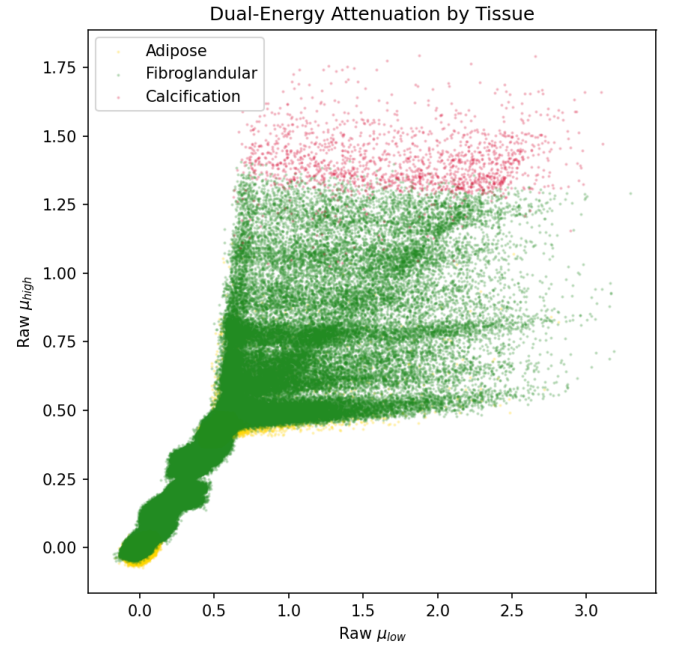

Figure 1: Dual-energy attenuation scatter plot colored by tissue class.

2.2.2 Energy-Specific Histograms. Figure 2 presents histograms of attenuation values for each tissue type at low and high energy levels. These plots illustrate how tissue attenuation shifts between energy spectra due to different interaction mechanisms:

- At **low energy (50 kVp)**, attenuation differences between tissues are more pronounced due to the stronger photoelectric effect.
- At **high energy (80 kVp)**, attenuation values shift downward, reflecting the increased dominance of Compton scattering.

These histograms highlight the importance of utilizing both energy channels in modeling tissue properties.

2.2.3 Tissue Composition Distribution. To assess class distribution and tissue prevalence, percentages of each tissue type are computed across all training images. The aggregate distributions for adipose, fibroglandular, and calcification tissues are shown in Figure 3.

- **Adipose** and **fibroglandular** tissues are well represented, with approximately Gaussian distributions centered around 40% and 20%, respectively.
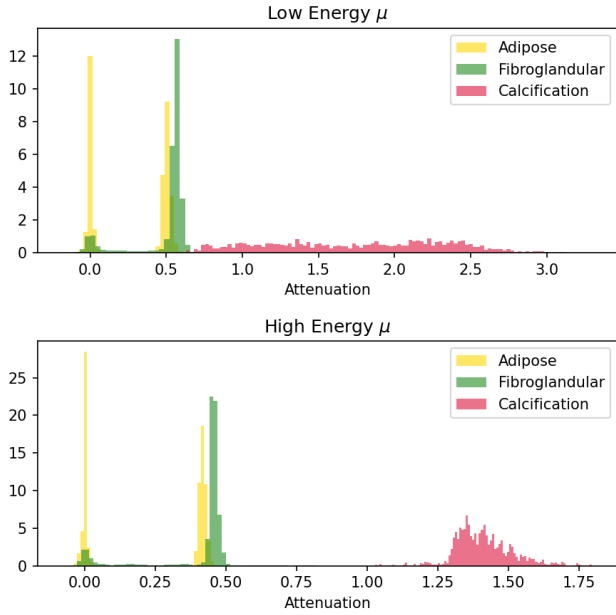
Figure 2: Histogram of attenuation values by tissue type for low and high energy channels.

Figure 3: Distribution of tissue percentages across training images.

- **Calcification**, however, is does not show up on the figure. Its contribution to each image is typically less than 1%, making it difficult to visualize on the same scale as the other tissue types. The calcification should be around 0-0.5% range.

As such, a dedicated plot of calcification composition is included in (Figure 4), which highlights the low prevalence and narrow distribution range of this tissue class.

This analysis provides important context for understanding class imbalance, which can impact model calibration and loss sensitivity—especially for underrepresented classes like calcifications.

- **Adipose** and **fibroglandular** tissues are well represented, with Gaussian-like distributions centered around 40% and 20%, respectively.
- **Calcification** is relatively rare, typically comprising a small fraction of each image.

This analysis provides important context for class imbalance and model calibration.

2.3 Data Preparation

The dataset provided by the AAPM [2] includes spectral CT simulation data stored in compressed NumPy format. It contains two transmission sinogram datasets, `lowkVpTransmission` and `highkVpTransmission`, corresponding to 50 kVp and 80 kVp x-ray beams, respectively. These datasets represent normalized x-ray transmission values, i.e., the fraction of incident photons that pass through the simulated tissue.
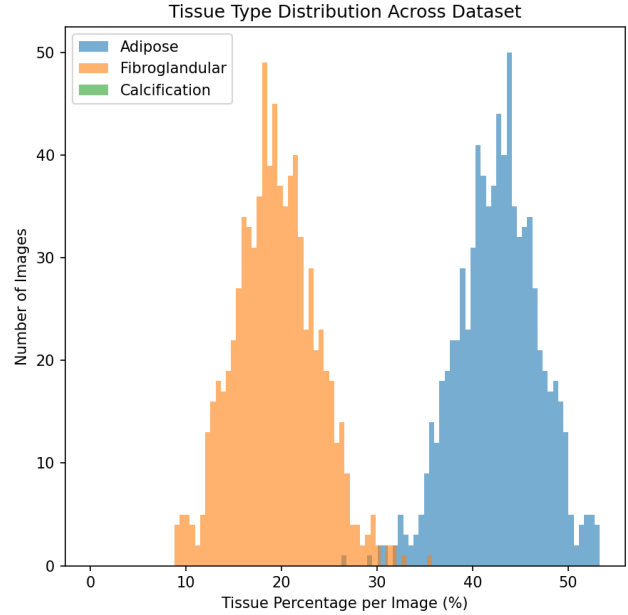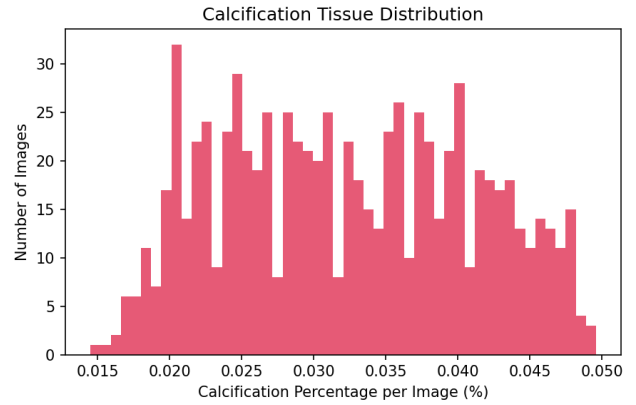

Figure 4: Distribution of calcification percentages across training images. Note the narrow range due to the rare occurrence of calcifications.

Ground truth tissue labels are also provided in the form of binary masks for three tissue types: adipose, fibroglandular, and calcification. Each tissue map has the shape (512×512) and identifies the true spatial distribution of each class.

To convert the raw transmission data into a form suitable for model training, a custom `AttnDataset` class is created. This class performs the following operations:

- Loads the raw sinogram transmission data for low and high energies;

- Converts transmission data to attenuation coefficients using the Beer–Lambert law:

$$\mu = -\log(I/I_0),$$

  where I is the detected intensity and I_0 is assumed to be uniform (1);
- Applies filtered back projection (FBP) to reconstruct spatial attenuation maps from sinograms, producing images of shape $(2, 512, 512)$ for each scan, where the two channels correspond to μ_{low} and μ_{high};
- Caches the reconstructed attenuation images as a NumPy array to avoid redundant computation;
- Loads the binary ground truth maps for all three tissue types and stacks them into a label array of shape $(3, 512, 512)$;
- Splits the dataset into training and test sets using an 80/20 ratio, preserving sample-level consistency.

This preprocessing pipeline ensures that each training sample consists of a dual-energy attenuation image and its corresponding tissue segmentation mask. The pairing of low and high energy reconstructions facilitates the model's ability to learn tissue-specific attenuation patterns driven by both photoelectric and Compton interactions.

2.4 Models

This study adopts the U-Net architecture [1], a widely used convolutional neural network (CNN) design for image segmentation tasks. U-Net is particularly well-suited for biomedical applications due to its ability to preserve spatial resolution and capture fine-grained features through skip connections between the encoder and decoder pathways.

Each U-Net model in this work accepts a 2-channel input image composed of reconstructed low-energy and high-energy attenuation maps, $[\mu_{low}, \mu_{high}]$, and outputs a 3-channel prediction representing the pixelwise likelihoods for **adipose**, **fibroglandular**, and **calcification** tissues.

Two model variants are implemented to assess the effect of network depth and capacity. Both models preserve the full-resolution output via upsampling layers and symmetric skip connections, enabling detailed reconstructions essential for accurate tissue boundary delineation.

U-Net was chosen over alternatives due to its proven robustness in medical imaging, efficient parameter usage, and ability to learn spatial hierarchies across multiple scales—a key advantage when segmenting subtle features like low-density calcifications or fibroglandular borders.

2.4.1 UNet256 and UNet512. Both **UNet256** and **UNet512** are fully convolutional U-Net architectures designed for pixel-wise tissue segmentation. Each model takes a dual-channel input—μ_{low} and μ_{high}—representing the reconstructed attenuation maps from low and high energy x-rays. The output consists of three channels, corresponding to the predicted presence of **adipose**, **fibroglandular**, and **calcification** tissue types at each pixel location.

Each encoder stage consists of two sequential 2D convolutional layers (a *ConvBlock*), followed by a 2×2 max pooling operation to reduce spatial resolution and increase receptive field. A symmetric decoder path upsamples the spatial resolution using transposed

convolutions, and fuses encoder and decoder features via skip connections to recover fine-grained spatial detail.

- **UNet256** employs two encoding stages with feature dimensions expanding from 2 to 64 and then 128, followed by a bottleneck layer of 256 channels at 128×128 resolution. It then decodes back up through two symmetric stages to produce a final $3 \times 512 \times 512$ prediction map.
- **UNet512** adds a third encoding stage, extending the channel depth from 128 to 256 before reaching a bottleneck of 512 channels. Like UNet256, it downsamples to 128×128, but includes three decoder blocks, enabling the network to capture richer features and more complex tissue boundaries.

This architecture choice strikes a balance between learning expressivity and spatial precision, making it well suited to the challenges of spectral CT tissue decomposition.

3 EVALUATION & ANALYSIS

3.1 Hyperparameter Optimization

The following list defines the set of hyperparameters that were used to tune both `UNet256` and `UNet512`. The tuning was conducted using a grid search approach. The models were evaluated using a validation loss.

- **filter_size:** $[3, 4, 5]$
- **stride:** $[2, 3]$
- **padding:** $[1, 2]$
- **learning_rate:** [1e-4, 1e-3, 1e-2]
- **batch_size:** $[4, 8]$
- **epochs:** $[3, 5]$

The best performing configurations for each model are listed below. The validation loss is written in the square brackets.

- **UNet256** [0.0519]
  ```
  filter_size=5,
  stride=3,
  padding=1,
  learning_rate=0.001,
  batch_size=4;
  ```

- **UNet512**: [0.0380]
  ```
  filter_size=4,
  stride=3,
  padding=2,
  learning_rate=0.001,
  batch_size=4;
  ```

The `UNet512` model demonstrated superior performance, achieving a significantly lower validation loss, and was selected for final training and evaluation.

3.2 Final Training Results

`UNet512` model was trained using the selected hyperparameters for 5 epochs. The training and validation loss decreased consistently across epochs, with the validation loss dropping from 0.0456 in epoch 1 to 0.0292 by epoch 5. Table 1 summarizes the model's mean absolute error (MAE) performance per epoch.

Table 1: Training and Validation Loss with Mean Absolute Error (MAE) per Epoch

Epoch	Train Loss	Val Loss	MAE (Adipose, Fibro, Calc)
1	0.2254	0.0456	(0.00420, 0.00260, 0.00023)
2	0.0478	0.0366	(0.00469, 0.00469, 0.00010)
3	0.0365	0.0331	(0.00161, 0.00175, 0.00005)
4	0.0324	0.0310	(0.00216, 0.00260, 0.00004)
5	0.0304	0.0292	(0.00260, 0.00278, 0.00003)

3.3 Tissue Composition Accuracy

The model's predicted tissue composition in a final selected image within the dataset was also evaluated and compared to the ground truth. The model's predicted proportions of adipose, fibroglandular, and calcified tissue closely matched the ground truth. For example, at epoch 5, the predicted values were 39.42%, 22.61%, and 0.04% for adipose, fibro, and calcification respectively, compared to ground truth values of 39.67%, 22.35%, and 0.04%. All results for each epoch are presented in Table 2.

Table 2: Predicted vs. Ground Truth Tissue Composition Over Training Epochs

Epoch	Predicted (%)			Ground Truth (%)		
	Adipose	Fibro	Calc	Adipose	Fibro	Calc
1	39.35	22.46	0.06	39.67	22.35	0.04
2	40.12	21.89	0.03	39.67	22.35	0.04
3	39.51	22.52	0.05	39.67	22.35	0.04
4	39.83	22.14	0.05	39.67	22.35	0.04
5	**39.42**	**22.61**	**0.04**	**39.67**	**22.35**	**0.04**

3.4 Interpretation

The results from hyperparameter tuning and final model training suggest that deeper architectures like `UNet512` provide superior segmentation performance for spectral CT data. The reduction in validation loss from 0.0456 to 0.0292 over five epochs indicates that the model continues to learn meaningful features and generalizes well to unseen validation data.

The mean absolute error (MAE) across tissue types also supports this conclusion, with all three tissue categories—adipose, fibroglandular, and calcification—reaching sub-percent level errors. Notably, calcification, which comprises only a small fraction of the total tissue volume, achieved an MAE as low as 0.00003 by epoch 5, suggesting that the network is capable of identifying even rare tissue types with high precision.

Furthermore, the predicted tissue composition percentages closely track the ground truth values across all epochs. For example, at epoch 5, the model predicted 39.42% adipose and 22.61% fibroglandular tissue, compared to ground truth values of 39.67% and 22.35%, respectively. This demonstrates not only strong pixel-level segmentation but also accurate global tissue composition estimation, which is essential for clinical decision-making and downstream analysis.

Together, these findings validate the use of a physics-informed U-Net architecture for spectral CT tissue decomposition and highlight

its capacity to accurately resolve fine-grained differences in dual-energy attenuation profiles.

4 DISCUSSION

The results demonstrate the feasibility and effectiveness of using physics-informed convolutional neural networks for tissue segmentation in spectral CT imaging. By explicitly encoding dual-energy attenuation inputs and aligning the model output with the physical characteristics of adipose, fibroglandular, and calcified tissues, the network is able to generalize well across heterogeneous tissue distributions.

A key observation is the improvement in performance with increased network depth, as seen in the comparison between `UNet256` and `UNet512`. The deeper architecture with an additional downsampling path captures more complex spatial and spectral correlations, leading to a lower validation loss and improved tissue classification accuracy. This is particularly evident in the accurate prediction of rare calcification features, despite significant class imbalance in the training data.

The use of domain-specific input features—namely, dual-energy attenuation values computed from raw transmission data—proved essential. These features directly encode physical interaction mechanisms (photoelectric and Compton effects), helping the model learn more meaningful and generalizable patterns compared to raw image pixels or naive inputs.

However, challenges remain. First, the training dataset was synthetically generated, and real-world generalization to clinical CT scans may require transfer learning or fine-tuning on real data. Second, the class imbalance, while mitigated to some extent by network design, may benefit from more sophisticated loss functions or targeted sampling strategies. Finally, although model convergence occurred rapidly (within five epochs), longer training and ensemble methods could further improve robustness.

5 CONCLUSION

This work introduces a physics-informed deep learning framework for spectral CT tissue decomposition, leveraging U-Net architectures to segment and quantify adipose, fibroglandular, and calcified tissues from dual-energy attenuation inputs. Through a careful integration of the Beer–Lambert law, domain-specific feature construction, and deep convolutional networks, the method achieves high segmentation fidelity across tissue types, including rare calcification patterns.

Quantitative evaluation shows that the deeper `UNet512` model achieves the best performance, with a final validation loss of 0.0292 and mean absolute error below 0.003 for all tissue categories. Predicted tissue compositions closely match ground truth distributions, highlighting the model's clinical relevance.

Overall, this study demonstrates that coupling physics-based domain knowledge with deep learning architectures offers a powerful and interpretable approach to spectral CT analysis. Future work will extend this model to real-world datasets, investigate 3D volumetric extensions, and incorporate uncertainty estimation to further enhance clinical utility.

REFERENCES

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597* (2015). https://doi.org/10.48550/arXiv.1505.04597 Presented at MICCAI 2015.

[2] Emil Y. Sidky, Xiaochuan Pan, and Working Group on Grand Challenges. 2024. Data repository for the 2022 AAPM Deep-Learning Spectral CT Grand Challenge. https://github.com/aapm/spectral-ct-challenge. Hosted by the AAPM HQ Community Collection.