-RESEARCH PAPER

International Journal of Mining and Geo-Engineering

IJMGE

IJMGE 59-4 (2025) 409-417

DOI: 10.22059/IJMGE.2025.399223.595283

Improving µCT image segmentation through architectural enhancements in the U-Net model

Shahin Mahmoudi [a], MirSaleh Mirmohammadi [a, *] and Omid Asghari [b]

[a] *School of Mining Engineering, College of Engineering, University of Tehran, Tehran, Iran,*

[b] *Civil and Environmental Engineering Department, School of Mining & Petroleum Engineering, Faculty of Engineering, University of Alberta, Edmonton, Alberta, Canada.*

Article History:

Received: 23 July 2025.

Revised: 09 September 2025.

Accepted: 01 November 2025.

ABSTRACT

Micro-Computed Tomography (µCT) is an indispensable non-destructive technique for characterizing the internal microstructures of materials like porous media, but any quantitative analysis hinges on accurate image segmentation—a task complicated by image noise, poor contrast, and intricate features. While deep learning, and the U-Net architecture in particular, has shown considerable promise in automating this process, this study presents a comparative analysis between the standard U-Net and a bespoke, modified architecture for segmenting multi-phase micro-computed tomography (µCT) images of Bentheimer sandstone. Our modified U-Net introduces several architectural enhancements: optimized convolutional blocks for superior feature extraction, spatial dropout for more effective regularization, and L2 weight regularization to mitigate overfitting. We trained both models on 2D slices from two core samples and subsequently evaluated them against an independent blind test set of 1872 slices from a third core, which contained four distinct phases: porosity, quartz, clay, and feldspar. The quantitative results reveal that the modified U-Net decidedly outperforms its standard counterpart, achieving a macro-averaged Dice Similarity Coefficient (DSC) of 0.92 versus 0.88, and a macro-averaged Intersection over Union (IoU) of 0.85 versus 0.80. Most notably, our model demonstrated substantial gains in segmenting the challenging minority phases; the DSC for clay surged from 0.71 to 0.85, and for feldspar, it rose from 0.85 to 0.87, all while maintaining stable performance on the majority phases of porosity and quartz. These statistical improvements are corroborated by qualitative visual assessments, which confirm superior boundary delineation and a reduction in misclassifications. Ultimately, our findings indicate that the proposed architectural refinements yield a more accurate and robust segmentation model for micro-computed tomography (µCT) imagery, providing a more reliable foundation for downstream Digital Rock Physics (DRP) applications critical to the mining and geo-engineering sectors, such as geomechanical stability assessment and mineral liberation analysis.

Keywords: *Deep learning, Digital rock physics, Image segmentation, Micro-computed tomography (µCT), Porous media characterization.*

1. Introduction

The characterization of porous media remains a cornerstone of geoscience and engineering, with profound implications for fields as diverse as hydrocarbon recovery, CO_2 sequestration, groundwater management, and geothermal energy [1, 2]. In the mining sector, this characterization is equally critical, forming the basis of geometallurgical models that predict ore processability, optimizing mineral liberation and recovery, and assessing the geomechanical and hydrogeological stability of mine structures from pit slopes to underground excavations [3, 4]. The ability to accurately predict fluid flow—a complex interplay of microstructural architecture, fluid properties, and interfacial phenomena—is of paramount importance [5, 6, 7]. In this context, Bentheimer sandstone serves as a ubiquitous model rock. Its relative homogeneity and well-documented properties make it an ideal benchmark for validating novel analytical techniques and conducting fundamental pore-scale investigations [8, 9, 10]. This standardization fosters comparability and ensures the robustness of research findings, as seen in studies where multi-phase distributions in Bentheimer have yielded critical insights into oil recovery mechanisms [5]. The advent of Micro-Computed Tomography (µCT) has revolutionized the study of porous media. By enabling non-destructive 3D imaging of internal structures and fluid transport at the pore scale, it has opened new windows into subsurface processes [1, 11, 12]. Capable of resolutions from the micron to sub-micron level, micro-computed tomography (µCT) preserves the integrity of the sample and uniquely allows for in-situ imaging under representative laboratory conditions [7, 13, 14]. The voluminous 3D datasets generated by this technique are indispensable inputs for Digital Rock Physics (DRP) analyses and advanced computational simulations [5, 15]. Crucially for the mining industry, micro-computed tomography (µCT) is emerging as a transformative tool for process mineralogy, enabling three-dimensional mineral liberation analysis (MLA). This technique provides quantitative data on the volumetric distribution and locking characteristics of valuable minerals within the ore matrix, information that is essential for optimizing comminution circuits and downstream separation processes like flotation and leaching [3, 4]. However, the sheer scale of micro-computed tomography (µCT) data necessitates the development of automated, fast, and highly reliable image processing and segmentation workflows [1, 13].

Digital Rock Physics (DRP) offers a computational pathway to predict petrophysical properties like porosity and permeability directly

* Corresponding author. *E-mail address:* m.mirmohammadi@ut.ac.ir (M. Mirmohammadi).

from high-resolution 3D micro-computed tomography (μCT) images, thereby complementing or even replacing traditional, often destructive, laboratory measurements [16, 17, 18]. The central aim of DRP is to forge a direct link between a rock's microstructure and its macroscopic behavior, which is essential for improving reservoir characterization and optimizing resource management [6]. Beyond fluid flow, the DRP workflow is now integral to geomechanics, where high-fidelity digital rock models are used to simulate mechanical properties and predict failure mechanisms [19]. By accurately representing the rock's microstructure, including mineralogy, porosity, and micro-fractures, these simulations provide critical insights into rock mass strength and deformability. This capability is paramount in mining engineering for applications such as analyzing the stability of underground openings, characterizing the excavation damaged zone (EDZ), and modeling the behavior of open-pit slopes [20]. A typical DRP workflow begins with image acquisition, followed by preprocessing, segmentation, and finally, the numerical computation of physical properties [16]. Within this workflow, image segmentation—the partitioning of an image into its distinct material phases—stands out as the most critical step [12, 17, 21]. The fidelity of the segmentation directly governs the reliability of all subsequently derived petrophysical properties and flow simulations [6, 18, 22].

Even subtle inaccuracies in delineating phase boundaries can cascade into significant errors in estimating pore volume, connectivity, and tortuosity, particularly in complex multi-phase systems [5, 23]. This 'segmentation bottleneck' has direct and severe consequences for engineering applications [3]. In mineral processing, for instance, inaccurate phase delineation leads to erroneous mineral liberation estimates, resulting in suboptimal grinding strategies and reduced metallurgical recovery [4]. In geomechanics, the failure to accurately segment micro-fractures and mineral boundaries can lead to unreliable predictions of rock strength and failure pathways, potentially compromising the safety and design of mining excavations [19, 20]. It is therefore no exaggeration to state that high-precision segmentation is a prerequisite for realizing the full potential of DRP [20, 24].

Traditional segmentation algorithms, such as thresholding and watershed methods, often struggle when applied to complex micro-computed tomography (μCT) images. They are easily confounded by imaging noise, artifacts, partial volume effects, and the overlapping grayscale intensities of different mineral phases [16, 17, 25, 26, 27]. These methods frequently depend on manual parameter tuning, which introduces user bias and can be computationally demanding. The resulting inaccuracies in segmentation can severely compromise the integrity of DRP analyses [21, 24, 28]. The growing complexity of micro-computed tomography (μCT) datasets only accentuates the urgent need for more sophisticated segmentation solutions. Deep learning, and Convolutional Neural Networks (CNNs) in particular, has ushered in a new era for computer vision and is finding increasingly powerful applications in the geosciences [6, 29, 30].

CNNs possess the remarkable ability to automatically learn discriminative, hierarchical features directly from raw data, thereby obviating the need for manual feature engineering and enabling the capture of highly complex patterns [16]. Architectures like Fully Convolutional Neural Networks (FCNNs) have proven exceptionally effective at segmenting intricate structures, even in the presence of significant image imperfections [16, 31, 32]. The U-Net model, first proposed by Ronneberger et al. (2015) [33], has become a de facto standard for semantic segmentation in biomedical imaging and is rapidly gaining traction in the geosciences and DRP [10, 34, 35, 36]. Its elegant U-shaped encoder-decoder structure, augmented with skip connections, masterfully fuses deep, semantic features with shallow, high-resolution spatial information.

This design enables both precise localization and the delineation of fine-grained boundaries [20, 30], making it particularly well-suited for resolving the complex inter-grain textures in multi-mineral ores and the subtle features of micro-fracture networks in geological materials [20, 21]. The proven effectiveness of U-Net, even with limited training data, and its inherent adaptability make it a robust foundation for DRP applications [34, 37, 38]. However, the standard U-Net architecture is not without its limitations. To achieve optimal performance on complex, multi-phase micro-computed tomography (μCT) images of geological materials—which present subtle intensity variations, noise, and diverse features—strategic enhancements are necessary [16, 20, 35, 39]. Architectural modifications are therefore aimed at improving accuracy, robustness, and efficiency by tailoring the model to the specific challenges posed by geological images [37, 40].

This study puts forth a modified U-Net architecture that incorporates several such enhancements: optimized convolutional blocks, spatial dropout, and L2 regularization. Depthwise separable convolutions are employed to reduce the parameter count and computational footprint compared to standard convolutions. This allows for faster training and inference and can lead to improved accuracy by mitigating overfitting, a crucial consideration for large datasets [35, 37, 41, 42]. We also integrate spatial dropout, a regularization technique that randomly nullifies entire feature maps during training. This forces the network to learn a more diverse and robust set of features, an approach that is particularly advantageous for segmentation tasks [22, 40, 43, 44, 45].

Finally, L2 regularization (or weight decay) is applied to penalize large model weights, which encourages simpler models that are less susceptible to noise in the training data and thus generalize better to unseen examples [40, 46]. By combining these improvements, we aim to develop a modified U-Net architecture that is not only computationally efficient and robust but also specifically tailored for the high-precision segmentation of complex scientific imagery.

To test these modifications, this paper presents a comparative evaluation of the standard U-Net against our proposed modified architecture for the task of segmenting multi-phase micro-computed tomography (μCT) images of Bentheimer sandstone. The primary contribution of this work is an empirical assessment of whether these targeted modifications translate into tangible improvements in segmentation performance within the context of DRP [20]. Our central hypothesis is that the modified U-Net will demonstrate superior computational efficiency due to its optimized convolutional blocks, while also achieving enhanced generalization and segmentation accuracy as a result of the spatial dropout and L2 regularization, ultimately leading to more reliable phase delineation. This increased reliability is particularly consequential for geo-engineering. For example, the ability to accurately segment challenging minority phases, such as clays, is critical as their presence can disproportionately govern the rock's hydro-mechanical behavior and the efficiency of mineral processing [3]. Therefore, our hypothesis is that these targeted architectural modifications will produce a segmentation model whose outputs can serve as more robust and trustworthy inputs for the DRP-based mineralogical and geomechanical simulations that inform modern mine design and operation.

The remainder of this paper is structured as follows: Section 2 details the methodology, including the dataset, model architectures, training procedures, and evaluation metrics. Section 3 presents and discusses the comparative results. Finally, Section 4 offers our conclusions, summarizing the key findings and their implications, and suggests avenues for future research.

2. Methodology

This section outlines the experimental framework used to compare the standard U-Net architecture against our proposed modified U-Net for segmenting multi-phase micro-computed tomography (μCT) images of Bentheimer sandstone. We describe the dataset, detail the preprocessing and data augmentation techniques, specify the architectures of both models, and explain the training protocol and quantitative evaluation metrics.

2.1. Dataset description

The dataset for this study comprises μCT images of Bentheimer sandstone, which were previously characterized by Liang et al. (2022) [47]. Each image slice contains four distinct phases: porosity, quartz, clay, and feldspar, as illustrated in Figure 1.

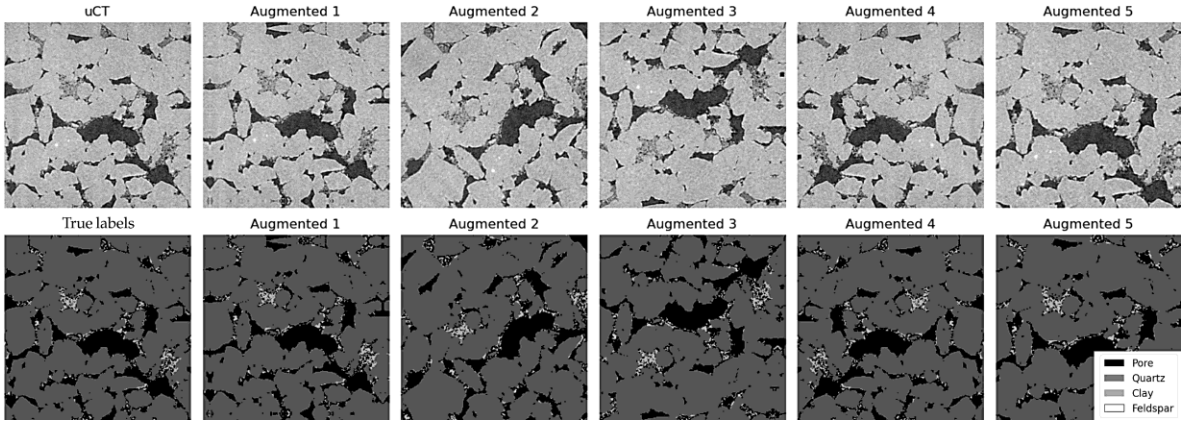

Figure 1. An example of the data augmentation process on a µCT image slice. The top row displays the original preprocessed slice from Sample A (far left) alongside five examples of randomly generated augmented images. The bottom row shows the corresponding ground-truth segmentation masks, which have undergone the exact same random transformations. This process was employed to expand the training dataset and enhance the model's generalization capabilities. The legend identifies the four distinct phases: Pore, Quartz, Clay, and Feldspar.

The dataset for this study comprises micro-computed tomography (µCT) images from three distinct Bentheimer sandstone core samples, designated A, B, and C, with each core containing 1872 2D image slices. The volumetric distribution of the four constituent phases for each core sample is detailed in Table 1. A key characteristic of this dataset is the significant class imbalance, with porosity and quartz representing the vast majority of the volume, while clay and feldspar constitute challenging minority phases, collectively accounting for less than 5% of the total volume in each sample. This low prevalence presents a significant challenge for segmentation algorithms, as these minority phases are critical for accurate geomechanical and mineralogical analysis. To ensure a rigorous and unbiased evaluation of the models' generalization capabilities, the entirety of core sample C (1872 slices) was held out as an independent blind test set.

For the development of the training and validation sets, data was sourced from core samples A and B. Given that the manual annotation of ground truth labels for micro-computed tomography (µCT) images is a meticulous and time-intensive process, we adopted a sparse sampling strategy to create a high-quality, manageable dataset. Specifically, we selected slices at 100-slice intervals from both core A and core B, yielding 20 unique slices from each. This resulted in a total pool of 40 annotated 2D images. This pool was then randomly partitioned, with 80% (32 slices) allocated to the training set and the remaining 20% (8 slices) designated for the validation set.

2.2. Preprocessing and data augmentation

To prepare the micro-computed tomography (µCT) images for the models, a standardized preprocessing workflow was applied to all 2D slices (Figure 2). Initially, the pixel intensities of the grayscale images were normalized to a [0, 1] range by dividing each pixel's value by the maximum possible intensity. Following normalization, all image slices and their corresponding labels were resized to a uniform dimension of 512x512 pixels using bilinear interpolation to ensure dimensional consistency and manage the computational load.

To enhance model robustness, improve generalization, and mitigate the risk of overfitting on the limited training set, a series of data augmentation techniques were applied on-the-fly during the training process. These transformations, which were applied randomly and identically to both the input images and their segmentation masks to maintain spatial correspondence, included random rotations within a ± 15 degree range and random horizontal and vertical flips, each with a 50% probability.

Additionally, the augmentation strategy involved random zooming, where images were scaled between 90% and 110% of their original size, followed by cropping or padding to restore the 512x512 dimension. The dataset was further diversified through random shearing with an intensity range of ±0.05 and random translations that shifted images horizontally and vertically by up to ±10% of their dimensions.

2.3. Model architectures

We implemented and compared two U-Net-based architectures: a standard U-Net serving as our baseline, and a modified U-Net incorporating our proposed enhancements.

2.3.1. Standard U-Net

The baseline model employed in this study is a classic U-Net architecture, characterized by its symmetric encoder-decoder structure and the use of skip connections to merge low-level and high-level feature maps.

The encoder, or contracting path, is composed of four blocks designed for feature extraction and downsampling. Each block contains two sequential 3x3 convolutional layers, with each layer followed by Batch Normalization and a ReLU activation function. To reduce spatial dimensions, a 2x2 Max Pooling operation with a stride of 2 is applied after each block. As the data progresses through the encoder, the number of feature channels doubles at each step, increasing from an initial 64 to 512. At the base of the "U" lies the bottleneck layer, which consists of two 3x3 convolutional layers (also with Batch Normalization and ReLU) and operates on 1024 feature channels, capturing the most abstract features.

The decoder, or expansive path, symmetrically mirrors the encoder to reconstruct the segmentation map. It consists of four blocks, where each block begins with a 2x2 transposed convolution for upsampling, a process that halves the number of feature channels. The upsampled feature map is then concatenated with the corresponding feature map from the encoder via a skip connection, a critical step that allows the network to recover fine-grained spatial information. This combined map is subsequently processed by two 3x3 convolutional layers with Batch Normalization and ReLU. Through the decoder, the feature channels are progressively halved from 1024 back down to 64. Finally, the output layer utilizes a 1x1 convolution to map the 64-channel feature map to the four output classes (porosity, quartz, clay, and feldspar), followed by a Softmax activation function to generate pixel-wise class probabilities.

2.3.2. Proposed modified u-net

Our modified U-Net architecture (Figure 3) preserves the core encoder-decoder structure and skip connections of the standard U-Net but integrates the following key enhancements to improve regularization and feature extraction:

Optimized Convolutional Blocks (Encoder): We replaced the standard pair of 3x3 convolutions in each encoder block. Our modified blocks consist of two sequential units, each comprising a 3x3 depthwise separable convolution (a depthwise followed by a 1x1 pointwise convolution), Batch Normalization, and a ReLU activation. This change is designed to improve both computational efficiency and feature representation.

Table 1. Volumetric phase distribution for the three Bentheimer sandstone core samples.

Phase	Sample A (%)	Sample B (%)	Sample C (Test Set) (%)
Porosity	19.6	23.4	25.3
Quartz	75.4	73.9	72.4
Clay	2.8	0.8	1.0
Feldspar	2.2	1.9	1.3

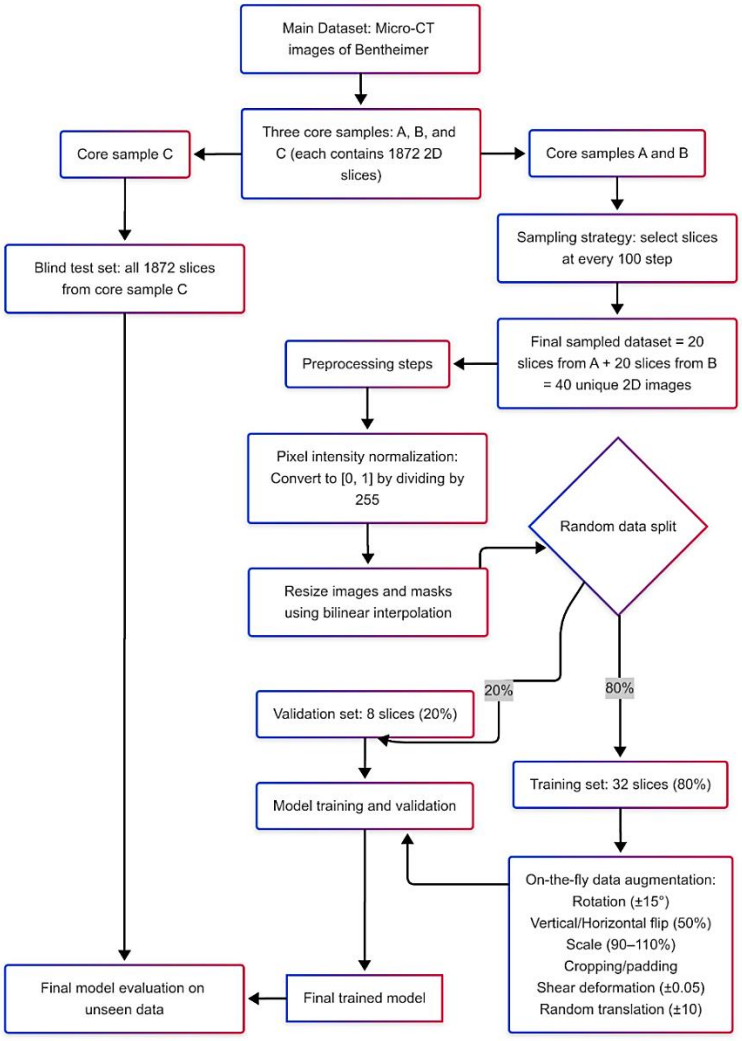

Figure 2. Flowchart of the complete data processing pipeline. The diagram illustrates the workflow from the initial splitting of core samples into training, validation, and blind test sets, through preprocessing and on-the-fly data augmentation, to the final model evaluation.

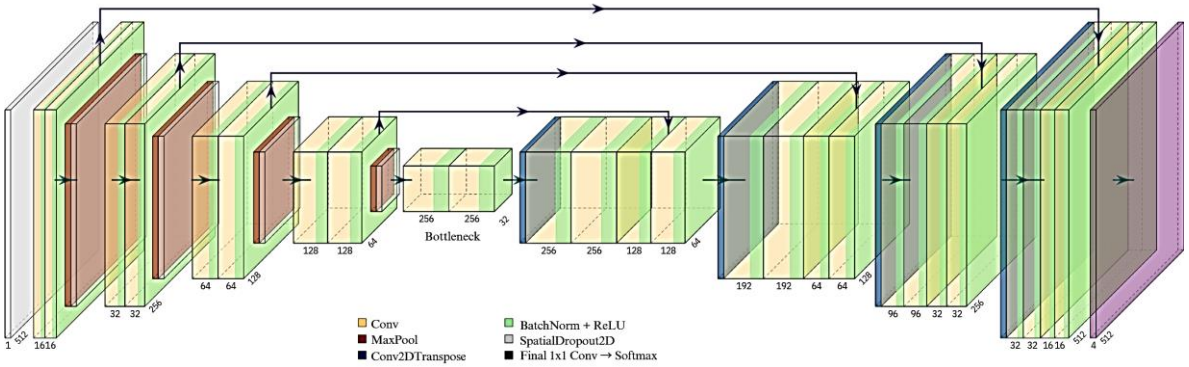

Figure 3. Architecture of the proposed modified U-Net model.

1. **Optimized Convolutional Blocks (Encoder):** We replaced the standard pair of 3x3 convolutions in each encoder block. Our modified blocks consist of two sequential units, each comprising a 3x3 depthwise separable convolution (a depthwise followed by a 1x1 pointwise convolution), Batch Normalization, and a ReLU activation. This change is designed to improve both computational efficiency and feature representation.
2. **Spatial Dropout:** After each max pooling operation in the encoder, we inserted a spatial dropout layer with a rate of 0.25. This form of dropout nullifies entire feature maps, providing a more structured regularization that is particularly effective for convolutional layers.
3. **L2 Regularization:** We applied L2 weight regularization (weight decay) with a penalty factor of 1×10^{-4} to the kernels of all convolutional layers across the entire network (encoder, bottleneck, and decoder). This discourages excessively large weights, promoting simpler models that are less prone to overfitting. The L2 penalty is added to the loss function as follows:

$$L_{\text{total}} = L_{\text{original}} + \frac{\lambda}{2N}\sum_i w_i^2 \quad (1)$$

where λ is the regularization coefficient and N is the total number of weights.

The remaining components of the modified U-Net—the bottleneck, decoder path, and output layer—follow the same structural principles as the standard model (e.g., standard 3x3 convolutions in the decoder), with the crucial difference that L2 regularization is also applied to their convolutional layers.

2.4. Training setup and procedure

To ensure a fair and direct comparison, the training protocol was kept identical for both the standard and modified U-Net models. The Adam optimizer [48] was employed for model training, with an initial learning rate set to 1×10^{-4}. To facilitate dynamic adjustments during training, a reduce learning rate on plateau scheduler was utilized, which decreased the learning rate by a factor of 0.1 whenever the validation loss stagnated for five consecutive epochs. The models were trained using the categorical cross-entropy loss function, which is highly suitable for multi-class semantic segmentation tasks. The loss for a single pixel is calculated using the categorical cross-entropy function, a standard choice for multi-class classification tasks in deep learning [49]:

$$L_{CCE} = -\sum_{c=1}^{M} y_c \log\left(p_c\right) \quad (2)$$

where M is the total number of classes; the summation is performed over each class index c; y_c is a binary indicator (1 if class c is the true class for the pixel, and 0 otherwise); and p_c is the model's predicted probability that the pixel belongs to class c.

The training was conducted with a batch size of 8, a choice dictated by the available 8 GB of VRAM on the NVIDIA RTX 3070 Ti GPU used for all experiments. Models were set to train for a maximum of 100 epochs. However, an early stopping mechanism was implemented to prevent overfitting; training was halted if the validation loss did not improve for 15 consecutive epochs, and the model weights from the best-performing epoch were restored for final evaluation.

2.5. Evaluation strategy

The segmentation performance of the trained models was quantitatively assessed on the independent blind test set (Sample C). We used the following standard pixel-wise metrics, calculated for each of the four phases and then macro-averaged for an overall score:

1. **Dice Similarity Coefficient (DSC):** Measures the overlap between prediction and ground truth, calculated as $2 \times \frac{|X \cap Y|}{|X|+|Y|}$.
2. **Intersection over Union (IoU) / Jaccard Index:** Quantifies the similarity between the predicted and ground truth areas, given by $\frac{|X \cap Y|}{|X \cup Y|}$.
3. **Pixel Accuracy (PA):** The fraction of correctly classified pixels. This metric is commonly used for evaluating semantic segmentation performance [30, 31] and is defined as:

where the summation is over all M classes, indexed by c; TP_c represents the number of true positives (pixels correctly classified) for class c; and FP_c represents the number of false positives (pixels incorrectly classified as class c).

Beyond these quantitative measures, we also performed a qualitative visual comparison of the segmentation outputs on representative slices from the test set. This allowed us to assess the practical quality of the boundary delineations, particularly for minority phases and at complex interfaces.

3. Results and evaluation

This section presents a comprehensive comparative analysis of the segmentation performance of the standard U-Net and our proposed modified U-Net. The evaluation was performed on the independent blind test set (Sample C), which contains 1872 µCT slices, to provide a robust measure of how well each model generalizes to new data. Our analysis integrates quantitative metrics, qualitative visual assessments, and a discussion of the practical implications and computational overhead.

3.1. Quantitative Performance Analysis

We quantified the segmentation performance of both models using macro-averaged Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and Pixel Accuracy across the four phases (porosity, quartz, clay, and feldspar). To dig deeper, we also computed these metrics on a per-phase basis, which is crucial for understanding how the models handle the challenging minority mineral phases.

3.1.1. Overall segmentation metrics

The macro-averaged performance metrics, summarized in Table 3, clearly demonstrate the superiority of the modified U-Net architecture. Compared to the standard U-Net, our modified model achieved a 4.25% increase in DSC, a 6.33% increase in IoU, and a 2.79% increase in pixel accuracy.

These aggregate results strongly suggest that the architectural enhancements—optimized convolutional blocks, spatial dropout, and L2 regularization—collectively contribute to a more accurate and reliable segmentation of these complex multi-phase microstructures.

3.1.2. Per-phase segmentation metrics

To gain a more granular understanding of model behavior, we analyzed the DSC and IoU scores for each individual phase, as detailed in Tables 4 and 5. The per-phase results are particularly revealing and underscore the modified U-Net's primary advantage. As established in Table 1, the dataset is characterized by a pronounced class imbalance, with clay and feldspar being significant minority phases. For instance, in the blind test set (Sample C), clay constitutes merely 1.0% of the total volume. This low prevalence is directly correlated with the standard U-Net's poor performance, which achieved a DSC of only 71.4% for this phase. In contrast, the modified U-Net achieved a DSC of 84.7% for clay, marking a remarkable 18.63% improvement.

The IoU score showed an even more dramatic gain of 32.32% for this challenging phase. This demonstrates that the architectural enhancements, particularly the stronger regularization, enabled the model to learn discriminative features from a very limited number of examples. Crucially, this substantial improvement for minority phases did not come at the expense of the majority phases (porosity and quartz), where performance remained stable and high. This capability to accurately identify low-abundance components is precisely what makes the model more robust for real-world DRP applications where such phases can govern the overall rock behavior.

Table 2. Summary of Key Differences Between Standard and Modified U-Net Models.

Feature	Standard U-Net	Modified U-Net
Encoder Convolution	Standard Convolution	Depthwise Separable Convolution
Kernel L2 Regularization	None	Yes (factor 1×10^{-4} in all sections)
Bias L2 Regularization	None	Yes (factor 1×10^{-4} in all sections)
Spatial Dropout	None	2D Spatial Dropout with rate 0.25 in encoder
Decoder Convolution	Standard Convolution	Standard Convolution
Optimizer & Learning Rate	Adam with rate 1×10^{-3}	Adam with rate 1×10^{-3}
Kernel Initializer	he_uniform	he_uniform

Table 3. Overall segmentation performance on the blind test set (sample c).

Improvement	Standard U-Net	Modified U-Net	Improvement
Dice Similarity Coefficient (%)	88.0	91.7	+ 4.25
Intersection over Union (%)	80.2	85.2	+ 6.33
Pixel Accuracy (%)	90.0	92.5	+ 2.79

Table 4. Per-phase dice similarity coefficient (dsc) on the blind test set (sample c).

Phase (%)	Standard U-Net	Modified U-Net	Improvement
Porosity	98.3	98.0	- 0.37
Quartz	97.2	96.9	- 0.23
Clay	71.4	84.7	+ 18.63
Feldspar	85.0	87.2	+ 2.62

Table 5. Per-phase intersection over union (iou) on the blind test set (sample c).

Phase (%)	Standard U-Net	Modified U-Net	Improvement
Porosity	96.7	96.0	- 0.72
Quartz	94.5	94.1	- 0.45
Clay	55.6	73.5	+ 32.32
Feldspar	73.9	77.3	+ 4.64

3.2. Qualitative visual evaluation

A visual inspection of the segmentation results offers crucial practical insights that complement the quantitative metrics. Figure 4 provides a side-by-side comparison of the original µCT image, the ground truth, the standard U-Net output, and the modified U-Net output for representative slices from the test set.

A qualitative analysis of these slices consistently underscores the superior performance of the modified U-Net. The key visual improvements directly translate to more reliable outcomes in mining and geo-engineering applications:

1. **Improved Delineation of Minority Phases:** The modified U-Net excels at identifying small, dispersed clay particles which the standard U-Net often misses or misclassifies (Figure 4). This is a direct reflection of its superior quantitative performance on this phase, despite clay only constituting 1.0% of the test sample's volume (Table 1). This capability is critically important in both mineral processing and geomechanics. In mineral processing, accurately identifying clay is vital because it can adversely affect flotation and dewatering processes. In geomechanics, the presence and distribution of clays significantly influence the rock's hydro-mechanical properties, such as strength and swelling potential, which are key inputs for assessing the stability of pit slopes and underground excavations.
2. **Sharper and More Accurate Inter-Phase Boundaries:** The boundaries between all phases are delineated with greater precision by the modified U-Net. This translates directly to more accurate 3D Mineral Liberation Analysis (MLA). Precise boundary definition is the foundation for calculating the degree to which valuable minerals are locked with gangue. An accurate MLA model, built on a high-fidelity segmentation like ours, allows engineers to optimize grinding energy and reagent dosage, thereby enhancing metallurgical recovery.
3. **Reduced Misclassification and Noise:** The cleaner segmentation maps from the modified U-Net, with fewer misclassified pixels, indicate greater robustness. For DRP-based simulations, this reduces uncertainty. When simulating fluid flow for mine dewatering studies or stress distribution for stability analysis, spurious misclassifications can introduce significant errors into the model, leading to unreliable predictions. Our model's output provides a more trustworthy digital twin of the rock fabric.
4. **Better Preservation of Fine Structures:** The modified U-Net does a better job of preserving subtle structural details, both within the pore space and in complex mineral arrangements, resulting in a more faithful representation of the true microstructure.

These visual improvements strongly corroborate the quantitative findings and highlight the practical value of the architectural enhancements for producing high-fidelity segmentations.

3.3. Training dynamics and model stability

To assess the learning behavior and stability of the models, we analyzed their training and validation loss curves over the course of training (Figure 5). These curves, especially the validation loss, offer insight into how well the models generalize and whether they are prone to overfitting.

The standard U-Net model (left, Figure 5) displayed considerable volatility in its validation loss, with prominent spikes (e.g., to values above 0.6) around epochs 5 and 15. This erratic behavior suggests instability in the learning process and a tendency to overfit the small training dataset, which can compromise the model's reliability on new data.

In stark contrast, the modified U-Net (right, Figure 5) demonstrated far greater stability. After a few initial epochs, its validation loss curve follows a much smoother, more consistent downward trend that closely tracks the training loss. This stability is a clear testament to the positive impact of the architectural modifications—spatial dropout and L2 regularization—in mitigating overfitting and enhancing the model's ability to generalize. This stable learning behavior aligns perfectly with the model's superior performance on the blind test set (Tables 2-4). Both models benefited from an early stopping mechanism, which ensured they were evaluated at their peak performance on the validation set.

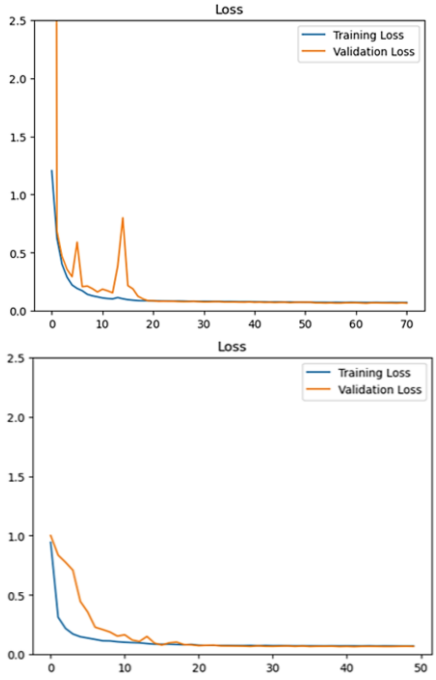

Figure 5. Training and validation loss curves for the standard (up) and modified (down) U-Net models during training.

3.4. Computational considerations

A brief comparison of the computational costs reveals a modest trade-off. The modified U-Net, with its depthwise separable convolutions and additional regularization steps, has a slightly larger number of trainable parameters (an increase of about 7-9%) compared to the standard U-Net. From a practical engineering standpoint, this marginal increase in computational cost is highly justified by the substantial gains in segmentation accuracy and model reliability. The cost of inaccurate geological or geomechanical models in mining—which can lead to inefficient mineral recovery or, in the worst case, catastrophic slope or tunnel failure—far outweighs the minor additional investment in computational resources during the model training phase. Therefore, this represents a highly favorable trade-off for any application where high-fidelity microstructural characterization is paramount to operational success and safety.

4. Conclusion

In this study, we conducted a systematic comparative evaluation of a standard U-Net architecture against a bespoke, modified U-Net for the challenging task of multi-phase segmentation of µCT images of Bentheimer sandstone. Our primary goal was to determine whether targeted architectural enhancements—specifically, the integration of optimized convolutional blocks, spatial dropout, and L2 regularization—could yield significant improvements in segmentation accuracy and robustness over a conventional baseline.

The comprehensive evaluation, conducted on a large, independent blind test set, unequivocally demonstrated the superiority of the modified U-Net architecture. Quantitatively, our model achieved higher macro-averaged scores across all key metrics. Most critically, the model delivered substantial improvements in the segmentation of challenging, low-abundance minority phases like clay—which comprised only 1.0% of the test data—an enhancement that is vital for the reliability of subsequent geomechanical and mineral processing simulations.

These quantitative gains were further validated by qualitative visual assessments, which revealed that the modified U-Net produced cleaner segmentation maps with sharper inter-phase boundaries and superior preservation of fine structural details. These significant performance improvements were realized with only a modest increase in model parameters and computational time, signaling a highly advantageous trade-off.

The principal contribution of this work is the demonstration that a synergistic combination of architectural modifications can markedly elevate the performance of U-Net models for complex µCT image segmentation, thereby providing a more trustworthy input for Digital Rock Physics (DRP) workflows. We posit that the optimized convolutional blocks enhanced feature extraction capabilities, while the dual regularization strategies effectively curtailed overfitting and promoted better generalization.

While these results are highly encouraging, our findings are based on a single, albeit standard, type of porous medium. Future research should extend this work along several promising avenues. Evaluating the proposed architecture on a wider array of geological samples, particularly ore-bearing rocks and samples from excavation-damaged zones, would be a valuable next step. Adapting these enhancements to fully 3D U-Net frameworks could unlock further benefits for volumetric segmentation.

In conclusion, the modified U-Net architecture presented here offers a tangible improvement over the standard U-Net, directly addressing a critical bottleneck in the DRP workflow. By providing a more accurate, robust, and reliable tool for segmenting complex microstructures, this work contributes to enhancing the fidelity of downstream models used in the mining and geo-engineering sectors. Ultimately, a more accurate segmentation foundation leads to more reliable predictions of rock mass behavior and mineral processing efficiency, contributing to safer and more economically viable engineering designs.

References

[1] Zou, S., & Sun, C. (2020). X-ray microcomputed imaging of wettability characterization for multiphase flow in porous media: A review. Capillarity, 3(3), 36–44. doi: https://doi.org/10.46690/capi.2020.03.01

[2] Li, X., Li, B., Liu, F., Li, T., & Nie, X. (2023). Advances in the application of deep learning methods to digital rock technology. Advances in Geo-Energy Research, 8(1), 5–18. doi: https://doi.org/10.46690/ager.2023.04.02

[3] Guntoro, P. I., Ghorbani, Y., Koch, P.-H., & Rosenkranz, J. (2019). X-ray Microcomputed Tomography (µCT) for Mineral Characterization: A Review of Data Analysis Methods. Minerals, 9(3), 183. https://doi.org/10.3390/min9030183

[4] Tung, Patrick, Halim, Amalia, Wang, Helen, Rich, Anne, Chen, Xiao, Regenauer-Lieb, Klaus, & Marjo, Christopher. (2024). Deep learning 3D-mineral liberation analysis with micro-X-ray fluorescence, micro-computed tomography, and deep learning segmentation. BIO Web Conf., 129, 27004. https://doi.org/10.1051/bioconf/202412927004

[5] Feali, M., Pinczewski, W. V., Cinar, Y., Arns, C., Arns, J.-Y., Turner, M., Senden, T., Francois, N., & Knackstedt, M. A. (2012). Qualitative and quantitative analyses of the three-phase distribution of oil, water, and gas in Bentheimer sandstone by use of micro-CT imaging. SPE Reservoir Evaluation & Engineering. doi: https://openresearch-repository.anu.edu.au/items/099f39fc-2db0-4964-913b-62ab8c499492

[6] Liu, M., & Mukerji, T. (2022). Digital transformation in rock physics: Deep learning and data fusion. The Leading Edge, 41(9), 591–598. doi: https://doi.org/10.1190/tle41090591.1

[7] Center of Innovation for Flow Through Porous Media (COIFPM). "Micro-scale Tomography." COIFPM, coifpm.org/index.php/research/micro-scale-tomography. Accessed 31 May 2025. *https://www.coifpm.org/index.php/research/micro-scale-tomography*

[8] Wetzel, M., Kempka, T., & Kühn, M. (2020). Hydraulic and mechanical impacts of pore space alterations within a sandstone quantified by a flow velocity-dependent precipitation approach. Materials, 13(14), 3100. doi: https://doi.org/10.3390/ma13143100

[9] Shah, S. M., Crawshaw, J. P., & Boek, E. S. (2014). Bentheimer Sandstone [Dataset]. figshare. doi: https://doi.org/10.6084/m9.figshare.1200096.v1

[10] Brondolo, F., & Beaussant, S. (2024). DINOv2 rocks geological image analysis: Classification, segmentation, and interpretability [Preprint]. arXiv. https://arxiv.org/abs/2407.18100

[11] Nickerson, S., Shu, Y., Zhong, D., Könke, C., & Tandia, A. (2019). Permeability of porous ceramics by X-ray CT image analysis. Acta Materialia, 172, 121–130. doi: https://doi.org/10.1016/j.actamat.2019.04.053

[12] Wildenschild, D., & Sheppard, A. P. (2013). X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. Advances in Water Resources, 51, 217–246. doi: https://doi.org/10.1016/j.advwatres.2012.07.018

[13] TESCAN. (n.d.). Micro-CT solutions. Retrieved June 4, 2025, from https://info.tescan.com/micro-ct

[14] Cobos, S. F., Norley, C. J., Pollmann, S. I., & Holdsworth, D. W. (2022). Cost-effective micro-CT system for non-destructive testing of titanium 3D printed medical components. PLOS ONE, 17(10), e0275732. doi: https://doi.org/10.1371/journal.pone.0275732

[15] Cengiz, I. F., Oliveira, J. M., & Reis, R. L. (2018). Micro-CT – A digital 3D microstructural voyage into scaffolds: A systematic review of the reported methods and results. Biomaterials Research, 22, 26. doi: https://doi.org/10.1186/s40824-018-0136-8

[16] Karimpouli, S., & Tahmasebi, P. (2019). Segmentation of digital rock images using deep convolutional autoencoder networks. Computers & Geosciences, 126, 142–150. doi: https://doi.org/10.1016/j.cageo.2019.02.003

[17] Madonna, C., Almqvist, B. S. G., & Saenger, E. H. (2012). Digital rock physics: Numerical prediction of pressure-dependent ultrasonic velocities using micro-CT imaging. Geophysical Journal International, 189(3), 1475–1482. doi: https://doi.org/10.1111/j.1365-246X.2012.05437.x

[18] Al-Marzouqi, H. (2018). Digital rock physics: Using CT scans to compute rock properties. IEEE Signal Processing Magazine, 35(2), 121–131. doi: https://doi.org/10.1109/MSP.2017.2784459

[19] Balcewicz, M., Siegert, M., Gurris, M., Ruf, M., Krach, D., Steeb, H., & Saenger, E. H. (2021). Digital rock physics: A geological driven workflow for the segmentation of anisotropic Ruhr sandstone. *Frontiers in Earth Science*, *9*, 673753. https://doi.org/10.3389/feart.2021.673753

[20] Wang, H., Guo, R., Dalton, L. E., Crandall, D., Hosseini, S. A., Fan, M., & Chen, C. (2024). Comparative assessment of U-Net-based deep learning models for segmenting microfractures and pore spaces in digital rocks. SPE Journal, 29, 5779–5791. doi: https://doi.org/10.2118/215117-PA

[21] Roslin, A., Marsh, M., Provencher, B., Mitchell, T. R., Onederra, I. A., & Leonardi, C. R. (2023). Processing of micro-CT images of granodiorite rock samples using convolutional neural networks (CNN), Part II: Semantic segmentation using a 2.5D CNN. Minerals Engineering, 195, 108027. doi: https://doi.org/10.1016/j.mineng.2023.108027

[22] Palacio-Mancheno, P. E., Larriera, A. I., Doty, S. B., Cardoso, L., & Fritton, S. P. (2014). 3D assessment of cortical bone porosity and tissue mineral density using high-resolution μCT: Effects of resolution and threshold method. Journal of Bone and Mineral Research, 29(1), 142–150. doi: https://doi.org/10.1002/jbmr.2012

[23] Li, Q., Ma, C., Zhang, C., Guo, Y., & Zhou, T. (2024). Study on the Microstructure and Permeability Characteristics of Tailings Based on CT Scanning Technology. Applied Sciences, 14(24), 12032. doi: https://doi.org/10.3390/app142412032

[24] Chawshin, K., Berg, C. F., Varagnolo, D., & et al. (2022). Automated porosity estimation using CT-scans of extracted core data. Computational Geosciences, 26, 595–612. doi: https://doi.org/10.1007/s10596-022-10143-9

[25] Kornilov, A., Safonov, I., & Yakimchuk, I. (2022). A Review of Watershed Implementations for Segmentation of Volumetric Images. Journal of Imaging, 8(5), 127. doi: https://doi.org/10.3390/jimaging8050127

[26] Zhao, L., Zhang, H., Sun, X., Ouyang, Z., Xu, C., & Qin, X. (2024). Application of ResUNet-CBAM in Thin-Section Image Segmentation of Rocks. Information, 15(12), 788. doi: https://doi.org/10.3390/info15120788

[27] Wang, H., Yang, X., Zhou, C., Yan, J., Yu, J., & Xie, K. (2025). Constructions of multi-scale 3D digital rocks by associated image segmentation method. Frontiers in Earth Science, 12. doi: https://doi.org/10.3389/feart.2024.1518561

[28] Kornilov, A. S., & Safonov, I. V. (2018). An Overview of Watershed Algorithm Implementations in Open Source Libraries. Journal of Imaging, 4(10), 123. doi: https://doi.org/10.3390/jimaging4100123

[29] Yu, Q., Wang, G., Cheng, H., Guo, W., & Liu, Y. (2024). The segmentation and intelligent recognition of structural surfaces in borehole images based on the U2-Net network. PLOS ONE, 19(3), e0299471. doi: https://doi.org/10.1371/journal.pone.0299471

[30] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7), 3523–3542. doi: https://doi.org/10.1109/TPAMI.2021.3059968

[31] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. arXiv. https://arxiv.org/abs/1411.4038

[32] Liu, X., Deng, Z., & Yang, Y. (2019). Recent progress in semantic image segmentation. Artificial Intelligence Review, 52, 1089–1106. doi: https://doi.org/10.1007/s10462-018-9641-3

[33] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (Vol. 9351, pp. 234–241). Springer. doi: https://doi.org/10.1007/978-3-319-24574-4_28

[34] Azad, R., Khodapanah Aghdam, E., Rauland, A., Jia, Y., Haddadi Avval, A., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical image segmentation review: The success of U-Net. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 10076–10095. doi: https://doi.org/10.1109/TPAMI.2024.3435571

[35] Aldi, F., Yuhandri, & Tajuddin, M. (2024). Enhanced U-Net architecture for glottis segmentation with VGG-16. Journal of Informatics and Visualization, 8(4), 2173–2180. doi: https://doi.org/10.62527/joiv.8.4.3088

[36] Safarov, F., Khojamuratova, U., Komoliddin, M., Kurbanov, Z., Tamara, A., Nizamjon, I., Muksimova, S., & Cho, Y. I. (2025). Lightweight evolving U-Net for next-generation biomedical imaging. Diagnostics, 15(9), 1120. doi: https://doi.org/10.3390/diagnostics15091120

[37] Alshawi, R., Hoque, M. T., & Flanagin, M. C. (2023). A Depth-Wise Separable U-Net Architecture with Multiscale Filters to Detect Sinkholes. Remote Sensing, 15(5), 1384. doi: https://doi.org/10.3390/rs15051384

[38] Chandra, N., Sawant, S., & Vaidya, H. (2023). An efficient U-Net model for improved landslide detection from satellite images. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 91, 13–28. doi: https://doi.org/10.1007/s41064-023-00232-4

[39] Mahmoudi, S., Asghari, O., & Boisvert, J. (2025). Addressing class imbalance in micro-CT image segmentation: A modified U-Net model with pixel-level class weighting. Computers & Geosciences, 196, 105853. doi: https://doi.org/10.1016/j.cageo.2025.105853

[40] Sarsembayeva, T., Mansurova, M., Abdildayeva, A., & Serebryakov, S. (2025). Enhancing U-Net Segmentation Accuracy Through Comprehensive Data Preprocessing. Journal of Imaging, 11(2), 50. doi: https://doi.org/10.3390/jimaging11020050

[41] Jin, Z., Li, X., Yang, H., Wu, B., & Zhu, X. (2023). Depthwise separable convolution U-Net for 3D seismic data interpolation. Frontiers in Earth Science, 10, 1005505. doi: https://doi.org/10.3389/feart.2022.1005505

[42] Cheng, H. (2024). Lightweight depthwise separable U-Net for laptop-based abdominal multi-organ segmentation. OpenReview. https://openreview.net/forum?id=R9MagNE8ub

[43] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

[44] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. arXiv. https://arxiv.org/abs/1411.4280

[45] Lee, S., & Lee, C. (2020). Revisiting spatial dropout for regularizing convolutional neural networks. Multimedia Tools and Applications, 79, 34195–34207. doi: https://doi.org/10.1007/s11042-020-09054-7

[46] Simplilearn. (2025, March 26). The best guide to regularization in machine learning. https://www.simplilearn.com/tutorials/machine-learning-tutorial/regularization-in-machine-learning

[47] Liang, J., Sun, Y., Lebedev, M., Gurevich, B., Nzikou, M., Vialle, S., & Glubokovskikh, S. (2022). Multi-mineral segmentation of micro-tomographic images using a convolutional neural network. *Computers & Geosciences*, *168*, 105217. https://doi.org/10.1016/j.cageo.2022.105217

[48] Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization [Preprint]. arXiv. https://arxiv.org/abs/1412.6980

[49] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/