ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF COMPUTER AND INFORMATICS

Segmenting Neonatal Brain Lesions from Low-Resolution Diffusion MRI

Graduation Project Final Report

Büşra Koyun, Ayça Pektaş
150180072, 150180108

Department: Computer Engineering
Division: Computer Engineering

Advisor: Assistant Professor Dr. Yusuf Hüseyin Şahin

June 2025

Statement of Authenticity

We hereby declare that in this study

1. all the content influenced from external references are cited clearly and in detail,
2. and all the remaining sections, especially the theoretical studies and implemented software that constitute the fundamental essence of this study is originated by our individual authenticity.

İstanbul, June 2025

Büşra Koyun, Ayça Pektaş

Acknowledgments

We sincerely thank our advisor, Assistant Professor Dr. Yusuf Hüseyin Şahin, for his continuous guidance, insightful feedback, and encouragement throughout the course of this project. His expertise and support were vital to the development of our work, both conceptually and technically.

We are especially grateful to him for providing access to laboratory computing resources, which enabled us to conduct extensive experiments and complete our research efficiently.

Segmenting Neonatal Brain Lesions from Low-Resolution Diffusion MRI

(SUMMARY)

Accurate segmentation of Hypoxic Ischemic Encephalopathy (HIE) lesions in neonatal diffusion MRI remains a significant challenge due to the small, diffuse nature of the lesions, as well as the constraints introduced by accelerated imaging techniques. Detecting these small lesions becomes especially difficult when MRI scans are acquired at low spatial resolutions, which is common in neonatal care where motion artifacts are prevalent and scanning times need to be minimized. The complexity of this task is further heightened by the presence of aliasing artifacts in the undersampled MRI data, which degrade image quality and complicate the segmentation process.

In this study, we tackle the issue of neonatal HIE lesion segmentation from low-resolution MR images by simulating realistic clinical conditions through the application of 4-fold equispaced undersampling of Apparent Diffusion Coefficient (ADC) and Z-score ADC (Z_{ADC}) volumes from the BONBID-HIE dataset. By simulating undersampling, we generate paired aliased images via inverse Fourier reconstruction, thereby replicating the data constraints typically encountered in clinical environments. We train three different segmentation models under these settings to evaluate their effectiveness in handling the challenges posed by the undersampled data. The first model is a baseline SwinUNETR architecture, which operates without the incorporation of frequency-domain information. The second model enhances the baseline by adding global frequency features derived from Discrete Cosine Transform (DCT), providing the model with additional spectral information. The third and most advanced model integrates localized frequency features by applying a Block DCT layer in the early encoder stages, allowing for a more spatially distributed representation of frequency information.

Our experimental results demonstrate that lesion segmentation from undersampled images is not only feasible but also significantly improved by the inclusion of frequency-domain priors. Specifically, the Block DCT model, which uses localized frequency features, achieved the highest Dice score, highlighting its ability to effectively capture the fine-grained structures of the lesions despite the reduced data quality. The incorporation of frequency-domain information improves segmentation robustness and accuracy, even when the data is heavily undersampled, as is the case in neonatal diffusion MRI.

This work underscores the potential benefits of combining transformer-based architectures, such as SwinUNETR, with frequency-aware augmentation techniques. By leveraging both spatial and frequency-domain information, our models demonstrate enhanced performance in the context of limited k-space sampling, which is common in clinical MRI acquisition protocols. The findings suggest that such hybrid approaches, which combine the strengths of both spatial and spectral feature representations, can lead to more robust segmentation models capable of handling the complex challenges of neonatal brain imaging.

Düşük Çözünürlüklü Difüzyon MRI'dan Yenidoğan Beyin Lezyonlarının Segmentasyonu

(ÖZET)

Neonatal hipoksik-iskemik ensefalopati (HIE) lezyonlarının difüzyon MRI üzerinde doğru şekilde segmentasyonu, lezyonların küçük ve yaygın yapısı ile hızlandırılmış görüntüleme yöntemlerinin kısıtlamaları nedeniyle oldukça zorlu bir görevdir. Bu tür küçük lezyonların tespiti, özellikle çok düşük çözünürlüklü MR görüntülerinin elde edilmesi durumunda daha da zorlaşır. Bu durum, neonatal bakımda hareket artefaktlarının yaygın olması ve tarama sürelerinin kısaltılmasının gerekli olması gibi klinik sınırlamalardan kaynaklanmaktadır. Ayrıca, kısmi örnekleme yapılan MR verilerinde aliasing artefaktlarının bulunması, görüntü kalitesini bozar ve segmentasyon işlemini karmaşıklaştırır.

Bu çalışmada, düşük çözünürlüklü MR görüntülerinden neonatal HIE lezyonlarının segmentasyonu görevini ele alıyoruz. Gerçek klinik koşulları simüle etmek için BONBID-HIE veri setinden alınan Apparent Diffusion Coefficient (ADC) ve Z-score ADC (Z_{ADC}) hacimlerinin 4 katlı eşit örnekleme ile alt örneklenmesi gerçekleştirilmiştir. Ardından, ters Fourier dönüşümü ile aliasing görüntüleri üretilmiş ve bu koşullarda üç farklı segmentasyon modeli eğitilmiştir. İlk model, temel bir SwinUNETR modelidir ve bu modelde frekans-domaini bilgisi kullanılmamaktadır. İkinci model, temel modelin üzerine global frekans özellikleri eklenmiş ve Discrete Cosine Transform (DCT) ile elde edilen frekans bilgileri modelin üçüncü giriş kanalı olarak dahil edilmiştir. Üçüncü ve en ileri model ise, yerel frekans özelliklerini erken encoder katmanlarına entegre eden bir Block DCT katmanı içermektedir. Bu model, daha lokalize frekans bilgileri ile segmentasyon yapılmasına olanak sağlar.

Deneysel sonuçlar, alt örneklenmiş görüntülerden lezyon segmentasyonunun mümkün olduğunu ve frekans-domaini önceliklerinin segmentasyon performansını önemli ölçüde iyileştirdiğini göstermektedir. Özellikle Block DCT modeli, yerel frekans özelliklerini kullandığında en yüksek Dice skoru elde edilmiştir, bu da lezyonların küçük yapılarının düşük çözünürlükteki verilere rağmen etkili bir şekilde yakalanabildiğini göstermektedir. Frekans-domaini bilgisinin eklenmesi, segmentasyonun dayanıklılığını ve doğruluğunu artırmakta, özellikle neonatal difüzyon MRI'da veri kısıtlamaları altında segmentasyon yapılabilmesini sağlamaktadır.

Bu çalışma, SwinUNETR gibi transformer tabanlı mimarilerin, frekans bilgisiyle artırılmış augmentasyon teknikleriyle birleşmesinin potansiyel faydalarını vurgulamaktadır. Hem mekansal hem de frekans-domaini bilgilerini kullanarak, bu modeller, klinik olarak kısıtlı k-space örnekleme koşullarında bile daha güçlü segmentasyon performansı göstermektedir. Çalışmamız, frekans-domaini bilgisi ile birleşen bu tür hibrit yaklaşımların, neonatal beyin görüntülemede karşılaşılan karmaşık zorluklarla başa çıkabilen daha güçlü segmentasyon modellerine yol açabileceğini ortaya koymaktadır.

Contents

1 Introduction and Project Summary

Magnetic Resonance Imaging (MRI) is an essential tool in neonatal neuroimaging, offering high soft-tissue contrast and non-invasive insights into early brain development. However, traditional MRI protocols require prolonged acquisition times, which pose difficulties in neonatal care due to increased risk of motion artifacts and the limited tolerance of newborns. To mitigate these constraints, acquisition acceleration techniques such as parallel imaging and compressed sensing have been widely adopted. These methods reduce acquisition time by undersampling the k-space domain, but often introduce aliasing artifacts that affect image quality and complicate downstream tasks such as lesion segmentation.

Neonatal Hypoxic Ischemic Encephalopathy (HIE) is a severe birth-related condition that leads to brain injury and long-term neurological deficits. Transformer-based models such as SwinUNETR [1] have recently emerged as strong candidates for medical image segmentation, offering long-range feature modeling and hierarchical representations that are well-suited to detecting small and scattered abnormalities. However, segmenting brain lesions in neonatal Hypoxic Ischemic Encephalopathy (HIE) is challenging due to their small size, diffuse nature, and low contrast in MRI. Results from the BONBID-HIE Challenge highlight this difficulty, with state-of-the-art models such as 3D U-Net and SwinUNETR achieving Dice scores of only around 0.5 on average across participants [2]. This is significantly lower than segmentation performance reported in other brain imaging tasks and underscores the need for approaches specifically tailored to the unique characteristics of neonatal lesions.

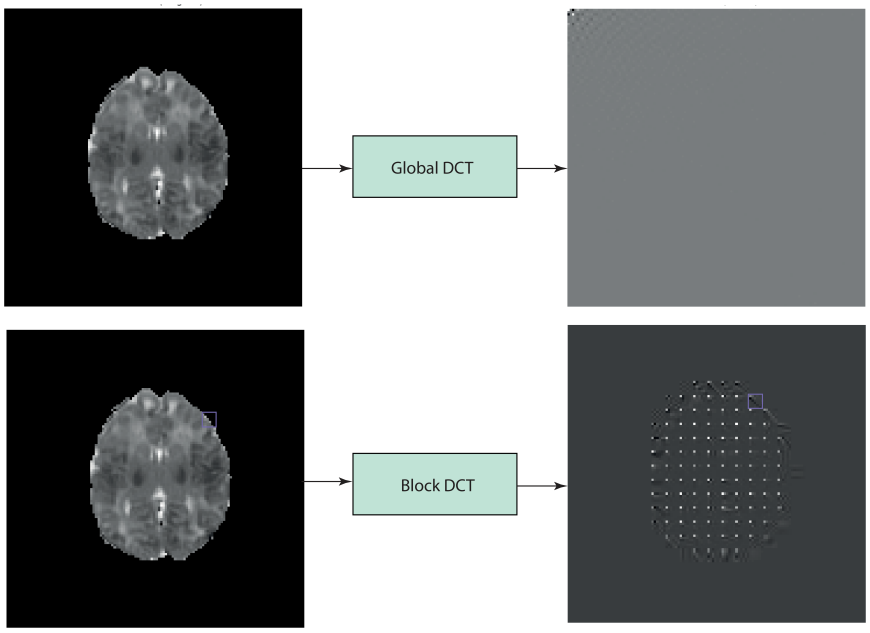


Figure 1.1: Comparison of global and block-wise DCT representations. Left side illustrates the transformation of the input image into its global frequency domain representation using Global DCT. Right side demonstrates localized feature extraction using Block DCT, where frequency information is preserved in spatially distributed patches.

In this work, we explore transformer-based segmentation models under realistically simulated acquisition conditions. This is the first study to perform segmentation directly on low-resolution MRI without reconstructing them into high-resolution counterparts. We construct a segmentation framework based on SwinUNETR and evaluate three variants of the progressively advanced model. The first is a dual-channel baseline that takes Apparent Diffusion Coefficient (ADC) and Z-score ADC (Z_{ADC}) maps as input. The second augments the baseline by appending global Discrete Cosine Transform (DCT) features, treated as a third input channel to introduce frequency-domain information. The third and most advanced model incorporates a Block DCT layer that divides the image into spatial patches and calculates local DCT for each patch. This Block DCT integration allows the network to learn regional frequency priors during training, enriching the internal representation of spatial and spectral lesion cues. An example of DCT and Block DCT calculation on the same image is given in Figure 1.1.

Through comparative experiments with undersampled diffusion MRI data from the BONBID-HIE dataset[3], we demonstrate that this architectural design enhances segmentation robustness and improves precision in detecting small neonatal lesions. The results highlight the benefit of combining hierarchical attention with embedded frequency priors for segmentation under clinically constrained imaging conditions. The remainder of this paper is organized as follows. Section 2 reviews prior work on neonatal MRI segmentation and frequency-domain learning in medical imaging. Section 3 describes our proposed models and the architecture of the learnable DCT layer. Section 4 details the dataset, preprocessing steps, and training configurations. Section 5 presents quantitative and qualitative evaluation results. Finally, Section 6 concludes the paper and outlines potential directions for future research.

2 Comparative Literatre Survey

Neonatal HIE lesion segmentation is particularly challenging due to the small size, diffuse nature, and subtle intensity of the lesions. Traditional methods based on handcrafted features and thresholding have shown limited success, especially under real-world conditions such as low signal-to-noise ratios and sparse annotations.

The BONBID-HIE Challenge has emerged as a comprehensive benchmark for this task, with transformer-based approaches—particularly SwinUNETR demonstrating strong performance. Participating teams applied strategies such as extensive data augmentation, multi-task learning, and post-processing with classical machine learning techniques like random forests, highlighting a shift toward hybrid approaches that combine anatomical priors with data-driven learning.

Transformer models have gained attention for their ability to model long-range spatial dependencies, crucial for detecting spatially sparse lesions. SwinUNETR [1], built on the Swin Transformer [4], achieves this through hierarchical attention mechanisms and UNet-like skip connections, enabling efficient multi-scale feature extraction.

2.1 Frequency-Domain Learning and Discrete Cosine Transform

Discrete Cosine Transform (DCT) is a classical frequency-domain method that compacts energy and decorrelates image data, making it valuable for image representation. In recent years, DCT has been successfully incorporated into deep learning pipelines to enhance both segmentation and reconstruction performance in medical imaging [5]. These approaches leverage the spectral priors introduced by DCT to improve model robustness, especially in scenarios with limited spatial information or undersampled data [6].

Recent studies have used DCT as a component in deep learning pipelines, especially in computer vision tasks that benefit from spectral representations. Lightweight DCT-based models improve accuracy and reduce input size without modifying CNNs [7]. Embedding DCT-derived features into hybrid CNN transformer architectures appears to improve attention alignment and feature expressiveness across scales [8]. Further, DCT has been utilized to initialize and compress attention matrices in vision transformers, leading to more efficient training and inference with preserved accuracy [9].

In medical imaging, DCT has proven useful both as a sparsity-inducing prior and a complementary frequency representation. DCT-Net [5], a dual-domain transformer that leverages DCT information alongside spatial features to reconstruct high-fidelity MR images from undersampled k-space acquisitions, exemplifies this integration. DCT-based features have also guided segmentation models in noisy MRI scans, showing increased robustness in localizing brain tumor regions in structurally complex or artifact-prone MRIs [10]. Super-resolution methods that fuse DCT with convolutional networks have demonstrated enhanced tumor classification accuracy on low-resolution MRIs [11].

Additionally, filtering strategies using DCT coefficients have been developed to denoise MRI volumes by selectively shrinking high-frequency components, effectively preserving structural detail while suppressing noise artifacts. Together, these recent developments highlight the growing value of DCT as a frequency-aware component in deep learning frameworks, enabling more accurate, efficient, and resilient MRI reconstruction and segmentation—particularly under clinically constrained conditions.

Recent work in image denoising, such as DCTNet [12], demonstrates that embedding Discrete Cosine Transform (DCT) filterbanks within neural networks can improve sensitivity to subtle image structures by leveraging the energy compaction properties of DCT. These Block-based DCT variants also allow spatially localized frequency representations, which are particularly effective in highlighting texture abnormalities in small regions.

However, most prior applications of DCT remain in denoising or global feature extraction and are not tightly coupled with segmentation backbones. Our work builds on this foundation by directly incorporating both global and Block DCT representations into SwinUNETR, evaluating their impact on lesion segmentation performance under clinically constrained (i.e., k-space undersampled) MRI scenarios.

Undersampling of k-space in MRI is a widely adopted strategy to reduce scan duration, particularly in time-sensitive or vulnerable populations such as neonates. Long acquisition times in conventional MRI not only increase the risk of motion artifacts but also limit accessibility and patient comfort [13]. In neonatal imaging, these concerns are especially critical due to the difficulty in keeping infants still and the clinical desire to avoid sedation. To simulate realistic clinical conditions, our study applies a $4\times$ equispaced undersampling strategy, consistent with modern accelerated imaging protocols. This approach has been validated in large-scale clinical studies showing that up to $8\times$ acceleration can maintain diagnostic image quality [14]. Deep learning-based sampling and reconstruction methods have also shown high-fidelity results with $8\times$ undersampling, demonstrating that aggressive acceleration is technically feasible [14]. Such techniques have also been employed in segmentation challenges using $8\times$ undersampled data to reflect practical acquisition settings [15]. These findings support our simulation design and reflect the growing clinical viability of accelerated MRI.

To the best of our knowledge, our study is the first to systematically evaluate an undersampling strategy in conjunction with dual- and tri-channel transformer segmentation models using k-space simulated neonatal MRI data. By comparing pure spatial-domain transformer models with frequency-augmented variants, we seek to bridge the gap between robust spatial representation and frequency-aware lesion enhancement under data-constrained settings.

3 Developed Approach and System Model

3.1 Dataset

Hypoxic Ischemic Encephalopathy (HIE) is a neonatal brain injury caused by perinatal oxygen deprivation and ischemia. It affects approximately 1 to 5 per 1,000 term-born babies and remains a major contributor to neonatal morbidity and mortality worldwide. Accurate detection of brain lesions associated with HIE on MRI is vital for early prognosis, risk assessment, and treatment evaluation. However, HIE lesions are typically small, diffuse, and multifocal. They occupy less than 1% of brain in more than half of affected neonates, which pose challenges for automated segmentation algorithms [3].

To address the lack of publicly available datasets for HIE lesion segmentation, the BOston Neonatal Brain Injury Dataset for Hypoxic Ischemic Encephalopathy (BONBID-HIE) was introduced [3]. This dataset includes diffusion-weighted MRI data from 133 term-born neonates diagnosed with HIE which was acquired within the first 14 days after birth. For each subject, the dataset provides skull-stripped Apparent Diffusion Coefficient (ADC) maps (ADC_{ss}), standardized Z-scored ADC maps (Z_{ADC}), and manually annotated binary lesion masks (LABEL) confirmed by expert pediatric neuroradiologists. The Z_{ADC} maps quantify voxel-wise deviations from a normative neonatal ADC atlas, helping to distinguish pathology from region-specific normal variation. These imaging components, along with clinical data, form a comprehensive resource aimed at improving segmentation performance for small and anatomically variable brain lesions in neonates.

We used the publicly available BONBID-HIE dataset [3], which was released as part of the Brain Oxygenation and Neonatal Brain Injury Detection (BONBID) challenge. However, access to the test data is restricted and requires challenge participation, which was not available at the time of our experiments. Consequently,we used the publicly available training and validation data.

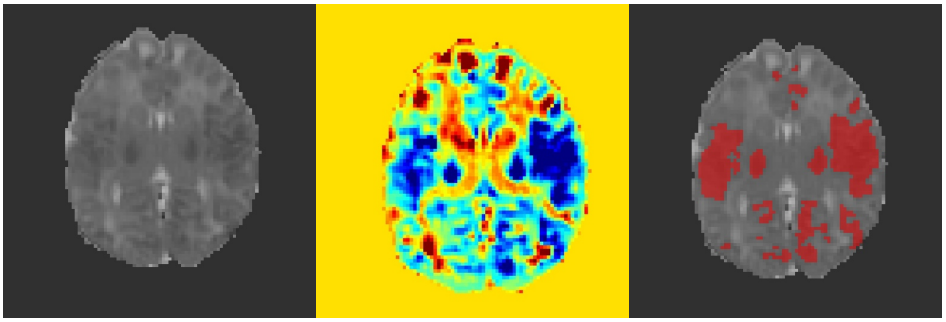
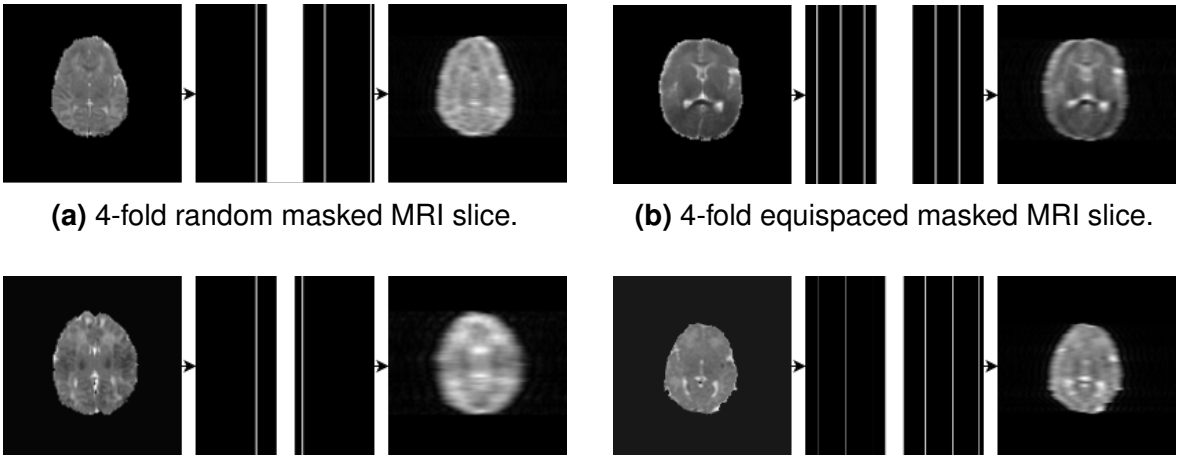


Figure 3.1: Example MRI maps from the BONBID-HIE dataset: ADC maps, Z-scored ADC maps (Z_{ADC}), and binary lesion masks (LABEL) for HIE lesion detection.

3.2 Preprocessing

The dataset consists of 3D diffusion-weighted MRI volumes from neonates diagnosed or suspected with HIE. Each case includes Apparent Diffusion Coefficient (ADC) maps, their z-score normalized counterparts (Z_{ADC}), and expert-annotated binary lesion masks.

Prior to model training, we simulated accelerated MRI acquisition by performing undersampling in the k-space domain. Specifically, slice-wise 2D equispaced undersampling masks were applied to both the Apparent Diffusion Coefficient (ADC) and Z-score ADC (Z_{ADC}) volumes. In addition, we implemented 4-fold and 8-fold random undersampling as well as 4-fold and 8-fold equispaced undersampling strategies to simulate realistic clinical acquisition conditions. For the 4-fold undersampling, a center fraction of 0.20 was selected, while for the 8-fold undersampling, a center fraction of 0.10 was chosen.



(a) 4-fold random masked MRI slice. **(b)** 4-fold equispaced masked MRI slice.

(c) 8-fold random masked MRI slice. **(d)** 8-fold equispaced masked MRI slice.

Figure 3.2: Visualization of the undersampling process with 4-fold and 8-fold random and equispaced masked MRI slices.

Each 3D volume was processed slice-by-slice:

1. The slice was converted to a complex-valued tensor and Fourier-transformed using a centered 2D FFT.
2. A binary undersampling mask was applied in the frequency domain.
3. The masked k-space was zero-filled and inverse-transformed to generate aliased images.
4. The magnitude of the resulting complex image was extracted to yield the final masked slice.

This process generated three output volumes per subject: masked ADC, masked Z_{ADC}, and the applied sampling mask. Example slices, sampling masks, and their corresponding results are shown in Fig.3.2.

3.3 Network Architecture

Our segmentation framework is built upon the SwinUNETR architecture, a hybrid vision transformer model that combines the hierarchical structure of U-Net with the self-attention mechanisms of Swin Transformers. This architecture is particularly effective for medical image segmentation tasks due to its capacity for capturing long-range dependencies and fine-grained spatial context in 3D volumes.

We explore two distinct architectural variants that integrate frequency-domain information to enhance lesion segmentation under undersampled diffusion MRI settings. The overall network architecture of the proposed segmentation framework is illustrated in Figure 3.3. As shown, the system leverages the SwinUNETR backbone to extract multiscale contextual features from the undersampled MRI input. The architecture incorporates a hierarchical transformer-based encoder and a convolutional decoder, allowing effective spatial and frequency-domain feature fusion.

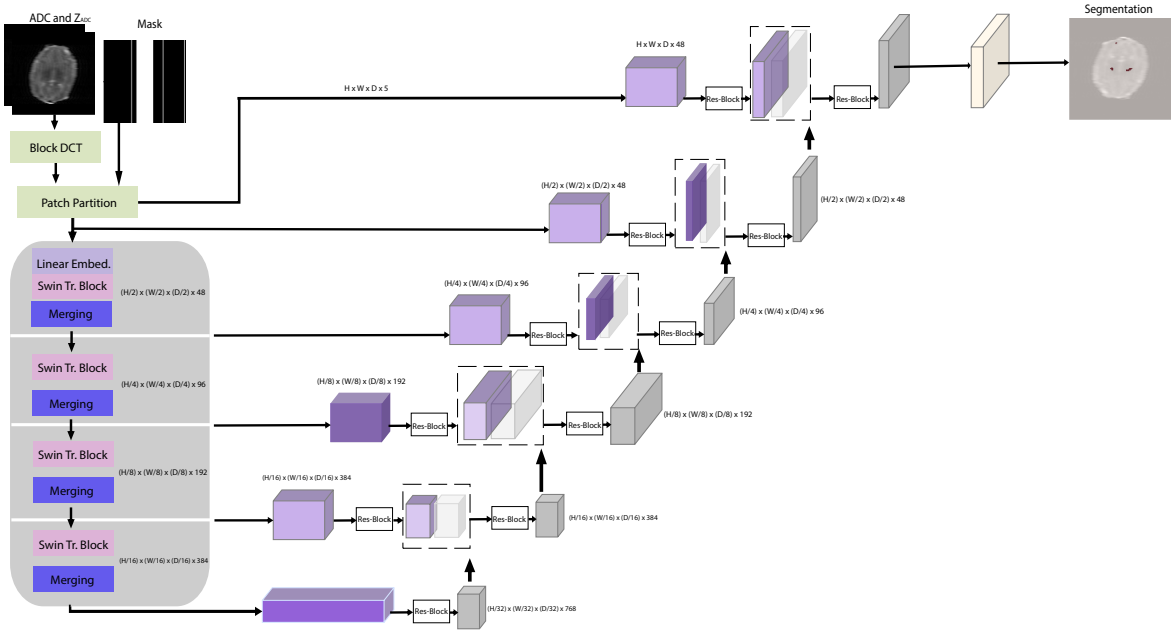


Figure 3.3: Network Architecture of the segmentation framework using SwinUNETR.

3.3.1 Baseline with Global DCT Features

The first variant introduces frequency information as an additional input channel to the model. A global Discrete Cosine Transform (DCT) is applied to the ADC volume, and the resulting frequency representation is concatenated with the original ADC and Z_{ADC} channels (see Figure 1.1):

$$I_{\text{input}} = \text{concat}(I_{\text{ADC}}, I_{Z_{\text{ADC}}}, \mathcal{D}(I_{\text{ADC}})) \tag{1}$$

where $\mathcal{D}(\cdot)$ denotes the DCT operation applied on each slice of the ADC input.

3.3.2 Frequency-Aware Encoding via Learnable DCT Blocks

The second variant incorporates a learnable frequency extraction module into the SwinUNETR encoder. This module, referred to as `VolumeDCTPatchLayer`, extracts localized frequency features from spatial regions of the input volume and fuses them with spatial features. These frequency descriptors are integrated into the early encoding layers of the model:

$$x_{\text{input}} = \text{concat}(I_{\text{ADC}}, I_{\text{Z}_{\text{ADC}}}, F_{\text{DCT}}) \tag{2}$$

where F_{DCT} represents the learnable frequency feature map computed from the ADC input. The combined representation is then processed by the transformer-based encoder. Additionally, a hybrid convolutional block is used in the early stage of the encoder to jointly process spatial and spectral features as shown in Figure 1.1:

$$x_{\text{enc}} = \phi(\mathcal{D}_{\text{local}}(x) + \text{Conv3D}(x)) \tag{3}$$

where ϕ is a non-linear activation function and $\mathcal{D}_{\text{local}}$ represents a local DCT operation.

To address class imbalance from sparse, small lesions, we use Focal Tversky Loss [16], which emphasizes hard-to-segment regions. For optimization, we adopt the AdamW optimizer [17].

4 Experimentation Environment and Experiment Design

To assess the effect of frequency-domain features under k-space undersampling, we designed three SwinUNETR-based setups.

In the first setup, we simulate a global frequency representation by applying slice-wise 2D Discrete Cosine Transform (DCT) to the ADC channel and appending it as a third input channel alongside ADC and Z_{ADC} volumes. In this case, the DCT is computed over the entire spatial extent of the image (128×128), effectively acting as a global frequency descriptor.

The second setup introduces localized frequency modeling through a learnable *VolumeDCTPatchLayer*, which applies patch-wise DCT operations within the network. Here, the patch size is set to 48, enabling the extraction of regional spectral patterns while preserving local anatomical context.

The third setup serves as a baseline and does not utilize any frequency-domain augmentation. The model is trained solely with dual-channel input consisting of ADC and Z_{ADC} volumes.

All three configurations are trained and validated under identical conditions, including data augmentations, loss functions, optimizer settings, and learning rate schedules. This controlled design allows for a focused comparison of the effects of global versus localized DCT features, as well as the overall contribution of frequency information to neonatal lesion segmentation in diffusion MRI.

5 Comparative Evaluation and Discussion

All models were trained on 4-fold an 8-fold undersampled data with equispaced and random sampling masks. The segmentation performance was evaluated using the Dice similarity coefficient. Table 5.1 summarizes the average Dice scores across all validation cases. The model utilizing Block DCT (patch size 48) consistently outperformed the others, highlighting the effectiveness of frequency-domain priors in low-data regimes. In contrast, the baseline model without any DCT features yielded the lowest segmentation accuracy. The final model checkpoint with the highest validation Dice score was obtained using the Block DCT configuration, confirming that leveraging frequency-based features improves segmentation robustness in neonatal diffusion MRI.

Model	**Best Dice**
Baseline (No DCT)	0.52
Global DCT (Patch 128)	0.55
Block DCT (Patch 48)	0.56

Table 5.1: Mean Dice scores of segmentation models trained on 4-fold undersampled data.

As shown in Table 5.2, the performance of Block DCT models trained on undersampled data with random and equispaced sampling masks across various patch sizes is summarized. The table presents the Mean Dice scores and Standard Deviations (Std) for the best 5 models, selected based on the highest performance over 300 epochs, for both 4-fold and 8-fold undersampling strategies, using patch sizes of 32, 48, and 64.

Mask Type	**Patch Size**	**4-fold**		**8-fold**	
		Mean	**Std**	**Mean**	**Std**
Random	**32**	0.5233	0.0138	0.457	0.0149
	48	0.4965	0.0174	0.4453	0.0077
	64	0.4942	0.0214	0.4364	0.0214
Equispaced	**32**	0.5269	0.0137	0.4583	0.0148
	48	0.5013	0.0175	0.4483	0.0076
	64	0.4991	0.0232	0.4383	0.0212

Table 5.2: Results of Block DCT models trained on undersampled data using random and equispaced sampling masks with different patch sizes. The table shows the Mean Dice scores and Standard Deviations (Std) for models trained with 4-fold and 8-fold undersampling strategies.

As seen in Figure 5.1, a qualitative comparison of lesion segmentation performance is presented for models trained with different patch sizes on two slices (Slice 8 and Slice 16). The first column shows the ground truth masks, followed by the predicted segmentations for patch sizes 32, 48, and 64, along with their corresponding Dice scores.

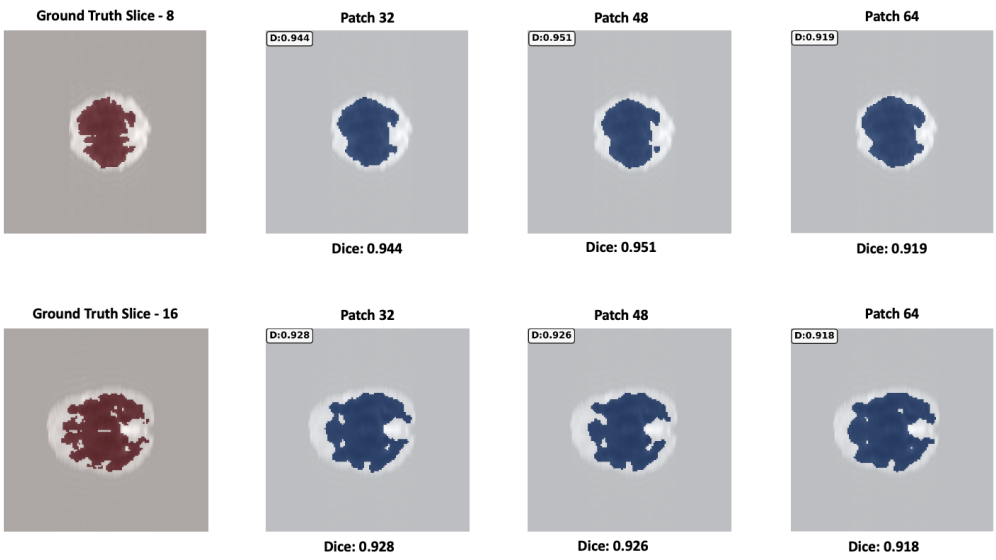


Figure 5.1: Qualitative comparison of lesion segmentation performance for models trained with different patch sizes on two slices (Slice 8 and Slice 16). Ground truth masks are on the first column, followed by predicted segmentations and corresponding Dice scores for patch sizes 32, 48, and 64.

In summary, the results demonstrate that the Block DCT model, especially with a patch size of 32, consistently outperforms other models across different undersampling strategies, including both 4-fold and 8-fold undersampling. The inclusion of frequency-domain priors, particularly through the Block DCT configuration, significantly improves the segmentation accuracy, as seen in the higher Dice scores. This highlights the effectiveness of incorporating both spatial and frequency-domain features, making the Block DCT model more robust in segmenting small neonatal brain lesions, even under challenging low-data and aliasing conditions. Additionally, the qualitative comparisons further underscore the model's superior performance in accurately detecting and segmenting lesions in low-resolution MRI scans.

6 Conclusion and Future Work

This study investigated frequency-aware segmentation models for neonatal diffusion MRI under simulated k-space undersampling conditions, with the aim of improving the detection of small and diffuse lesions. Three SwinUNETR-based variants—a baseline model, a model augmented with global DCT features, and a third integrating localized Block DCT into early encoder layers—were comparatively evaluated.

Experimental results demonstrated that the **Block DCT model (patch size: 48)** achieved the highest segmentation accuracy, with an average Dice score of **0.56**. This indicates that localized frequency representations can significantly improve lesion segmentation performance even under aliasing and low-resolution conditions. The Global DCT variant also outperformed the baseline model (Dice: 0.55 vs. 0.52), yet the spatially distributed spectral cues offered by Block DCT appear to be more effective for detecting small lesion structures.

In addition, experiments across different undersampling strategies and patch sizes revealed that **32-sized patches combined with equispaced sampling masks** provided the most stable and accurate results (4-fold: 0.5269 ± 0.0137). Larger patch sizes (e.g., 64) or more aggressive sampling rates (e.g., 8-fold) were associated with reduced performance, suggesting a trade-off between spatial resolution and spectral granularity in frequency-domain representations. This finding underscores the importance of selecting appropriate patch sizes and sampling patterns when integrating frequency-based priors into segmentation networks.

Despite these promising outcomes, the study has several limitations. First, the evaluation was limited to internal validation due to the unavailability of an external test set. As such, the generalizability of the proposed models across different clinical institutions, MRI scanners, and acquisition protocols remains to be verified. Future work should include *multi-center validation* to assess model robustness in real-world clinical environments.

Additionally, the current frequency integration strategy relies on simple *channel-wise concatenation*, which may not fully exploit the complex relationships between spatial and spectral features. More sophisticated fusion techniques—such as *attention-based spectral-spatial fusion* or *learnable frequency weighting mechanisms*—may further enhance model performance.

In future work, we plan to investigate hybrid spatial-frequency fusion mechanisms beyond simple concatenation, and explore unsupervised pretraining techniques that exploit frequency characteristics for representation learning in low-data regimes. We also aim to validate our models on external datasets and explore extensions to multimodal neonatal MRI.

In conclusion, this work demonstrates the potential of frequency-domain augmentation in improving transformer-based segmentation of neonatal brain lesions. Our results suggest that embedding spectral features into segmentation pipelines can enhance robustness under clinically realistic, undersampled acquisition settings.

References

[1] A. Hatamizadeh *et al.*, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," *arXiv preprint arXiv:2201.01266*, 2022.

[2] I. E. Toubal, E. S. Kazemi, G. Rahmon, T. Kucukpinar, M. Almansour, M.-L. Ho, and K. Palaniappan, "Fusion of deep and local features using random forests for neonatal hie segmentation," in *AI for Brain Lesion Detection and Trauma Video Action Recognition: First BONBID-HIE Lesion Segmentation Challenge and First Trauma Thompson Challenge, MICCAI 2023*, ser. Lecture Notes in Computer Science, vol. 14567. Springer, 2025, pp. 3–13.

[3] R. Bao, Y. Song, S. Bates, R. Weiss, A. Foster, C. Jaimes Cobos, S. Sotardi, Y. Zhang, R. Gollub, P. Grant, and Y. Ou, "Boston neonatal brain injury data for hypoxic ischemic encephalopathy (bonbid-hie): I. mri and lesion labeling," *Scientific Data*, vol. 12, no. 1, p. 53, 2025.

[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," *arXiv preprint arXiv:2103.14030*, 2021, available at: https://doi.org/10.48550/arXiv.2103.14030. [Online]. Available: https://arxiv.org/abs/2103.14030

[5] Y. Sun, X. Zhang, Y. Xu, and D. Zhang, "Dct-net: A dual-domain transformer for fast and accurate mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 5, pp. 1213–1225, 2022.

[6] E. Hot and P. Sekulić, "Compressed sensing mri using masked dct and dft measurements," in *4th Mediterranean Conference on Embedded Computing (MECO)*, 2015, available at: https://arxiv.org/abs/1503.07384.

[7] K. Xu, M. Qin, F. Sun, Y. Wang, Y.-K. Chen, and F. Ren, "Learning in the frequency domain," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1740–1749.

[8] Z. Li *et al.*, "Frequency-guided attention networks for medical image segmentation," *Medical Image Analysis*, vol. 85, p. 102764, 2023.

[9] H. Pan, E. Hamdan, X. Zhu, A. Cetin, and U. Bagci, "Discrete cosine transform based decorrelated attention for vision transformers," *arXiv preprint arXiv:2405.13901*, 2024, available at: https://arxiv.org/abs/2405.13901.

[10] F. Liu *et al.*, "Dct-enhanced u-net for robust mri brain tumor segmentation," *Computers in Biology and Medicine*, vol. 153, p. 106404, 2023.

[11] A. Deshpande, V. Estrela, and P. Patavardhan, "The dct-cnn-resnet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the resnet50," *Neuroscience Informatics*, vol. 1, p. 100013, 2021.

[12] K. Karaoğlu *et al.*, "Dctnet: Discrete cosine transform-enhanced network for robust image denoising," *Signal, Image and Video Processing*, 2023.

[13] A. Rastogi, G. Brugnara, M. Foltyn-Dumitru, M. A. Mahmutoglu, C. J. Preetha, E. Kobler, I. Pflüger, M. Schell, K. Deike-Hofmann *et al.*, "Deep-learning-based reconstruction of undersampled mri to reduce scan times: a multicentre, retrospective, cohort study," *The Lancet Oncology*, vol. 25, no. 3, pp. 400–410, 2024.

[14] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in mri," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1139–1152, 2020.

[15] A. A. Tolpadi, J. Thayer, H. Haacke, S. Kazemifar *et al.*, "K2s challenge: A public evaluation of knee mri segmentation from undersampled data," *Bioengineering*, vol. 10, no. 2, p. 267, 2023.

[16] S. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *MLMI*, ser. LNCS, vol. 10541. Springer, 2017, pp. 379–387.

[17] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations (ICLR)*, 2019, arXiv preprint arXiv:1711.05101.