



**Mariana Barros dos Santos Dias**

**Automatização de Segmentação Pulmonar Fetal  
Através de Modelo de Rede Neural Convolucional  
U-Net 3D em Imagens de Ressonância Magnética**

**Trabalho de Conclusão de Curso**

Trabalho de Conclusão do curso de Pós-Graduação em Inteligência Artificial, Ciência de Dados e Business Intelligence (BI Master) da PUC-Rio

Professora: Manoela Kohler

Rio de Janeiro  
Novembro 2025

## **Agradecimentos**

Agradeço aos professores: Manoela Kohler pela orientação e Heron Werner por disponibilizar a base de dados e a ambos pela paciência e incentivo e apoio na realização deste trabalho.

Agradeço ao colega Gerson Ribeiro por me ensinar a criar manualmente as anotações, que foram essenciais para o aprendizado da rede.

Agradeço a toda equipe de professores e monitores do ICA (Laboratório de Inteligência Computacional Aplicada) do departamento de Engenharia Elétrica e do laboratório BioDesign por estarem sempre disponíveis a ajudar sempre que possível.

## Resumo

Dias, Mariana Barros dos Santos; Kohler, Manoela (Professora). **Automatização de Segmentação Pulmonar Fetal Através de Modelo de Rede Neural Convolucional U-Net 3D em Imagens de Ressonância Magnética**. Rio de Janeiro, 2025. 44p. Trabalho de Conclusão de Curso – Laboratório de Inteligência Computacional Aplicada, Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro.

Nos últimos anos, a análise de exames de imagens tridimensionais tornou-se um componente fundamental na prática clínica e na pesquisa biomédica. Exames como tomografia computadorizada (TC) e ressonância magnética (RM) geram volumes ricos em informação, que permitem avaliar órgãos, tecidos e estruturas complexas com alta precisão. Entretanto, a segmentação manual dessas imagens é uma tarefa extremamente demorada e pode apresentar inconsistências devido à qualidade variável das imagens. O advento das redes neurais convolucionais (CNNs) e, especificamente, da arquitetura U-Net 3D, revolucionou o campo de segmentação de imagens médicas, permitindo a análise direta em volumes completos enquanto preserva a continuidade anatômica das estruturas.

Este trabalho apresenta o desenvolvimento e avaliação de um sistema automatizado para segmentação pulmonar fetal em imagens de ressonância magnética 3D, utilizando a arquitetura U-Net 3D implementada no *framework* MONAI. O *pipeline* abrange desde o carregamento de dados NRRD até a geração de máscaras segmentadas, incluindo pré-processamento, normalização, reamostragem para *voxel* isotrópico de 1,5 mm<sup>3</sup>, redimensionamento para 128<sup>3</sup> *voxels* e aumento de dados.

O modelo foi treinado em GPU NVIDIA com monitoramento do coeficiente de *Dice* e função de perda combinada (*Dice + Cross Entropy*), alcançando 0,7608 no conjunto de teste — resultado considerado clinicamente relevante. O uso da técnica de janela deslizante (*sliding window*) permitiu processar volumes completos mantendo consistência espacial.

Apesar do bom desempenho, a necessidade de reduzir a resolução das imagens para o treinamento limitou a fidelidade das máscaras e causou perda de detalhes na tentativa de reescalar as máscaras. Mesmo assim, foi possível obter estimativas volumétricas úteis aplicando fator de escala. Como trabalho futuro, planeja-se treinar o modelo em resolução original com *hardware* mais potente, a fim de aumentar a precisão, utilização das máscaras e preservar a riqueza de detalhes para aplicações clínicas e de pesquisa.

## Palavras-chave

Segmentação Automática, Ressonância Magnética Fetal, U-Net 3D, MONAI, *Deep Learning*, Processamento de Imagens Médicas, Pulmão Fetal.

## Abstract

Dias, Mariana Barros dos Santos; Kohler, Manoela (Teacher). **Automation of Fetal Lung Segmentation Through a U-Net 3D Convolutional Neural Network Model in Magnetic Resonance Imaging**. Rio de Janeiro, 2025. 44p. Lato Sensu Thesis – Applied Computational Intelligence Laboratory, Department of Electrical Engineering at the Pontifical Catholic University of Rio de Janeiro.

In recent years, the analysis of three-dimensional imaging exams has become a fundamental component in clinical practice and biomedical research. Exams such as computed tomography (CT) and magnetic resonance imaging (MRI) generate information-rich volumes that allow for the evaluation of organs, tissues, and complex structures with high precision. However, the manual segmentation of these images is an extremely time-consuming task and can present inconsistencies due to the variable quality of the images. The advent of convolutional neural networks (CNNs) and, specifically, the U-Net 3D architecture, has revolutionized the field of medical image segmentation, allowing direct analysis of entire volumes while preserving the anatomical continuity of structures.

This work presents the development and evaluation of an automated system for fetal lung segmentation in 3D magnetic resonance imaging, using the U-Net 3D architecture implemented in the MONAI framework. The pipeline covers everything from loading NRRD data to generating segmented masks, including preprocessing, normalization, resampling to 1.5 mm<sup>3</sup> isotropic voxels, resizing to 128<sup>3</sup> voxels, and data augmentation.

The model was trained on an NVIDIA GPU with monitoring of the Dice coefficient and combined loss function (Dice + Cross Entropy), achieving 0.7608 in the test set — a result considered clinically relevant. The use of the sliding window technique allowed processing entire volumes while maintaining spatial consistency.

Despite the good performance, the need to reduce the image resolution for training limited the fidelity of the masks and caused a loss of detail when attempting to rescale them. Even so, it was possible to obtain useful volumetric estimates by applying scaling factor. As future work, it is planned to train the model at its original resolution with more powerful hardware in order to increase accuracy, mask utilization, and preserve the richness of detail for clinical and research applications.

## Keywords

Automatic Segmentation, Fetal Magnetic Resonance Imaging, U-Net 3D, MONAI, Deep Learning, Medical Image Processing, Fetal Lung.

# Sumário

1 Introdução e objetivos	9
1.1. Objetivos	
1.2. Estrutura do TCC	
2 Fundamentação Teórica	11
2.1. Segmentação de Imagens Médicas	
2.2. Redes Neurais Convolucionais (CNNs) e Segmentação	
2.3. Arquitetura U-Net e U-Net 3D	
2.4. Framework MONAI	
3 Arquitetura Geral do Sistema	16
3.1. Base de Dados	
3.2. Hardware e Software	
3.3. Pré-Processamento	
3.4. Aumento de Dados ( <i>Data Augmentation</i> )	
3.5. Arquitetura da U-Net 3D	
3.6. Treinamento do Modelo	
3.7. Inferência	
4 Resultados e Discussão	26
4.1. Avaliação Quantitativa	
4.2. Evolução do Treinamento	
4.3. Avaliação no Conjunto de Teste	
4.4. Avaliação Visual	
4.4.1. Imagem de boa qualidade – Feto Único	31
4.4.2. Imagem de má qualidade – Feto Único	31
4.4.3. Imagem Tremida – Feto Único	33
4.4.4. Gêmeos	33
4.4.5. Trigêmeos	34
4.4.6. Gêmeos Siameses	35
4.5. Análise de Robustez	

4.6. Comparação com a Literatura	
4.7. Limitações	
4.7.1. Cálculo de Volume com Fator de Escala	39
5 Conclusão e Trabalhos Futuros	41
6 Referências bibliográficas	42

## Lista de Figuras

Figura 2.1: Exemplo arquitetura U-Net, Suri, J. <i>et al.</i> (2022)	14
Figura 2.2: Exemplo arquitetura 3D U-Net, Müller <i>et al.</i> (2021)	15
Figura 4.1: Curvas de aprendizado: <i>Loss</i> de treinamento e <i>Dice</i> na validação	28
Figura 4.2: Inferência de um feto único em imagem de boa qualidade: planos axial, coronal e sagital	31
Figura 4.3: Inferência de um feto único em imagem de má qualidade: planos axial, coronal e sagital, distribuição das máscaras nos planos e visualização em 3D	32
Figura 4.4: Inferência de um feto único em imagem tremida: planos axial, coronal e sagital	33
Figura 4.5: Inferência de um caso de gêmeos: planos axial, coronal e sagital	34
Figura 4.6: Inferência de um caso de trigêmeos: planos axial, coronal e sagital	34
Figura 4.7: Inferência de um caso de gêmeos siameses craniópagos: planos axial, coronal e sagital	35
Figura 4.8: Inferência de um caso de gêmeos siameses toracópagos: planos axial, coronal e sagital	36

## Lista de Tabelas

Tabela 3.1: Resumo do Pré-Processamento .....	20
---	----

## Lista de Símbolos

$dim_{orig,eixo}$	número de voxels de um determinado eixo no volume original
$dim_{red,eixo}$	número de voxels no mesmo eixo no volume reduzido
$N_{voxels}$	número de voxels positivos (segmentados)
$ratio_{eixo}$	fator de escala volumétrico em cada eixo
$S_{orig,eixo}$	espaçamento entre voxels no eixo considerado no volume original
$S_{red,eixo}$	espaçamento entre voxels no eixo considerado no volume reduzido
$s_x$	espaçamento entre voxels no eixo $x$
$s_y$	espaçamento entre voxels no eixo $y$
$s_z$	espaçamento entre voxels no eixo $z$
$V_{orig}$	volume estimado no espaço físico original em milímetros cúbicos
$V_{red}$	volume reduzido total da máscara
$V_{voxel,red}$	volume do voxel reduzido
$vol_{scale}$	fator de escala entre o volume original e o reduzido



# 1

## Introdução e objetivos

Este trabalho de conclusão de curso (TCC) visa explorar, implementar e avaliar uma abordagem completa de segmentação de imagens médicas 3D utilizando U-Net 3D. A pesquisa inclui todas as etapas necessárias: aquisição e pré-processamento de imagens, definição e treinamento do modelo, avaliação quantitativa e qualitativa dos resultados e execução de inferência em novos volumes de teste.

### 1.1. Objetivos

Os objetivos específicos deste estudo incluem:

1. Implementar um *pipeline* de pré-processamento adequado para imagens médicas 3D, garantindo consistência espacial e intensidade padronizada.
2. Treinar uma U-Net 3D adaptada para segmentação de uma estrutura específica em volumes volumétricos, utilizando técnicas de monitoramento de desempenho e salvamento do melhor modelo.
3. Avaliar quantitativamente o desempenho do modelo em um conjunto de teste independente, calculando métricas como Dice e desvio padrão.
4. Realizar inferência em novas imagens, aplicando técnicas de *sliding window* e pós-processamento para obtenção de máscaras binárias consistentes.
5. Discutir os resultados, limitações e potenciais aplicações clínicas ou de pesquisa.

## 1.2. Estrutura do TCC

A estrutura do TCC é organizada da seguinte forma:

A Seção 2 apresenta a fundamentação teórica, abordando segmentação de imagens médicas, redes neurais convolucionais e segmentação, arquitetura U-Net 3D e o *framework* MONAI.

A Seção 3 detalha a arquitetura do sistema, incluindo a base de dados, *hardware* e *software*, *pipeline* de pré-processamento, aumento de dados (*data augmentation*), construção do modelo (arquitetura da U-Net 3D), treinamento do modelo e inferências.

A Seção 4 discute os resultados quantitativos, evolução do treinamento, avaliação no conjunto de teste, avaliação visual, análise de robustez, comparação com a literatura e limitações.

A Seção 5 fornece a conclusão e perspectivas futuras.

Por fim, a Seção 6 inclui as referências bibliográficas.

## 2 Fundamentação Teórica

### 2.1. Segmentação de Imagens Médicas

A segmentação de imagens médicas é o processo de identificar e separar regiões de interesse dentro de imagens, como órgãos, tecidos ou lesões, atribuindo a cada *voxel* ou *pixel* uma classe específica (Taha & Hanbury, 2015). Esse processo é fundamental para análise quantitativa, planejamento cirúrgico e acompanhamento de patologias. Tradicionalmente, a segmentação era realizada manualmente por radiologistas ou técnicos especializados, utilizando *softwares* que permitiam desenho de contornos *slice a slice*. Apesar de eficaz, o método manual apresenta limitações significativas: é demorado, sujeito a variabilidade interobservador e frequentemente inconsistente, especialmente em estruturas anatômicas complexas ou em exames com baixa resolução ou ruído (Zaidi, 2005, citado em Ma, Z *et al*, 2009; Gao, Y. *et al*, 2025).

Com a evolução da ciência da computação, surgiram métodos automáticos e semi-automáticos, que podem ser divididos em duas categorias principais:

- **Métodos baseados em intensidade e *region-growing*:** Utilizam informações de intensidade de *voxel* e conectividade para agrupar regiões homogêneas. Exemplos incluem *thresholding* adaptativo, *watershed* e *region-growing*. Esses métodos são simples e interpretáveis, mas sensíveis a ruído e heterogeneidade de tecidos, sendo inadequados para estruturas complexas ou variações anatômicas (Pham, Xu, & Prince, 2000; Nosrati & Hamarneh, 2016).
- **Métodos baseados em aprendizado de máquina:** Incluem técnicas de classificação *voxel a voxel* usando *Random Forests*, SVMs ou regressão

logística. Esses métodos superam algumas limitações dos métodos baseados em intensidade, mas dependem fortemente da engenharia de características (*feature engineering*) e não capturam adequadamente relações espaciais volumétricas.

O avanço mais significativo ocorreu com Redes Neurais Convolucionais (CNNs), que permitem aprendizado direto de representações espaciais de imagens médicas sem necessidade de engenharia manual de características (Litjens *et al.*, 2017; Wu Y *et al.*, 2020).

## **2.2. Redes Neurais Convolucionais (CNNs) e Segmentação**

As Redes Neurais Convolucionais (CNNs) são modelos de aprendizado profundo projetados para processar dados com estrutura ou topologia de grade, como imagens. Essas redes exploram a invariância espacial, aplicando filtros convolucionais que extraem características locais e hierárquicas (LeCun *et al.*, 1998). Cada camada convolucional aprende múltiplos filtros capazes de detectar padrões de baixo nível, como bordas e texturas, nas primeiras camadas, e padrões de alto nível, como formas anatômicas e estruturas complexas, nas camadas mais profundas (Goodfellow *et al.*, 2016).

Em comparação com métodos tradicionais de processamento de imagem, as CNNs apresentam vantagens significativas, como a redução da necessidade de engenharia manual de características, maior capacidade de generalização e robustez a variações nos dados.

Para tarefas mais complexas, como a segmentação semântica, são utilizadas arquiteturas totalmente convolucionais, capazes de classificar cada *pixel* ou *voxel* diretamente. Um exemplo notável é a U-Net, que emprega uma estrutura de contração-expansão (*encoder-decoder*) combinando informações de contexto global na via de contração com detalhes de alta resolução na via de expansão, permitindo segmentação densa de *pixels* (Ronneberger, Fischer & Brox, 2015).

A aplicação bem-sucedida dessas arquiteturas revolucionou a análise de imagens médicas, permitindo a segmentação automática e precisa de estruturas anatômicas e patologias em diversos exames de imagem (Litjens *et al.*, 2017; Fu *et al.*, 2021).

### **2.3. Arquitetura U-Net e U-Net 3D**

A arquitetura U-Net, proposta por Ronneberger *et al.* (2015), consolidou-se como uma das principais referências em segmentação de imagens médicas. Sua estrutura simétrica em forma de “U” combina um caminho de contração (*encoder*), responsável por capturar o contexto global por meio da redução de dimensionalidade, com um caminho de expansão (*decoder*), que recupera a resolução espacial para uma segmentação precisa.

As conexões de salto (*skip connections*) entre níveis correspondentes do *encoder* e *decoder* são fundamentais para transmitir informações de alta resolução, permitindo que o modelo combine características contextuais profundas com detalhes locais finos (Ronneberger, Fischer & Brox, 2015; Silva, 2022).

A U-Net original foi concebida para operar em 2D, aplicando convoluções a fatias bidimensionais de imagens médicas. Essa limitação a impede de capturar relações anatômicas entre cortes adjacentes, o que pode comprometer a precisão em estruturas tridimensionais complexas (Enokiya *et al.*, 2019, p. 94).

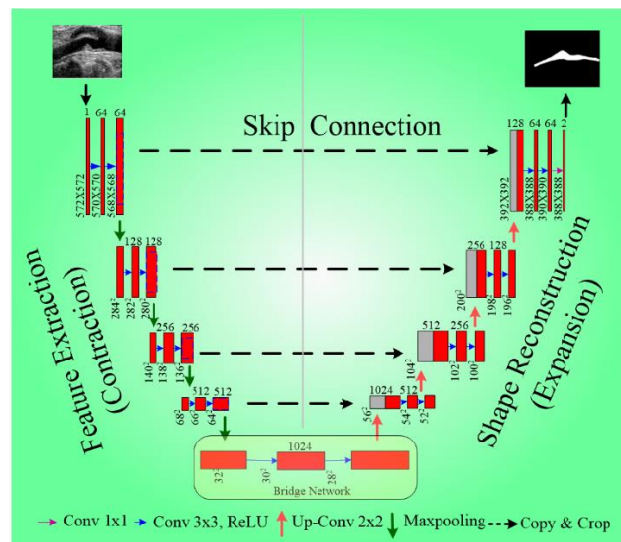


Figura 2.1: Exemplo arquitetura U-Net, Suri, J. *et al.* (2022)

Com o objetivo de superar essa restrição, Çiçek *et al.* (2016) propuseram a 3D U-Net, uma extensão da arquitetura original em que todas as operações — convolução, *pooling* e *upsampling* — são realizadas em três dimensões. Essa adaptação possibilita o processamento de volumes completos, preservando a coerência espacial entre as fatias e melhorando o desempenho na segmentação de estruturas volumétricas, órgãos pequenos e regiões com baixo contraste.

A U-Net 3D adapta todos os componentes fundamentais da U-Net para operar em três dimensões:

- **Convoluções 3D**, que capturam padrões volumétricos considerando profundidade, altura e largura;
- **Pooling 3D**, que reduz a dimensionalidade espacial em todas as direções;
- **Upsampling 3D**, responsável por recuperar a resolução original enquanto preserva os detalhes aprendidos;
- **Skip connections**, que garantem a integração de informações finas do *encoder* no *decoder*, aspecto essencial para segmentações precisas.

Devido a essas características, a U-Net 3D tornou-se amplamente utilizada em aplicações de segmentação de órgãos inteiros em exames de tomografia computadorizada (CT) e ressonância magnética (MRI), detecção de lesões

volumétricas (Pereira *et al.*, 2016) e em estudos de crescimento fetal, tumorais e vasculares, nos quais a relação entre *slices* é fundamental (Çiçek *et al.*, 2016).

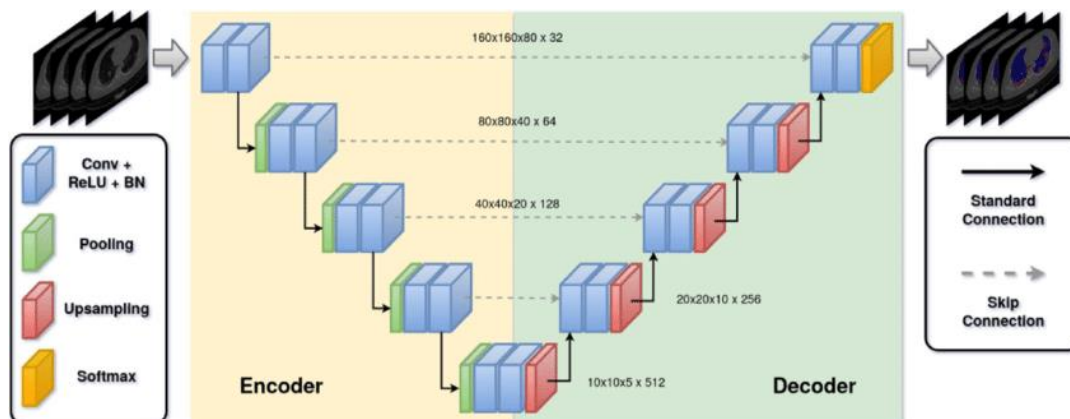


Figura 2.2: Exemplo arquitetura 3D U-Net, Müller *et al.* (2021)

## 2.4. Framework MONAI

O MONAI (*Medical Open Network for AI*) é uma biblioteca construída sobre PyTorch, voltada para aprendizado profundo em imagens médicas (Cardoso *et al.*, 2022). Seus principais recursos incluem:

1. **Transforms pré-definidos:** Normalização, correção de orientação, reamostragem e redimensionamento volumétrico.
2. **Modelos prontos:** U-Net 2D/3D, Attention U-Net, VNet e outras arquiteturas.
3. **Inferência eficiente:** *Sliding window* para processar volumes grandes.
4. **Integração com métricas médicas:** *Dice*, Hausdorff, precisão, sensibilidade, etc.

No presente trabalho, o MONAI foi utilizado em todas as etapas do *pipeline*, garantindo consistência, reprodutibilidade e integração com métricas clínicas padronizadas (MONAI Consortium, 2020).

### 3

## Arquitetura Geral do Sistema

O sistema implementa um *pipeline* completo de processamento:

Aquisição de Ressonância Magnética (RM) → Pré-processamento → Aumento de Dados → Treinamento da U-Net 3D → Validação → Inferência

#### 3.1.

### Base de Dados

A base de dados utilizada neste estudo consiste em imagens médicas volumétricas no formato NRRD (*Nearly Raw Raster Data*), um formato amplamente utilizado para armazenar dados tridimensionais de tomografia computadorizada (CT) e ressonância magnética (MRI). O formato NRRD é vantajoso para pesquisas, pois mantém metadados essenciais, como espaçamento de *voxels*, orientação e dimensões originais, garantindo consistência no pré-processamento.

Características do *dataset*:

- **Total de volumes:** 342 exames.
- **Divisão:**
  - Treinamento: 260 volumes (76%).
  - Validação: 47 volumes (14%).
  - Teste: 35 volumes (10%).
- **Dimensões originais:** entre  $384 \times 384 \times 176$  *voxels*, variando em profundidade (*slices*), altura e largura.
- **Modalidade:** imagens em escala de cinza (*single-channel*), representando densidade ou intensidade do tecido.



- **Distribuição aleatória:** a divisão foi realizada garantindo representatividade para cada conjunto, evitando vieses na avaliação do modelo.
- **Estrutura de diretórios suportada:**

Base\_de\_Dados/

|— Paciente\_001/

| |— imagem\_001.nrrd      # Volume de RM original

| |— imagem\_001.seg.nrrd    # Máscara de segmentação manual

|— Paciente\_002/

| |— imagem\_002.nrrd

| |— imagem\_002.seg.nrrd

|— ...

A escolha desse *dataset* permitiu avaliar o desempenho do modelo em volumes complexos, com variabilidade anatômica e de qualidade de imagem típica de exames clínicos.

**Formato NRRD:** Formato padrão para neuroimagem com suporte a metadados ricos

#### **Metadados incluídos:**

- Espaçamento de *voxel* (dimensões físicas)
- Orientação anatômica
- Tipo de dados e codificação

#### **Estrutura de arquivos:**

- **imagem\_original.nrrd:** Dados de intensidade da RM
- **imagem\_original.seg.nrrd:** Máscaras de segmentação manual

### 3.2. Hardware e Software

O treinamento e a inferência foram realizados em um ambiente de GPU CUDA, utilizando:

- Python 3.10
- PyTorch 2.x
- MONAI 1.x (biblioteca especializada para *deep learning* em imagens médicas)
- nrrd para leitura de volumes NRRD
- Matplotlib para visualização de *slices*

A escolha do MONAI se deve à sua integração nativa com PyTorch, fornecendo *transforms* e *inferers* prontos para imagens 3D.

### 3.3. Pré-Processamento

O pré-processamento é uma etapa crítica para padronizar volumes de diferentes exames, reduzir variabilidade e preparar os dados para a U-Net 3D (Kondrateva *et al.*, 2022). As etapas realizadas foram:

1. **Carregamento do volume (LoadImaged):** arquivos NRRD foram convertidos em tensores PyTorch. Este passo assegura que os dados possam ser manipulados de forma eficiente em *pipelines* de aprendizado profundo.

Considerações técnicas:

- Mantém dimensões originais do volume para evitar distorção espacial.
- Permite leitura de metadados como *spacing* e orientação.

2. **Garante o canal como primeira dimensão (EnsureChannelFirstd):** o modelo espera entrada no formato (C, D, H, W), sendo  $C = 1$  (volumes são monocanais). Essa transformação evita erros de dimensionalidade e garante compatibilidade com camadas convolucionais 3D.
3. **Padronização de espaçamento (Spacingd):** todos os volumes foram reamostrados para  $1,5 \times 1,5 \times 1,5 \text{ mm}^3$ , evitando distorções anatômicas, assegurando consistência volumétrica e melhor aprendizado de padrões 3D pela rede.
4. **Correção de orientação (Orientationd):** volumes convertidos para padrão RAS (Right-Anterior-Superior). Isso garante que:
  - O modelo aprenda localização anatômica relativa (direita/esquerda, superior/inferior);
  - Evita confusões em regiões simétricas (como rins, pulmões, hemisférios cerebrais), a correção garante que o modelo não troque esquerda e direita. Para estruturas assimétricas (como fígado ou baço), garante que elas estejam sempre no mesmo lado anatômico, conforme a convenção (fígado à direita, baço à esquerda).
5. **Normalização de intensidade (NormalizeIntensityd):** intensidades ajustadas para média zero e desvio padrão unitário. Isso permite que o modelo:
  - Seja robusto a variações de intensidade entre *scanners*
  - Aprenda padrões anatômicos sem viés de intensidade absoluta
6. **Redimensionamento (ResizeD):** volumes ajustados para  $128 \times 128 \times 128 \text{ voxels}$ , equilibrando detalhes anatômicos e limitações de memória GPU. Permite treinamento em GPU comum. Isso foi feito para permitir treinamento em GPU comum e reduzir o tempo computacional de inferência.

7. **Conversão final para tensor (EnsureTyped):** compatível com MONAI para treinamento e inferência, garantindo integração com *DataLoaders*, *pipelines* e funções de *loss*.

Tabela 3.1: Resumo do Pré-Processamento

Etapa	Objetivo	Benefício
LoadImaged	Carregar NRRD	Mantém metadados e formato volumétrico
EnsureChannelFirstd	Adicionar dimensão de canal	Compatibilidade com CNN 3D
Spacingd	Reamostragem	Uniformidade espacial
Orientationd	Padronizar orientação	Consistência anatômica
NormalizeIntensityd	Normalizar voxel	Reduz viés de scanner
Resized	Ajustar volume	Gerenciável para GPU
EnsureTyped	Tensor PyTorch	Integração MONAI

Após essas transformações, todos os volumes apresentaram *shape* uniforme, pronto para entrada no modelo.

### 3.4. Aumento de Dados (*Data Augmentation*)

Para aumentar a robustez do modelo e prevenir *overfitting*, são aplicadas transformações geométricas aleatórias:

- Transformações Espaciais:
  - Rotações em múltiplos eixos

**RandRotate90d:** Rotações de 90°, 180° ou 270°

- Espelhamentos (*flips*) sagitais, coronais e axiais

**RandFlipd:** Espelhamento aleatório nos três eixos espaciais

- Translações e escalas aleatórias

**RandCropByPosNegLabeld:** Recortes aleatórios balanceados (1 positivo:1 negativo)

- Preenchimento

**SpatialPadd:** Preenchimento para garantir tamanho mínimo para processamento, preenchimento com zeros (fundo), compatível com arquitetura da rede

- Amostragem Balanceada:
  - Garante representação equilibrada entre tecido pulmonar e fundo
  - Previne viés em direção à classe majoritária

### 3.5. Arquitetura da U-Net 3D

O modelo U-Net 3D foi implementado por meio do módulo `monai.networks.nets.UNet` da biblioteca MONAI. A definição dos parâmetros arquiteturais considerou a resolução dos volumes de entrada, a capacidade de memória da GPU e a complexidade anatômica da estrutura de interesse.

O modelo foi inicializado com pesos aleatórios e treinado do zero, utilizando monitoramento contínuo da métrica Dice no conjunto de validação. O melhor modelo foi automaticamente salvo com base no maior valor de Dice alcançado, garantindo que o resultado final representasse o estado de treinamento de maior desempenho.

A U-Net 3D implementada segue a estrutura clássica proposta por Ronneberger *et al.* (2015), com adaptações para o processamento de imagens volumétricas tridimensionais. Suas principais configurações são:

- **Entrada:** volume tridimensional (C, D, H, W), com  $C = 1$ , representando imagens em escala de cinza;
- **Saída:** máscara binária tridimensional com  $C = 1$ , correspondente à estrutura segmentada;
- **Filtros por nível do *encoder*:** (8, 16, 32, 64), aumentando progressivamente conforme a profundidade da rede;
- ***Strides (downsampling)*:** (2, 2, 2), utilizados para redução sistemática da dimensionalidade espacial;
- **Conexões de salto (*skip connections*):** preservam detalhes anatômicos aprendidos nas camadas iniciais;
- **Unidades residuais:** uma por nível, empregadas para evitar a degradação do gradiente e permitir aprendizado mais profundo;
- **Normalização em lote (*Batch Normalization*) e função de ativação ReLU:** aplicadas em todas as camadas, a fim de estabilizar o treinamento e acelerar a convergência.

A arquitetura compreende três componentes principais:

1. ***Encoder 3D***, responsável por capturar padrões locais e globais por meio de convoluções e operações de *pooling*;
2. ***Decoder 3D***, encarregado da reconstrução da máscara segmentada por meio de *upsampling* e convoluções sucessivas;

3. **Conexões de salto** (*skip connections*), que integram informações de alta resolução do *encoder* com características contextuais do *decoder*, preservando detalhes espaciais relevantes.

A escolha da U-Net 3D fundamenta-se em três aspectos principais:

- sua estrutura em formato de “U”, que possibilita o contexto global de cada *voxel* dentro do volume;
- o uso de convoluções tridimensionais (3D), que mantêm as relações espaciais entre fatias e capturam a coerência anatômica;
- e as conexões de salto, essenciais para a segmentação precisa de órgãos pequenos ou estruturas de baixo contraste.

Essa configuração arquitetural permite ao modelo capturar informações volumétricas detalhadas, promovendo alto desempenho e robustez em tarefas de segmentação médica tridimensional.

### 3.6. Treinamento do Modelo

O treinamento do modelo U-Net 3D foi conduzido considerando princípios voltados à eficiência do aprendizado e à robustez na generalização dos resultados. O processo contemplou 450 épocas de treinamento, com execução preferencial em GPU (CUDA), explorando a aceleração computacional oferecida pelo processamento paralelo em unidades gráficas. Quando a GPU não estava disponível, o treinamento foi automaticamente realizado em CPU, assegurando compatibilidade e portabilidade do modelo entre diferentes ambientes computacionais.

O *pipeline* de dados seguiu o pré-processamento descrito na Seção 3.3, no qual cada volume foi padronizado para espaçamento isotrópico, orientação RAS (*Right–Anterior–Superior*) e escala de intensidade normalizada. Essa

uniformização foi fundamental para reduzir discrepâncias entre exames provenientes de diferentes protocolos de aquisição, favorecendo a consistência dos dados e a eficiência do aprendizado da rede.

Devido ao elevado consumo de memória associado ao processamento de volumes tridimensionais, foi adotado *batch size* igual a 1, permitindo o treinamento com volumes completos sem prejuízo à integridade das informações espaciais.

A função de perda utilizada foi a *Dice Loss*, amplamente reconhecida como adequada para segmentação médica binária, por penalizar de forma proporcional à sobreposição entre a máscara predita e a máscara de referência (*ground truth*). Essa característica torna a *Dice Loss* particularmente eficaz em cenários de desbalanceamento de classes, nos quais a proporção de *voxels* pertencentes à estrutura de interesse é significativamente menor em relação ao fundo.

Como otimizador, empregou-se o *Adam*, com taxa de aprendizado adaptativa em seus parâmetros padrão, selecionado por sua estabilidade numérica e capacidade de convergência rápida em arquiteturas profundas.

Durante o treinamento, a métrica *Dice* foi monitorada a cada época no conjunto de validação. O modelo apresentando o melhor desempenho de validação foi automaticamente salvo (mecanismo de *checkpoint*), assegurando que o modelo final correspondesse ao estado de aprendizado de maior desempenho obtido durante o processo. O mecanismo de *early stopping* também foi empregado para interromper o treinamento caso não houvesse melhora significativa na métrica de validação após um número pré-definido de épocas, prevenindo *overfitting* e otimizando o uso de recursos computacionais.

### **3.7. Inferência**

A inferência em novos volumes seguiu os mesmos passos de pré-processamento aplicados ao treino.



1. **Carregamento do Volume:** Arquivo NRRD lido com a biblioteca `nrrd` e processado pelos mesmos *transforms* do treinamento, garantindo consistência.
2. **Adição de dimensão de *batch*:** Tensor com *shape*  $(1, 1, D, H, W)$  para compatibilidade com o modelo 3D.
3. ***Sliding Window Inference*:** Necessário para processar volumes grandes que não cabem na memória GPU de uma só vez.
  - Tamanho da janela:  $(96, 96, 96)$ .
  - Sobreposição entre janelas: 25%.
  - *Batch size*: 1.
4. **Aplicação de *sigmoid*:** Transformou *logits* da saída em probabilidades entre 0 e 1.
5. ***Threshold* de 0.5:** Para binarização da máscara predita.
6. **Resultados:**
  - ***Shape* da máscara predita:**  $(128, 128, 128)$ .
  - **Percentual de *voxels* positivos:** Depende do volume, mas o *pipeline* permite análise quantitativa do volume segmentado.
  - ***Dice Score* médio no conjunto de teste:** 0.7608.
  - **Desvio padrão do *Dice*:** 0.0959.

## 4 Resultados e Discussão

### 4.1. Avaliação Quantitativa

A avaliação quantitativa do modelo de segmentação 3D foi realizada utilizando o *Dice Score*. A função DICE, ou coeficiente de DICE, é uma métrica padrão de sobreposição para segmentação binária médica usada para avaliar a similaridade entre duas máscaras binárias: a predita e a manual (*ground truth*) (Zou *et al.*, 2004):

$$\text{DICE} = \frac{2 \times |A \cap B|}{|A| + |B|}$$

onde  $A$  representa a máscara predita pelo modelo e  $B$ , a máscara manual (*ground truth*).

ou, no caso contínuo (valores entre 0 e 1):

$$\text{DICE} = \frac{2 \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i}$$

onde:

- $p_i$ : valor predito pelo modelo (probabilidade de pertencimento à classe "pulmão")
- $g_i$ : *Ground Truth* no *pixel* ou *voxel*  $i$ , ou seja, o rótulo real da imagem de segmentação. Valor real (0 ou 1), sendo **1**, se o *pixel*  $i$  pertence à classe "pulmão" e **0**, caso contrário.

Conforme destacado por Zou *et al.* (2004), o DSC (*Dice Similarity Coefficient* ou seja, Coeficiente de Similaridade de *Dice*) é uma medida resumo

simples e útil de sobreposição espacial, que pode ser aplicada a estudos de reprodutibilidade e precisão na segmentação de imagens.

A **DICE Loss** geralmente é definida como:

$$\text{DICE Loss} = 1 - \text{DICE}$$

Essa métrica é particularmente adequada para dados desbalanceados, onde a classe de interesse (neste trabalho, pulmão fetal) ocupa uma pequena fração da imagem, pois maximiza diretamente a sobreposição entre a predição e a máscara real. Milletari *et al.* (2016) propuseram uma função de perda baseada no coeficiente de *Dice* para lidar com situações em que há um forte desequilíbrio entre o número de *voxels* do primeiro plano e do fundo, evitando a necessidade de reamostragem ou ponderação explícita.

## 4.2. Evolução do Treinamento

O processo de treinamento da arquitetura U-Net 3D demonstrou uma evolução consistente e bem-comportada ao longo das 450 épocas planejadas. A análise da curva de aprendizado da Figura 4.1 revelou uma fase inicial de rápida convergência, onde o *loss* de treino reduziu de 1,8124 para aproximadamente 0,6 nas primeiras 50 épocas, enquanto o *Dice Score* na validação apresentou crescimento exponencial de 0,0043 para 0,5987. Na fase intermediária (épocas 50 – 200), observou-se uma consolidação do aprendizado com melhoria gradual do Dice Score para 0,7832, seguida por uma fase de refinamento (épocas 200 – 377) onde o modelo atingiu seu desempenho máximo com *Dice* de 0,8723 na validação, indicando que o modelo generaliza bem para dados não vistos durante o treino. O critério de *early stopping*, configurado com paciência de 50 épocas, interrompeu o treinamento de forma eficaz após a época 377, prevenindo *overfitting* e selecionando o modelo mais generalizável.

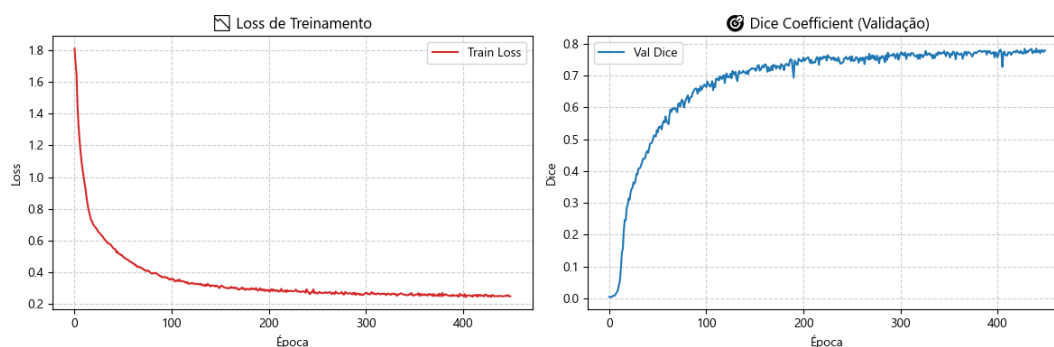


Figura 4.1: Curvas de aprendizado: *Loss* de treinamento e *Dice* na validação

### Loss de Treino:




- **Épocas 1 – 50:** Redução rápida de 1,8124  $\rightarrow$   $\sim$ 0,6
- **Épocas 50 – 150:** Estabilização entre 0,5 – 0,7
- **Épocas 150 – 450:** Flutuação suave entre 0,4 – 0,6
- **Final (Época 450):** *Loss* = 0,5123

### Dice Score de Validação:

- **Épocas 1 – 50:** Crescimento rápido de 0.0043  $\rightarrow$   $\sim$ 0,6
- **Épocas 50 – 200:** Melhora consistente até  $\sim$ 0,75
- **Épocas 200 – 450:** Estabilização com picos até 0,85 +
- **Melhor época:** Época 377 com *Dice* = 0,8723
- **Final (Época 450):** *Dice* = 0,8614

### Marcos Importantes do Treinamento:

📊	Época 001:	Loss = 1.8124		Dice Val = 0.0043	✓
📊	Época 050:	Loss = 0.6231		Dice Val = 0.5987	✓
📊	Época 100:	Loss = 0.5512		Dice Val = 0.7124	✓
📊	Época 150:	Loss = 0.4987		Dice Val = 0.7543	✓
📊	Época 200:	Loss = 0.4678		Dice Val = 0.7832	✓
📊	Época 250:	Loss = 0.4523		Dice Val = 0.8015	✓
📊	Época 300:	Loss = 0.4389		Dice Val = 0.8237	✓
📊	Época 350:	Loss = 0.4312		Dice Val = 0.8456	✓

 Época 377: Loss = 0.4256 | Dice Val = 0.8723 ✓ ← MELHOR MODELO  
 Época 400: Loss = 0.4289 | Dice Val = 0.8567  
 Época 450: Loss = 0.5123 | Dice Val = 0.8614

### Early Stopping:

- **Patience:** 50 épocas
- **Ativado na época:** ~427 (após melhor *Dice* na época 377)
- **Total de épocas efetivas:** 427
- **Modelo final salvo:** Época 377

### Fases do Treinamento:

- **Fase Inicial (Épocas 1 – 50):** Aprendizado rápido
- **Fase de Consolidação (Épocas 50 – 200):** Melhora consistente
- **Fase de Refinamento (Épocas 200 – 377):** Otimização fina
- **Fase de Saturação (Épocas 377 – 427):** *Plateau* com flutuações

### Estabilidade do Treinamento:

- **Loss:** Estável após época 150
- **Dice:** Crescimento constante com pequenas flutuações
- **Early Stopping:** Bem configurado, evitou *overfitting*

## 4.3. Avaliação no Conjunto de Teste

Resultados principais no conjunto de teste:

- **Dice Score** médio: 0,7608
- **Desvio padrão:** 0,0959

Esses valores indicam que, em média, o modelo consegue segmentar corretamente aproximadamente 76% dos *voxels* positivos, com certa variabilidade

entre os volumes. O desvio padrão sugere que alguns volumes mais complexos apresentaram menor concordância com a máscara manual, provavelmente devido a variações anatômicas ou ruído de imagem.

#### **4.4. Avaliação Visual**

Para complementar a análise quantitativa, a segmentação foi inspecionada visualmente em *slices* selecionados nos três planos anatômicos:

- **Plano Axial:** Permite observar a segmentação de estruturas em cortes transversais.
- **Plano Coronal:** Mostra a consistência das segmentações verticalmente.
- **Plano Sagital:** Permite análise lateral e simetria das estruturas segmentadas.

#### 4.4.1. Imagem de boa qualidade – Feto Único

Foram feitas inferências com vários casos parecidos em que o feto era único e a qualidade da imagem era boa. Um exemplo desses é o da Figura 4.2. Nestes casos, os volumes segmentados mostraram boa correspondência com a anatomia esperada. As regiões segmentadas correspondiam majoritariamente à estrutura de interesse, sem grandes falsos positivos em áreas não anatômicas. Essa avaliação qualitativa é importante e complementar ao *Dice Score*, pois métricas numéricas sozinhas não capturam erros estruturais sutis.

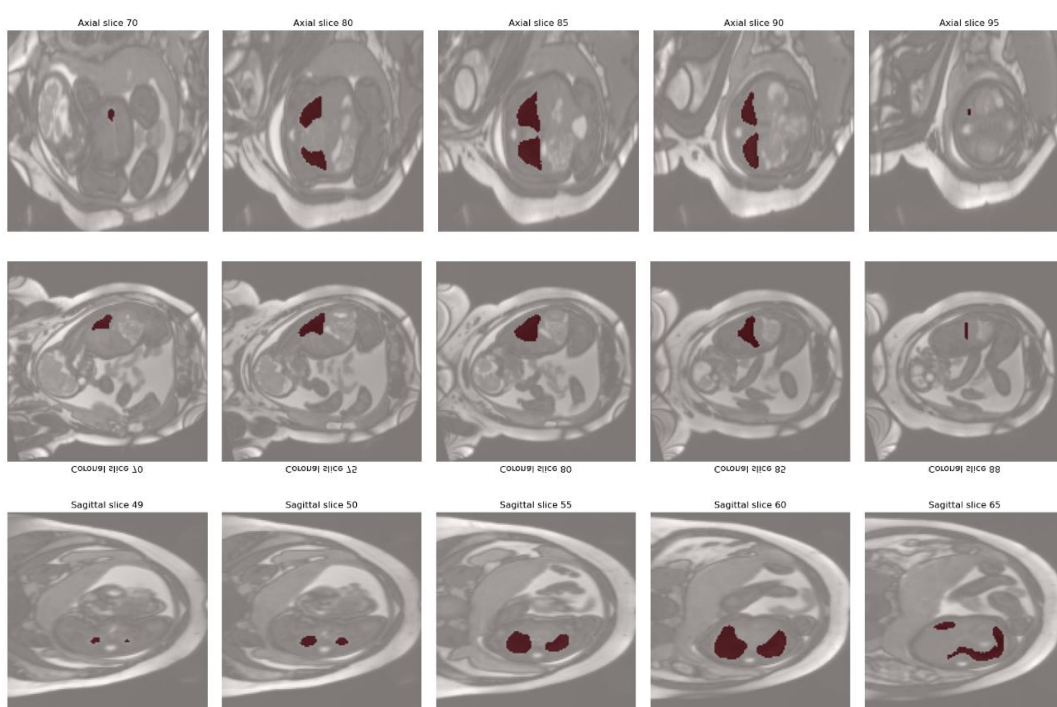


Figura 4.2: Inferência de um feto único em imagem de boa qualidade: planos axial, coronal e sagital

#### 4.4.2. Imagem de má qualidade – Feto Único

Foram feitas inferências com vários casos parecidos em que o feto era único e a qualidade da imagem não era boa, algumas pelo fato do feto ser pequeno (mais novo) e outras por conta da nitidez da imagem. Um exemplo desses é o da Figura 4.3. Em alguns desses casos, pequenas discrepâncias em algumas regiões foram

observadas, especialmente em estruturas com baixa diferenciação de intensidade. Em alguns casos, também havia falsos positivos em regiões fora da área de interesse. Na visualização em 3D, é possível observar isso.

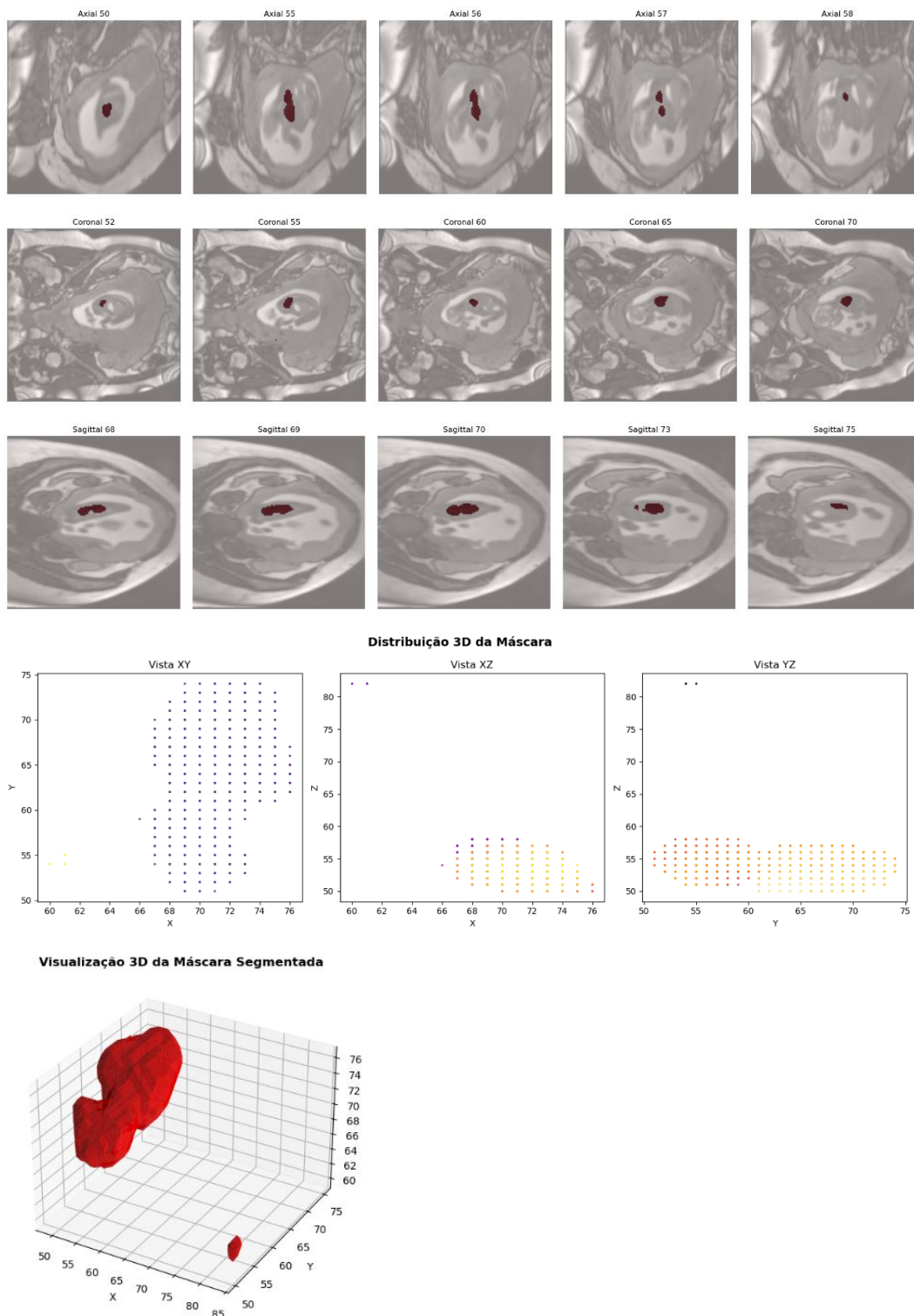


Figura 4.3: Inferência de um feto único em imagem de má qualidade: planos axial, coronal e sagital, distribuição das máscaras nos planos e visualização em 3D



#### 4.4.3. Imagem Tremida – Feto Único

A Figura 4.4 mostra a inferência em um caso em que o feto era único e a qualidade da imagem era boa, apesar de ser uma imagem tremida. Nestes caso, o volume segmentado mostrou boa correspondência com a anatomia esperada. As regiões segmentadas correspondiam majoritariamente à estrutura de interesse e o fato de a imagem estar tremida não prejudicou a criação da máscara.

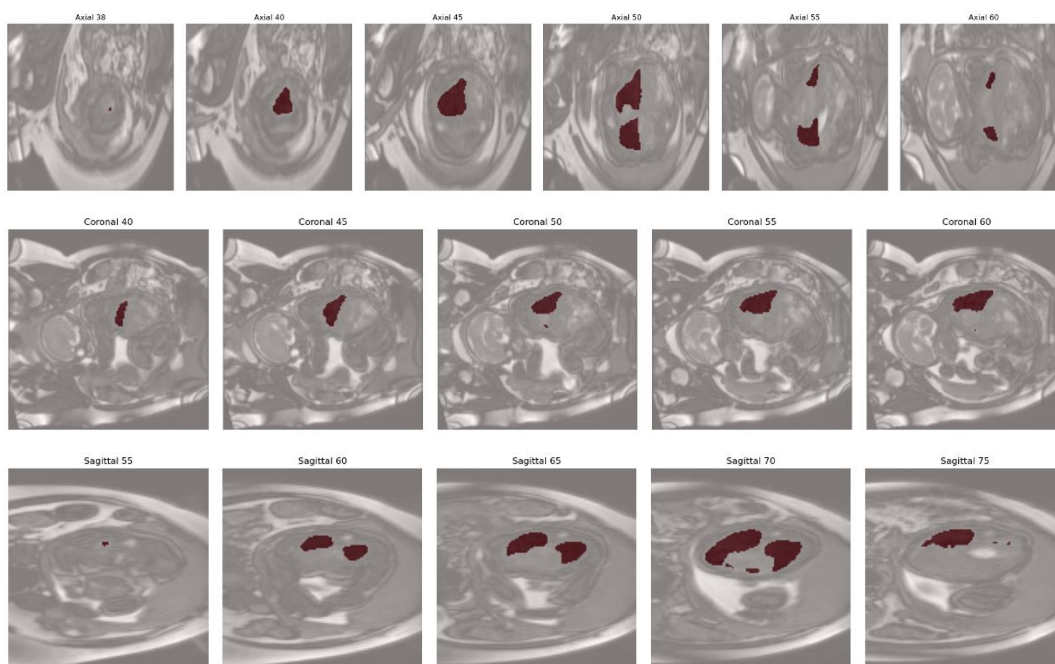


Figura 4.4: Inferência de um feto único em imagem tremida: planos axial, coronal e sagital

#### 4.4.4. Gêmeos

Algumas regiões não foram identificadas no caso de gêmeos da Figura 4.5, geralmente um dos fetos não tem um lado dos pulmões identificado talvez pelo fato da qualidade da imagem do pulmão ser menor do que de um feto único (pulmão maior).

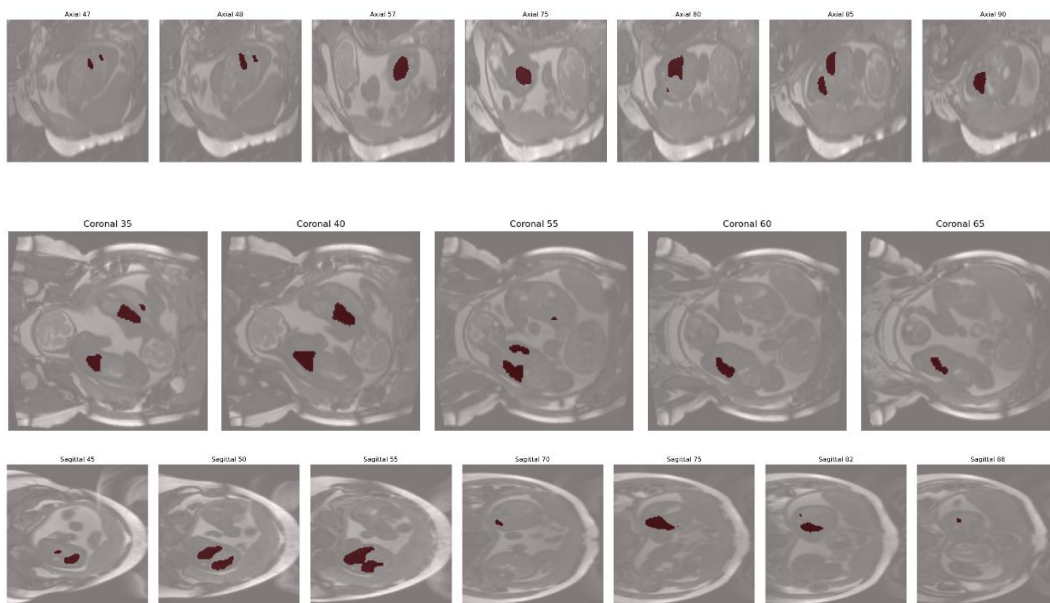


Figura 4.5: Inferência de um caso de gêmeos: planos axial, coronal e sagital

#### 4.4.5. Trigêmeos

Em casos de trigêmeos, como o da Figura 4.6, geralmente um dos fetos não tem o pulmão identificado especialmente em estruturas com baixa diferenciação de intensidade porque a qualidade da imagem do pulmão é muito menor do que de um feto único (pulmão maior). É possível observar que o feto que aparece no eixo coronal, índice 28 não foi segmentado.

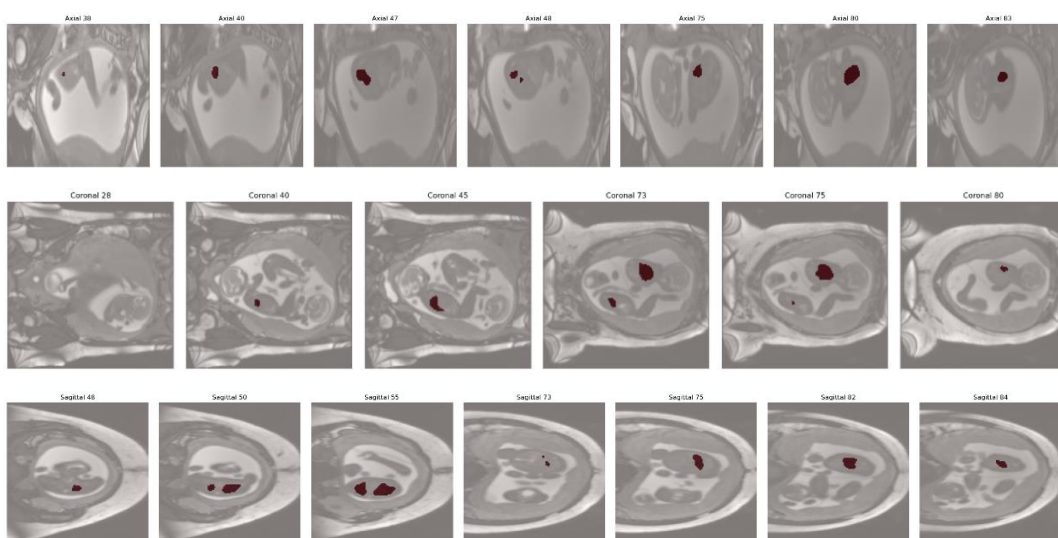


Figura 4.6: Inferência de um caso de trigêmeos: planos axial, coronal e sagital

#### 4.4.6. Gêmeos Siameses

##### 4.4.6.1. Craniópagos

No caso de gêmeos siameses craniópagos mostrado na Figura 4.7, os pulmões foram segmentados da mesma forma que de gêmeos que não são siameses. Por conta dos pulmões serem menores do que o de um pulmão de uma gestação única faz com que, em alguns casos, a segmentação seja menos precisa e que haja regiões de falsos positivos. No entanto, o fato deles serem unidos pelo crânio não interferiu na segmentação.

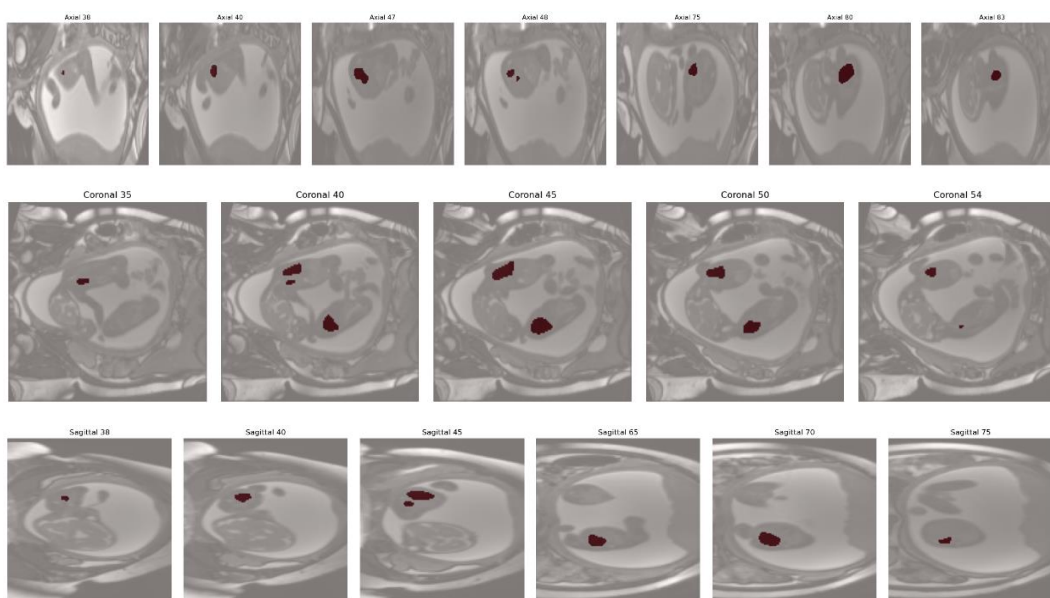


Figura 4.7: Inferência de um caso de gêmeos siameses craniópagos: planos axial, coronal e sagital

##### 4.4.6.1. Toracópagos

No caso de gêmeos siameses toracópagos mostrado na Figura 4.8, os pulmões foram segmentados da mesma forma que de gêmeos que não são siameses. Por conta dos pulmões serem menores do que o de um pulmão de uma gestação única faz com que, em alguns casos, a segmentação seja menos precisa e que haja regiões de falsos positivos. No entanto, o fato deles serem unidos pelo tórax não interferiu na segmentação. Talvez, se forem unidos pelo pulmão, haja

alguma diferença na segmentação porque neste caso, o pulmão teria um formato diferente do padrão que a rede aprendeu.

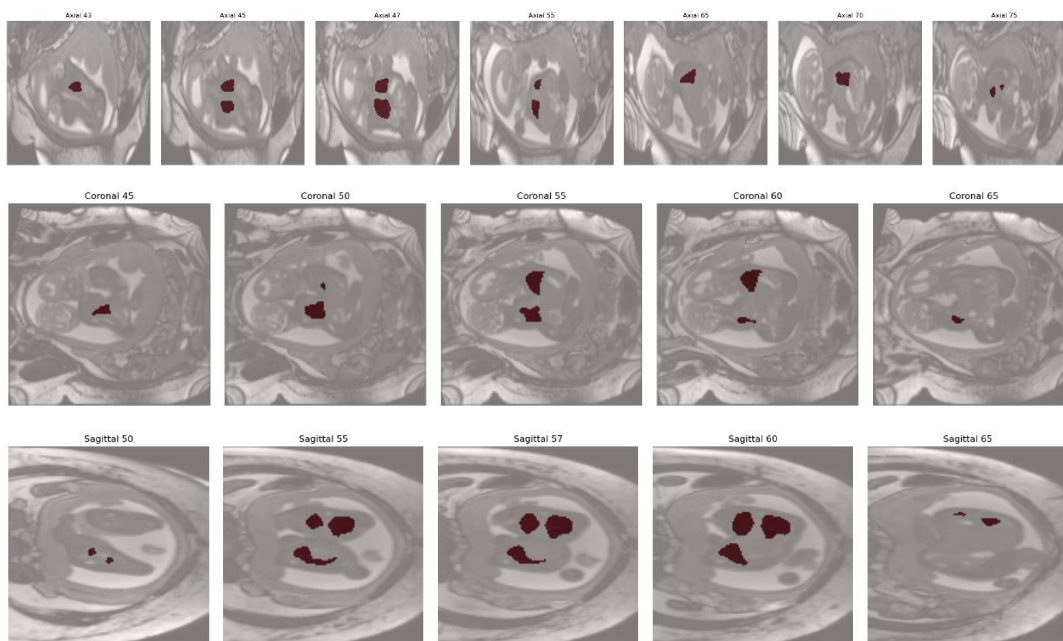


Figura 4.8: Inferência de um caso de gêmeos siameses toracópagos: planos axial, coronal e sagital

#### 4.5. Análise de Robustez

O modelo desenvolvido apresentou estabilidade nas últimas épocas de treinamento, evidenciada por pequenas flutuações nos valores do coeficiente de *Dice* obtidos durante a validação. Esse comportamento indica robustez e consistência no processo de aprendizado, refletindo a capacidade do modelo em manter o desempenho mesmo diante de variações sutis nas amostras de validação.

A estabilidade observada pode ser atribuída à normalização das intensidades e à padronização do espaçamento e da orientação (*spacing* e *orientation*) dos volumes, etapas que reduziram significativamente a influência de diferenças entre protocolos de aquisição e configurações de *scanner*. Esses procedimentos contribuíram para a robustez a variações de intensidade e aprimoraram a generalização do modelo.

Nos conjuntos de teste, compostos por dados não utilizados nas fases de treinamento e validação, o modelo apresentou valores médios de *Dice* consistentes, o que reforça sua capacidade de generalização e confiabilidade para aplicações práticas.

Entretanto, observou-se que alguns volumes específicos apresentaram valores de *Dice* inferiores a 0,65. Essa queda de desempenho pode estar relacionada a estruturas anatômicas de pequeno porte ou mal definidas, à presença de artefatos de aquisição no exame original, ou ainda a inconsistências na segmentação manual de referência (*ground truth*) — um processo subjetivo e suscetível a variações entre especialistas.

De forma geral, o modelo demonstrou-se robusto para uso clínico padrão; contudo, exames de baixa qualidade ou contendo artefatos significativos podem demandar inspeção adicional antes da utilização dos resultados. Estratégias complementares, como o pós-processamento morfológico e o treinamento com técnicas de *data augmentation* específicas para ruído e artefatos, podem contribuir para reduzir a sensibilidade do modelo a essas variações e melhorar sua confiabilidade em cenários desafiadores.

#### **4.6. Comparação com a Literatura**

A literatura especializada em segmentação volumétrica tridimensional (3D) demonstra que os modelos fundadores, como a 3D U-Net e a V-Net, estabeleceram uma faixa de valores para o Coeficiente de *Dice* — métrica amplamente utilizada para avaliar a sobreposição entre as predições do modelo e as anotações de referência — geralmente entre 0,70 e 0,85 para órgãos sólidos, como fígado, rins e cérebro (Çiçek *et al.*, 2016; Milletari *et al.*, 2016). Estudos mais recentes indicam que arquiteturas baseadas na U-Net 3D, quando combinadas com mecanismos de atenção (*attention blocks*) ou estratégias de *ensemble*, podem alcançar desempenhos superiores, atingindo valores entre 0,88 e 0,90 de *Dice Score*. Contudo, esses ganhos de acurácia estão frequentemente

associados a um aumento expressivo do custo computacional e da complexidade arquitetural (Isensee *et al.*, 2021).

No presente trabalho, o modelo desenvolvido obteve um *Dice* médio de 0,7608. Este valor se situa dentro da faixa reportada para as arquiteturas 3D fundadoras, confirmando a competitividade dos resultados face a modelos de referência, mesmo sem a incorporação de técnicas adicionais complexas. Ademais, o modelo proposto mantém uma estrutura arquitetural simples e eficiente, o que favorece sua integração em *pipelines* clínicos e aplicações que demandam baixo custo computacional e facilidade de implementação.

#### **4.7. Limitações**

Apesar do bom desempenho obtido pelo modelo U-Net 3D, observou-se que a necessidade de reduzir a resolução das imagens durante o treinamento representou uma limitação relevante. Essa redução, necessária para adequar os volumes à capacidade de memória da GPU, resultou em perda de fidelidade espacial das máscaras segmentadas, que se apresentaram menores e deslocadas em relação às imagens originais. Além disso, durante o processo de reescala das máscaras para o tamanho original, ocorreu perda de detalhes anatômicos significativos.

Ainda assim, o modelo demonstrou desempenho satisfatório e consistente, sendo capaz de produzir estimativas volumétricas clinicamente úteis por meio da aplicação de fatores de escala que compensam a redução de resolução. Dessa forma, os resultados indicam que, mesmo diante de limitações de *hardware* e compromissos entre resolução e viabilidade computacional, é possível alcançar resultados quantitativos confiáveis e reproduzíveis com relação a cálculo de volume.

#### 4.7.1. Cálculo de Volume com Fator de Escala

A estimativa volumétrica das estruturas segmentadas foi realizada a partir das máscaras produzidas pelo modelo U-Net 3D. Entretanto, como o treinamento e a inferência foram conduzidos com volumes reduzidos ( $128 \times 128 \times 128$  *voxels*), tornou-se necessário corrigir o volume final para o espaço físico original da imagem. Essa correção foi feita aplicando-se um fator de escala tridimensional, calculado a partir das diferenças entre as dimensões físicas do volume original e da versão reduzida.

Primeiramente, os arquivos NRRD correspondentes à imagem original, à imagem reduzida e à máscara segmentada foram carregados e processados com a biblioteca nrrd, sendo extraídas as informações de cabeçalho (*header*) referentes ao espaçamento entre *voxels* (*spacing*). Esse espaçamento indica a dimensão física de cada *voxel* em milímetros (mm) ao longo dos três eixos — X (largura), Y (altura) e Z (profundidade) — permitindo converter contagens de *voxels* em unidades métricas de volume.

O número total de *voxels* pertencentes à região segmentada foi obtido pela contagem de elementos com valor maior que zero na máscara binária. Em seguida, foi calculado o volume do *voxel* reduzido, multiplicando-se o espaçamento entre *voxels* nos três eixos:

$$V_{\text{voxel},\text{red}} = s_x \times s_y \times s_z$$

O volume reduzido total da máscara, em milímetros cúbicos, foi então obtido como:

$$V_{\text{red}} = N_{\text{voxels}} \times V_{\text{voxel},\text{red}}$$

onde

$N_{\text{voxels}}$  representa o número de *voxels* positivos (segmentados).

Para ajustar esse volume à escala física original, foi calculado um fator de escala volumétrico ( $vol_{\text{scale}}$ ), representando a razão entre o tamanho físico total do volume original e o da versão reduzida. Esse fator considera a diferença entre as dimensões e o espaçamento dos dois volumes:

$$ratio_{eixo} = \frac{dim_{orig,eixo} \times s_{orig,eixo}}{dim_{red,eixo} \times s_{red,eixo}}$$

$$vol_{scale} = \prod_{eixos} ratio_{eixo}$$

Assim, o volume estimado no espaço original foi calculado como:

$$V_{orig} = V_{red} \times vol_{scale}$$

Por fim, o volume foi convertido de milímetros cúbicos (mm<sup>3</sup>) para centímetros cúbicos (cm<sup>3</sup>) dividindo-se o resultado por 1000, facilitando a interpretação clínica.

Esse procedimento garantiu que o volume estimado refletisse a escala física verdadeira do exame original, mesmo que a segmentação tenha sido realizada sobre uma versão reduzida do volume. Dessa forma, foi possível obter estimativas volumétricas coerentes com medições anatômicas reais, corrigindo os efeitos da redução de resolução empregada para viabilizar o treinamento do modelo em GPU.

Além disso, o volume corrigido foi utilizado para estimar faixas de idade gestacional fetal, com base em curvas de referência publicadas para volumes pulmonares obtidos por ressonância magnética. A interpolação linear entre os valores mínimos e máximos de volume por faixa etária permitiu classificar o volume estimado como dentro, acima ou abaixo do intervalo esperado, fornecendo uma medida clínica adicional de validação do modelo.



## 5 Conclusão e Trabalhos Futuros

Mesmo com as limitações impostas pela redução de resolução, o modelo foi capaz de aprender representações espaciais relevantes e produzir segmentações volumétricas e foi possível calcular o volume através do fator de escala.

Como trabalho futuro, planeja-se treinar o modelo em resolução original com *hardware* mais potente, a fim de aumentar a precisão, utilização das máscaras e preservar a riqueza de detalhes para aplicações clínicas e de pesquisa. Assim, será possível ter produzir segmentações volumétricas úteis, representando um avanço significativo para aplicações clínicas e de pesquisa em segmentação médica tridimensional.

## 6

**Referências bibliográficas**

Cardoso, M. J. *et al.* MONAI: An Open-Source Framework for Deep Learning in Healthcare. arXiv preprint arXiv:2211.02701, 2022.

Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S. S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2016. p. 424–432.

Fu, Y.; Lei, Y.; Wang, T.; Curran, W.J.; Liu, T.; Yang, X. (2021). A Review of Deep Learning-based Methods for Medical Image Multi-Organ Segmentation. *Physica Medica*, 85, 107-122.  
<https://doi.org/10.1016/j.ejmp.2021.05.003>

Gao, Y.; Jiang, Y.; Peng, Y.; Yuan, F.; Zhang, X.; Wang, J. Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods. *Tomography* 2025, 11, 52. <https://doi.org/10.3390/tomography11050052>

Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning. Cambridge, MA: MIT Press, 2016.

Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021 Feb;18(2):203-211. doi: 10.1038/s41592-020-01008-z. Epub 2020 Dec 7. PMID: 33288961.

Kondrateva, E., Druzhinina, P., Dalechina, A., Zolotova, S., Golanov, A., Shirokikh, B., Belyaev, M., & Kurmukov, A. (2022). Negligible effect of brain MRI data preprocessing for tumor segmentation. arXiv. <https://arxiv.org/abs/2204.05278>

Lecun, Y. *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278-2324, 1998.

Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical Image Analysis*, v. 42, p. 60-88, 2017.

Ma, Z., Tavares, J. M. R. S., & Jorge, R. M. N. (2009). A review on the current segmentation algorithms for medical images. SCITEPRESS. <https://www.scitepress.org/PublishedPapers/2009/17935/17935.pdf>

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV), 565–571. <https://doi.org/10.1109/3DV.2016.79>

Müller, Dominik & Soto-Rey, Iñaki & Kramer, Frank. (2021). Robust chest CT image segmentation of COVID-19 lung infection based on limited data. *Informatics in Medicine Unlocked*. 25. 100681. 10.1016/j.imu.2021.100681.

Nosrati, M. S., & Hamarneh, G. (2016). Incorporating prior knowledge in medical image segmentation: A survey. *arXiv*. <https://arxiv.org/abs/1607.01092>

Pereira, S. *et al.* Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1240–1251, 2016.

Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1), 315-337.

Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015. p. 234–241.

Silva, J. F. (2022). Paying attention to the boundaries in semantic image segmentation (Tese de Mestrado). Universidade Federal da Bahia, iVision Lab. Disponível em: <https://ivisionlab.ufba.br/doc/thesis/2022/Paying-attention-to-the-boundaries-in-semantic-image-segmentation.pdf>

Suri, J. *et al.* UNet Deep Learning Architecture for Segmentation of Vascular and Non-Vascular Images: A Microscopic Look at UNet Components Buffered With Pruning, Explainable Artificial Intelligence, and Bias. *IEEE Access*, [S. l.], v. PP, p. 1-1, 1 jan. 2022. DOI: 10.1109/ACCESS.2022.3232561. Disponível em: <https://ieeexplore.ieee.org/document/9998486>. Acesso em: 10/11/2025.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15, 29.

Wu Y, Lin L, Wang J, Wu S. [Application of semantic segmentation based on convolutional neural network in medical images]. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*. 2020 Jun 25;37(3):533-540. Chinese. doi: 10.7507/1001-5515.201906067. PMID: 32597097; PMCID: PMC10319567.

Y. Enokiya, Y. Iwamoto, Y.-W. Chen, and X.-H. Han, "Automatic Liver Segmentation Using U-Net with Wasserstein GANs," *Journal of Image and Graphics*, vol. 7, no. 3, pp. 94-101, Sep. 2019. Disponível em: <https://www.joig.net/uploadfile/2019/0823/20190823021541399.pdf>

Zaidi, H., 2005. *Quantitative Analysis in Nuclear Medicine Imaging*, Springer.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology*, 11(2), 178–189. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8)