

Estephane Mendes Nascimento

**Desenvolvimento de método em cascata para  
realização do cálculo do ângulo de Cobb em  
imagens de raio-x da coluna vertebral**

São Luís - MA

Março de 2026

Estephane Mendes Nascimento

Universidade Federal do Maranhão – UFMA

Centro de Ciências Exatas e Tecnologias

Departamento de Engenharia Elétrica

Programa de pós-graduação em Engenharia Elétrica

**Desenvolvimento de método em cascata para realização  
do cálculo do ângulo de Cobb em imagens de raio-x da  
coluna vertebral**

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

São Luís - MA

Março de 2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Nascimento, Estephane Mendes.

Desenvolvimento de método em cascata para realização do cálculo do ângulo de Cobb em imagens de raio-x da coluna vertebral / Estephane Mendes Nascimento. - 2026.

86 f.

Coorientador(a) 1: João Dallyson Sousa de Almeida.

Orientador(a): Aristófanês Corrêa Silva.

Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, São Luís, 2026.

1. Coluna Vertebral. 2. Raio-x. 3. Escoliose. 4. Ângulo de Cobb. 5. Segmentação. I. Almeida, João Dallyson Sousa de. II. Silva, Aristófanês Corrêa. III. Título.

Estephane Mendes Nascimento

## **Desenvolvimento de método em cascata para realização do cálculo do ângulo de Cobb em imagens de raio-x da coluna vertebral**

Dissertação de mestrado apresentada ao programa de pós-graduação em Engenharia Elétrica da Universidade Federal do Maranhão na área de Ciência da Computação, como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

Trabalho apresentado em: São Luís - MA, 27 de Março de 2026.

**Prof. Dr. Aristófanés Corrêa Silva**  
Orientador

**Prof. Dr. João Dallyson Sousa de Almeida**  
Coorientador

**Prof. Dr. Anselmo Cardoso de Paiva**  
Universidade Federal do Maranhão  
Banca Examinadora

**Profa. Dra. Aura Conci**  
Universidade Federal Fluminense  
Banca Examinadora

São Luís - MA  
Março de 2026

*Este trabalho é dedicado a minha família*

# Agradecimentos

Gostaria de agradecer, primeiramente, a Deus, por estar sempre presente em minha vida e por ter me concedido forças para concluir esta etapa.

Aos meus pais, Bernardete Lourdes e José Pedro, pelo incentivo à continuidade dos meus estudos, pelo apoio incondicional e por todos os sacrifícios realizados ao longo dessa caminhada. A minha irmã, Sofia, que esteve ao meu lado durante todo esse processo. A minha tia Ivone e a minha avó Lourdes, por todo o apoio, carinho e preocupação.

Aos meus professores da Universidade Federal do Maranhão (UFMA), em especial ao meu orientador, Aristófanés, pelas valiosas contribuições ao longo desta pesquisa, bem como pelo apoio, paciência e compreensão. Ao meu coorientador, João Dallyson, por estar sempre disposto a ajudar, pela compreensão, apoio e acolhimento que se estendem desde a graduação. A professora Simara, pelo incentivo e compreensão em todos os momentos em que precisei. Ao professor Anselmo, pelo acolhimento, disponibilidade e por ter possibilitado minha entrada no Núcleo de Computação Aplicada (NCA). Ao professor Ítalo, por toda a ajuda fornecida durante o mestrado.

Aos órgãos de fomento FAPEMA, CAPES e CNPq, pelo apoio financeiro e pelos recursos concedidos durante o mestrado.

Aos meus amigos Alexandre Pessoa, Gabriel, Kelson, Daniel, Alexandre Araújo, Alan, Alison, Maurício, Marcus Vinícius, Saulo, Pedro, João Pedro, Victor e Mário, por toda a ajuda oferecida e pela amizade.

Aos amigos que fiz no grupo de jovens da igreja, Brenna, Gabriel, Alice, Aline, Isadora e Caio.

Agradeço também a minha amiga Mackele, Arnaud, Márcio, Joana, Anderson, Sabrina, Max, Sara, Marcos, Lucas, Marcos Vinícius e Celso por tornarem meus dias mais alegres e por estarem sempre dispostos a ajudar.

*Tudo o que vale a pena ter,  
vale a pena se esforçar para ter.  
(Andrew Carnegie)*

# Resumo

A escoliose é uma deformidade da coluna vertebral que afeta cerca de 2 a 4% da população mundial, e seu diagnóstico precoce é fundamental para conter a progressão da doença e evitar casos graves, que podem ser reversíveis apenas por meio de cirurgia corretiva. O diagnóstico é geralmente realizado por meio de imagens de raio-x, nas quais o especialista mede o ângulo de Cobb, parâmetro utilizado para determinar a gravidade da deformidade. Com o objetivo de auxiliar os especialistas nesse processo, este trabalho apresenta um método para a estimativa automática do ângulo de Cobb com base na segmentação das vértebras. Inicialmente, realiza-se a extração da região de interesse (ROI) por meio da segmentação da coluna, seguida pela segmentação das vértebras a partir da imagem delimitada pela ROI e, por fim, estima-se o ângulo de Cobb para cada máscara vertebral obtida. O método proposto alcançou coeficientes Dice de 92,14% e 83,23% para a segmentação da coluna e das vértebras, respectivamente, e um erro absoluto médio (MAE) de 8,72° no cálculo do ângulo de Cobb.

**Palavras-chave:** Coluna vertebral, Raio-x, Escoliose, Ângulo de Cobb, Segmentação, Unet 3+.

# Abstract

Scoliosis is a spinal deformity that affects approximately 2 to 4% of the global population, and its early diagnosis is essential to prevent disease progression and avoid severe cases, which may only be reversible through corrective surgery. The diagnosis is usually performed using x-ray images, in which the specialist measures the Cobb angle, a parameter used to determine the severity of the deformity. In order to assist specialists in this process, this work presents a method for the automatic estimation of the Cobb angle based on vertebral segmentation. Initially, the region of interest (ROI) is extracted through spine segmentation, followed by vertebrae segmentation from the image delimited by the ROI and, finally, the Cobb angle is estimated for each obtained vertebral mask. The proposed method achieved Dice coefficients of 92.14% and 83.23% for spine and vertebrae segmentation, respectively, and a mean absolute error (MAE) of 8.72° in Cobb angle calculation.

**Keywords:** Spine, X-ray, Scoliosis, Cobb angle, Segmentation, UNet 3+.

# Lista de ilustrações

Figura 1 – Coluna vertebral. . . . .	30
Figura 2 – Paciente com escoliose. . . . .	31
Figura 3 – Tipos de curvas geradas pela escoliose na coluna vertebral. . . . .	33
Figura 4 – Ângulo de Cobb. . . . .	34
Figura 5 – Goniômetro . . . . .	34
Figura 6 – Arquitetura de uma rede neural com uma camada oculta . . . . .	37
Figura 7 – Modelo matemático do neurônio de McCulloch-Pitts . . . . .	37
Figura 8 – Rede Neural Convolutiva. . . . .	41
Figura 9 – Aplicação de zero padding. . . . .	42
Figura 10 – Operação de convolução. . . . .	42
Figura 11 – Operação de Max pooling. . . . .	42
Figura 12 – Vetorização da última camada de pooling da rede neural convolutiva. . . . .	43
Figura 13 – Arquitetura U-Net. . . . .	44
Figura 14 – Arquitetura U-Net 3+. . . . .	45
Figura 15 – Representação do mapa de características agregado das <i>skip connections</i> . . . . .	46
Figura 16 – Esquema do Attention Gate. . . . .	47
Figura 17 – Bloco Squeeze-and-Excitation. . . . .	48
Figura 18 – Composição camada do codificador U-Net 3+ com adição dos mecanismos de atenção SE e CBAM. . . . .	48
Figura 19 – Visão geral do bloco CBAM. . . . .	49
Figura 20 – Fluxograma do método proposto e suas etapas: (1) Extração da ROI, (2) Segmentação das vértebras, (3) Estimação dos ângulos de Cobb. . . . .	53
Figura 21 – Pré-processamento da imagem de entrada para extração da ROI . . . . .	53
Figura 22 – U-Net 3+ modificada com adição de <i>Attention Gate</i> . . . . .	54
Figura 23 – Pós-processamento: (a) Saída da rede de segmentação, (b) Aplicação de erosão e dilatação, (c) Identificação da estrutura da coluna e (d) Imagem intermediária do pós-processamento . . . . .	56
Figura 24 – Aplicação do fecho convexo após o processamento da máscara . . . . .	56
Figura 25 – Recorte região da coluna vertebral com base na ROI . . . . .	56
Figura 26 – Segmentação das vértebras . . . . .	57
Figura 27 – U-Net 3+ modificada com adição dos mecanismos de atenção SE e CBAM . . . . .	58
Figura 28 – Etapas da estimação do ângulo de Cobb: (a) Centróides vertebrais, (b) Pontos terminais de cada vértebra, (c) Linhas terminais de cada vértebra, (d) Linhas que mostram a inclinação de cada vértebra e (e) Placas terminais das vértebras usadas no cálculo dos ângulos de Cobb encontrados, com os respectivos valores . . . . .	62

Figura 29 – Máscara real das vértebras e da coluna vertebral. . . . .	64
Figura 30 – Curva de loss de treinamento e validação . . . . .	67
Figura 31 – Pós-processamento: (a) Segmentação da coluna, (b) Resultado pós- processamento e (c) Máscara real . . . . .	69
Figura 32 – Curva de Loss de treinamento e validação . . . . .	71
Figura 33 – Máscara predita com junção de vértebras e máscara anotada . . . . .	75
Figura 34 – Pior caso de estimação do ângulo de Cobb . . . . .	76
Figura 35 – Melhor caso de estimação do ângulo de Cobb . . . . .	77

# Lista de tabelas

Tabela 1 – Estimativa do Ângulo de Cobb sem Segmentação . . . . .	24
Tabela 2 – Estimativa do Ângulo de Cobb com Segmentação . . . . .	27
Tabela 3 – Definição do ângulo de Cobb . . . . .	35
Tabela 4 – Comparação de arquiteturas para extração da ROI . . . . .	66
Tabela 5 – Validação cruzada da Unet 3+ com Attention Gate para segmentação da coluna . . . . .	67
Tabela 6 – Pós-processamento segmentação da coluna . . . . .	68
Tabela 7 – Aplicação do fecho convexo após o pós-processamento segmentação da coluna . . . . .	69
Tabela 8 – Comparação de arquiteturas para segmentação das vértebras com base na ROI gerada na etapa de extração feita por cada modelo . . . . .	70
Tabela 9 – Comparação de arquiteturas para segmentação das vértebras com base na ROI gerada pelo modelo Unet 3+ com Attention Gate . . . . .	71
Tabela 10 – Validação cruzada da Unet 3+ com Bloco SE e CBAM para segmentação das vértebras . . . . .	72
Tabela 11 – Comparação da segmentação sem extração da ROI em diferentes arquiteturas . . . . .	72
Tabela 12 – Comparação da segmentação sem extração da ROI em diferentes arquiteturas . . . . .	73
Tabela 13 – Comparação da segmentação da coluna com os trabalhos relacionados .	74
Tabela 14 – Comparação da segmentação das vértebras com os trabalhos relacionados	74
Tabela 15 – Comparação do cálculo do ângulo de Cobb com os trabalhos relacionados	75

# Lista de abreviaturas e siglas

- AASCE2019** Estimativa Automatizada Precisa da Curvatura Espinhal (do inglês, *Accurate Automated Spinal Curvature Estimation*). 23
- ACF** Canal de Características Agregadas (do inglês, *Aggregated Channel Features*). 25
- AG** Portão de Atenção (do inglês, *Attention Gate*). 46, 47
- AP** ântero-posterior. 19, 26, 64
- ASPP** *Atrous Spatial Pyramid Pooling*. 65
- CACM** Método de Cálculo do Ângulo de Cobb (do inglês, *Cobb Angle Calculation Method*). 23
- CBAM** *Convolutional Block Attention Module*. 26, 48, 57, 70, 71, 78
- CD** Distância de Chebyshev (do inglês, *Chebyshev distance*). 23, 26
- CLAHE**) Equalização adaptativa do Histograma com Contraste Limitado (do inglês, *Contrast Limited Adaptive Histogram Equalization*). 78
- CMAE** Erro Médio Absoluto Circular (do inglês, *Circular Mean Absolute Error*). 21, 23, 26, 27
- CNNs** Redes Neurais Convolucionais (do inglês, *Convolutional Neural Network*). 40, 41
- DL** Aprendizado profundo (do inglês, *Deep Learning*). 18
- ED** Distância Euclidiana (do inglês, *Euclidean distance*). 23, 26
- EI** Escoliose Idiopática. 17
- EIA** Escoliose idiopática do Adolescente. 17
- EM** Erro Médio. 22
- FCN** Redes Totalmente Conectadas (do inglês, *Fully Connected Network*). 43
- FN** Falso Negativo. 49
- FP** Falso Positivo. 49
- FREM** Módulo de Aprimoramento da Robustez de Recursos (do inglês, *Feature Robustness Enhancement Module*). 23

**HRNET** Rede de Alta Resolução (do inglês, *High Resolution Network*). 23

**IA** inteligência artificial. 18

**ICC** Coeficiente de Correlação Intraclasse (do inglês, *Intraclass Correlation Coefficient*). 22

**IOU** Intersecção sobre União (do inglês, *Intersection over Union*). 40, 49, 50

**LaNet** Rede Sensível a Pontos de Referência (do inglês, *Landmark-aware Network*). 23

**LocNet** Rede de Localização (do inglês, *Localization Network*). 25, 26

**LOF** Função Objetivo Orientada a Pontos de Referência (do inglês, *Landmark-aware Objective Function*). 23

**MAD** Desvio Absoluto Médio (do inglês, *Mean Absolute Deviation*). 22

**MAE** Erro Absoluto Médio (do inglês, *Mean Absolute Error*). 25, 26, 50, 73–75

**Mask R-CNN** Rede Neural Convolutiva Baseada em Regiões com Máscara (do inglês, *Mask Region-based Convolutional Neural Network*). 21, 22

**MD** Distância de Manhattan (do inglês, *Manhattan distance*). 23, 26

**mIOU** Média da Intersecção sobre a União (do inglês, *Mean Intersection Over Union*). 26

**ML** aprendizado de máquina (do inglês, *Machine Learning*). 18

**MS-SSIM** Índice de similaridade estrutural multiescala (do inglês, *Multi-scale structural similarity index*). 39

**MT** Torácica Principal (do inglês, *Main Thoracic*). 24, 25, 59, 60

**MWCE-CE** estrutura de Estimação do Ângulo de Cobb Aprimorada por Convolução Wavelet Multiescala (do inglês, *Multi-scale Wavelet Convolution-Enhanced Cobb angle Estimation framework*). 26

**OMS** Organização Mundial da Saúde. 17

**PLN)** Processamento de linguagem natural. 40

**PT** Torácica Proximal (do inglês, *Proximal Thoracic*). 24, 25, 59, 60

**RegNet)** Rede de Regressão (do inglês, *Regression Network*). 25

**ReLU** Unidade Linear Retificadora (do inglês, *Rectified Linear Unit*). 38, 42, 43, 45, 46

**RGB** Vermelho, Verde e Azul (do inglês, *Red, Green and Blue*). 22

**RNA** Redes Neurais Artificiais. 35

**ROI** Região de Interesse (do inglês, *Region of Interest*). 19, 22, 26, 28, 52, 56, 57, 63–65, 68–70, 72, 73, 76, 78

**SE** *Squeeze-and-Excitation*. 47, 57, 70, 71, 78

**SegNet** Rede de Segmentação (do inglês, *Segmentation Network*). 25, 26

**SMAPE** Erro Percentual Absoluto Médio Simétrico (do inglês, *Symmetric Mean Absolute Percentage Error* ). 21, 23, 25, 26, 50, 73–75

**TL/L** Toracolombar/Lombar (do inglês, *Thoracolumbar/Lumbar*). 24, 25, 59, 60

**VFLDNet** Rede de Detecção de Pontos de Referência Focada na Vértebra (do inglês, *Vertebra-Focused Landmark Detection Network*). 23

**VLTENet** Rede de Localização e Estimativa de Inclinação de Vértebras (do inglês, *Vertebra Localization and Tilt Estimation Network*). 23

**VN** Verdadeiro Negativo. 49

**VP** Verdadeiro Positivo. 49

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
1.1	<b>Objetivos</b>	<b>19</b>
1.1.1	Objetivo Geral	19
1.1.2	Objetivos Específicos	19
1.2	<b>Organização do trabalho</b>	<b>20</b>
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>21</b>
2.1	<b>Estimação do ângulo de Cobb sem segmentação</b>	<b>21</b>
2.2	<b>Estimação do ângulo de Cobb com uso de segmentação</b>	<b>24</b>
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>29</b>
3.1	<b>Coluna vertebral</b>	<b>29</b>
3.2	<b>Escoliose</b>	<b>30</b>
3.2.1	Ângulo de Cobb	33
3.3	<b>Aprendizado Profundo</b>	<b>35</b>
3.3.1	Redes neurais artificiais	35
3.3.2	Neurônio artificial	36
3.3.3	Função de ativação	38
3.3.4	Função de perda	39
3.3.5	Rede Neural Convolucional	40
3.3.6	U-Net	43
3.3.7	U-Net 3+	44
3.3.8	Atenção	46
3.3.9	<i>Attention Gate</i>	46
3.3.10	<i>Squeeze-and-Excitation</i>	47
3.3.11	<i>Convolutional Block Attention Module</i>	48
3.3.12	Métricas de avaliação	49
3.4	<b>Considerações Finais</b>	<b>51</b>
<b>4</b>	<b>MÉTODO</b>	<b>52</b>
4.1	<b>Extração da ROI</b>	<b>52</b>
4.1.1	Pré-processamento	52
4.1.2	Segmentação da coluna vertebral	52
4.1.3	Pós-processamento	54
4.2	<b>Recorte da imagem de raio-x com base na ROI</b>	<b>56</b>
4.3	<b>Segmentação das vértebras</b>	<b>57</b>

4.4	<b>Estimação automática do ângulo de Cobb</b>	<b>58</b>
4.4.1	Construção das retas médias vertebrais	59
4.4.2	Detecção automática das vértebras	59
4.4.3	Cálculo do ângulo de Cobb	60
4.5	<b>Considerações Finais</b>	<b>63</b>
5	<b>RESULTADOS E DISCUSSÃO</b>	<b>64</b>
5.1	<b>Base de imagens</b>	<b>64</b>
5.2	<b>Configuração dos experimentos</b>	<b>65</b>
5.3	<b>Experimentos</b>	<b>65</b>
5.3.1	Extração da ROI	65
5.3.2	Aplicação do fecho convexo	68
5.3.3	Segmentação das vértebras com base na ROI	69
5.4	<b>Comparação segmentação das vértebras sem a extração da ROI</b>	<b>72</b>
5.4.1	Comparação do cálculo do ângulo de Cobb para segmentação direta e método proposto	73
5.4.2	Comparação com os trabalhos relacionados	73
6	<b>CONCLUSÃO</b>	<b>78</b>
	<b>REFERÊNCIAS</b>	<b>80</b>

# 1 Introdução

A escoliose é uma condição caracterizada por alterações morfológicas tridimensionais da coluna vertebral, afetando os planos coronal, sagital e transversal. No plano coronal, manifesta-se por uma inclinação lateral cuja gravidade é definida por um ângulo de Cobb superior a  $10^\circ$  (FEITOSA; FRANÇA; FISIOTERAPIA, 2023). De acordo com a Organização Mundial da Saúde (OMS), essa patologia acomete entre 2% e 4% da população mundial, o que corresponde a aproximadamente 200 milhões de pessoas; no Brasil, estima-se que cerca de 6 milhões de indivíduos apresentem esse diagnóstico (FEITOSA; FRANÇA; FISIOTERAPIA, 2023). A deformidade pode ter diferentes etiologias, incluindo malformações congênitas de uma ou mais vértebras, causas neuromusculares associadas a doenças do sistema neuromuscular, origem idiopática, além de fatores pós-traumáticos, degenerativos, infecciosos ou tumorais (CREFITO-14, 2022; FEITOSA; FRANÇA; FISIOTERAPIA, 2023).

O tipo mais frequente na população é a Escoliose Idiopática (EI). Conforme a própria denominação sugere, essa forma não apresenta uma causa definida (FEITOSA; FRANÇA; FISIOTERAPIA, 2023). A EI é classificada em infantil, juvenil e do adolescente, sendo a Escoliose idiopática do Adolescente (EIA) responsável por aproximadamente 90% dos casos. A prevalência global da EIA varia entre 0,5% e 5,2%, com maior risco de progressão observado em indivíduos do sexo feminino, possivelmente em decorrência de fatores genéticos (FERREIRA, 2025). Essa condição tem início, em geral, durante o período conhecido como “estirão do crescimento”, ocorrendo normalmente entre 11 e 13 anos em meninas e entre 14 e 16 anos em meninos. Nessa fase, o crescimento acelerado da coluna ocorre de forma desorganizada, resultando em alterações morfológicas que tendem a persistir na vida adulta e nas fases subsequentes do desenvolvimento (FEITOSA; FRANÇA; FISIOTERAPIA, 2023).

O diagnóstico da EIA é realizado com base na anamnese, no exame físico e na avaliação por imagem radiológica. A confirmação ocorre por meio de radiografia, quando o ângulo de Cobb é superior a  $10^\circ$ , na ausência de distúrbios neuromusculares, malformações vertebrais ou outras patologias associadas (HORNE; FLANNERY; USMAN, 2014).

Para a quantificação do grau de desvio da coluna vertebral, utiliza-se o ângulo de Cobb, considerado o padrão ouro para a avaliação da deformidade escoliótica. A mensuração é realizada por um especialista a partir de uma radiografia, na qual são traçadas linhas ao longo das placas terminais superior e inferior da curva escoliótica. Em seguida, são construídas linhas perpendiculares a essas placas, sendo o ângulo formado entre essas perpendiculares definido como o ângulo de Cobb (AMES et al., 2019).

A identificação precoce e o tratamento adequado da escoliose, especialmente em estágios iniciais, podem retardar ou até interromper a progressão da deformidade antes da maturidade esquelética, contribuindo para melhores desfechos a longo prazo na vida adulta (LUK et al., 2010). As abordagens terapêuticas incluem fisioterapia, uso de coletes ortopédicos e intervenção cirúrgica, sendo a escolha do tratamento determinada principalmente pelo grau de deformidade apresentado.

Nos últimos anos, a **inteligência artificial (IA)** e o **aprendizado de máquina (do inglês, *Machine Learning*) (ML)** têm se consolidado como ferramentas fundamentais para a resolução de problemas complexos em diferentes domínios (RICHENS; LEE; JOHRI, 2020). A IA tem como objetivo dotar os computadores de capacidades inteligentes, sendo o aprendizado um requisito essencial para a manifestação de qualquer comportamento inteligente (KONONENKO, 2001). Nesse contexto, o aprendizado de máquina surge como um dos principais paradigmas da IA, permitindo que sistemas computacionais aprendam padrões a partir de dados sem a necessidade de programação explícita.

O aprendizado de máquina explora a crescente disponibilidade de dados, especialmente no contexto médico, para fornecer diagnósticos mais precisos e personalizados (RICHENS; LEE; JOHRI, 2020). Em particular, algoritmos de ML e de **Aprendizado profundo (do inglês, *Deep Learning*) (DL)** têm demonstrado elevada eficácia na identificação de padrões sutis e anomalias em dados de imagem, frequentemente imperceptíveis à análise humana (SULTANA et al., 2023). Essas abordagens têm impulsionado avanços significativos em tarefas como detecção, segmentação e classificação de imagens médicas.

Dessa forma, o aprendizado de máquina mostra-se especialmente adequado para a análise de dados médicos, possibilitando que o conhecimento diagnóstico seja automaticamente extraído a partir de casos previamente resolvidos. Os modelos derivados podem então ser utilizados como ferramentas de apoio à decisão clínica, auxiliando o médico no diagnóstico de novos pacientes e contribuindo para o aumento da velocidade, da precisão e da confiabilidade do processo diagnóstico (KONONENKO, 2001).

No que se refere à estimação do ângulo de Cobb, estudos como o de (SARDJONO et al., 2013) evidenciam que métodos automáticos apresentam desempenho superior quando comparados às abordagens manuais tradicionalmente descritas na literatura. Além de proporcionar maior consistência e precisão nas medições, a automatização do cálculo do ângulo de Cobb contribui significativamente para a redução do tempo necessário ao diagnóstico, bem como para a diminuição da carga cognitiva e do cansaço do especialista, decorrentes da repetição contínua das medições em múltiplos pacientes. Dessa forma, os métodos automáticos configuram-se como ferramentas promissoras de apoio à prática clínica, potencializando a eficiência e a confiabilidade do diagnóstico.

Diante do exposto, este trabalho tem como objetivo propor um método para o cálculo automático do ângulo de Cobb, com a finalidade de auxiliar profissionais da área

ortopédica na realização dessa medição. Busca-se, assim, reduzir o esforço associado ao processo manual, minimizar a variabilidade interobservador, considerando as diferenças nos resultados quando a mensuração é realizada por distintos especialistas, e contribuir para maior agilidade e eficiência no diagnóstico.

## 1.1 Objetivos

Nesta seção serão descritos o objetivo geral e os objetivos específicos.

### 1.1.1 Objetivo Geral

O objetivo deste trabalho é propor um método automático para o cálculo do ângulo de Cobb em imagens de raio-x **ântero-posterior (AP)** da coluna vertebral de pacientes com escoliose em diferentes estágios de progressão da doença. Para isso, são empregadas técnicas de aprendizado profundo, processamento de imagens e cálculos geométricos, com o intuito de fornecer uma ferramenta de apoio ao diagnóstico e uma primeira opinião aos profissionais da área ortopédica.

### 1.1.2 Objetivos Específicos

Para alcançar o objetivo principal deste trabalho, foi proposto o desenvolvimento de um método estruturado em etapas, visando atender ao objetivo geral. Dessa maneira, fez-se necessário o cumprimento dos seguintes objetivos específicos.

- Desenvolver um método computacional para extração de uma **Região de Interesse** (do inglês, *Region of Interest*) (ROI);
- Desenvolver um método computacional para a geração de mapas de segmentação das vértebras;
- Desenvolver um método computacional para a estimativa do ângulo de Cobb a partir dos mapas de segmentação das vértebras;
- Comparar os resultados obtidos pelo método proposto com as medições realizadas por especialistas e com os resultados presentes na literatura.

## 1.2 Organização do trabalho

Este trabalho está organizado da seguinte forma:

**Capítulo 2 – Trabalhos Relacionados** – capítulo dedicado à apresentação dos trabalhos relacionados ao tema investigado, descrevendo os métodos propostos e os resultados obtidos na literatura.

**Capítulo 3 – Fundamentação Teórica** – capítulo que reúne os conceitos necessários para a compreensão do trabalho desenvolvido. Nele são abordadas a estrutura da coluna vertebral, a escoliose, seus processos de desenvolvimento e tratamento, bem como o conceito do ângulo de Cobb utilizado para a mensuração do grau de escoliose. Adicionalmente, são apresentados os fundamentos de redes neurais, redes neurais convolucionais, os principais termos associados a essas técnicas e as métricas empregadas para validação e comparação dos resultados deste trabalho.

**Capítulo 4 – Método** – capítulo no qual são apresentadas e detalhadas todas as etapas necessárias para o desenvolvimento da pesquisa, incluindo a descrição do método em cascata utilizado para a segmentação das vértebras e do método para a estimativa do ângulo de Cobb.

**Capítulo 5 – Resultados e Discussão** – capítulo destinado à apresentação dos resultados obtidos com a metodologia desenvolvida, bem como à análise e discussão crítica desses resultados.

**Capítulo 6 – Conclusão** – capítulo que apresenta uma visão geral do trabalho desenvolvido, destacando a eficácia da metodologia proposta, suas limitações, as contribuições alcançadas e as possibilidades de melhorias, além de sugestões para trabalhos futuros.

## 2 Trabalhos Relacionados

Considerando o esforço significativo despendido por profissionais da área ortopédica na marcação manual dos ângulos pelo método de Cobb em imagens de raio-x de pacientes com escoliose, diversos estudos têm sido desenvolvidos com o objetivo de automatizar esse processo e auxiliar o diagnóstico clínico. Os métodos propostos podem ser classificados em duas categorias principais: aqueles que estimam o ângulo de Cobb diretamente, sem a necessidade de segmentação prévia, e aqueles que incluem uma etapa inicial de segmentação das estruturas vertebrais para, posteriormente, realizar a mensuração. A busca por trabalhos relacionados a este estudo foi realizada por meio do Google Acadêmico, abrangendo publicações no período de 2018 a 2026. Neste capítulo, são apresentados alguns dos principais trabalhos voltados à mensuração automática do ângulo de Cobb em imagens de raio-x da coluna vertebral.

### 2.1 Estimação do ângulo de Cobb sem segmentação

No que se refere aos trabalhos que não utilizam segmentação, a imagem de raio-x é empregada diretamente como entrada do método, sendo nela estimadas as terminações das vértebras e calculados os ângulos de Cobb, sem a geração de imagens intermediárias. Nesse contexto, [Kim et al. \(2020\)](#) desenvolveram um método composto por três etapas. Na primeira, os centróides de cada vértebra são localizados por meio da Centroid-Net; na segunda, a inclinação vertebral é estimada pela M-Net; por fim, o ângulo de Cobb é calculado a partir da combinação do mapa de confiança gerado pela Centroid-Net com as inclinações fornecidas pela M-Net. Os resultados obtidos indicaram um [Erro Percentual Absoluto Médio Simétrico](#) (do inglês, *Symmetric Mean Absolute Percentage Error*) (SMAPE) de 7,84% e um [Erro Médio Absoluto Circular](#) (do inglês, *Circular Mean Absolute Error*) (CMAE) de 3,51.

No mesmo segmento, [Yi et al. \(2020\)](#) empregaram a ResNet34 ([HE et al., 2016a](#)) para a extração de características no *encoder* da rede. As *skip connections* combinam as características extraídas pela ResNet34 ([HE et al., 2016a](#)) com aquelas provenientes de camadas profundas. A partir da saída da última camada, são construídos *heatmaps* por meio de camadas convolucionais, permitindo a localização dos pontos terminais de cada vértebra. Como métrica de validação, foi utilizado o [SMAPE](#), cujo valor obtido foi de 10,81%.

De forma distinta, [Suri et al. \(2023\)](#) propuseram a rede SpineTK para estimar o ângulo de Cobb, uma modificação da [Rede Neural Convolutiva Baseada em Regiões com Máscara](#) (do inglês, *Mask Region-based Convolutional Neural Network*) (Mask R-CNN)

(HE et al., 2017). A SpineTK utiliza imagens de canal único como entrada, em contraste com as imagens Vermelho, Verde e Azul (do inglês, *Red, Green and Blue*) (RGB) de três canais para as quais a arquitetura Mask R-CNN foi originalmente projetada. Além disso, as camadas finais da rede foram adaptadas para produzir seis pontos de referência. O método foi aplicado a um conjunto privado de imagens radiográficas coletadas entre 2005 e 2020 em centros de radiologia da Universidade da Pensilvânia, sendo avaliado tanto nesse conjunto quanto no conjunto de dados público proposto por (FRAIWAN; AUDAT; MANASREH, 2022), alcançando um Erro Médio (EM) inferior a 2° em ambos os cenários.

Outra abordagem para a estimação do ângulo de Cobb foi apresentada por Zhou, Zhu e Yao (2023), na qual é proposta uma estrutura de rede para a detecção dos pontos centrais vertebrais, utilizando a MobileNetV3 (HOWARD et al., 2019) como *backbone* para extração de características. A arquitetura emprega a U-Net (RONNEBERGER; FISCHER; BROX, 2015) como módulo intermediário de refinamento, enquanto a etapa de previsão final adota o conceito de detecção de pontos-chave da CenterNet (DUAN et al., 2019), utilizando convoluções  $3 \times 3$  para ajuste da dimensionalidade do mapa de características e convoluções  $1 \times 1$  para a geração de mapas de ativação dos pontos centrais das vértebras e de seus deslocamentos. Para o cálculo do ângulo de Cobb, um polinômio de sétimo grau é ajustado aos pontos centrais estimados pela rede, sendo o ângulo determinado a partir da inclinação da reta normal à curva ajustada nas vértebras terminais superior e inferior. O desempenho do método foi avaliado em relação às medições de dois especialistas, obtendo Coeficiente de Correlação Intraclasse (do inglês, *Intraclass Correlation Coefficient*) (ICC) de 0,897 e 0,901 e Desvio Absoluto Médio (do inglês, *Mean Absolute Deviation*) (MAD) de 3,13 e 3,04, respectivamente, destacando-se que o método não realiza segmentação, sendo a U-Net (RONNEBERGER; FISCHER; BROX, 2015) utilizada apenas para extração e refinamento de características.

O trabalho de Maeda et al. (2023) explorou uma abordagem baseada na detecção de vértebras utilizando uma base privada composta por 1.021 imagens radiográficas de corpo inteiro. Inicialmente, foi realizada a extração da ROI, correspondente à porção da coluna vertebral formada por 12 vértebras torácicas e 5 lombares, por meio de aprendizado por transferência com a ResNet34 (HE et al., 2016a), na qual a camada de saída foi substituída por uma camada totalmente conectada com quatro neurônios, responsáveis por prever valores normalizados das coordenadas  $(x, y)$  dos cantos superior esquerdo e inferior direito da ROI. Em seguida, a detecção das vértebras foi realizada utilizando novamente a ResNet34 pré-treinada, sendo empregados mapas de calor com quatro canais para a localização vertebral e geradas três saídas: mapas de calor, deslocamento do centro e deslocamento do centro vertebral. Por fim, o ângulo de Cobb foi calculado a partir das inclinações estimadas com base nos quatro cantos de cada vértebra, identificando-se as vértebras terminais de inclinação máxima e mínima. Os resultados obtidos indicaram erro médio entre 1,7° e 2,2°, e ICC de 0,973.

A Rede Sensível a Pontos de Referência (do inglês, *Landmark-aware Network*) (LaNet), proposta por Yang, Wang e Meng (2024), adota uma abordagem baseada em pontos de referência, na qual quatro pontos são previstos diretamente em cada vértebra e os pontos esquerdo e direito são conectados para formar uma linha reta, dispensando a etapa de segmentação. O método é composto por três módulos: o Módulo de Aprimoramento da Robustez de Recursos (do inglês, *Feature Robustness Enhancement Module*) (FREM), responsável pela extração de características globais, locais e contextuais; a Função Objetiva Orientada a Pontos de Referência (do inglês, *Landmark-aware Objective Function*) (LOF), utilizada como função de perda; e o Método de Cálculo do Ângulo de Cobb (do inglês, *Cobb Angle Calculation Method*) (CACM), que determina o ângulo a partir da identificação das vértebras de inflexão com base na variação da inclinação vertebral. Os experimentos realizados no conjunto de dados do desafio Estimativa Automatizada Precisa da Curvatura Espinhal (do inglês, *Accurate Automated Spinal Curvature Estimation*) (AASCE2019) (WU et al., 2017) apresentaram desempenho de 8,99 no SMAPE, 3,63 no CMAE, 7,78 na Distância Euclidiana (do inglês, *Euclidean distance*) (ED), 11,84 na Distância de Manhattan (do inglês, *Manhattan distance*) (MD) e 6,29 na Distância de Chebyshev (do inglês, *Chebyshev distance*) (CD).

Lu et al. (2024) propuseram a Rede de Localização e Estimativa de Inclinação de Vértebras (do inglês, *Vertebra Localization and Tilt Estimation Network*) (VLTENet), com o objetivo de superar o desempenho da Rede de Detecção de Pontos de Referência Focada na Vértebra (do inglês, *Vertebra-Focused Landmark Detection Network*) (VFLDNet) (YI et al., 2020) na mensuração do ângulo de Cobb. Diferentemente da VFLDNet, que se baseia na detecção de pontos-chave, a VLTENet adota uma abordagem focada nos pontos centrais das vértebras, combinando informações de alta resolução com mecanismos de atenção de canal para a extração de características e a geração de mapas dos centros vertebrais e de suas inclinações. O ângulo de Cobb é então estimado por meio do traçado de retas paralelas às inclinações vertebrais que passam pelos pontos centrais. Durante o treinamento e a validação, foi utilizado o conjunto de dados do desafio AASCE2019 (WU et al., 2017), enquanto os testes foram realizados em uma base privada, obtendo-se um erro médio absoluto de 2,15 em relação às medições manuais realizadas por especialistas, valor inferior ao obtido pela VFLDNet, que apresentou erro médio absoluto de 2,80 na mesma métrica.

No trabalho desenvolvido por Huang et al. (2024), foram utilizadas radiografias de adolescentes com escoliose idiopática provenientes de uma base de dados construída pelos próprios autores. O método inicia com a identificação da maior área conectada por meio de operações morfológicas na imagem de raio-x, seguida do recorte da região corporal e do cálculo do maior retângulo externo da máscara vertebral, cujo resultado é utilizado como entrada para a Rede de Alta Resolução (do inglês, *High Resolution Network*) (HRNET) (WANG et al., 2020), com *backbone* ResNet101 (HE et al., 2016a),

responsável por localizar os pontos terminais de cada vértebra. No módulo final, são detectados quatro vértices de contorno das vértebras torácicas e lombares, sendo o ângulo de Cobb estimado a partir de um oscilograma<sup>1</sup> que relaciona a sequência vertebral às inclinações das placas terminais. O método foi avaliado nas curvas *Torácica Proximal* (do inglês, *Proximal Thoracic*) (PT), *Torácica Principal* (do inglês, *Main Thoracic*) (MT) e *Toracolombar/Lombar* (do inglês, *Thoracolumbar/Lumbar*) (TL/L), apresentando erros absolutos médios de  $3,04 \pm 1,94$ ,  $2,72 \pm 1,74$  e  $2,53 \pm 1,60$ , respectivamente.

A Tabela 1 sumariza os trabalhos que abordam a estimação do ângulo de Cobb sem a necessidade de segmentação.

Tabela 1 – Estimativa do Ângulo de Cobb sem Segmentação

<b>Autores</b>	<b>Abordagem</b>	<b>Base de Imagens</b>	<b>Qtd. Imagens</b>
(KIM et al., 2020)	Centroid-Net e M-Net	Base privada e AASCE2019	Treino: 431, Validação: 50, Teste: 148
(YI et al., 2020)	ResNet34 com localização de pontos terminais	AASCE2019	Treino: 348, validação: 116, Teste: 116
(SURI et al., 2023)	SpineTK	Base privada e Base Fraiwan	Treino: 509, Validação: 180, Teste: 621
(ZHOU; ZHU; YAO, 2023)	MobileNetV3, U-Net e CenterNet	Base privada	Treino: 323, Validação: 57, Teste: 40
(MAEDA et al., 2023)	ResNet34 com Heatmaps e cálculo do deslocamento	Base privada	Treino: 915, Validação: 155, Teste: 106
(YANG; WANG; MENG, 2024)	LaNet	AASCE2019 e Base privada	Treino: 707, Validação: 128, Teste: 98
(LU et al., 2024)	VLTENet	AASCE2019 e Base privada	–
(HUANG et al., 2024)	HRNet com backbone da ResNet101	Base privada	Imagens de 500 pacientes

Fonte: AUTORAL

## 2.2 Estimação do ângulo de Cobb com uso de segmentação

No que se refere aos métodos baseados em segmentação, isto é, aqueles que realizam segmentação em pelo menos uma de suas etapas, no trabalho desenvolvido por Tu et al.

<sup>1</sup> gráfico que representa a variação de uma grandeza física ao longo do tempo.

(2019), os autores propuseram a rede de detecção e segmentação DU-Net, destinada à segmentação dos contornos da coluna vertebral. A detecção da região da coluna foi realizada por meio do recurso de [Canal de Características Agregadas \(do inglês, \*Aggregated Channel Features\*\) \(ACF\)](#) (DOLLÁR et al., 2014), em conjunto com a arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015), utilizada na etapa de segmentação. O cálculo do ângulo de Cobb foi efetuado a partir da identificação dos pontos centrais do contorno da coluna vertebral, seguida do ajuste da curva correspondente, sendo a mensuração realizada com base na inclinação dessa curva. O conjunto de dados empregado foi obtido junto ao Departamento de Radiologia do Primeiro Hospital Afiliado da Universidade Médica de Anhui, bem como a partir do conjunto público de radiografias de tórax do NIH (WANG et al., 2017). Como resultados, o método alcançou um MAE de  $2,9^\circ$ , precisão de 86,30%, coeficiente Dice de 90,28% e índice IoU de 82,29%.

Para a estimação do ângulo de Cobb, no trabalho desenvolvido por Maguire (2020), foi realizada a segmentação das vértebras com base na arquitetura U-Net (RONNEBERGER; FISCHER; BROX, 2015), empregando a função de perda *Tversky loss* (SALEHI; ERDOGMUS; GHOLIPOUR, 2017), bem como a técnica de *dropout* e aumento de dados aplicado à base de treinamento do AASCE2019 (WU et al., 2017). Para a mensuração do ângulo de Cobb, o método de estimação recebe a máscara resultante da segmentação das vértebras e realiza uma etapa de pós-processamento para a remoção de ruídos. A imagem resultante é então utilizada no cálculo do ângulo de Cobb. Como resultado, o método alcançou um coeficiente Dice de 74,90% na segmentação com a U-Net (RONNEBERGER; FISCHER; BROX, 2015), o qual foi elevado para 84,69% após o pós-processamento, além de um MAE de 4,87, 5,22 e 6,55 para os ângulos estimados nas regiões PT, MT e TL/L, respectivamente.

O trabalho de Cui et al. (2021) teve como objetivo desenvolver um método para a detecção dos quatro cantos de cada vértebra e o posterior cálculo do ângulo de Cobb. Para isso, foi empregada a rede U-Net (RONNEBERGER; FISCHER; BROX, 2015) na etapa de segmentação vertebral, cuja saída binária serviu como entrada para o processo de identificação dos quatro cantos de cada vértebra. As coordenadas dos retângulos ajustados às vértebras segmentadas foram então utilizadas para a estimação do ângulo de Cobb. Como resultados, o método alcançou um coeficiente Dice de 78% na etapa de segmentação e um valor de 21,675% para a métrica SMAPE na avaliação da precisão do cálculo do ângulo de Cobb.

Liang et al. (2022) propuseram um método fundamentado na predição das direções intermediárias de flexão vertebral. A abordagem integra uma [Rede de Localização \(do inglês, \*Localization Network\*\) \(LocNet\)](#) a uma [Rede de Segmentação \(do inglês, \*Segmentation Network\*\) \(SegNet\)](#) da coluna vertebral, cujo resultado é posteriormente utilizado por uma [Rede de Regressão \(do inglês, \*Regression Network\*\) \(RegNet\)](#) responsável por

estimar as direções de flexão das vértebras localizadas. Os resultados reportados indicaram um coeficiente Dice de 90% para a [SegNet](#), enquanto a [LocNet](#) obteve precisão de 97,15%, sensibilidade de 96,69% e Dice de 96,91%. Para a mensuração do ângulo de Cobb, foram alcançados valores de [CMAE](#) de 2,92°, [SMAPE](#) de 6,87%, [MD](#) de 8,76°, [ED](#) de 5,95° e [CD](#) de 5,01°.

De forma semelhante, [Zhao et al. \(2022\)](#) apresentaram a U-shaped network empregada para a segmentação das vértebras. Os experimentos foram conduzidos utilizando uma base privada, composta por 185 imagens (110 para treinamento e 75 para teste), além do conjunto de teste do AASCE2019 ([WU et al., 2017](#)), com 98 imagens. Para o cálculo do ângulo de Cobb, é determinado o menor retângulo que envolve o conjunto de pontos bidimensionais de entrada, a partir do qual são extraídas as coordenadas dos vértices e o ângulo de deflexão em relação ao plano horizontal; em seguida, são identificadas as vértebras terminais superior e inferior. O método obteve coeficiente Dice de 81,46% na etapa de segmentação e, quanto à estimação do ângulo de Cobb, apresentou [CMAE](#) de 2,45° e [SMAPE](#) de 6,63%.

O trabalho de [Peng et al. \(2024\)](#) propôs a rede ATT U-Net 3+, baseada na arquitetura U-Net 3+ ([HUANG et al., 2020](#)), para a segmentação de vértebras e posterior estimação do ângulo de Cobb. O conjunto experimental foi composto por 155 imagens de raio-x da coluna vertebral em incidência [ântero-posterior \(AP\)](#) da porção superior do corpo. Inicialmente, a região da coluna foi extraída a partir da imagem completa por meio de uma [ROI](#) definida com base no histograma de intensidades em escala de cinza. Para a segmentação vertebral, a ATT U-Net 3+ incorpora um módulo de atenção à U-Net 3+ ([HUANG et al., 2020](#)), no qual as *skip connections* de cada nível do *decoder* são combinadas em um mecanismo de atenção baseado no módulo [Convolutional Block Attention Module \(CBAM\)](#) ([WOO et al., 2018](#)), sendo o resultado encaminhado à respectiva camada de decodificação. Na etapa subsequente, os pontos do contorno das vértebras são determinados e uma elipse é ajustada iterativamente para aproximar os cantos vertebrais, permitindo o cálculo das inclinações. O ângulo de Cobb é então obtido pela diferença entre as inclinações máxima e mínima. Em termos de desempenho, a ATT U-Net 3+ alcançou valores de  $95,11 \pm 0,716$ ,  $96,11 \pm 0,620$ ,  $92,37 \pm 0,761$ ,  $89,51 \pm 0,661$  e  $86,21 \pm 0,571$  para acurácia, precisão, sensibilidade, Dice e [Média da Intersecção sobre a União \(do inglês, Mean Intersection Over Union\) \(mIOU\)](#), respectivamente, além de apresentar coeficiente de correlação de Pearson de 0,983 e [Erro Absoluto Médio \(do inglês, Mean Absolute Error\) \(MAE\)](#) de 1,6° na estimação do ângulo de Cobb.

Por fim, [Liu et al. \(2026\)](#) propôs uma [estrutura de Estimação do Ângulo de Cobb Aprimorada por Convolução Wavelet Multiescala \(do inglês, Multi-scale Wavelet Convolution-Enhanced Cobb angle Estimation framework\) \(MWCE-CE\)](#), utilizando a rede WA-SEUNet para a segmentação das vértebras. A abordagem incorpora um módulo de

extração de características residuais baseado na Transformada *Wavelet* (GONZALEZ; WOODS, 2010), com o objetivo de capturar informações de textura vertebral, além de um módulo de atenção de canal duplo no *decoder*, responsável por integrar características locais e globais. Os experimentos foram conduzidos utilizando o conjunto de dados do desafio AASCE2019 (WU et al., 2017), alcançando coeficiente Dice de 0,9432 e precisão de 0,9833 na etapa de segmentação. A qualidade da estimativa do ângulo de Cobb foi avaliada por meio do *CMAE*, obtendo valor de 2,24°.

A Tabela 2 sumariza os trabalhos que abordam a estimativa do ângulo de Cobb com uso de segmentação.

Tabela 2 – Estimativa do Ângulo de Cobb com Segmentação

<b>Autores</b>	<b>Abordagem</b>	<b>Base de Imagens</b>	<b>Qtd. Imagens</b>
(TU et al., 2019)	DU-Net	Base privada	Treino: 460, Validação: 20, Teste: 81
(MAGUIRE, 2020)	Unet com refinamento da segmentação	AASCE2019	Treino: 431, Validação: 50, Teste: 128
(CUI et al., 2021)	U-Net	AASCE2019	609 imagens
(LIANG et al., 2022)	LocNet, SegNet e RegNet	AASCE2019	Treino: 481, Validação: 128, Teste: 98
(ZHAO et al., 2022)	U-shaped network	Base privada e Imagens de teste do AASCE2019	283 imagens
(PENG et al., 2024)	ATT U-Net 3+	Base privada	155 imagens
(LIU et al., 2026)	MWCE-CE e WA-SEUNet	AASCE2019	609 imagens
<b>Método proposto</b>	Método em cascata com U-Net 3+ modificada	AASCE2019	Treino: 431, Validação: 50, Teste: 128

Fonte: AUTORAL

Neste trabalho, apresenta-se um método automático para o cálculo do ângulo de Cobb baseado em técnicas de segmentação de imagens. A adoção dessa abordagem justifica-se pelo fato de que estudos que utilizam segmentação (Tu et al. (2019), Maguire (2020), Cui et al. (2021), Liang et al. (2022), Zhao et al. (2022), Peng et al. (2024), Liu et al. (2026)) têm alcançado resultados tão promissores quanto aqueles que não a empregam (Huang et al. (2024), Lu et al. (2024), Yang, Wang e Meng (2024), Maeda et al. (2023), Zhou, Zhu e Yao (2023), Suri et al. (2023), Yi et al. (2020), Kim et al. (2020)).

O método proposto é composto por duas etapas de segmentação. A primeira tem como objetivo segmentar a coluna vertebral para a extração da ROI, enquanto a segunda realiza a segmentação individual das vértebras, cuja saída é utilizada como entrada no processo de estimação do ângulo de Cobb. Diferentemente das abordagens propostas por Maeda et al. (2023) e Peng et al. (2024), nas quais a extração da ROI é realizada a partir de pontos de referência da localização da coluna na radiografia utilizando, respectivamente, a ResNet34 e histogramas de intensidade em escala de cinza, o método proposto baseia-se na segmentação direta da coluna vertebral. Dessa forma, obtém-se com maior precisão a região correspondente exclusivamente à coluna, ao passo que as abordagens mencionadas delimitam essa região por meio de um retângulo, o que pode incluir estruturas que não pertencem à coluna vertebral.

Para delimitar a região da coluna vertebral, de modo geral, emprega-se inicialmente a etapa de localização das vértebras, seguida de sua segmentação, conforme apresentado por Tu et al. (2019) e Liang et al. (2022). Diferentemente dessas abordagens, o presente trabalho propõe uma estratégia baseada em dupla segmentação, a qual permite delimitar tanto a coluna vertebral como as vértebras. Por fim, no que se refere ao cálculo do ângulo de Cobb, foi adotado o método de estimação proposto por Maguire (2020), semelhante ao utilizado por Cui et al. (2021). Esses métodos apresentam menor complexidade computacional quando comparados a abordagens baseadas em redes de regressão específicas para essa finalidade, como as propostas por Liang et al. (2022) e Kim et al. (2020).

## 3 Fundamentação Teórica

Este capítulo apresenta os principais fundamentos teóricos que sustentam o desenvolvimento deste trabalho. Inicialmente, descreve-se a estrutura da coluna vertebral, seguida da definição e caracterização da escoliose, patologia abordada neste estudo, bem como do ângulo de Cobb, método empregado para a mensuração do grau de escoliose. Na sequência, são discutidos os conceitos relacionados ao aprendizado profundo, com ênfase em redes neurais, redes neurais convolucionais e suas aplicações na segmentação de imagens médicas. Por fim, são apresentados os principais componentes desses modelos, incluindo funções de perda, funções de ativação e métricas de avaliação, elementos essenciais para a compreensão da metodologia adotada e dos resultados obtidos.

### 3.1 Coluna vertebral

A coluna vertebral corresponde a cerca de dois quintos da altura total do corpo, sendo composta por uma série de ossos chamados vértebras. Essa estrutura funciona como uma haste forte e flexível com elementos que podem se mover para frente, para trás e para os lados, além de girar. Ela envolve e protege a medula espinhal, sustenta a cabeça e serve como ponto de fixação para as costelas, a cintura pélvica e os músculos das costas e dos membros superiores (TORTORA; NIELSEN, 2017).

O número total de vértebras durante o desenvolvimento inicial é 33. Em seguida, várias vértebras nas regiões sacral e coccígea se fundem. Como resultado, a coluna vertebral adulta normalmente contém 26 vértebras. Sendo 7 vértebras cervicais (na região do pescoço), 12 torácicas (envoltas pelas 12 costelas), 5 vértebras lombares (sustentam a parte inferior das costas), sacro (constituído por cinco vértebras sacrais fundidas) e coccígeo (constituído por quatro vértebras coccígeas fundidas) (TORTORA; NIELSEN, 2017).

A coluna vertebral quando vista de frente, não apresenta nenhuma curvatura. Vista de perfil ou lateralmente apresenta curvaturas normais, chamadas de fisiológicas, são elas a lordose e a cifose. Essas curvaturas estão presentes em diferentes partes da coluna, apresentando-se como lordose cervical, cifose dorsal ou torácica, lordose lombar e cifose sacral. Essas curvas ajudam a manter o equilíbrio e a postura ereta. Tanto a perda quanto um grau exagerado de lordose ou cifose podem prejudicar significativamente o equilíbrio de um indivíduo (NEUWIRTH; OSBORN, 2001).

A Figura 1 ilustra a coluna vertebral nas vistas frontal e lateral. Em ambas as representações, observa-se a divisão da coluna em seus diferentes segmentos anatômicos, que compreendem as regiões cervical, torácica, lombar, sacral e coccígea, identificadas

por cores distintas. Na vista frontal, a coluna vertebral apresenta alinhamento retilíneo, sem a presença de curvaturas. Em contraste, na vista lateral, evidenciam-se as curvaturas fisiológicas da coluna: a lordose cervical, formada pelas vértebras cervicais (em azul); a cifose torácica, composta pelas vértebras torácicas (em roxo); a lordose lombar, constituída pelas vértebras lombares (em verde); e, por fim, a cifose sacral, representada em amarelo.

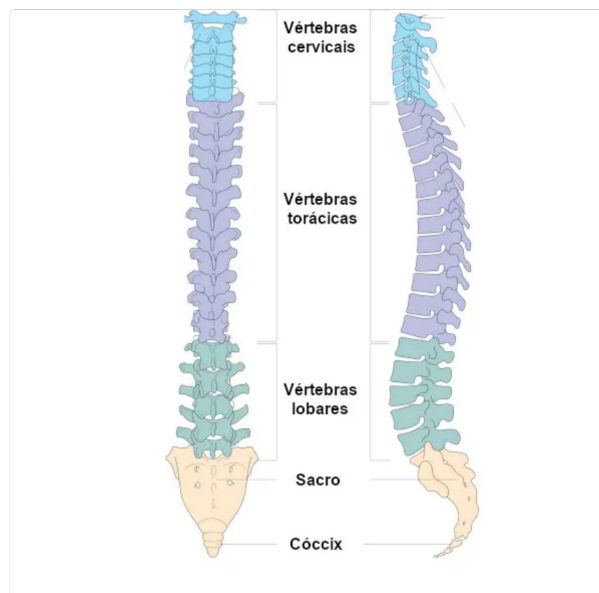


Figura 1 – Coluna vertebral.

Fonte: Adaptada de (SANTOS, 2026) .

## 3.2 Escoliose

Quando a coluna vista de frente apresenta alguma curvatura, geralmente há uma patologia ou desvio anatômico. Dentre as possíveis anomalias está a escoliose.

A escoliose é uma deformidade espinhal tridimensional que pode ocorrer em qualquer idade. Pode ser idiopática <sup>1</sup> ou secundária em crianças, idiopática e degenerativa em adultos (LACROIX et al., 2023). A medida que a coluna se curva, desviando-se para o lado, ela também tende a girar, com isso as vértebras giram em seu próprio eixo. Essa rotação das vértebras gera a característica física mais perceptível da escoliose, a gibosidade <sup>2</sup> costal, sendo essa a característica mais perceptível. Além dela, outros sinais físicos como uma escápula <sup>3</sup> mais alta que a outra e cintura assimétrica, também estão associados a essa deformidade. Dessa forma, essa deformação é clinicamente definida como a presença de um desvio lateral de 10 graus ou mais na coluna, frequentemente associado à rotação das vértebras (NEUWIRTH; OSBORN, 2001).

<sup>1</sup> possui origem desconhecida.

<sup>2</sup> protuberância ou “corcunda” nas costas.

<sup>3</sup> osso par, plano e triangular localizado na parte posterior superior do tórax.

A Figura 2 apresenta a coluna vertebral de um paciente diagnosticado com escoliose, na qual é possível observar características típicas dessa deformidade. Nota-se, ao longo do eixo anatômico da coluna vertebral, a presença de uma curvatura lateral, onde normalmente se esperaria um alinhamento retilíneo. Adicionalmente, observam-se o desnivelamento dos ombros e da bacia, bem como a assimetria das escápulas e da cintura, os quais constituem sinais característicos da escoliose.

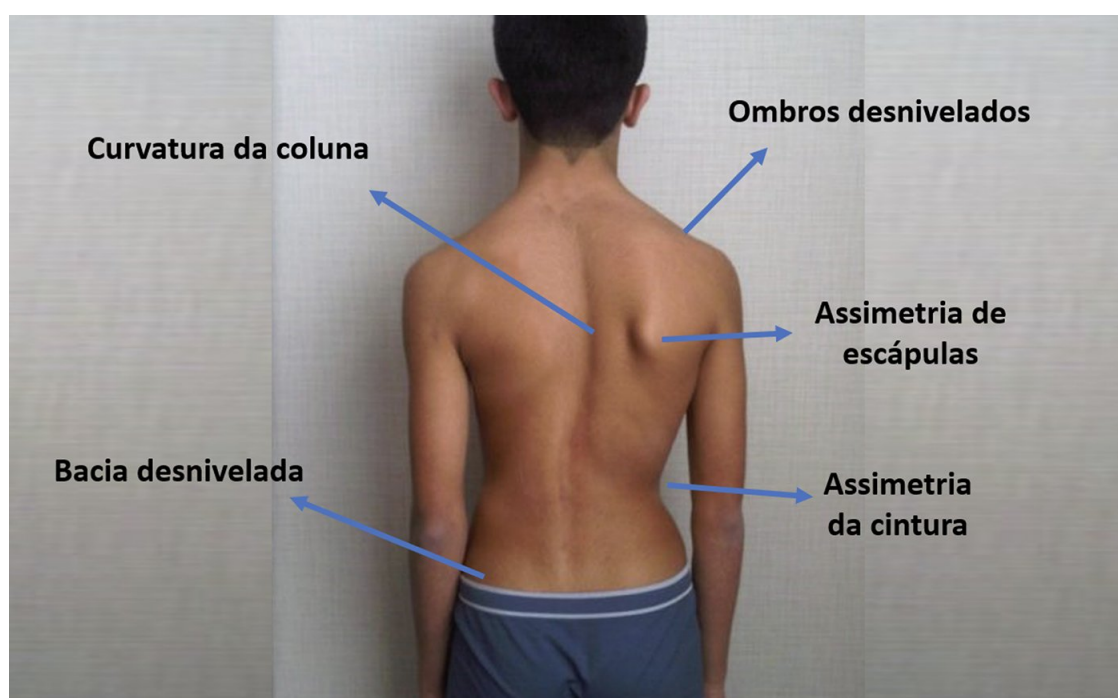


Figura 2 – Paciente com escoliose.

Fonte: Adaptada de (ORTOPEDIA, 2026).

A escoliose pode ser adquirida de diversas formas, como postura, hormônios, lesões e genética. Ela pode ser assintomática durante meses ou anos, até que se torne grave o suficiente para ser detectada. Na maioria dos casos, quando é diagnosticada, a causa já não é aparente. Portanto, grande parte dos casos de escoliose é classificada, por padrão, como sendo de "origem desconhecida" ou "idiopática" (HAWES; O'BRIEN, 2006).

Em aproximadamente 80% dos casos de escoliose, a causa não é conhecida, sendo classificados como escoliose idiopática (American Association of Neurological Surgeons, 2024). Os 20% restantes correspondem a casos com causa identificável, incluindo escoliose congênita, neuromuscular e degenerativa (NEUWIRTH; OSBORN, 2001).

A escoliose congênita, presente no nascimento, resulta de defeitos congênitos que criam anormalidades na forma como a coluna vertebral é formada, resultando no crescimento assimétrico da coluna, produzindo assim a escoliose. Ela é frequentemente associada a outros problemas, como disfunções renais, anormalidades do trato urinário ou defeitos cardíacos congênitos (NEUWIRTH; OSBORN, 2001).

A escoliose neuromuscular é causada por uma ou mais doenças neuromusculares resultantes do funcionamento inadequado dos nervos ou músculos (ou ambos) ao redor da coluna vertebral. Dentre as doenças que podem causar a escoliose neuromuscular estão a poliomielite, a paralisia cerebral e a distrofia muscular (NEUWIRTH; OSBORN, 2001).

O último tipo, a escoliose degenerativa, se forma durante a vida adulta em uma coluna vertebral previamente reta como resultado da degeneração dos discos, artrite e/ou perda de suporte na coluna vertebral. A escoliose degenerativa está mais comumente associada à dor e ao desconforto do que a escoliose idiopática (NEUWIRTH; OSBORN, 2001).

A escoliose pode gerar diferentes tipos de curvas na coluna vertebral. São elas:

- Curvas torácicas: Ocorrem na coluna torácica, 95% delas têm convexidade à direita.
- Curvas toracolombares: Abrangem a coluna torácica e lombar, com o ápice localizado na junção entre a coluna torácica e lombar, 85% dessas curvas demonstram convexidade à esquerda.
- Curvas lombares: Estão situadas inteiramente dentro das cinco vértebras da coluna lombar, a maioria (70%) das curvas lombares exibe convexidade à esquerda.
- Curvaturas duplas principais: Desviam-se da coluna reta em forma de S. Duas curvaturas se formam, uma com o ápice à direita do alinhamento natural da coluna e a outra com o ápice à esquerda. Mais tipicamente, a curvatura da direita ocorre na coluna torácica e a da esquerda na região lombar.

Esses quatro tipos provavelmente representam 95% de todas as curvas escolióticas (NEUWIRTH; OSBORN, 2001).

Os quatro tipos mencionados são apresentados na Figura 3. Inicialmente, observa-se a coluna vertebral em sua formação normal, sem deformidades. Em seguida, são ilustradas as principais variações da escoliose: a curvatura lombar, formada pelas vértebras lombares; a curvatura torácica, composta pelas vértebras torácicas; a curvatura tóraco-lombar, envolvendo vértebras das regiões torácica e lombar; e, por fim, a escoliose de dupla curva, caracterizada por curvaturas torácica e lombar distintas.

Os pacientes com escoliose compartilham geralmente de sintomas comuns, como dor nas costas, e à medida que essa deformação se agrava, podem apresentar transtornos respiratórios e cardíacos e/ou deformidades (MINGHELLI, 2008). Para tratar cada grau dessa deformação, são usados diferentes tipos de tratamento, sendo eles o uso de colete ortopédico, gesso, fisioterapia e, em casos graves, a correção cirúrgica.

Para acelerar o processo de detecção da escoliose, foram criados programas de triagem em escolas. Visto que o diagnóstico precoce pode remediar a progressão dessa

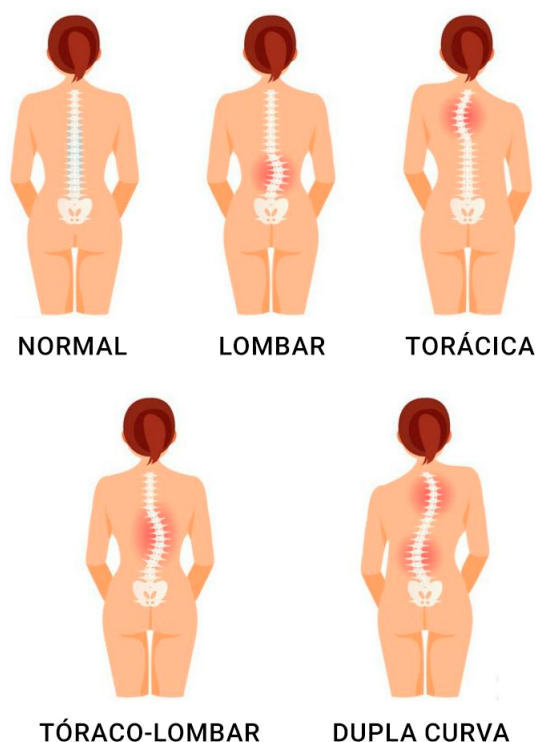


Figura 3 – Tipos de curvas geradas pela escoliose na coluna vertebral.

Fonte: Adaptada de (YUNES, 2026).

deformação (HAWES; O'BRIEN, 2006). O método mais usado para mensurar a gibosidade (curvatura) escoliótica nas triagens escolares é o teste de Adams. No entanto, esse teste apresenta limitações por fornecer informações subjetivas que podem diferir entre dois avaliadores (MINGHELLI, 2008). Dessa forma, para confirmar o diagnóstico de escoliose, é solicitada a radiografia pósterio-anterior da coluna vertebral, que permite determinar a localização da curvatura e o grau de deformidade. Além do exame de raio-x, também são usados a tomografia computadorizada e a ressonância magnética, sendo estes últimos meios utilizados principalmente nas alterações congênicas, nas associadas a quadros algícos e nas que apresentam alterações neurológicas (MINGHELLI, 2008).

### 3.2.1 Ângulo de Cobb

Geralmente, os especialistas solicitam um exame de raio-x da coluna vertebral do paciente para mensurar o grau de desvio da coluna causado pela escoliose. Para isso, utiliza-se como métrica o ângulo de Cobb.

Para medir o ângulo de Cobb, linhas são traçadas em uma radiografia da coluna vertebral, paralelas às vértebras mais alta e mais baixa envolvidas na curvatura escoliótica. Em uma coluna reta, ambas as linhas seriam horizontais. Em seguida, uma linha é traçada perpendicularmente a cada uma dessas linhas. O ângulo de interseção dessas duas

perpendiculares corresponde à magnitude da curvatura. Em uma coluna reta, essas perpendiculares seriam linhas verticais paralelas e o ângulo de Cobb seria zero grau (NEUWIRTH; OSBORN, 2001).

A Figura 4 apresenta, à esquerda, a marcação das retas paralelas traçadas nas vértebras selecionadas pelo especialista para cada curvatura identificada, bem como as retas perpendiculares a essas paralelas, cujo cruzamento define o ângulo de Cobb. À direita, observa-se a marcação e o cálculo do ângulo de Cobb em uma imagem de radiografia da coluna vertebral. Nessa abordagem, a métrica é obtida a partir do ângulo formado pelas retas concorrentes derivadas das retas traçadas nas vértebras escolhidas pelo especialista, constituindo uma forma alternativa de mensuração do mesmo ângulo. Esse método é geralmente empregado quando o ponto de interseção das retas concorrentes é visível na imagem radiográfica. Em ambiente clínico, o ângulo de Cobb é comumente medido com o auxílio de um goniômetro, utilizado tanto para o traçado das retas quanto para a mensuração do ângulo. O goniômetro está ilustrado na Figura 5.

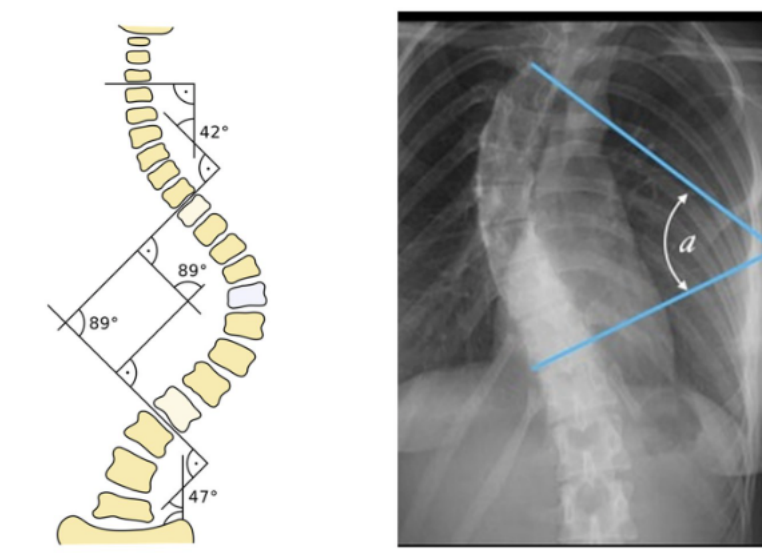
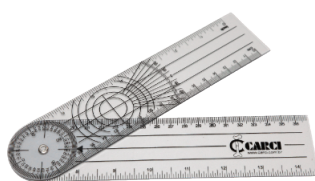


Figura 4 – Ângulo de Cobb.

Fonte: Adaptada de (YUNES, 2026).

Figura 5 – Goniômetro



Fonte: Adaptada de (CARCI, 2026).

O nível de gravidade da escoliose é determinado com base no valor do ângulo de Cobb. Entretanto, não há consenso absoluto entre os profissionais quanto aos limites exatos

de cada categoria, sendo observadas divergências médias de até 5° em relação aos valores de referência, quando comparamos os valores de classificação usados neste trabalho em relação aos usados por [Koumbourlis \(2006\)](#), [Baron e Vaccaro \(2016\)](#), [Vialle \(2017\)](#) e [Association et al. \(2003\)](#). A Tabela 3 apresenta a classificação adotada neste trabalho, baseada nos intervalos estabelecidos por [\(HORNG et al., 2019\)](#), os quais relacionam o grau de deformidade, quantificado pelo ângulo de Cobb, ao respectivo diagnóstico clínico. De acordo com essa classificação, ângulos inferiores a 10° não caracterizam escoliose; valores entre 10° e 20° indicam deformidade leve; entre 20° e 40° caracterizam deformidade moderada; e ângulos superiores a 40° estão associados a casos graves da patologia.

Tabela 3 – Definição do ângulo de Cobb

Grau do ângulo de Cobb	Diagnóstico
0-10°	Não possui escoliose
10-20°	Escoliose leve
20-40°	Escoliose moderada
>40°	Escoliose grave

Fonte: Adaptada de [\(HORNG et al., 2019\)](#)

### 3.3 Aprendizado Profundo

O aprendizado profundo constitui uma classe de métodos pertencentes ao campo do aprendizado de máquina, caracterizando-se pelo desenvolvimento de modelos de redes neurais profundas capazes de realizar inferências precisas a partir de grandes volumes de dados. Nos métodos tradicionais de aprendizado de máquina, a seleção e a extração de atributos (características) a serem fornecidos ao modelo geralmente demandam significativo esforço humano e conhecimento especializado. Em contraste, o aprendizado profundo busca aprender automaticamente, a partir dos dados brutos, representações e atributos relevantes para a execução da tarefa proposta. Dessa forma, o aprendizado profundo é particularmente adequado para contextos onde os dados são complexos e onde existem grandes conjuntos de dados disponíveis. No setor de saúde, o aprendizado profundo é usado para processar imagens médicas (raios-x, tomografias computadorizadas e ressonâncias magnéticas) e diagnosticar condições de saúde ([KELLEHER, 2019](#)). Logo, esta seção apresenta os principais conceitos teóricos necessários à compreensão das técnicas e dos métodos empregados no desenvolvimento deste trabalho.

#### 3.3.1 Redes neurais artificiais

Uma [Redes Neurais Artificiais \(RNA\)](#) é um sistema projetado para modelar a maneira como o cérebro realiza uma tarefa particular ([FLECK et al., 2016](#)). Uma rede

neural usa a estratégia de dividir para conquistar para aprender uma função. Para isso, ela é formada por neurônios que tem como objetivo aprender uma função simples, e a função mais complexa definida pela rede é criada pela combinação das funções simples (KELLEHER, 2019). É por meio dessa técnica que é escolhida qual função mais se adequa ao problema tratado.

Uma rede neural artificial é estruturada em camadas. A primeira delas, denominada camada de entrada, é responsável por receber os dados brutos. Esses dados são então processados pelos neurônios das camadas intermediárias, ou camadas ocultas. Cada neurônio em uma camada oculta recebe as saídas da camada anterior, ponderadas por coeficientes denominados pesos, além de um termo adicional chamado viés. Esses dois elementos fazem parte dos parâmetros da rede. Os pesos representam a importância relativa atribuída a cada característica de entrada, enquanto o viés atua no ajuste do limiar de ativação, possibilitando a ativação do neurônio mesmo quando os sinais de entrada são reduzidos. Por fim, a camada de saída aplica uma função de ativação ao resultado proveniente da última camada oculta, produzindo a predição final do modelo.

O processo de aprendizado da rede neural ocorre por meio da propagação direta (*forward propagation*) dos dados da camada de entrada até a camada de saída, na qual é obtida a predição do modelo. Em seguida, o erro entre a predição e o valor esperado é calculado e propagado de volta pelas camadas da rede por meio do algoritmo de retropropagação do erro (*backpropagation*), permitindo a atualização iterativa dos parâmetros do modelo, como pesos e vieses. Esse procedimento possibilita que a rede aprenda representações progressivamente mais adequadas à tarefa considerada.

A Figura 6 ilustra a estrutura de uma rede neural artificial composta por uma camada de entrada, uma camada oculta e uma camada de saída, responsável pela predição do modelo. Na representação, as conexões entre os neurônios são indicadas por setas em cinza, correspondendo aos pesos que interligam as unidades da camada de entrada aos neurônios da camada oculta e, subsequentemente, estes ao neurônio da camada de saída.

### 3.3.2 Neurônio artificial

Um neurônio é uma unidade básica de processamento em uma rede neural. É um nó que processa todos os sinais de entrada (sinapses) de outros nós e gera uma saída de acordo com uma função de transferência chamada função de ativação (DU; SWAMY, 2013). O conceito de neurônio artificial foi criado por Warren McCulloch e Walter Pitts (MCCULLOCH; PITTS, 1943) em 1943, e posteriormente usado para o desenvolvimento do perceptron por Frank Rosenblatt (ROSENBLATT, 1958).

A Figura 7 apresenta o modelo matemático do neurônio desenvolvido por McCulloch-Pitts. Onde  $(x_1, x_2, \dots, x_{j_i})$  são as entradas,  $(w_1, w_2, \dots, w_{j_i})$  os pesos,  $\phi(\cdot)$  a função de ativação

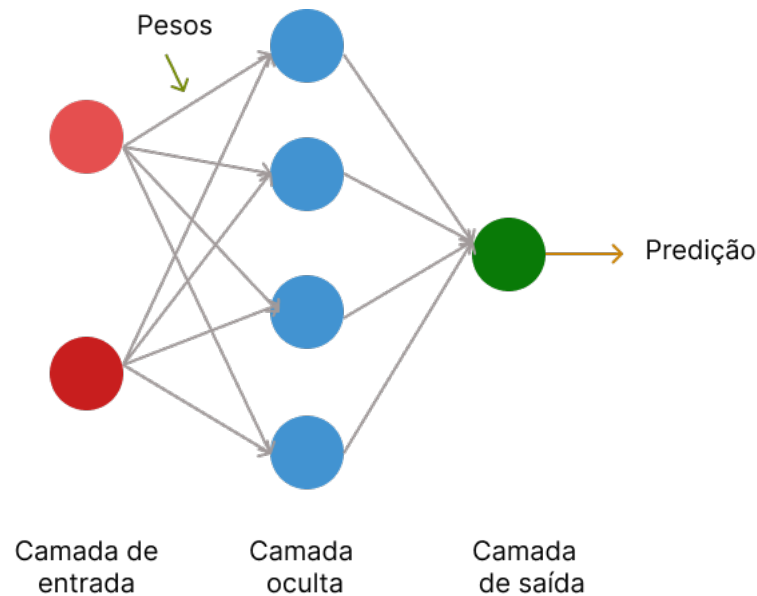


Figura 6 – Arquitetura de uma rede neural com uma camada oculta

Fonte: AUTORAL

aplicada ao neurônio,  $\theta$  o limiar ou viés e  $y$  o valor de saída do nó.

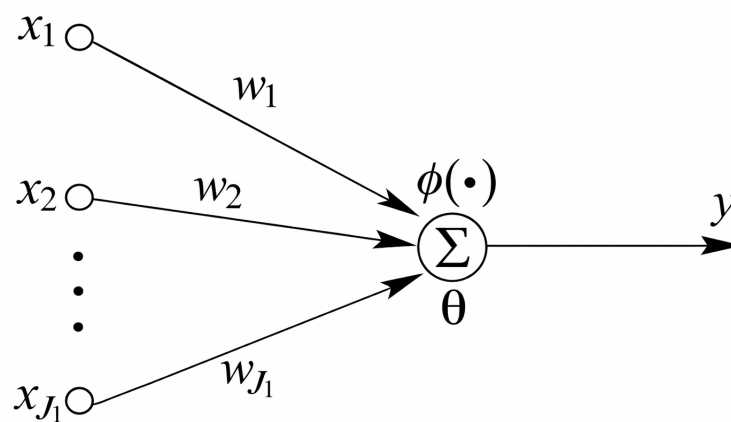


Figura 7 – Modelo matemático do neurônio de McCulloch-Pitts

Fonte: Adaptada de (DU; SWAMY, 2013)

O cálculo realizado pelo neurônio consiste na soma ponderada das entradas pelos respectivos pesos, acrescida do termo de viés, conforme apresentado na Equação 3.1. Sobre o resultado dessa operação, é aplicada a função de ativação, como descrito na Equação 3.2. Nessas equações,  $x_i$  representa as entradas do modelo, enquanto  $w_i$  corresponde aos pesos associados a cada uma dessas entradas, e  $b$  refere-se ao viés adicionado à combinação linear. A variável  $z$  denota a soma ponderada resultante, caracterizando a transformação linear. Em seguida, aplica-se a função de ativação  $\sigma$ , responsável por introduzir não linearidade ao

modelo. Por fim,  $a$  representa a saída do neurônio após a aplicação da função de ativação.

$$z = \sigma\left(\sum_{i=1}^N x_i w_i + b\right) \quad (3.1)$$

$$a = \sigma(z) \quad (3.2)$$

### 3.3.3 Função de ativação

A função de ativação realiza um mapeamento linear ou não linear entre a entrada e a saída do neurônio. Sua principal finalidade é restringir a amplitude do sinal de saída, normalizando o valor resultante do somatório para um intervalo limitado, o que permite, em alguns casos, interpretá-lo como uma probabilidade. Com base no valor produzido pela função de ativação, determina-se se o neurônio será considerado ativo ou inativo dentro da rede neural.

Existem diversos tipos de funções de ativação, entre as quais destacam-se a função linear, a tangente hiperbólica, a *sigmoid* e a **Unidade Linear Retificadora** (do inglês, *Rectified Linear Unit*) (ReLU), entre outras. Neste trabalho, foram empregadas as funções ReLU e *sigmoid*: a ReLU foi utilizada nas camadas internas do modelo, enquanto a *sigmoid* foi aplicada na camada de saída, com a finalidade de realizar a predição final. A função ReLU já é originalmente adotada na arquitetura utilizada e, por esse motivo, foi mantida. Por sua vez, a função *sigmoid* foi escolhida em razão do problema envolver apenas duas classes (fundo e coluna), sendo mais adequada para classificação binária.

A função ReLU retorna o próprio valor da entrada quando este é positivo e zero quando assume valores negativos. A Equação 3.3 apresenta a formulação matemática que descreve esse comportamento. Alternativamente, a função ReLU pode ser interpretada como uma operação de limiarização, na qual os valores positivos são preservados integralmente, enquanto os valores negativos são mapeados para zero, conforme expresso na Equação

$$f(x) = \max(0, x) \quad (3.3)$$

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (3.4)$$

A função de ativação *sigmoid* é comumente empregada nos neurônios de saída de redes neurais multicamadas. Ela converte valores reais de entrada em saídas no intervalo entre 0 e 1, podendo ser interpretadas como probabilidades. Ao introduzir não linearidade ao modelo, a função *sigmoid* possibilita o aprendizado de padrões mais complexos e a definição de fronteiras de decisão mais elaboradas. Quando aplicada ao neurônio de saída,

sua resposta é expressa pela Equação 3.5, em que  $x$  representa a entrada da função e  $e$  corresponde à base do logaritmo natural.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.5)$$

### 3.3.4 Função de perda

O processo de aprendizado de uma rede neural ocorre por meio da atualização dos pesos durante a etapa de *backpropagation*, conforme descrito na Seção 3.3.1. Para isso, é necessário calcular a diferença entre a saída predita pela rede em cada ciclo de treinamento e a resposta correta, denominada *ground truth*. Nesse contexto, a função de perda, ou função de erro, é empregada para quantificar o desvio entre a predição e o *ground truth*. Quanto maior o valor da perda, menor é a assertividade do modelo. Assim, o objetivo do treinamento consiste em minimizar a função de perda.

Neste trabalho, foi adotada a função de perda *Hybrid loss* (HUANG et al., 2020), definida como a soma de três diferentes funções de perda: *Focal Loss* (LIN et al., 2017) ( $l_{fl}$ ), *MS-SSIM loss* (HUANG et al., 2020) ( $l_{ms-ssim}$ ) e *IOU loss* (MÁTTYUS; LUO; URTASUN, 2017) ( $l_{iou}$ ). Essa combinação é apresentada na Equação 3.6, na qual a Hybrid Loss é representada por ( $l_{seg}$ ).

$$l_{seg} = l_{fl} + l_{ms-ssim} + l_{iou} \quad (3.6)$$

Conforme apresentado em (LIN et al., 2017), a Focal Loss foi desenvolvida para lidar com cenários de detecção de objetos em estágio único, nos quais existe um forte desbalanceamento entre as classes de primeiro plano e de fundo. Essa função de perda é baseada na Cross-Entropy Loss e tem como objetivo priorizar os exemplos que o modelo classifica incorretamente, em detrimento daqueles previstos com alta confiança. Para isso, os autores propõem a inclusão de um fator modulador  $(1-pt)^\gamma$  na *Cross-Entropy Loss*, em que o parâmetro de focalização ajustável  $\gamma \geq 0$  controla a taxa de atenuação dos exemplos considerados fáceis. Adicionalmente, a função incorpora um parâmetro de ponderação responsável por ajustar a importância relativa de cada amostra. A formulação matemática da Focal Loss é apresentada na Equação 3.7.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3.7)$$

Huang et al. (2020) propôs o uso da função de perda baseada no Índice de similaridade estrutural multiescala (do inglês, *Multi-scale structural similarity index*) (MS-SSIM) (WANG; SIMONCELLI; BOVIK, 2003) a qual atribui maior peso às regiões com contornos difusos. Essa função é apresentada na Equação 3.8, na qual  $M$  representa o

número total de escalas;  $\mu_p, \mu_g$  bem como  $\sigma_p, \sigma_g$  correspondem, respectivamente, às médias e aos desvios padrão de  $p$  e  $g$ ;  $\sigma_{pg}$  denota a covariância entre essas variáveis.  $\beta_m$  e  $\gamma_m$  determinam a importância relativa dos dois componentes em cada escala. Além disso, duas constantes positivas,  $C_1 = 0,01^2$  e  $C_2 = 0,03^2$  são adicionadas para evitar divisão por zero.

$$\ell_{ms-sim} = 1 - \prod_{m=1}^M \left( \frac{2\mu_p\mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1} \right)^{\beta_m} \left( \frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2} \right)^{\gamma_m} \quad (3.8)$$

A última função que compõe o conjunto da *Hybrid loss* é a *IOU loss*, baseada na métrica [Intersecção sobre União](#) (do inglês, *Intersection over Union*) (IOU), a qual será detalhada na Seção 3.3.12. O objetivo da IoU Loss é maximizar a sobreposição entre a predição do modelo e o *ground truth*. Essa função é apresentada na Equação 3.9, em que  $p_{ic}$  representa o valor predito no pixel  $i$  para a classe  $c$ , e  $p_{ic}^*$  corresponde à distribuição real, modelada como uma função delta em  $y_i^*$ , o rótulo correto ([MÁTTYUS; LUO; URTASUN, 2017](#)).

$$\ell_{\text{soft-IoU}} = \frac{1}{|C|} \sum_c \frac{\sum_i p_{ic} p_{ic}^*}{\sum_i p_{ic} + p_{ic}^* - p_{ic} p_{ic}^*} \quad (3.9)$$

### 3.3.5 Rede Neural Convolutacional

As [Redes Neurais Convolucionais](#) (do inglês, *Convolutional Neural Network*) (CNNs) constituem uma classe específica de redes neurais artificiais. Inspiradas em mecanismos biológicos, essas redes foram desenvolvidas para reproduzir padrões de conectividade observados no córtex visual humano. Em comparação com métodos tradicionais de classificação de imagens, que dependem de etapas extensas de pré-processamento e da definição manual de filtros, as [CNNs](#) demandam significativamente menos preparação dos dados. Devido a essas características, as [CNNs](#) são amplamente empregadas em aplicações como reconhecimento de imagens e vídeos, sistemas de recomendação, classificação de imagens, análise de imagens médicas e [Processamento de linguagem natural \(PLN\)](#) ([PEDRYCZ; CHEN, 2020](#)).

As [CNNs](#) diferenciam-se das redes neurais artificiais convencionais por realizarem operações de convolução, em vez de simples multiplicações matriciais, em pelo menos uma de suas camadas ([PEDRYCZ; CHEN, 2020](#)). Essa característica permite que as [CNNs](#) explorem de forma mais eficiente a estrutura espacial dos dados, sendo especialmente adequadas para tarefas envolvendo imagens.

As [CNNs](#) são compostas, de forma geral, por quatro componentes principais: camadas convolucionais, aplicação de funções de ativação não lineares, camadas de *pooling* (ou agrupamento) e camadas totalmente conectadas. Essa arquitetura pode ser observada na

Figura 8. A camada convolucional é responsável pela aplicação da operação de convolução, na qual um filtro, também chamado de *kernel*, é aplicado sobre os dados de entrada. Esse *kernel* possui dimensões pré-definidas e seus pesos são ajustados ao longo do processo de treinamento. Cada aplicação do filtro resulta em uma matriz denominada mapa de características (*feature map*), que representa os padrões extraídos da entrada.

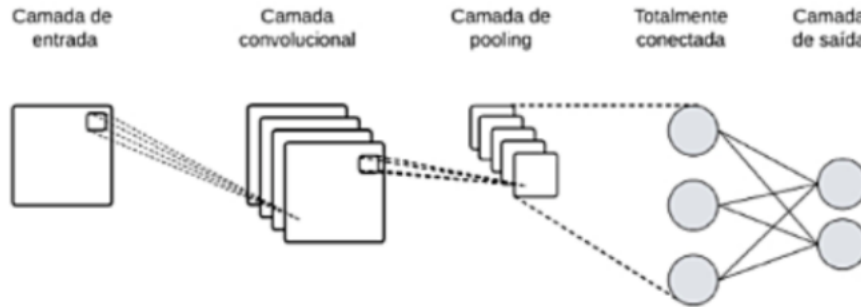


Figura 8 – Rede Neural Convolucional.

Fonte: Adaptada de (CATARINO, 2025).

No contexto de imagens, o *kernel* pode ser entendido como uma pequena matriz que se desloca sobre a imagem, calculando, em cada posição, a soma do produto escalar entre seus valores e os valores correspondentes da imagem. Esse processo é ilustrado na Equação 3.10, em que  $I$  representa a imagem de entrada,  $K$  corresponde ao *kernel* de convolução,  $S(i, j)$  representa o valor da imagem resultante na posição  $(i, j)$ ,  $i$  e  $j$  indicam as coordenadas espaciais da saída e  $m$  e  $n$  são os índices utilizados no cálculo da soma. A quantidade de pixels que o filtro avança a cada deslocamento é denominada *stride*. Além disso, antes da convolução, pode-se aplicar o *padding*, que consiste em adicionar pixels nas bordas da imagem. Um exemplo comum é o *zero padding*, no qual as bordas são preenchidas com o valor zero. O uso de *padding* evita problemas relacionados às bordas da imagem e reduz a perda excessiva de informação espacial quando múltiplas camadas convolucionais são utilizadas. A Figura 9 exemplifica esse processo, comparando uma matriz de imagem sem *padding* com outra submetida ao *zero padding*.

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n) \quad (3.10)$$

Durante o treinamento, os *kernels* são ajustados para capturar características relevantes da imagem, como bordas, contrastes e padrões locais, que são altamente informativos para a identificação de objetos (PEDRYCZ; CHEN, 2020). Esse mecanismo de extração automática de atributos é um dos principais fatores que tornam as CNNs eficazes em tarefas de reconhecimento e classificação de imagens. A Figura 10 ilustra o processo de convolução de uma imagem de entrada utilizando um *kernel* de dimensão  $2 \times 2$ , evidenciando a multiplicação elemento a elemento entre o filtro e a região correspondente

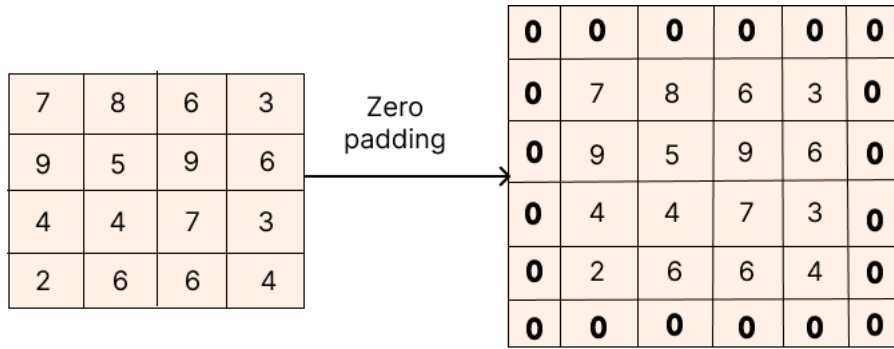


Figura 9 – Aplicação de zero padding.

Fonte: AUTORAL.

da imagem, seguida da soma dos resultados para compor cada posição do mapa de características.

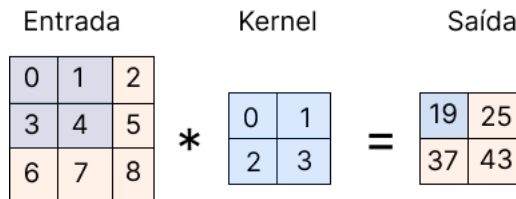


Figura 10 – Operação de convolução.

Fonte: AUTORAL.

Após cada camada convolucional, aplica-se uma função de ativação não linear, como a **ReLU**, descrita na Seção 3.3.3. Em seguida, é comum a utilização de uma camada de *pooling*, cujo objetivo é reduzir a dimensionalidade dos mapas de características, diminuindo o custo computacional e tornando a rede mais robusta a pequenas variações espaciais. O *pooling* atua de forma semelhante a um filtro deslizante, porém sem pesos treináveis. Entre as operações mais utilizadas está o *Max Pooling*, que seleciona o maior valor dentro da região coberta pelo filtro. Esse procedimento é apresentado na Figura 11, na qual o maior valor de cada região é preservado para formar o mapa resultante do agrupamento.

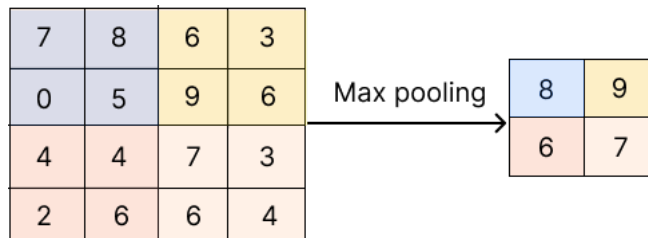


Figura 11 – Operação de Max pooling.

Fonte: AUTORAL.

Por fim, o mapa de características obtido após a última camada de *pooling* é convertido em um vetor unidimensional, que serve como entrada para a camada totalmente

conectada, operação ilustrada na Figura 12. Essa camada é composta por neurônios interligados por pesos e vieses e tem a função de combinar as características extraídas nas etapas anteriores, estabelecendo a relação entre essas informações e a saída final do modelo, sendo posicionada imediatamente antes da camada de saída.

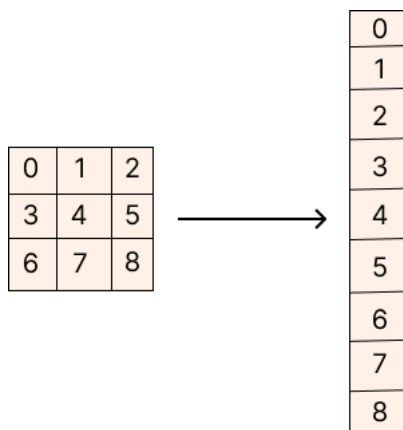


Figura 12 – Vetorização da última camada de pooling da rede neural convolucional.

Fonte: AUTORAL.

Além da operação de convolução mencionada anteriormente, destaca-se também a convolução transposta, amplamente utilizada em redes neurais convolucionais com a finalidade de aumentar a resolução espacial dos mapas de características. Essa operação permite reconstruir representações com maiores dimensões espaciais a partir de mapas reduzidos, sendo particularmente empregada em arquiteturas que realizam processos de decodificação ou reconstrução de imagens.

### 3.3.6 U-Net

A U-Net é uma arquitetura de rede neural convolucional proposta por (RONNEBERGER; FISCHER; BROX, 2015), amplamente empregada em tarefas de segmentação de imagens. Essa arquitetura representa uma evolução em relação às **Redes Totalmente Conectadas** (do inglês, *Fully Connected Network*) (FCN) apresentadas por (LONG; SHENHAMER; DARRELL, 2015). Seu nome deriva do formato em “U” de sua estrutura, ilustrada na Figura 13, a qual é composta por dois caminhos principais: um caminho de contração, ou *encoder*, e um caminho de expansão, ou *decoder*.

O caminho de contração segue a configuração típica de redes convolucionais, sendo formado pela aplicação repetida de duas convoluções  $3 \times 3$ , cada uma seguida pela função de ativação **ReLU**, e por uma operação de *max pooling*  $2 \times 2$  com *stride* igual a 2. Na região de ligação entre o *encoder* e o *decoder*, são aplicadas duas convoluções  $3 \times 3$ , ambas seguidas da ativação **ReLU**. Já no caminho de expansão, utiliza-se uma convolução transposta  $2 \times 2$  para aumento da resolução espacial, seguida por duas convoluções  $3 \times 3$ , também acompanhadas da função de ativação **ReLU**. Na camada final do *decoder*, uma convolução

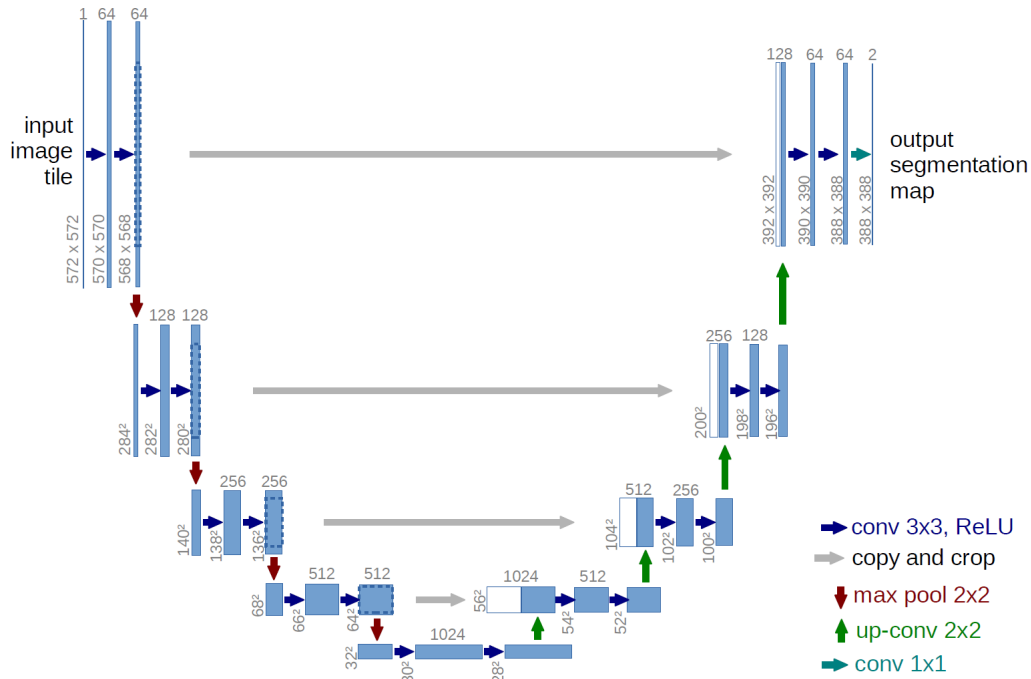


Figura 13 – Arquitetura U-Net.

Fonte: Adaptada de (RONNEBERGER; FISCHER; BROX, 2015).

$1 \times 1$  é empregada para mapear cada vetor de características para o número desejado de classes.

No total, a U-Net é composta por 23 camadas convolucionais. Além disso, a arquitetura incorpora *skip connections* entre camadas correspondentes do *encoder* e do *decoder*, permitindo a transferência direta dos mapas de características. Essas conexões têm como objetivo fornecer informações espaciais adicionais ao *decoder*, contribuindo para a geração de representações semânticas mais precisas.

### 3.3.7 U-Net 3+

A U-Net 3+ (HUANG et al., 2020) é uma rede neural convolucional baseada na U-net (RONNEBERGER; FISCHER; BROX, 2015). Logo, ela possui um caminho de contração e expansão, que serve para extrair características da imagem e reconstruí-la respectivamente. O diferencial dessa rede encontra-se no redesenho das *skip connections* que, diferente da U-net (RONNEBERGER; FISCHER; BROX, 2015), não vincula somente os mapas de características de mesmo nível, mas une tanto do mesmo nível quanto de níveis anteriores. Essa abordagem foi realizada para unir tanto características locais quanto globais da imagem e assim obter mais informações de contexto, evitando a perda de informações importantes.

A Figura 14 apresenta a arquitetura da U-Net 3+ proposta por (HUANG et al., 2020), na qual se observa uma reformulação das *skip connections*. Nessa arquitetura, o

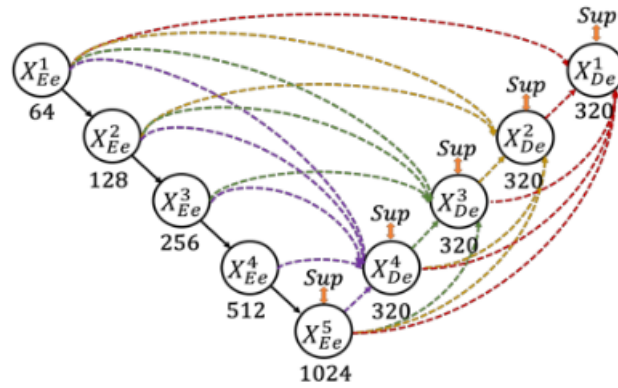


Figura 14 – Arquitetura U-Net 3+.

Fonte: Adaptada de (HUANG et al., 2020).

decodificador passa a receber informações não apenas de múltiplas camadas do *encoder*, mas também de diferentes níveis internos do próprio *decoder*. Para ilustrar essa reorganização das *skip connections*, a Figura 15 mostra o processo de construção do mapa de características agregado a partir dessas conexões. A seguir, descreve-se em detalhes o processo de construção do mapa de características agregado ilustrado na Figura 15.

Tomando como exemplo a camada  $X^3_{De}$  do *decoder*, observa-se que, semelhantemente a U-Net (RONNEBERGER; FISCHER; BROX, 2015), o mapa de características de mesma escala proveniente da camada  $X^3_{En}$  do *encoder* é diretamente incorporado ao decodificador. Entretanto, diferentemente da U-Net tradicional, o decodificador também recebe os mapas de características das camadas anteriores do *encoder*, especificamente  $X^1_{En}$  e  $X^2_{En}$ , os quais são ajustados à escala correspondente por meio de operações de *max pooling*. Adicionalmente, são integrados mapas de características oriundos de camadas mais profundas do *decoder*, como  $X^4_{De}$  e  $X^5_{De}$ , que contêm informações semânticas de alto nível e são redimensionados por interpolação bilinear.

Para combinar eficientemente as informações de baixo nível, mais detalhadas, com as representações semânticas profundas, a U-Net 3+ emprega um mecanismo de agregação de características aplicado ao mapa concatenado. Esse mecanismo é composto por 320 filtros convolucionais de tamanho  $3 \times 3$ , seguidos por uma etapa de normalização e pela aplicação da função de ativação ReLU.

Neste trabalho, foram realizadas duas etapas de segmentação da coluna vertebral. Para esse propósito, foram propostas duas modificações na arquitetura da U-Net 3+ (HUANG et al., 2020). As Seções 4.1.2 e 4.3 apresentam, respectivamente, cada uma das modificações introduzidas em relação à arquitetura original da U-Net 3+ (HUANG et al., 2020).

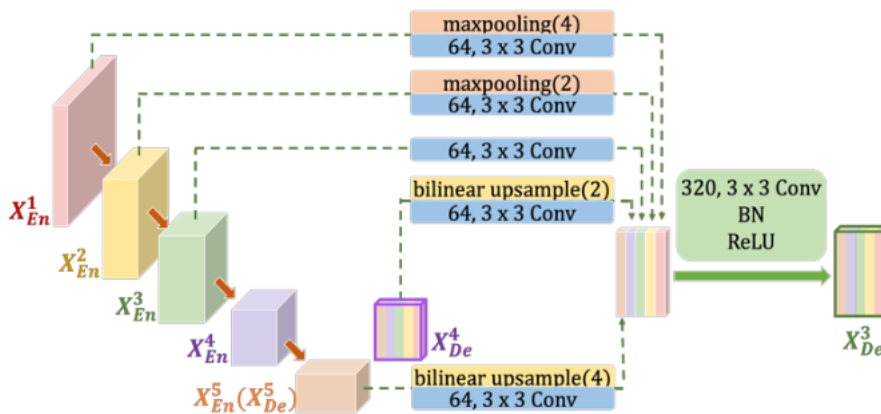


Figura 15 – Representação do mapa de características agregado das *skip connections*.

Fonte: Adaptada de (HUANG et al., 2020).

### 3.3.8 Atenção

Neste estudo, foram empregados diferentes mecanismos de atenção, os quais são detalhados nesta seção.

### 3.3.9 Attention Gate

Oktay et al. (2018) propuseram o mecanismo de **Portão de Atenção** (do inglês, *Attention Gate*) (AG), cujo objetivo é permitir que a rede neural foque automaticamente em estruturas-alvo com diferentes tamanhos e formatos. Nesse mecanismo, regiões consideradas mais relevantes recebem pesos maiores, enquanto áreas menos informativas são atenuadas por meio de pesos reduzidos. À medida que o modelo é treinado, o AG aprende a intensificar progressivamente a atenção nas regiões associadas a coeficientes mais elevados, contribuindo para uma segmentação mais precisa.

Conforme ilustrado na Figura 16, o AG pode ser descrito da seguinte forma: dois vetores são utilizados como entrada, o mapa de características  $x$ , proveniente do *encoder*, e o sinal de controle  $g$ , oriundo do *decoder*. Esses vetores são projetados para um mesmo espaço e combinados por meio de uma soma, resultando em um mapa de ativação alinhado. O vetor resultante é submetido à função de ativação **ReLU**, seguida por uma convolução  $1 \times 1$ . Em seguida, aplica-se a função *sigmoid*, que normaliza os valores para o intervalo  $[0,1]$ , gerando os coeficientes de atenção, nos quais valores mais próximos de 1 indicam maior relevância das características.

Posteriormente, os coeficientes de atenção são redimensionados para as dimensões originais do vetor  $x$  e, então, multiplicados elemento a elemento pelo mapa de características original. Esse processo escala o vetor  $x$  de acordo com a importância espacial aprendida pelo AG, e o resultado é transmitido normalmente pela conexão de salto para o *decoder*.

Por fim, destaca-se que o sinal de *gating* não corresponde a um único vetor global

aplicado a todos os pixels da imagem, mas sim a um sinal em forma de grade, condicionado as informações espaciais. Além disso, cada **AG** agrega informações provenientes de múltiplas escalas da imagem, conforme apresentado em (OKTAY et al., 2018), o que aumenta a resolução espacial do sinal de atenção e contribui para um melhor desempenho do modelo.

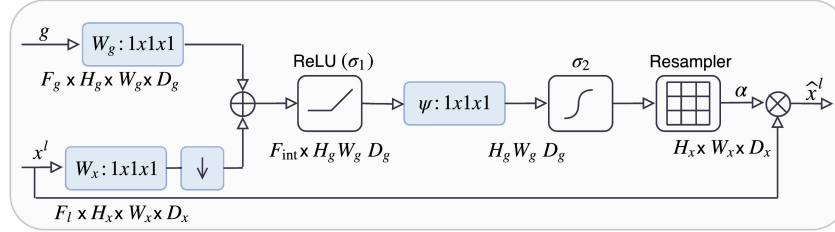


Figura 16 – Esquema do Attention Gate.

Fonte: Adaptada de (OKTAY et al., 2018).

### 3.3.10 Squeeze-and-Excitation

O Bloco *Squeeze-and-Excitation* (SE) proposto por (HU; SHEN; SUN, 2018) compreende um mecanismo de controle leve que se concentra em aprimorar o poder de representação da rede, modelando relacionamentos entre canais de maneira computacionalmente eficiente. Esse mecanismo que permite a rede realizar a recalibração de características, por meio do qual ela pode aprender a usar informações globais para enfatizar seletivamente características informativas e suprimir as menos úteis (HU; SHEN; SUN, 2018).

A estrutura do bloco SE é representada na Figura 17. Ele pode ser descrito como para qualquer transformação  $F_{tr}$  que mapeie a entrada  $X$  para os mapas de características  $U$ , onde  $U \in R^{H \times W \times C}$ . As características  $U$  são primeiramente passadas por uma operação de compressão, que produz um descritor de canal agregando mapas de características em suas dimensões espaciais ( $H \times W$ ). A função desse descritor é produzir uma incorporação da distribuição global de respostas de características por canal, permitindo que informações do campo receptivo global da rede sejam usadas por todas as suas camadas. A agregação é seguida por uma operação de excitação, que assume a forma de um mecanismo de autogateamento simples que recebe a incorporação como entrada e produz uma coleção de pesos de modulação por canal. Esses pesos são aplicados aos mapas de características  $U$  para gerar a saída do bloco SE, que pode ser alimentada diretamente em camadas subsequentes da rede (HU; SHEN; SUN, 2018).

Considerando que os blocos SE podem ser empregados em diferentes profundidades na arquitetura da rede, esse conhecimento foi explorado para adicionar um bloco SE após cada bloco convolucional do codificador na rede apresentada na Seção 4.3. A Figura 18 ilustra a configuração de cada camada do codificador da referida rede, na qual a estrutura é composta por um bloco convolucional original da U-Net 3+ (HUANG et al., 2020), seguido por um bloco SE.

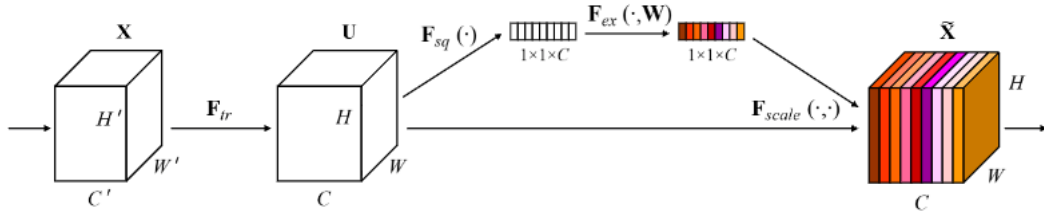


Figura 17 – Bloco Squeeze-and-Excitation.

Fonte: Adaptada de (HU; SHEN; SUN, 2018).

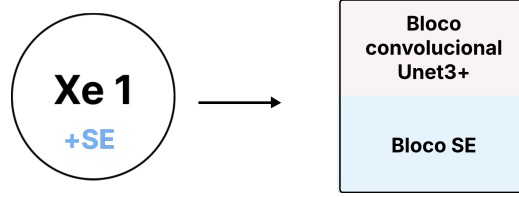


Figura 18 – Composição camada do codificador Unet 3+ com adição dos mecanismos de atenção SE e CBAM.

Fonte: Autoral.

### 3.3.11 Convolutional Block Attention Module

O mecanismo de atenção *Convolutional Block Attention Module (CBAM)*, proposto por (WOO et al., 2018), foi desenvolvido com o objetivo de enfatizar características relevantes tanto no domínio dos canais quanto no domínio espacial. Para esse fim, são aplicados de forma sequencial módulos de atenção de canal e de atenção espacial, permitindo que cada ramo da arquitetura aprenda, respectivamente, “o que” e “onde” deve concentrar sua atenção.

A Figura 19 apresenta uma visão geral do bloco *CBAM*. Na ilustração, é representado um mapa de características de entrada ( $F$ ) (*Input Feature*), a partir do qual é gerado um mapa de atenção de canal unidimensional ( $M_c$ ) por meio do módulo de atenção de canal (do inglês, *Channel Attention Module*). O mapa ( $M_c$ ) é então multiplicado ( $\otimes$ ) pelo mapa de características de entrada ( $F$ ), resultando em  $F'$ . Em seguida,  $F'$  é utilizado como entrada para o módulo de atenção espacial (do inglês, *Spatial Attention Module*), que produz um mapa de atenção espacial bidimensional ( $M_s$ ). Esse mapa é multiplicado por  $F'$ , originando  $F''$ , que corresponde a saída refinada do bloco *CBAM*. A operação descrita é formalizada na Equação 3.11.

$$\begin{aligned} \mathbf{F}' &= \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}, \\ \mathbf{F}'' &= \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}'. \end{aligned} \tag{3.11}$$

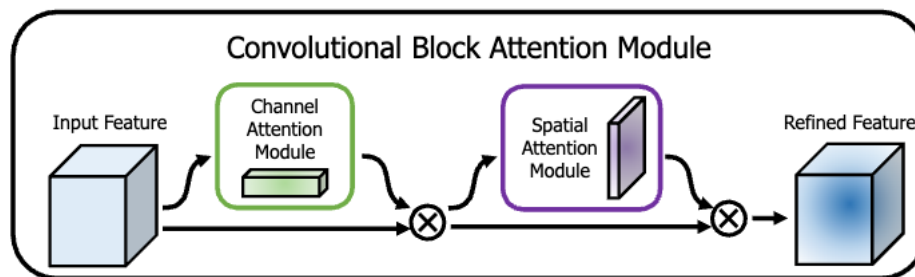


Figura 19 – Visão geral do bloco CBAM.

Fonte: Adaptada de (WOO et al., 2018).

### 3.3.12 Métricas de avaliação

Para a avaliação do desempenho do método proposto, foram consideradas métricas tanto para a qualidade da segmentação quanto para a estimativa do ângulo de Cobb. As métricas adotadas permitem uma análise quantitativa da concordância entre os resultados preditos pelo modelo e as anotações realizadas por especialistas.

O desempenho da segmentação da coluna vertebral e das vértebras foi avaliado por meio das métricas *Intersecção sobre União* (do inglês, *Intersection over Union*) (IOU), coeficiente Dice, sensibilidade e precisão. Considerando que o problema é formulado como uma segmentação binária, cada pixel da imagem é classificado como pertencente ao objeto de interesse (coluna vertebral) ou ao fundo. A partir dessa classificação, os pixels podem ser categorizados como *Verdadeiro Positivo* (VP), quando corretamente identificados como parte da coluna; *Falso Positivo* (FP), quando incorretamente classificados como coluna; *Verdadeiro Negativo* (VN), quando corretamente identificados como fundo; e *Falso Negativo* (FN), quando pixels da coluna são erroneamente classificados como fundo.

O coeficiente Dice é uma métrica que avalia a similaridade entre dois conjuntos de dados, variando no intervalo  $[0,1]$ , em que valores mais próximos de 1 indicam maior sobreposição entre eles. No contexto deste trabalho, valores elevados do Dice refletem uma alta concordância entre a máscara de segmentação predita pelo modelo e a segmentação manual realizada pelo especialista, também denominada *ground truth*. O coeficiente Dice é definido pela Equação 3.12.

$$\text{DICE} = \frac{2 \cdot VP}{(VP + FP) + (VP + FN)} \quad (3.12)$$

A sensibilidade mede a capacidade do modelo em identificar corretamente todos os pixels pertencentes a classe positiva, isto é, avalia a proporção de verdadeiros positivos em relação ao total de pixels que realmente fazem parte do objeto segmentado. Essa métrica é expressa pela Equação 3.13.

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \quad (3.13)$$

A precisão, por sua vez, indica a proporção de pixels corretamente classificados como positivos em relação ao total de pixels que o modelo previu como pertencentes à classe positiva. Essa métrica é definida pela Equação 3.14.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.14)$$

A métrica **IOU** corresponde à razão entre a área de interseção e a área de união das máscaras de segmentação predita e de referência. Assim como o coeficiente Dice, seu valor varia entre 0 e 1, sendo que valores mais elevados indicam melhor desempenho do modelo na tarefa de segmentação. A **IOU** é definida pela Equação 3.15.

$$\text{IoU} = \frac{VP}{VP + FP + FN} \quad (3.15)$$

Para avaliar a proximidade entre o ângulo de Cobb estimado pelo modelo e aquele medido pelo especialista, foram utilizadas métricas de erro amplamente empregadas na comparação entre valores preditos e valores de referência, a saber: o **Erro Absoluto Médio** (do inglês, *Mean Absolute Error*) (**MAE**) e o **Erro Percentual Absoluto Médio Simétrico** (do inglês, *Symmetric Mean Absolute Percentage Error*) (**SMAPE**).

O **MAE** representa a média das diferenças absolutas entre os valores reais e os valores preditos, fornecendo uma medida direta da magnitude média do erro. Quanto menor o valor do **MAE**, melhor é a qualidade das previsões realizadas pelo modelo. Essa métrica é expressa pela Equação 3.16, em que  $n$  representa o número total de observações,  $y_i$  o valor real da  $i$ -ésima observação e  $p_i$  o valor predito correspondente.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - p_i| \quad (3.16)$$

A métrica **Erro Percentual Absoluto Médio Simétrico** (do inglês, *Symmetric Mean Absolute Percentage Error*) (**SMAPE**) é amplamente utilizada para avaliar o desempenho de modelos de regressão, expressando o erro em termos percentuais de forma simétrica em relação aos valores reais e estimados. Essa métrica utiliza como fator de normalização a média entre o valor real e o valor predito, tornando-a menos sensível a valores próximos de zero e proporcionando uma avaliação mais equilibrada do erro. O valor do **SMAPE** varia entre 0% e 100%, sendo que valores mais próximos de 0% indicam maior precisão do modelo. O **SMAPE** é definido conforme a Equação 3.17, onde  $y_i$  representa o valor real,  $\hat{y}_i$  corresponde ao valor estimado pelo modelo e  $n$  é o número total de amostras avaliadas.

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \quad (3.17)$$

### 3.4 Considerações Finais

Neste capítulo, foram apresentados os principais conceitos teóricos que fundamentam o desenvolvimento deste trabalho. A descrição da escoliose permitiu contextualizar a patologia estudada, enquanto a abordagem sobre aprendizado profundo e redes neurais forneceu a base conceitual necessária para a compreensão das técnicas empregadas. Além disso, foram discutidos elementos essenciais dos modelos de aprendizado profundo, como redes neurais convolucionais, funções de perda, funções de ativação e métricas de avaliação, os quais desempenham papel fundamental no desempenho e na análise de sistemas de segmentação de imagens médicas. Dessa forma, este capítulo estabelece o embasamento teórico necessário para a compreensão das metodologias e experimentos apresentados nos capítulos subsequentes.

## 4 Método

Neste capítulo, é apresentado o método desenvolvido para a estimação do ângulo de Cobb, bem como o detalhamento de suas etapas.

Nesta pesquisa, foi adotada uma abordagem de segmentação em cascata seguida da estimação do ângulo de Cobb. O método é dividido em três etapas. Primeiramente, realiza-se a extração da ROI, que corresponde a primeira etapa de segmentação do método em cascata, cujo objetivo é segmentar a coluna vertebral. Em seguida, é realizada a segmentação das vértebras com base na ROI obtida na etapa anterior. Por fim, a partir do resultado da segmentação das vértebras, é feita a estimação do ângulo de Cobb utilizando o método desenvolvido por Maguire (2020), sendo estimados três ângulos para cada imagem de raio-x, o método descrito está ilustrado da Figura 20

### 4.1 Extração da ROI

A primeira etapa do método consiste na extração da *Região de Interesse* (do inglês, *Region of Interest*) (ROI), utilizada para delimitar a região da coluna vertebral, uma vez que as imagens de raio-x da coluna frequentemente incluem outras estruturas anatômicas, como costelas e parte do assoalho pélvico, além de, em alguns casos, objetos não removidos pelos pacientes no momento do exame. Dessa forma, o objetivo dessa etapa é realizar a segmentação da coluna vertebral, sem a delimitação individual das vértebras.

#### 4.1.1 Pré-processamento

Antes do processamento das imagens de raio-x pela rede selecionada, foi necessária a aplicação de um conjunto de etapas de pré-processamento. Inicialmente, realizou-se a aplicação de *padding* lateral (direita e esquerda), conforme o conceito apresentado na Seção 3.3.5, ajustado de acordo com a altura de cada imagem, uma vez que elas não apresentam dimensões uniformes. Essa etapa teve como objetivo tornar as imagens quadradas, isto é, com valores iguais de altura e largura. Em seguida, as imagens foram redimensionadas para  $320 \times 320$  pixels, de modo a atender ao tamanho de entrada requerido pela rede utilizada na segmentação da coluna vertebral. Essas etapas de pré-processamento encontram-se ilustradas na Figura 21.

#### 4.1.2 Segmentação da coluna vertebral

Para a realização da segmentação da coluna vertebral, foi empregada a arquitetura da rede U-Net 3+ (Seção 3.3.7), proposta por (HUANG et al., 2020), com a introdução de

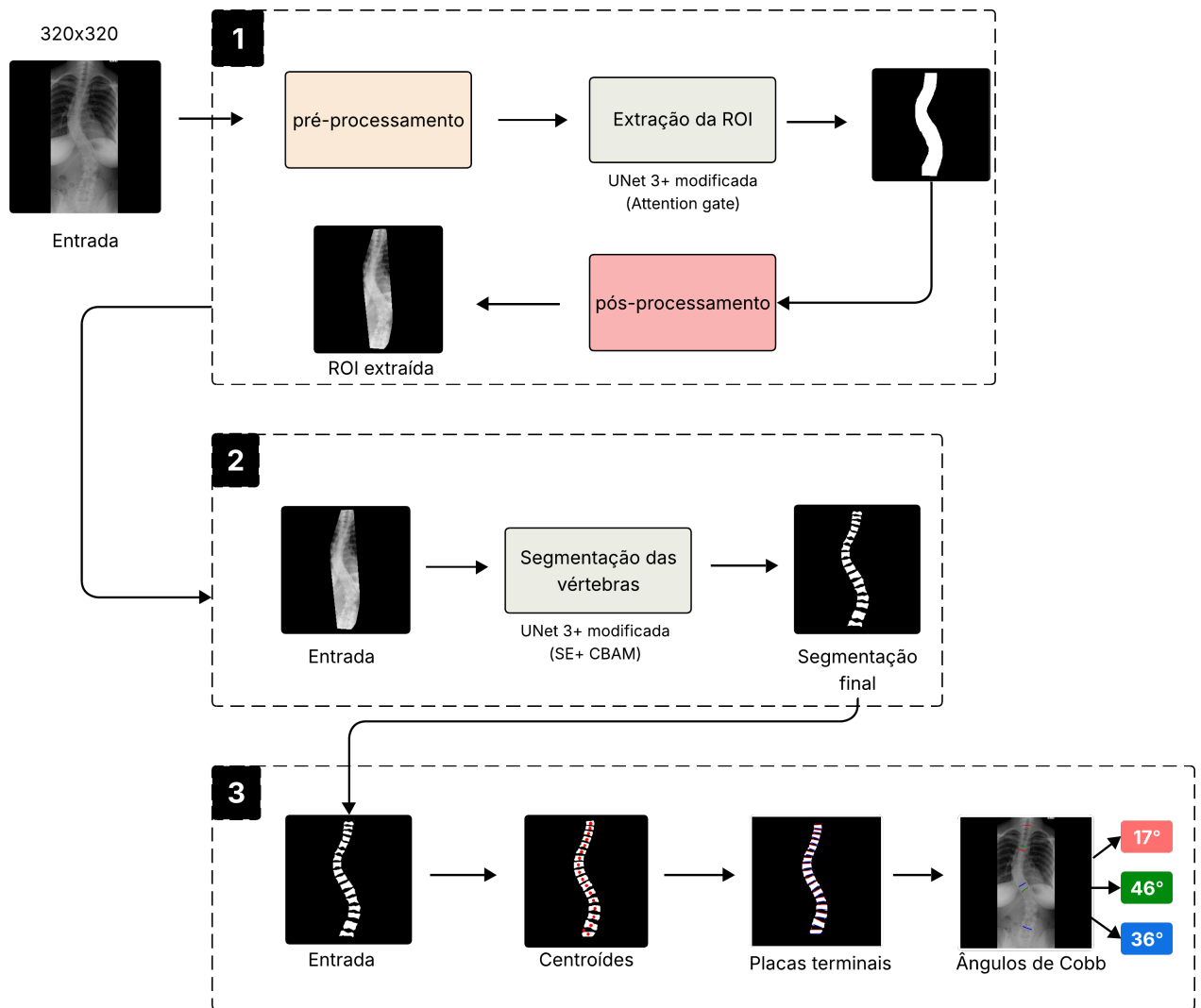


Figura 20 – Fluxograma do método proposto e suas etapas: (1) Extração da ROI, (2) Segmentação das vértebras, (3) Estimação dos ângulos de Cobb.

Fonte: AUTORAL.

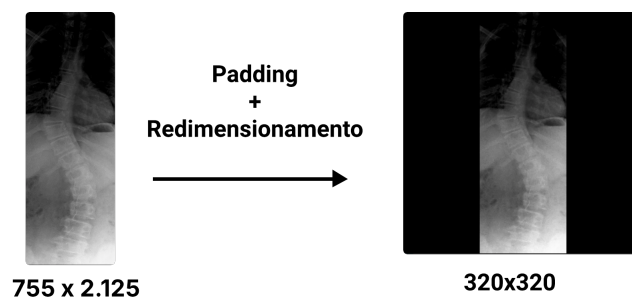


Figura 21 – Pré-processamento da imagem de entrada para extração da ROI

Fonte: AUTORAL.

modificações. Especificamente, foi adicionado um bloco de *Attention Gate* em cada *skip connection* que conecta o codificador ao decodificador. A Figura 22 ilustra a arquitetura modificada, na qual os blocos de atenção mencionados são representados por círculos na cor laranja, identificados pela sigla **AT**.

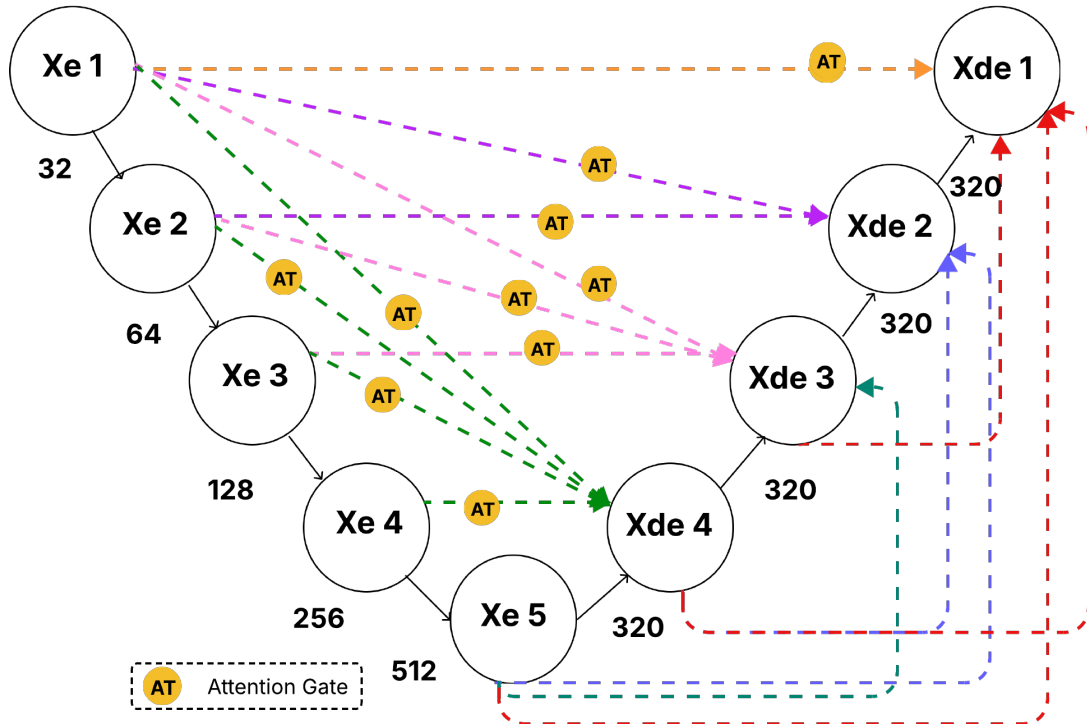


Figura 22 – U-Net 3+ modificada com adição de *Attention Gate*.

Fonte: Autoral.

Essa modificação foi usada, pois apresentou melhores resultados em relação a U-Net 3+ (HUANG et al., 2020) em sua forma original e ao desempenho de outros modelos como U-net (RONNEBERGER; FISCHER; BROX, 2015), Deeplabv3+ (CHEN et al., 2018), U-net++ (ZHOU et al., 2018), Swin-U-net (CAO et al., 2022) e  $U^2$ -Net (QIN et al., 2020).

### 4.1.3 Pós-processamento

Após a segmentação da região da coluna vertebral, a máscara resultante foi submetida a uma etapa de pós-processamento com o objetivo de refinar a segmentação obtida. Essa etapa consistiu na aplicação de operações morfológicas clássicas, especificamente erosão e dilatação, conforme descrito em (GONZALEZ, 2009).

A operação de erosão tem como finalidade a remoção de pixels localizados nas bordas dos objetos segmentados. Essa operação utiliza um elemento estruturante  $5 \times 5$  (Seção 3.3.5), cujo tamanho foi definido empiricamente, que percorre toda a imagem, preservando um pixel com valor 1 apenas quando todos os pixels cobertos pelo elemento estruturante também apresentam valor 1. No contexto deste trabalho, a erosão foi empregada para separar regiões que não pertenciam a coluna vertebral, mas que se encontravam conectadas

à região principal da máscara, bem como para remover pequenos objetos ou pontos indevidamente segmentados como coluna.

Em seguida, foi aplicada a operação de dilatação, cujo objetivo é a adição de pixels às bordas dos objetos segmentados. Nessa operação, o valor de um pixel é definido como 1 quando pelo menos um dos pixels abrangidos pelo elemento estruturante apresenta esse valor. A dilatação foi utilizada para preencher pequenos buracos e descontinuidades presentes na máscara após a etapa de erosão.

Além das operações morfológicas, foi realizada a remoção de regiões de pequena área que não pertenciam a coluna vertebral. Para isso, identificou-se a região de maior área na máscara segmentada, a qual foi considerada como correspondente a coluna, e eliminaram-se as demais regiões externas a essa área. Essa etapa mostrou-se necessária, uma vez que a operação de erosão, isoladamente, não é capaz de remover todos os objetos espúrios presentes na segmentação.

Após essa etapa, foi aplicado o algoritmo de fecho convexo, proposto por (BARBER; DOBKIN; HUHDANPAA, 1996), com o objetivo de regularizar a forma da máscara segmentada. Esse algoritmo recebe como entrada o conjunto de pontos correspondente ao contorno da máscara e determina o menor conjunto convexo que os envolve, resultando em um polígono convexo. A aplicação desse método contribuiu para a obtenção da máscara final, a qual foi utilizada nas etapas subsequentes do processamento.

A Figura 23 ilustra as etapas de pós-processamento realizadas até a identificação da maior região da máscara, correspondente a coluna vertebral. Nessa figura, observa-se inicialmente a máscara gerada pela etapa de segmentação (Figura 23a), que é utilizada como entrada para o pós-processamento. Em seguida, são aplicadas as operações morfológicas de erosão e dilatação. Sobre o resultado dessas operações, procede-se à identificação da região de maior área, mantendo-a como representação da coluna vertebral e eliminando-se as demais regiões de menor área.

A imagem final apresentada nessa figura evidencia que essas etapas de pós-processamento foram eficazes na remoção de falsos positivos, isto é, regiões indevidamente segmentadas como parte da coluna vertebral. No entanto, em alguns casos, mesmo após a aplicação dessas operações, ainda foram observados falsos positivos conectados à região principal da máscara.

Diante desse cenário, optou-se pela aplicação do algoritmo de fecho convexo, com o objetivo de regularizar e unificar o formato da máscara, evitando deformações indesejadas. A Figura 24 apresenta um exemplo em que, mesmo após as etapas iniciais de pós-processamento, ainda permanecia uma região agrupada à área principal (Figura 24a). Após a aplicação do fecho convexo (Figura 24b), obteve-se a máscara final, a qual foi utilizada nas etapas subsequentes do método.

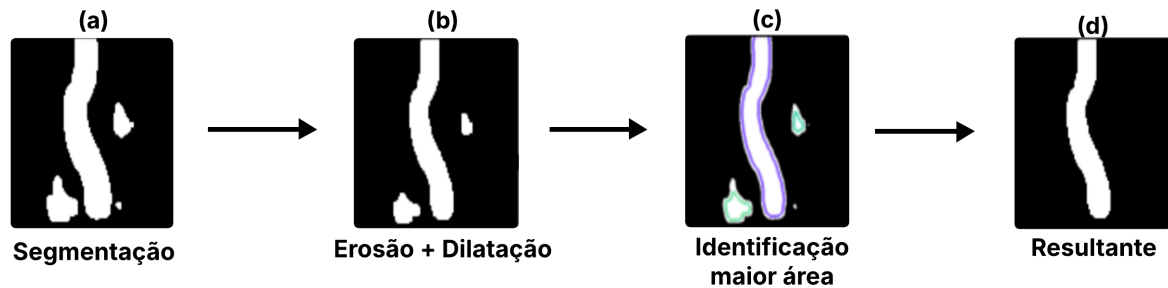


Figura 23 – Pós-processamento: (a) Saída da rede de segmentação, (b) Aplicação de erosão e dilatação, (c) Identificação da estrutura da coluna e (d) Imagem intermediária do pós-processamento

Fonte: Autoral.

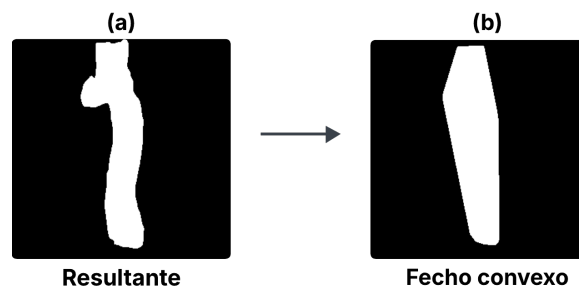


Figura 24 – Aplicação do fecho convexo após o processamento da máscara

Fonte: Autoral.

## 4.2 Recorte da imagem de raio-x com base na ROI

Após a segmentação da coluna vertebral, realizou-se o recorte da imagem de raio-x utilizada como entrada nessa etapa, com o objetivo de preservar exclusivamente a região correspondente a coluna. A Figura 25 apresenta a imagem de raio-x empregada como entrada no processo de segmentação (Figura 25a), a máscara resultante do pós-processamento (Figura 25b), correspondente a ROI, e o resultado final do recorte (Figura 25c).

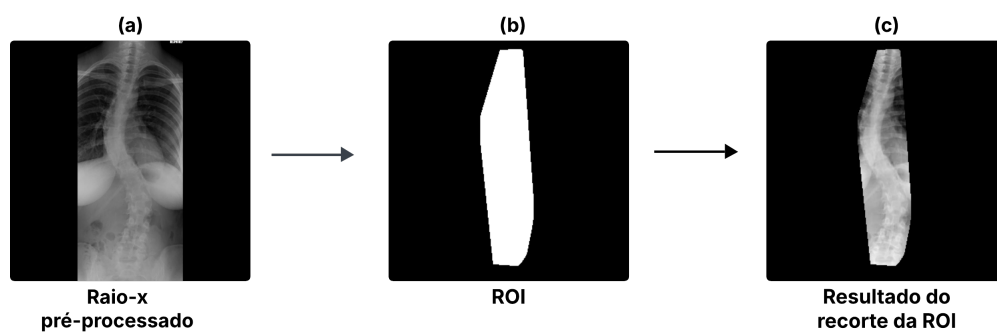


Figura 25 – Recorte região da coluna vertebral com base na ROI

Fonte: AUTORAL.

### 4.3 Segmentação das vértebras

Para essa etapa do método, foi utilizada como entrada da rede de segmentação das vértebras a imagem resultante do recorte da ROI (Figura 26c). A Figura 26 apresenta a imagem de entrada da etapa de segmentação (Figura 26a), seguida do respectivo resultado da segmentação vertebral (Figura 26b) e a máscara construída com a marcação do especialista (Figura 26c). Ressalta-se que não foi aplicado pós-processamento após a etapa de segmentação, uma vez que os métodos avaliados incluindo a aplicação de operações morfológicas, como erosão e dilatação, e o algoritmo *Watershed* (BARNES; LEHMAN; MULLA, 2014) não apresentaram melhorias significativas em comparação aos resultados obtidos sem a utilização dessas técnicas.

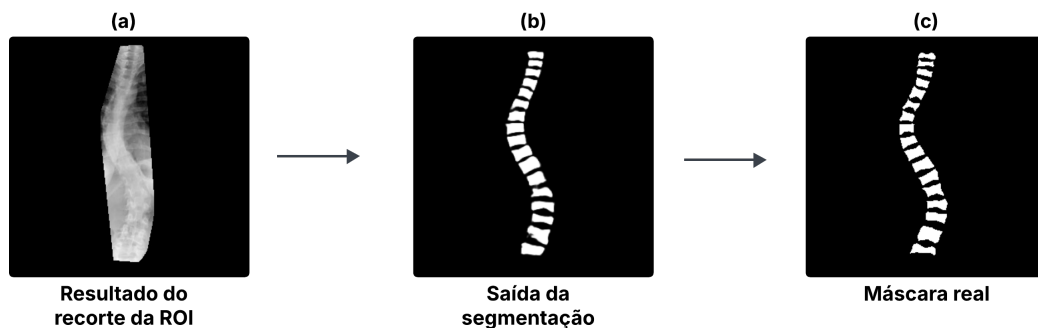


Figura 26 – Segmentação das vértebras

Fonte: Autoral.

Para a realização da segmentação das vértebras, foi empregada a arquitetura da rede U-Net 3+ (Seção 3.3.7), proposta por (HUANG et al., 2020), com a introdução de modificações. Foram incorporados blocos de atenção do tipo *Squeeze-and-Excitation* (SE) (Seção 3.3.10) e *Convolutional Block Attention Module* (CBAM) (Seção 3.3.11). O bloco SE foi aplicado imediatamente após cada bloco convolucional no codificador da U-Net 3+ (HUANG et al., 2020), enquanto o bloco CBAM foi inserido nas *skip connections*, de forma análoga ao bloco de *Attention Gate* apresentado na Seção 4.1.2. A Figura 27 ilustra a arquitetura modificada, na qual os blocos de atenção CBAM são representados por círculos na cor amarela, identificados pela sigla C, enquanto os blocos SE são indicados pelo símbolo de adição associado às camadas do codificador. A composição detalhada das camadas do codificador é apresentada na Figura 18, na Seção 3.3.10.

Essa segunda modificação da U-Net 3+ foi adotada por ter apresentado desempenho superior em comparação com a U-Net 3+ (HUANG et al., 2020) em sua forma original e ao desempenho de outros modelos como Unet (RONNEBERGER; FISCHER; BROX, 2015), Deeplabv3+ (CHEN et al., 2018), Unet++ (ZHOU et al., 2018), Swin-Unet (CAO et al., 2022) e  $U^2$ -Net (QIN et al., 2020).

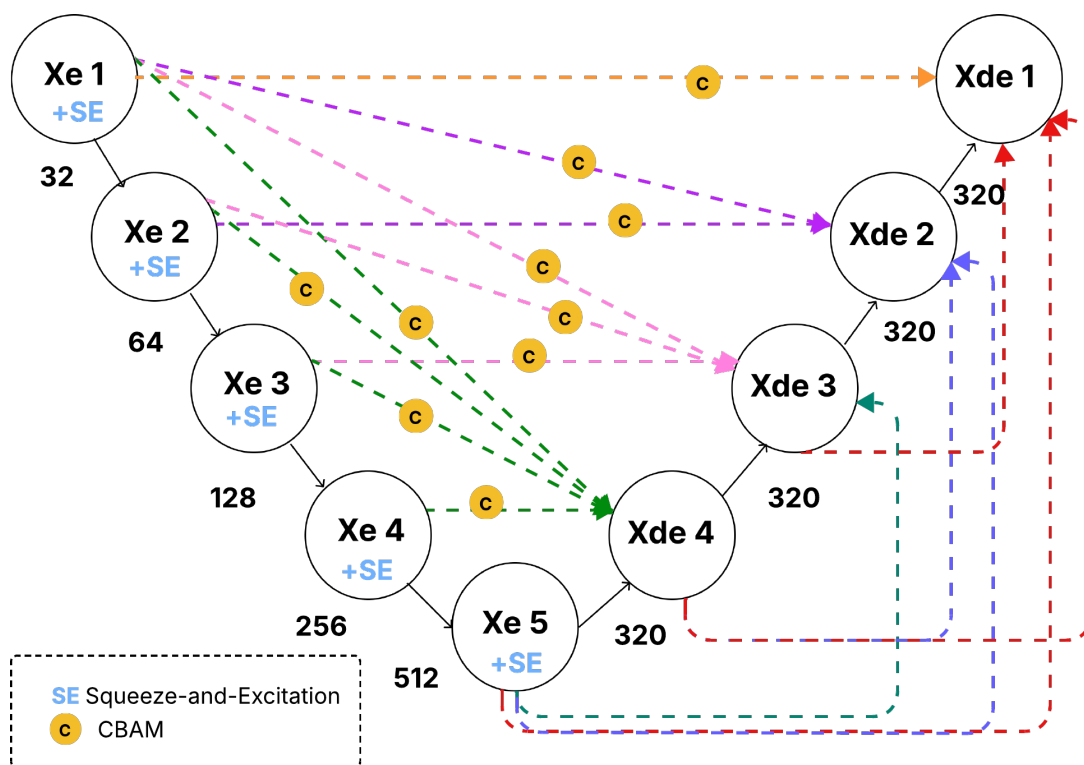


Figura 27 – U-Net 3+ modificada com adição dos mecanismos de atenção SE e CBAM

Fonte: Autoral.

#### 4.4 Estimação automática do ângulo de Cobb

Na etapa de estimação dos ângulos de Cobb foi usado o método desenvolvido por Maguire (2020).

A primeira etapa do método consiste na extração automática das placas terminais superiores e inferiores de cada vértebra, a partir da máscara binária da coluna vertebral segmentada.

Inicialmente, a máscara da coluna é rotulada por componentes conexos, permitindo a identificação individual de cada vértebra. Para cada componente, são calculadas propriedades geométricas, em particular o centróide, que é utilizado posteriormente para ordenar as vértebras no sentido crânio-caudal (de cima para baixo na imagem).

Para cada vértebra válida (regiões pequenas ou degeneradas são descartadas), é extraído o seu contorno externo. A partir desse contorno, calcula-se o retângulo mínimo que envolve cada vértebra. Com base nesse retângulo, é determinado o ângulo de inclinação da vértebra em relação ao eixo horizontal da imagem.

Em seguida, a máscara da vértebra é rotacionada para alinhamento horizontal, eliminando o efeito da inclinação global e permitindo uma análise mais precisa das suas bordas superior e inferior. Após a rotação, recalcula-se o centróide da vértebra e extrai-se uma linha horizontal próxima ao seu centro.

Ao longo de um intervalo central da largura da vértebra, são identificados, coluna a coluna, os pixels mais superiores e mais inferiores pertencentes à máscara rotacionada. Esses pontos representam, respectivamente, as bordas superior e inferior da vértebra. Sobre esses conjuntos de pontos é ajustado um modelo linear por mínimos quadrados, resultando em duas retas que aproximam as placas terminais superior e inferior.

Os extremos dessas retas são então utilizados como marcos anatômicos (*landmarks*) das placas terminais. Por fim, esses pontos são transformados de volta para o sistema de coordenadas original da imagem por meio da rotação inversa, preservando sua posição anatômica correta. Para cada vértebra, são obtidos quatro pontos: dois da placa superior e dois da placa inferior.

#### 4.4.1 Construção das retas médias vertebrais

Com os *landmarks* das placas terminais extraídos para todas as vértebras, o método calcula as retas médias vertebrais. Para isso, os pontos terminais (cantos) das placas superior e inferior de cada vértebra são usados para encontrar o ponto médio do lado esquerdo e direito. Esses pontos médios definem uma reta que representa a orientação global da vértebra, sendo utilizada como base para o cálculo de inclinações e para a identificação automática das vértebras limites das curvas escolióticas. A inclinação de cada vértebra é obtida a partir do coeficiente angular dessa reta média.

#### 4.4.2 Detecção automática das vértebras

O método realiza uma detecção automática das regiões [Torácica Proximal](#) (do inglês, *Proximal Thoracic*) (PT), [Torácica Principal](#) (do inglês, *Main Thoracic*) (MT) e [Toracolombar/Lombar](#) (do inglês, *Thoracolumbar/Lumbar*) (TL/L).

Inicialmente, a vértebra apical da curva torácica principal é estimada a partir do deslocamento lateral dos centróides vertebrais em relação a média global da coluna. Picos positivos ou negativos nesse deslocamento indicam o ponto de maior desvio lateral, caracterizando a vértebra apical.

A partir dessa vértebra apical, são identificadas as vértebras mais inclinadas acima e abaixo, com base na variação das inclinações das retas médias. Essas vértebras correspondem às vértebras limites superior e inferior de cada curva, conforme o critério clássico do método de Cobb.

De forma análoga, são determinadas as vértebras apicais e limites das curvas [PT](#) e [TL/L](#).

### 4.4.3 Cálculo do ângulo de Cobb

Uma vez definidas as vértebras limites de cada curva, o ângulo de Cobb é calculado a partir das inclinações das placas terminais:

- Utiliza-se a placa terminal superior da vértebra limite superior;
- Utiliza-se a placa terminal inferior da vértebra limite inferior.

O ângulo de Cobb é então definido como o valor absoluto da diferença entre os ângulos dessas duas retas em relação ao eixo horizontal, conforme apresentado na Equação 4.1:

$$\theta_{\text{Cobb}} = |\arctan(m_{\text{sup}}) - \arctan(m_{\text{inf}})| \quad (4.1)$$

Onde  $m_{\text{sup}}$  e  $m_{\text{inf}}$  representam os coeficientes angulares das placas terminais superior e inferior, respectivamente, calculados por meio da equação da reta definida por dois pontos, apresentada na Equação 4.2. Nessa equação,  $x$  e  $y$  correspondem as coordenadas vertical e horizontal de cada vértice da vértebra. Os dois vértices superiores são utilizados para determinar a inclinação da placa terminal superior, enquanto os dois vértices inferiores são empregados para determinar a inclinação da placa terminal inferior.

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (4.2)$$

Esse procedimento é repetido para as três curvas analisadas (PT, MT e TL/L), resultando em três ângulos de Cobb independentes.

Um pseudo-código é apresentado pelo algoritmo 1, no qual são descritas as etapas detalhadas nesta seção para a estimação automática do ângulo de Cobb.

**Algoritmo 1:** Estimativa Automática do Ângulo de Cobb

---

**Input:** Máscara binária da coluna segmentada

**Output:** Ângulos de Cobb das curvas PT, MT e TL/L

Rotular componentes conexos da máscara;

Calcular centróides e ordenar vértebras (crânio-caudal);

**foreach** *vértebra válida* **do**

- Extrair contorno externo;
- Calcular retângulo mínimo envolvente;
- Estimar ângulo de inclinação;
- Rotacionar máscara para alinhamento horizontal;
- Identificar pontos superiores e inferiores (intervalo central);
- Ajustar retas por mínimos quadrados (placas terminais);
- Transformar *landmarks* para coordenadas originais;
- Calcular reta média vertebral e sua inclinação;

Detectar vértebras apicais (PT, MT, TL/L) via deslocamento lateral;

Identificar vértebras limites superior e inferior;

**foreach** *curva (PT, MT, TL/L)* **do**

- Calcular coeficientes angulares das placas terminais limites;
- $\theta_{\text{Cobb}} \leftarrow |\arctan(m_{\text{sup}}) - \arctan(m_{\text{inf}})|$ ;

**return**  $\theta_{PT}, \theta_{MT}, \theta_{TL/L}$

---

Com o objetivo de representar as principais etapas do método proposto, a Figura 28 ilustra o fluxo de processamento adotado. Inicialmente, são identificados os centróides na imagem da máscara binária fornecida como entrada do método (Figura 28a). Em seguida, a partir do menor retângulo que delimita cada vértebra, são extraídos os pontos terminais correspondentes a cada estrutura vertebral (Figura 28b).

Posteriormente, são apresentadas as linhas terminais superior e inferior de cada vértebra (Figura 28c), obtidas com base nos pontos extraídos na etapa anterior. A partir dessas linhas, são determinadas as inclinações de cada vértebra, representadas por linhas indicativas da orientação vertebral (Figura 28d).

Por fim, é apresentada a imagem final (Figura 28e), na qual são destacadas as placas terminais superior e inferior das vértebras empregadas no cálculo de cada ângulo de Cobb. Essas marcações são sobrepostas à imagem de raio-x utilizada como entrada do método desenvolvido neste trabalho, juntamente com os valores dos ângulos estimados, associados às respectivas curvas às quais pertencem.

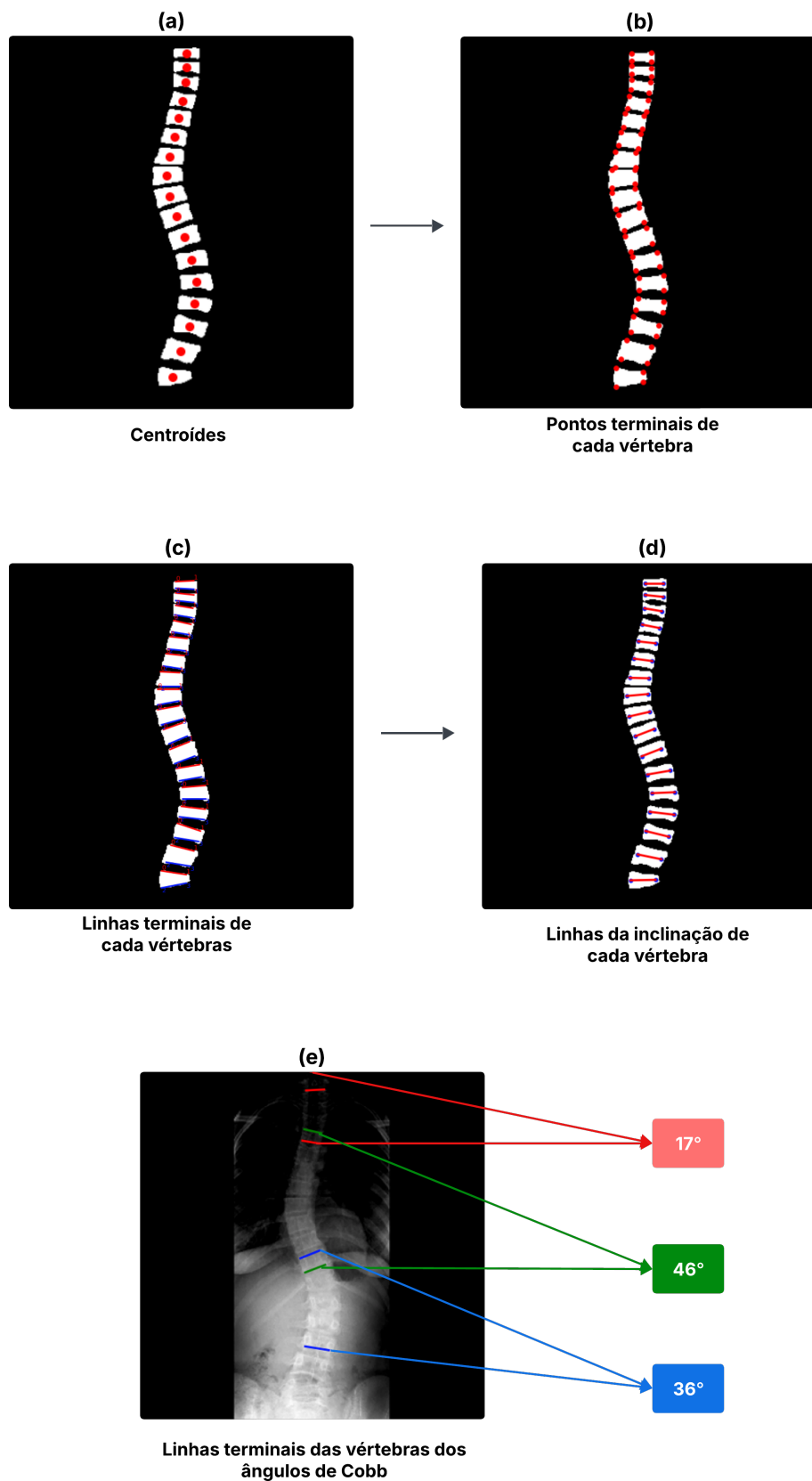


Figura 28 – Etapas da estimação do ângulo de Cobb: (a) Centróides vertebrais, (b) Pontos terminais de cada vértebra, (c) Linhas terminais de cada vértebra, (d) Linhas que mostram a inclinação de cada vértebra e (e) Placas terminais das vértebras usadas no cálculo dos ângulos de Cobb encontrados, com os respectivos valores

Fonte: Autoral.

## 4.5 Considerações Finais

Neste capítulo, foram apresentadas as etapas do método desenvolvido para a estimação do ângulo de Cobb, o qual se baseia em uma abordagem de segmentação em cascata das vértebras. Inicialmente, realiza-se a segmentação da coluna vertebral por meio da rede descrita na Seção 4.1.2, seguida de uma etapa de pós-processamento para a remoção de regiões erroneamente segmentadas e aplicação do fecho convexo, resultando na máscara final da ROI. Essa máscara é então utilizada para o recorte da região da coluna na imagem de raio-x.

Em seguida, a imagem recortada é empregada como entrada para a segmentação das vértebras, utilizando a rede apresentada na Seção 4.3. Por fim, a estimação do ângulo de Cobb é realizada com base no método proposto por Maguire (2020), que envolve a localização dos centróides vertebrais, o cálculo das inclinações a partir das placas terminais superior e inferior e a determinação dos ângulos correspondentes as curvas escolióticas identificadas.

## 5 Resultados e Discussão

Neste capítulo, são apresentados e discutidos os resultados obtidos nos experimentos realizados para a avaliação da metodologia proposta. Para isso, são analisados os resultados de cada etapa do método, iniciando pela segmentação da coluna vertebral (identificação da ROI), seguida pela avaliação da segmentação das vértebras e pela estimação do ângulo de Cobb. Além disso, é realizada a comparação entre os resultados obtidos neste estudo e aqueles reportados por outros métodos da literatura.

### 5.1 Base de imagens

Para a realização dos treinamentos e testes dos modelos empregados neste trabalho, foi utilizada a base de dados do desafio AASCE2019 (WU et al., 2017), a qual contém 609 imagens de raio-x da coluna vertebral em projeção AP, provenientes de pacientes com diferentes níveis de escoliose, sendo 481 destinadas ao treinamento e 128 ao teste. Nos experimentos realizados, o conjunto de treinamento foi subdividido em 431 imagens para treinamento e 50 para validação.

A base de dados disponibiliza, para cada imagem de raio-x, um arquivo associado contendo as coordenadas dos cantos de cada vértebra, bem como os valores dos três ângulos de Cobb mensurados por um especialista. No entanto, como não são fornecidas máscaras das vértebras, mas apenas as marcações correspondentes, as máscaras vertebrais foram geradas por meio do algoritmo proposto por Maguire (2020).

No que se refere as máscaras utilizadas como referência para a segmentação da coluna vertebral, empregou-se a união dos pontos correspondentes a cada vértebra, formando assim o contorno da coluna. A Figura 29 apresenta exemplos das máscaras utilizadas como referência tanto para as vértebras quanto para a coluna vertebral.

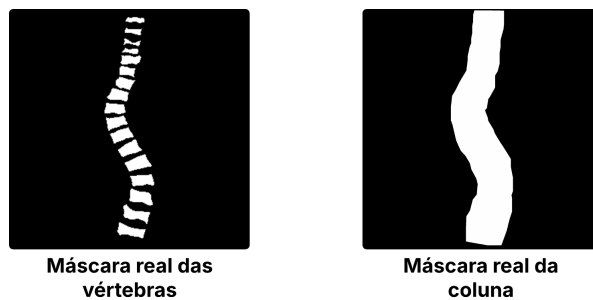


Figura 29 – Máscara real das vértebras e da coluna vertebral.

Fonte: Autoral.

## 5.2 Configuração dos experimentos

Os experimentos foram realizados em um *hardware* com as seguintes especificações: Processador Intel Core i5-9600KF de 3.70 GHz, 8 GB de memória RAM, placa gráfica NVIDIA GeForce GTX 1660 Super com 6 GB de VRAM e sistema operacional Windows 10.

Para as etapas de extração da ROI e de segmentação das vértebras, foi utilizada a base de dados descrita na Seção 5.1, diferenciando-se apenas as máscaras de referência. No caso da extração da ROI, empregou-se o contorno da coluna vertebral, enquanto, para a segmentação das vértebras, foi utilizada a máscara contendo as vértebras individualmente delimitadas.

Em ambas as etapas de segmentação, os modelos foram treinados por até 400 épocas, com *batch size* igual a 2, taxa de aprendizado de  $1e-4$ , *early stopping* com paciência de 20 épocas e função de perda Hybrid loss (HUANG et al., 2020), a qual apresentou melhor desempenho em ambos os cenários.

## 5.3 Experimentos

Os experimentos realizados foram comparados por meio das métricas apresentadas na Seção 3.3.12, de acordo com a etapa do método à qual se aplicam. Além disso, realiza-se uma análise visual dos resultados nas etapas de segmentação. As comparações dos resultados obtidos são apresentadas de forma progressiva, acompanhando cada etapa do método até a avaliação do sistema completo. Assim, inicialmente são apresentados os resultados da segmentação da coluna vertebral, correspondente a extração da ROI, seguidos pelos resultados da segmentação das vértebras e, por fim, pela análise do método completo com a estimativa dos ângulos de Cobb.

### 5.3.1 Extração da ROI

Para a etapa de extração da ROI, foram comparadas algumas arquiteturas com o objetivo de identificar aquela que apresentasse melhor desempenho na segmentação da coluna vertebral. A seleção das redes foi baseada nas características que as diferenciam e que poderiam contribuir para a melhoria dos resultados. A DeepLabv3+ (CHEN et al., 2018) emprega convoluções dilatadas e um módulo *Atrous Spatial Pyramid Pooling (ASPP)*, capaz de capturar informações contextuais em múltiplas escalas. Essas características são utilizadas no decodificador e combinadas com atributos de baixo nível extraídos das camadas iniciais do *backbone* neste experimento, a ResNet50 (HE et al., 2016b), que preservam maior detalhamento espacial.

A U-Net (RONNEBERGER; FISCHER; BROX, 2015) (Seção 3.3.8), por sua vez, foi incluída por ser uma arquitetura de referência em tarefas de segmentação. A Unet++ (ZHOU et al., 2018) estende esse modelo ao introduzir caminhos intermediários entre o codificador e o decodificador, promovendo múltiplos níveis de convolução e melhor fusão de características. A  $U^2$ -Net (QIN et al., 2020) utiliza blocos residuais em formato de U, formando mini U-Nets (RONNEBERGER; FISCHER; BROX, 2015), o que possibilita a captura simultânea de contexto local e global. Já a Swin-UNet (CAO et al., 2022) combina a estrutura da U-Net (RONNEBERGER; FISCHER; BROX, 2015) com o uso do Swin Transformer (LIU et al., 2021), explorando mecanismos de atenção hierárquica baseados em janelas deslocadas.

Por fim, a Unet 3+ (HUANG et al., 2020) se destaca pelo redesenho das *skip connections*, que passam a integrar informações provenientes de múltiplos níveis do codificador e do decodificador. A partir da análise dos resultados apresentados na Tabela 4, observa-se que as arquiteturas com melhor desempenho foram a DeepLabv3+ (CHEN et al., 2018) e a Unet 3+ (HUANG et al., 2020). Embora a DeepLabv3+ (CHEN et al., 2018) tenha apresentado os melhores resultados sem modificações arquiteturais, optou-se por aprofundar a análise da Unet 3+ (HUANG et al., 2020) com a incorporação de mecanismos de atenção. Essa escolha foi motivada pelo maior custo computacional da DeepLabv3+ (CHEN et al., 2018), que possui 26.809.616 parâmetros a mais em comparação com a Unet 3+ (HUANG et al., 2020) com atenção e 30.899.457 parâmetros adicionais em relação à Unet 3+ (HUANG et al., 2020) em sua configuração original.

Adicionalmente, a avaliação do desempenho da DeepLabv3+ em outras etapas do método indicou menor eficiência quando comparada à Unet 3+, o que reforçou essa escolha. Dessa forma, a rede que apresentou melhor desempenho foi a modificação proposta (Seção 4.1.2).

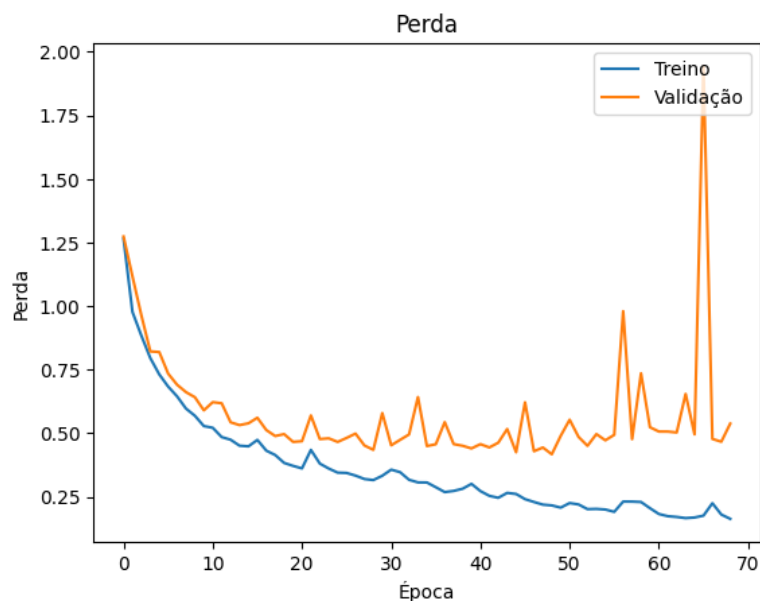
Tabela 4 – Comparação de arquiteturas para extração da ROI

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Deeplabv3+	91,58	87,28	<b>96,75</b>
Unet	70,74	67,25	76,09
Unet++	71,84	74,75	72,32
Swin-UNet	86,04	82,16	91,02
$U^2$ -Net	64,53	50,97	90,19
Unet 3+	89,38	82,78	97,61
<b>Unet 3+ com Attention Gate</b>	<b>92,14</b>	<b>88,45</b>	96,52

A curva de *loss* do modelo utilizado é apresentada na Figura 30. Observa-se que a *loss* de validação acompanha de forma consistente a *loss* de treinamento até aproximadamente a época 30. A partir desse ponto, nota-se uma maior oscilação na curva de validação, indicando possível início de instabilidade no processo de generalização.

Ressalta-se que o treinamento foi conduzido com a técnica de *early stopping*, permitindo a preservação do modelo com melhor desempenho na validação. Entre as possíveis causas para a variação observada nas últimas épocas destacam-se o tamanho reduzido do *batch size*, composto por apenas duas imagens, bem como o tamanho limitado da base de dados, fatores que podem contribuir para maior variabilidade nas atualizações dos gradientes e, conseqüentemente, nas métricas de validação.

Figura 30 – Curva de loss de treinamento e validação



Fonte: Autoral.

O resultado da aplicação da validação cruzada com 5 *folders* para a segmentação das vértebras utilizando a Unet 3+ com Attention Gate está apresentado na Tabela 5. Observa-se um baixo desvio padrão em todas as métricas avaliadas, indicando elevada estabilidade e consistência do modelo entre os diferentes *folders* executados.

Tabela 5 – Validação cruzada da Unet 3+ com Attention Gate para segmentação da coluna

Número do fold	Dice (%)	Precisão (%)	Sensibilidade (%)
1	97,34	87,37	96,88
2	91,38	86,95	96,70
3	91,22	85,72	97,87
4	90,75	85,32	97,49
5	91,03	86,22	96,76
Média	92,34	86,31	97,13
Desvio padrão	2,80	0,84	0,51

Fonte: Autoral.

Após a segmentação da coluna, foi aplicado às imagens resultantes do modelo com melhor desempenho um pós-processamento, conforme descrito na Seção 4.1.3, cujos

resultados são apresentados na Tabela 6. Observa-se uma redução no valor do coeficiente Dice após essa etapa, atribuída principalmente às operações de erosão empregadas para separar regiões indevidamente conectadas à coluna. Em contrapartida, verifica-se um aumento na precisão, indicando uma diminuição no número de falsos positivos detectados.

Tabela 6 – Pós-processamento segmentação da coluna

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Máscara com pós-processamento rede posto	88,70	<b>97,26</b>	81,75
Máscara sem pós-processamento rede posto	<b>92,14</b>	88,45	<b>96,52</b>

Fonte: Autoral.

Na Figura 31 são apresentados os resultados do pós-processamento: a primeira imagem corresponde à saída bruta da segmentação, a segunda ao resultado após a aplicação do pós-processamento e a terceira à máscara de referência (imagem anotada). Observa-se que, após o processamento, há uma redução da região segmentada.

O método de pós-processamento possui uma etapa que remove pequenas regiões desconectadas e preserva a maior componente conexa, presumivelmente correspondente a coluna vertebral. Esse procedimento elimina áreas espúrias não associadas a coluna, como a pequena região presente na Figura 31a.

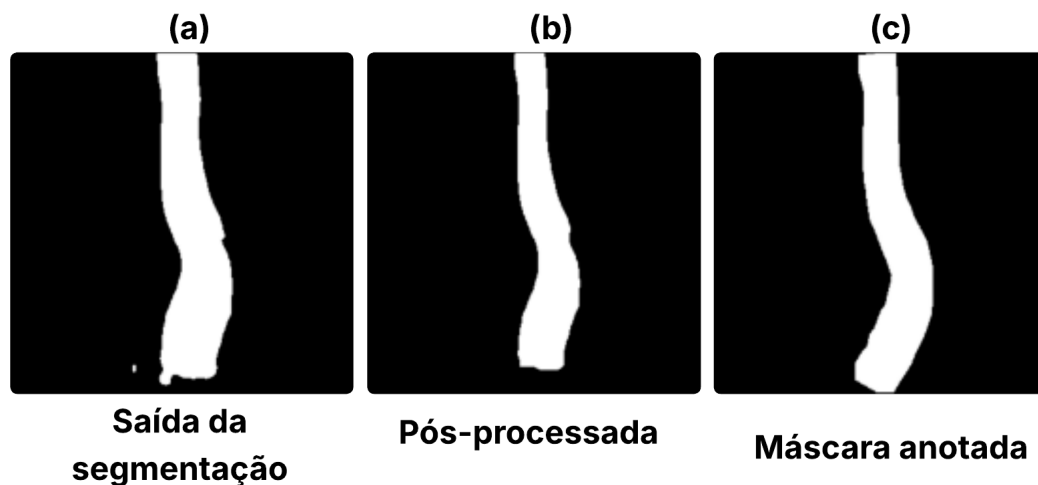
Entretanto, ressalta-se que a simples manutenção da maior área conectada não é suficiente para remover regiões indevidamente aderidas à coluna. Esse comportamento pode ser observado, por exemplo, na porção final da máscara gerada diretamente pela segmentação, na qual um pequeno fragmento permanece conectado à estrutura principal, não sendo eliminado apenas pelo critério de conectividade, sendo aplicado as outras etapas do pós-processamento explicadas na Seção 4.1.3.

### 5.3.2 Aplicação do fecho convexo

A aplicação do fecho convexo foi realizada após a etapa de pós-processamento da segmentação da coluna. Para analisar o impacto desse procedimento, avaliou-se o desempenho da U-Net 3+ (HUANG et al., 2020) sem modificações, uma vez que a utilização do fecho convexo foi incorporada ainda no início da pesquisa, no momento em que passou-se a aplicar o pós-processamento (Seção 4.1.3).

As imagens de raio-x dos conjuntos de treino e validação foram recortadas com base na ROI da coluna gerada a partir da máscara anotada, com posterior aplicação do fecho convexo na máscara. Para o conjunto de teste, utilizou-se o resultado da segmentação da coluna feita pela U-Net 3+ (HUANG et al., 2020), a qual foi aplicado o pós-processamento e, em seguida, o fecho convexo. Assim, a entrada da rede responsável pela segmentação

Figura 31 – Pós-processamento: (a) Segmentação da coluna, (b) Resultado pós-processamento e (c) Máscara real



Fonte: Autoral.

das vértebras passou a ser a imagem de raio-x recortada com base na ROI obtida após a aplicação do fecho convexo.

A Tabela 7 apresenta os resultados da segmentação das vértebras considerando apenas o pós-processamento e a combinação do pós-processamento com o fecho convexo. Observa-se que a aplicação do fecho convexo proporcionou melhorias tanto no coeficiente Dice quanto na precisão, sendo esta última indicativa da redução no número de falsos positivos durante a segmentação das vértebras.

Tabela 7 – Aplicação do fecho convexo após o pós-processamento segmentação da coluna

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Pós-processamento e fecho convexo	<b>82,48</b>	<b>78,60</b>	87,16
Somente pós-processamento	79,86	70,78	<b>92,02</b>

Fonte: Autoral.

### 5.3.3 Segmentação das vértebras com base na ROI

Na etapa de segmentação das vértebras utilizando a ROI, foram empregadas as mesmas arquiteturas descritas na Seção 5.3.1, com o objetivo de selecionar a rede mais adequada para essa tarefa. Todos os treinamentos foram realizados a partir da ROI extraída com base na máscara anotada, após a aplicação do fecho convexo.

Na fase de teste, foram realizadas duas análises distintas. A Tabela 8 apresenta os resultados obtidos a partir da ROI extraída da segmentação da coluna produzida por cada arquitetura na etapa anterior, após a aplicação do pós-processamento e fecho convexo. Já a Tabela 9 apresenta os resultados da segmentação das vértebras utilizando, para todos

os modelos avaliados, a ROI extraída pela U-Net 3+ com Attention Gate, que obteve o melhor desempenho na etapa anterior. Dessa forma, o resultado do melhor modelo da etapa de segmentação da coluna foi utilizado como entrada comum nos testes de segmentação das vértebras para todas as arquiteturas apresentadas na Tabela 9.

A partir dos resultados apresentados na Tabela 8, observa-se que a arquitetura com melhor desempenho foi a U-Net 3+ com *Attention Gate*, seguida pela U-Net 3+ (HUANG et al., 2020), considerando o coeficiente Dice como métrica de avaliação.

Tabela 8 – Comparação de arquiteturas para segmentação das vértebras com base na ROI gerada na etapa de extração feita por cada modelo

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Deeplabv3+	76,32	65,03	<b>92,85</b>
Unet	61,74	64,73	59,84
Unet++	60,89	61,34	62,67
Swin-Unet	63,75	50,80	86,14
U <sup>2</sup> -Net	42,40	28,65	82,87
Unet 3+	82,48	78,60	87,16
<b>Unet 3+ com Attention Gate</b>	<b>83,23</b>	<b>81,97</b>	84,80

Fonte: Autoral.

Na Tabela 9, é apresentado o resultado de outra modificação proposta com base na arquitetura U-Net 3+ (HUANG et al., 2020), denominada U-Net 3+ com blocos SE e CBAM, descrita na Seção 4.3. Essa arquitetura foi empregada exclusivamente na segmentação das vértebras, motivo pelo qual seus resultados são apresentados apenas na Tabela 9. Os resultados indicam que a U-Net 3+ com blocos SE e CBAM obteve o melhor desempenho na segmentação das vértebras.

Dessa forma, com base nos experimentos realizados, o método de segmentação em cascata adotado consistiu na utilização da U-Net 3+ com *Attention Gate* para a segmentação da coluna e da U-Net 3+ com blocos SE e CBAM para a segmentação das vértebras.

A curva de *loss* do modelo utilizado é apresentada na Figura 32. Observa-se que a *loss* de validação acompanha de forma consistente a *loss* de treinamento até aproximadamente a época 30. A partir desse ponto, nota-se uma maior oscilação na curva de validação, indicando possível início de instabilidade no processo de generalização.

Ressalta-se que o treinamento foi conduzido com a técnica de *early stopping*, permitindo a preservação do modelo com melhor desempenho no conjunto de validação. Observa-se que as curvas de *loss* de treinamento e validação iniciam com valores elevados e apresentam redução ao longo das épocas, indicando o aprendizado progressivo do modelo. A partir de determinado ponto, a *loss* de validação tende à estabilização, enquanto a *loss* de treinamento continua a decrescer. As oscilações observadas na curva de validação

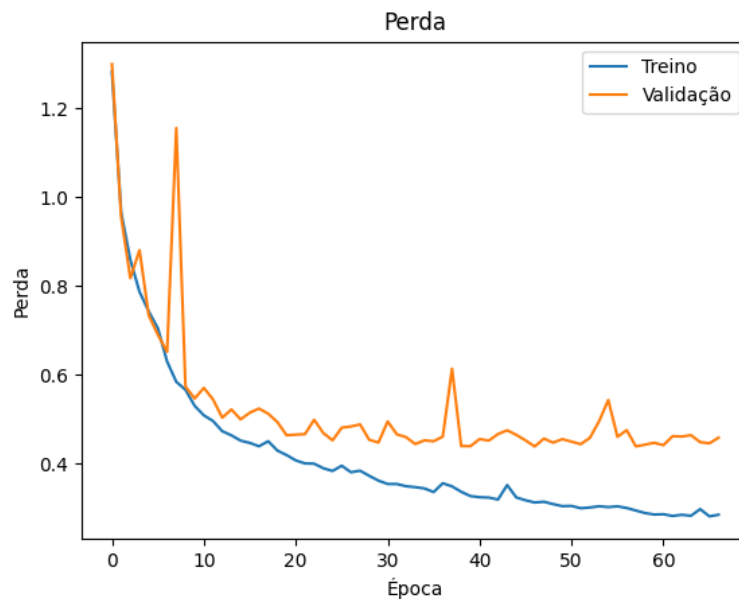
Tabela 9 – Comparação de arquiteturas para segmentação das vértebras com base na ROI gerada pelo modelo Unet 3+ com Attention Gate

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Deeplabv3+	67,20	55,92	85,09
Unet	79,56	77,26	82,25
Unet++	70,35	65,42	76,76
Swin-Unet	59,23	48,91	76,03
$U^2$ -Net	59,25	43,47	<b>93,88</b>
Unet 3+	81,11	73,36	91,01
Unet 3+ com Attention Gate	83,23	81,97	84,80
<b>Unet 3+ com Bloco SE + CBAM</b>	<b>83,55</b>	<b>82,40</b>	85,08

Fonte: Autoral.

podem estar associadas ao reduzido tamanho do *batch size*, composto por apenas duas imagens, bem como ao número limitado de amostras na base de dados, fatores que podem aumentar a variabilidade das atualizações dos gradientes e, conseqüentemente, da *loss*.

Figura 32 – Curva de Loss de treinamento e validação



Fonte: Autoral.

O resultado da aplicação da validação cruzada com 5 *folds* para a segmentação das vértebras utilizando a U-Net 3+ com blocos SE e CBAM está apresentado na Tabela 10. Observa-se um baixo desvio padrão em todas as métricas avaliadas, indicando elevada estabilidade e consistência do modelo entre os diferentes *folds* executados.

Tabela 10 – Validação cruzada da Unet 3+ com Bloco SE e CBAM para segmentação das vértebras

Número do fold	Dice (%)	Precisão (%)	Sensibilidade (%)
1	82,84	82,84	83,13
2	83,06	82,89	83,50
3	82,78	82,21	83,64
4	83,30	81,58	85,57
5	83,74	82,44	85,35
Média	83,14	82,39	84,23
Desvio padrão	0,39	0,53	1,13

Fonte: Autoral.

## 5.4 Comparação segmentação das vértebras sem a extração da ROI

Nesta etapa, foi realizada a segmentação direta das vértebras, sem a extração da ROI, utilizando as mesmas arquiteturas apresentadas na Seção 5.3.1, além das duas modificações propostas. A entrada da rede consistiu na imagem de raio-x original, após redimensionamento e aplicação de padding (Seção 4.1.1), e a saída correspondeu à máscara das vértebras.

A Tabela 11 apresenta os resultados dos testes desse experimento. Observa-se que as arquiteturas que se destacaram em relação as demais foram as duas modificações propostas baseadas na U-Net 3+ (HUANG et al., 2020) e a U-Net (RONNEBERGER; FISCHER; BROX, 2015). No entanto, para o cálculo do ângulo de Cobb, os resultados obtidos por essas redes foram inferiores quando comparados ao método de segmentação em cascata desenvolvido neste trabalho, conforme será apresentado na Seção 5.4.1.

Tabela 11 – Comparação da segmentação sem extração da ROI em diferentes arquiteturas

Arquitetura	Dice (%)	Precisão (%)	Sensibilidade (%)
Deeplabv3+	67,20	55,92	85,09
Unet	82,44	<b>81,55</b>	83,62
Unet++	80,68	74,64	88,07
Swin-Unet	65,87	54,40	84,07
$U^2$ -Net	59,11	42,08	<b>99,94</b>
Unet 3+	81,11	73,36	91,01
Unet 3+ com Attention Gate	82,87	76,61	90,55
<b>Unet 3+ com Bloco SE + CBAM</b>	<b>84,17</b>	80,59	88,35

Fonte: Autoral.

### 5.4.1 Comparação do cálculo do ângulo de Cobb para segmentação direta e método proposto

Considerando como entrada as imagens segmentadas pelo método proposto e aquelas obtidas diretamente pelas redes com melhor desempenho na segmentação sem a utilização da ROI, o cálculo do ângulo de Cobb foi realizado no conjunto composto por 128 imagens de teste.

A Tabela 12 apresenta a média das medições correspondentes a cada um dos três ângulos avaliados por imagem (MAE/ SMAPE 1, 2 e 3). Já os valores indicados como (MAE/ SMAPE médio) representam a média dos valores médios obtidos para os três ângulos. Para essa avaliação, foram empregadas as métricas MAE e SMAPE, descritas na Seção 3.3.12. A partir dos resultados, observa-se que o método proposto apresentou os melhores desempenhos em praticamente todas as medições, sendo superado apenas na métrica MAE do primeiro ângulo quando comparado à U-Net (RONNEBERGER; FISCHER; BROX, 2015). Considerando que valores menores de MAE e SMAPE indicam melhor desempenho.

Tabela 12 – Comparação da segmentação sem extração da ROI em diferentes arquiteturas

Arquitetura	MAE 1	MAE 2	MAE 3	SMAPE 1	SMAPE 2	SMAPE 3	MAE Médio	SMAPE Médio
Unet	<b>8.98</b>	8.44	12.01	53.57	29.00	49.24	9,81	49,24
Unet 3+ com Attention Gate	21.02	17.91	19.73	73.24	60.61	100.87	19,55	73,24
Unet 3+ com Bloco SE + CBAM	9.64	10.52	12.25	56.53	32.87	52.45	10,80	52,45
<b>Método proposto</b>	9.19	<b>7.67</b>	<b>9.29</b>	<b>48.64</b>	<b>27.81</b>	<b>44.80</b>	<b>8,72</b>	<b>44,80</b>

Fonte: Autoral.

### 5.4.2 Comparação com os trabalhos relacionados

Nesta seção, são comparados os resultados obtidos com o método proposto e aqueles reportados na literatura.

Considerando que alguns trabalhos da literatura realizaram a segmentação da coluna, a Tabela 13 apresenta a comparação entre os resultados obtidos pelo método proposto e aqueles reportados no trabalho de Liang et al. (2022). Observa-se que o método proposto superou, em termos de coeficiente Dice, a segmentação obtida pela SegNet apresentada em Liang et al. (2022).

No que se refere à segmentação das vértebras, a Tabela 14 apresenta a comparação entre o método proposto e trabalhos da literatura que abordam essa tarefa. Observa-se que, embora o método proposto não tenha alcançado os melhores resultados, ele apresenta desempenho superior aos métodos descritos por Zhao et al. (2022), Cui et al. (2021) e Maguire (2020).

Ressalta-se, contudo, que apenas a comparação com o trabalho desenvolvido por Maguire (2020) pode ser considerada equivalente, uma vez que ambos utilizam a mesma

Tabela 13 – Comparação da segmentação da coluna com os trabalhos relacionados

Trabalhos	Dice (%)	Precisão (%)	Sensibilidade (%)	Base de Imagens
(LIANG et al., 2022)	90	-	-	AASCE2019
<b>Método proposto</b>	92,14	88,45	96,52	AASCE2019

Fonte: Autoral.

base de teste. Os demais trabalhos empregam bases de dados privadas ou utilizam, no conjunto de testes, imagens que também foram consideradas neste estudo, o que limita a comparabilidade direta dos resultados.

Tabela 14 – Comparação da segmentação das vértebras com os trabalhos relacionados

Trabalhos	Dice (%)	Precisão (%)	Sensibilidade (%)	Base de Imagens
(TU et al., 2019)	90,28	86,3	-	Base privada
(CUI et al., 2021)	78	-	-	AASCE2019
(ZHAO et al., 2022)	81,46	-	-	Base privada e Imagens de teste do AASCE2019
(PENG et al., 2024)	89,51	96,11	92,37	Base privada
(LIU et al., 2026)	94,32	98,33	-	AASCE2019
(MAGUIRE, 2020)	74,9	-	-	AASCE2019
<b>Método proposto</b>	83,55	82,40	85,08	AASCE2019

Fonte: Autoral.

Na comparação entre os resultados obtidos para o cálculo do ângulo de Cobb pelo método proposto e os trabalhos relacionados apresentados na Tabela 15, observa-se que o desempenho do método não foi tão satisfatório quanto o da maioria dos estudos comparados. Em relação a métrica MAE, grande parte dos trabalhos reporta erro médio entre 1° e 5° em relação ao ângulo real, enquanto o método proposto apresentou erro próximo de 9°. Vale ressaltar que muitos trabalhos utilizaram bases de dados privadas e, mesmo entre aqueles que empregaram a mesma base de imagens, não foi adotada a mesma divisão do conjunto de dados utilizada neste estudo, com exceção de Maguire (2020).

Quanto a métrica SMAPE, o valor obtido também se mostrou significativamente superior aos demais. Esse comportamento pode ser explicado pela maior discrepância

Tabela 15 – Comparação do cálculo do ângulo de Cobb com os trabalhos relacionados

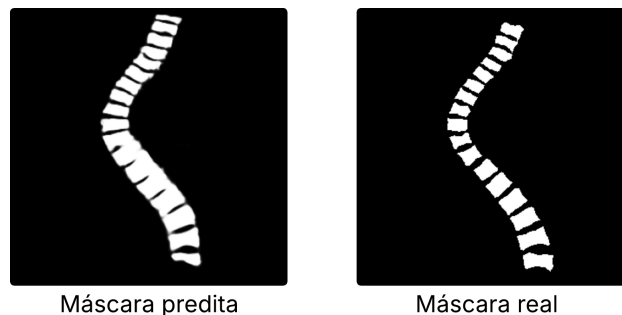
Trabalhos	MAE (°)	SMAPE (%)	CMAE	ME (°)	MAD	Base de Imagens
Tu et al. (2019)	2,9	-	-	-	-	Base privada
(KIM et al., 2020)	-	7,84	-	-	-	Base privada e AASCE2019
(YI et al., 2020)	-	10,81	-	-	-	AASCE2019
(MAGUIRE, 2020)	5,54	32,05	-	-	-	AASCE2019
Cui et al. (2021)	-	21,675	-	-	-	AASCE2019
Zhao et al. (2022)	-	6,63	-	-	-	Base privada e Imagens de teste do AASCE2019
Liang et al. (2022)	-	6,87	-	-	-	AASCE2019
(SURI et al., 2023)	-	-	-	2	-	Base privada e Base Fraiwan
(ZHOU; ZHU; YAO, 2023)	-	-	-	-	3,13	Base privada
(MAEDA et al., 2023)	-	-	-	1,7	-	Base privada
(YANG; WANG; MENG, 2024)	-	8,99	-	-	-	AASCE2019 e Base privada
(LU et al., 2024)	2,15	-	-	-	-	AASCE2019 e Base privada
(HUANG et al., 2024)	-	-	-	-	2,76	Base privada
(PENG et al., 2024)	1,6	-	-	-	-	Base privada
(LIU et al., 2026)	-	-	2,24	-	-	AASCE2019
<b>Método proposto</b>	<b>8,72</b>	<b>48</b>	-	-	-	<b>AASCE2019</b>

Fonte: Autoral.

observada na medição de dois dos três ângulos avaliados, conforme apresentado na Tabela 12. Como o **SMAPE** penaliza proporcionalmente diferenças relativas entre os valores preditos e reais, variações maiores impactam de forma mais expressiva o resultado final.

Cabe destacar que o método proposto empregou a mesma estratégia de cálculo do ângulo de Cobb utilizada em Maguire (2020), apresentando diferença de  $3.18^\circ$  na métrica **MAE** e aproximadamente 16% a mais no **SMAPE**. Essa discrepância pode ser atribuída principalmente a falhas na segmentação em algumas imagens, nas quais duas ou mais vértebras foram identificadas como uma única estrutura, conforme ilustrado na Figura 33. Como o cálculo do ângulo depende da identificação dos cantos de cada vértebra, essa união reduz o número de vértebras detectadas e compromete a estimativa da inclinação, uma vez que estruturas distintas passam a ser tratadas como uma única vértebra.

Figura 33 – Máscara predita com junção de vértebras e máscara anotada



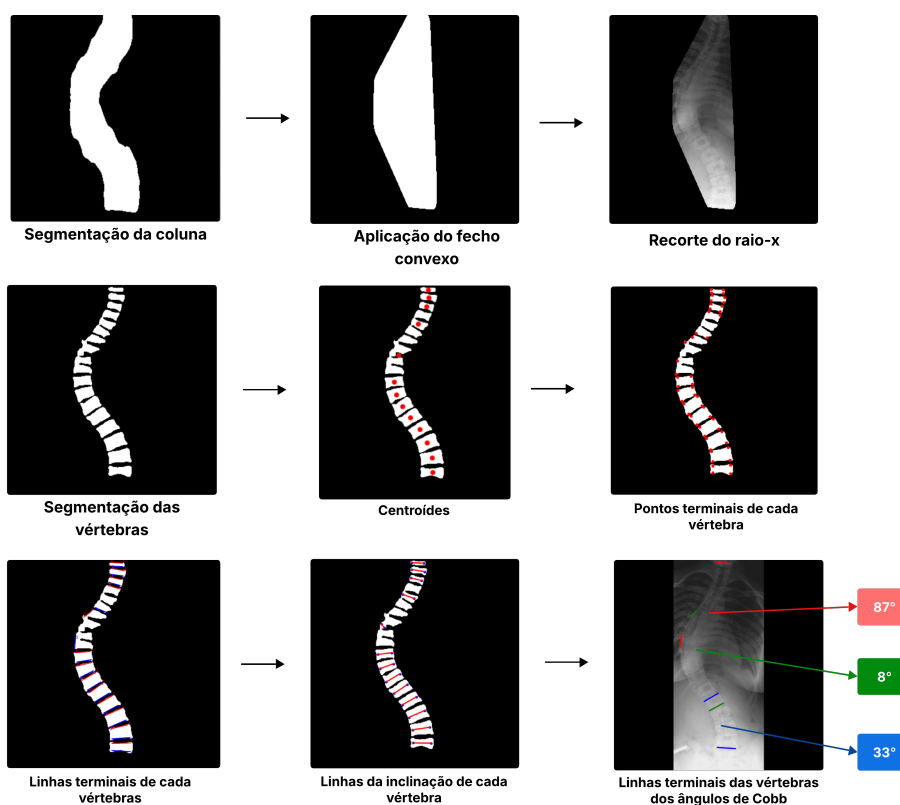
Fonte: Autoral.

Na Figura 34 apresenta-se o pior caso de estimação do ângulo de Cobb. A segmentação da coluna, após a aplicação do fecho convexo, identificou corretamente a região correspondente a coluna vertebral. Entretanto, na etapa de segmentação das vértebras, observa-se que algumas delas foram identificadas de forma unificada, como um único objeto.

Essa fusão compromete diretamente a estimação do ângulo de Cobb, pois impede o cálculo adequado da inclinação individual de cada vértebra, tratando vértebras adjacentes como se fossem uma única estrutura. Como consequência, a determinação dos centróides foi prejudicada, impactando também as etapas subsequentes do método.

Além disso, as vértebras unificadas encontram-se na região de uma das curvaturas da coluna, o que agrava o erro. Como os pontos terminais dessas vértebras não foram corretamente identificados, os dois primeiros ângulos que dependem diretamente dessas estruturas tiveram seus valores significativamente distorcidos, resultando em estimativas de 87°, 8° e 33°, enquanto as medições realizadas pelo especialista foram de 17°, 67° e 38°, respectivamente.

Figura 34 – Pior caso de estimação do ângulo de Cobb



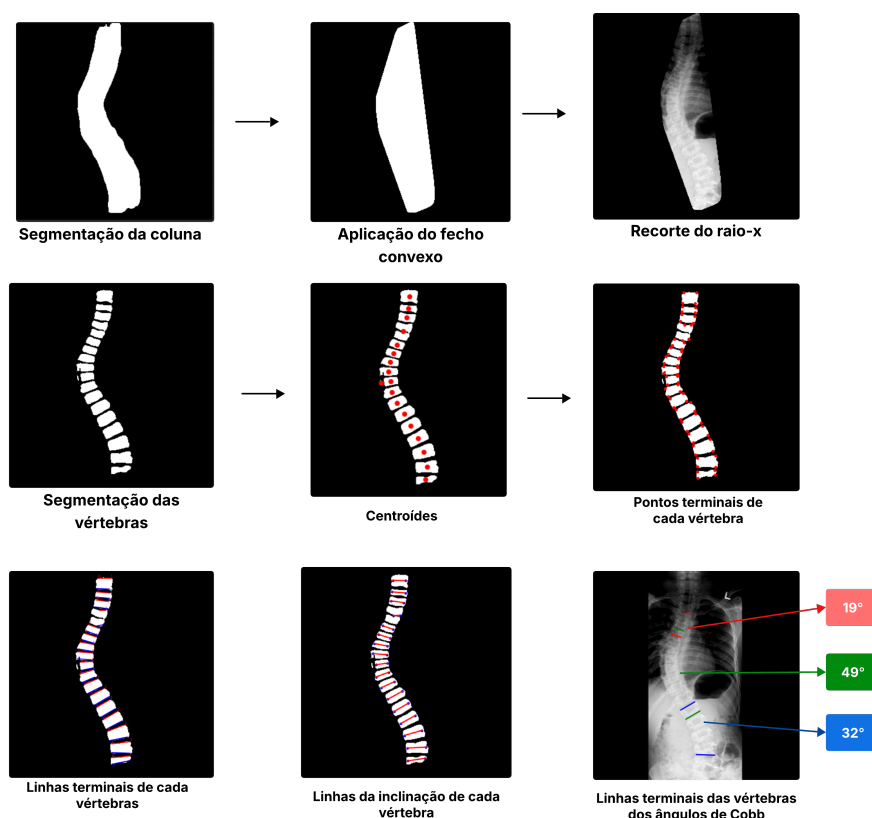
Fonte: Autoral.

Na Figura 35 apresenta-se o melhor caso de estimação do ângulo de Cobb. Observa-se que a extração da ROI foi realizada de forma adequada, delimitando corretamente a região da coluna vertebral. A segmentação das vértebras apresentou desempenho satisfatório, identificando-as individualmente e preservando suas características anatômicas, apenas uma pequena falha foi identificada na lateral de uma das vértebras, porém, isso não interferiu na estimação do ângulo de Cobb, pois esse tipo de falha é resolvido pelo método de estimação na etapa de encontrar os pontos terminais de cada vértebra.

Esse resultado refletiu diretamente na precisão do cálculo dos centróides, na deter-

minação dos pontos terminais das vértebras e nas etapas subsequentes do método. Como consequência, os valores estimados para o ângulo de Cobb mostraram-se praticamente coincidentes com aqueles obtidos pelo especialista. O método apresentou como resultado 19°, 49° e 32°, enquanto as medições realizadas pelo especialista foram de 19°, 49° e 33°, evidenciando elevada concordância entre as abordagens.

Figura 35 – Melhor caso de estimação do ângulo de Cobb



Fonte: Autoral.

A partir da análise dos resultados obtidos, considerou-se a inclusão de uma etapa adicional de pós-processamento com o objetivo de promover a separação das vértebras. Contudo, verificou-se que tal procedimento poderia comprometer o desempenho na maioria das imagens de teste, uma vez que o problema identificado ocorre apenas em casos específicos. Dessa forma, optou-se por não incorporar essa etapa ao método proposto, preservando o desempenho global do modelo.

Assim, embora os resultados obtidos para a medição do ângulo de Cobb não tenham superado o estado da arte, foram observados desempenhos satisfatórios tanto na segmentação da coluna quanto na segmentação das vértebras. Entretanto, esta última ainda demanda aprimoramentos, uma vez que sua precisão impacta diretamente a etapa de estimação do ângulo, influenciando o desempenho final do método.

## 6 Conclusão

A escoliose é uma deformidade da coluna vertebral que acomete indivíduos em diferentes faixas etárias. Seu diagnóstico pode ser realizado por meio de exames de raio-x da coluna vertebral, utilizando a medição do ângulo de Cobb. No entanto, essa medição é uma tarefa repetitiva e cansativa para o especialista, o que motivou o desenvolvimento de pesquisas voltadas a sua estimativa automática.

Este trabalho apresentou um método para a medição automática do ângulo de Cobb, dividido nas seguintes etapas: (1) pré-processamento das imagens de raio-x utilizadas como entrada para a rede de segmentação da coluna; (2) segmentação da coluna, etapa definida como extração da ROI; (3) pós-processamento para remoção de falsos positivos e aplicação do fecho convexo; (4) segmentação das vértebras a partir da ROI gerada; e (5) cálculo do ângulo de Cobb com base na máscara obtida pela segmentação das vértebras.

O método desenvolvido apresentou bom desempenho na segmentação da coluna, alcançando 92,14% de Dice com a adição de *attention gates* nas *skip-connections*, modificação proposta aplicada à Unet 3+, superando outros métodos empregados na mesma tarefa. Além disso, a extração da ROI contribuiu para a redução de falsos positivos observados quando a segmentação das vértebras era realizada sem essa etapa. Para a segmentação das vértebras, foi obtido um coeficiente Dice de 83,55%, valor próximo aos reportados em trabalhos relacionados, resultado obtido com a inclusão do bloco SE no *encoder* e do CBAM nas *skip-connections* da Unet 3+, correspondente a segunda modificação proposta nessa arquitetura. Entretanto, os resultados obtidos no cálculo do ângulo de Cobb indicam que ainda são necessários aprimoramentos, na qualidade da segmentação das vértebras, para que os resultados se tornem equivalentes aos apresentados na literatura especializada.

Vale ressaltar que, ao longo desta pesquisa, alguns experimentos adicionais não apresentados neste trabalho foram conduzidos na etapa de pré-processamento das imagens de raio-x. Entre eles, destacam-se a aplicação do algoritmo *Equalização adaptativa do Histograma com Contraste Limitado* (do inglês, *Contrast Limited Adaptive Histogram Equalization*) (CLAHE) e a especificação de histograma, ambos com o objetivo de melhorar o contraste das imagens. Essas técnicas foram aplicadas diretamente nas imagens originais; contudo, não resultaram em melhorias significativas no desempenho da segmentação da coluna, motivo pelo qual não foram incorporadas ao método proposto.

Adicionalmente, foram realizados testes voltados ao refinamento da segmentação das vértebras. Entre as abordagens avaliadas, destaca-se a aplicação de operações morfológicas, como erosão com o intuito de separar vértebras conectadas seguida de dilatação, visando restaurar a estrutura original das regiões segmentadas. Posteriormente, empregou-se o

algoritmo *Watershed* com o objetivo de separar componentes ainda aderidos.

Outra estratégia investigada consistiu na identificação de vértebras potencialmente unidas com base em suas propriedades geométricas, especialmente a área. Nesse contexto, regiões cuja área excedesse o valor médio esperado para uma vértebra foram consideradas como vértebras unidas. Para essas regiões, aplicaram-se operações morfológicas de erosão e dilatação, sendo o resultado reintegrado à máscara original das vértebras segmentadas. Entretanto, nenhuma das abordagens de refinamento propostas apresentou ganhos relevantes no desempenho final, não sendo, portanto, incorporadas ao método.

Como trabalhos futuros, sugere-se a investigação de estratégias para realizar o refinamento da segmentação das vértebras, como o uso de redes neurais dedicadas a essa tarefa, a exemplo da SegFormer, ou a adaptação de arquiteturas consolidadas, como a ResNet-50 e suas variações, para realizar uma segunda etapa de segmentação das vértebras. Outra possibilidade consiste na identificação de regiões com vértebras fundidas, seguida de uma ressegmentação local, utilizando como entrada apenas a região de interesse correspondente.

Além disso, acredita-se que a melhoria do contraste das imagens pode contribuir significativamente para o desempenho do método, uma vez que foi observado empiricamente que imagens com melhor contraste apresentaram maiores valores de Dice. Nesse sentido, recomenda-se a investigação de técnicas como correção gama e métodos baseados em Retinex. Por fim, as abordagens discutidas representam possíveis direções para aprimoramento dos resultados obtidos neste trabalho.

# Referências

- American Association of Neurological Surgeons. *Scoliosis*. 2024. Acessado em: 20 fev. 2026. Disponível em: <<https://www.aans.org/patients/conditions-treatments/scoliosis/>>. Citado na página 31.
- AMES, C. P. et al. *Cervical spine deformity surgery*. [S.l.]: Georg Thieme Verlag, 2019. Citado na página 17.
- ASSOCIATION, A. O. et al. *Foundations for Osteopathic Medicine*. [S.l.]: Lippincott Williams & Wilkins, 2003. Citado na página 35.
- BARBER, C. B.; DOBKIN, D. P.; HUHDANPAA, H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, Acm New York, NY, USA, v. 22, n. 4, p. 469–483, 1996. Citado na página 55.
- BARNES, R.; LEHMAN, C.; MULLA, D. Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, Elsevier, v. 62, p. 117–127, 2014. Citado na página 57.
- BARON, E. M.; VACCARO, A. R. *Operative Techniques: Spine Surgery E-Book*. [S.l.]: Elsevier Health Sciences, 2016. Citado na página 35.
- CAO, H. et al. Swin-unet: Unet-like pure transformer for medical image segmentation. In: SPRINGER. *European conference on computer vision*. [S.l.], 2022. p. 205–218. Citado 3 vezes nas páginas 54, 57 e 66.
- CARCI. *Goniômetro Grande*. 2026. Disponível em: <<https://www.carcioficial.com.br/goniometro-grande/>>. Citado na página 34.
- CATARINO, M. H. *Redes Neurais*. [S.l.]: Freitas Bastos, 2025. Citado na página 41.
- CHEN, L.-C. et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 801–818. Citado 4 vezes nas páginas 54, 57, 65 e 66.
- CREFITO-14. *Mês de conscientização da escoliose*. 2022. <<https://www.crefito14.org.br/noticias/noticias-mes-de-conscientizacao-da-escoliose-2022-06-01>>. Acesso em: 27 jan. 2026. Citado na página 17.
- CUI, J.-L. et al. Cobb angle measurement method of scoliosis based on u-net network. 2021. Citado 6 vezes nas páginas 25, 27, 28, 73, 74 e 75.
- DOLLÁR, P. et al. Fast feature pyramids for object detection. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 36, n. 8, p. 1532–1545, 2014. Citado na página 25.
- DU, K.-L.; SWAMY, M. N. *Neural networks and statistical learning*. [S.l.]: Springer Science & Business Media, 2013. Citado 2 vezes nas páginas 36 e 37.

- DUAN, K. et al. Centernet: Keypoint triplets for object detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2019. Citado na página 22.
- FEITOSA, G. L.; FRANÇA, F. J. R.; FISIOTERAPIA, C. de. Existe correlação entre a idade e a qualidade de vida de adultos com escoliose idiopática? 2023. Citado na página 17.
- FERREIRA, M. E. C. *Programa de intervenção conservadora em adolescentes com escoliose idiopática moderada e grave*. Tese (Doutorado) — Universidade de São Paulo, 2025. Citado na página 17.
- FLECK, L. et al. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, v. 1, n. 13, p. 47–57, 2016. Citado na página 35.
- FRAIWAN, M.; AUDAT, Z.; MANASREH, T. A dataset of scoliosis, spondylolisthesis, and normal vertebrae x-ray images. *Mendeley Data*, v. 1, p. 2022, 2022. Citado na página 22.
- GONZALEZ, R.; WOODS, R. Processamento digital de imagens. [sl]: Pearson prentice hall. *São Paulo*, 2010. Citado na página 27.
- GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009. Citado na página 54.
- HAWES, M. C.; O'BRIEN, J. P. The transformation of spinal curvature into spinal deformity: pathological processes and implications for treatment. *Scoliosis*, Springer, v. 1, n. 1, p. 3, 2006. Citado 2 vezes nas páginas 31 e 33.
- HE, K. et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2961–2969. Citado na página 22.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. Citado 3 vezes nas páginas 21, 22 e 23.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 65.
- HORNE, J. P.; FLANNERY, R.; USMAN, S. Adolescent idiopathic scoliosis: diagnosis and management. *American family physician*, v. 89, n. 3, p. 193–198, 2014. Citado na página 17.
- HORNG, M.-H. et al. Cobb angle measurement of spine from x-ray images using convolutional neural network. *Computational and mathematical methods in medicine*, Wiley Online Library, v. 2019, n. 1, p. 6357171, 2019. Citado na página 35.
- HOWARD, A. et al. Searching for mobilenetv3. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2019. p. 1314–1324. Citado na página 22.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 7132–7141. Citado 2 vezes nas páginas 47 e 48.

HUANG, H. et al. Unet 3+: A full-scale connected unet for medical image segmentation. In: IEEE. *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. [S.l.], 2020. p. 1055–1059. Citado 14 vezes nas páginas 26, 39, 44, 45, 46, 47, 52, 54, 57, 65, 66, 68, 70 e 72.

HUANG, X. et al. The comparison of convolutional neural networks and the manual measurement of cobb angle in adolescent idiopathic scoliosis. *Global Spine Journal*, SAGE Publications Sage CA: Los Angeles, CA, v. 14, n. 1, p. 159–168, 2024. Citado 4 vezes nas páginas 23, 24, 27 e 75.

KELLEHER, J. D. *Deep learning*. [S.l.]: MIT press, 2019. Citado 2 vezes nas páginas 35 e 36.

KIM, K. C. et al. Automation of spine curve assessment in frontal radiographs using deep learning of vertebral-tilt vector. *Ieee Access*, IEEE, v. 8, p. 84618–84630, 2020. Citado 5 vezes nas páginas 21, 24, 27, 28 e 75.

KONONENKO, I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, Elsevier, v. 23, n. 1, p. 89–109, 2001. Citado na página 18.

KOUMBOURLIS, A. C. Scoliosis and the respiratory system. *Paediatric respiratory reviews*, Elsevier, v. 7, n. 2, p. 152–160, 2006. Citado na página 35.

LACROIX, M. et al. Scoliosis. In: THIEME MEDICAL PUBLISHERS, INC. *Seminars in Musculoskeletal Radiology*. [S.l.], 2023. v. 27, n. 05, p. 529–544. Citado na página 30.

LIANG, Y. et al. Accurate cobb angle estimation on scoliosis x-ray images via deeply-coupled two-stage network with differentiable cropping and random perturbation. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 27, n. 3, p. 1488–1499, 2022. Citado 6 vezes nas páginas 25, 27, 28, 73, 74 e 75.

LIN, T.-Y. et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2980–2988. Citado na página 39.

LIU, S. et al. Automatic estimation of cobb angle based on wavelet convolution. *Biomedical Signal Processing and Control*, v. 113, p. 109092, 2026. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1746809425016039>>. Citado 4 vezes nas páginas 26, 27, 74 e 75.

LIU, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 10012–10022. Citado na página 66.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 3431–3440. Citado na página 43.

- LU, Q. et al. Superior performance of a center-point ai model over vfdnet in automated cobb angle estimation for scoliosis assessment. *European Spine Journal*, Springer, v. 33, n. 12, p. 4710–4719, 2024. Citado 4 vezes nas páginas 23, 24, 27 e 75.
- LUK, K. D. et al. Clinical effectiveness of school screening for adolescent idiopathic scoliosis: a large population-based retrospective cohort study. *Spine*, LWW, v. 35, n. 17, p. 1607–1614, 2010. Citado na página 18.
- MAEDA, Y. et al. Automatic measurement of the cobb angle for adolescent idiopathic scoliosis using convolutional neural network. *Scientific reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 14576, 2023. Citado 5 vezes nas páginas 22, 24, 27, 28 e 75.
- MAGUIRE, D. *Darraghmaguire*. 2020. <<https://github.com/darraghmaguire/automatic-scoliosis-assessment>>. [Online; accessed 18-Aug-2023]. Citado 10 vezes nas páginas 25, 27, 28, 52, 58, 63, 64, 73, 74 e 75.
- MÁTTYUS, G.; LUO, W.; URTASUN, R. Deeproadmapper: Extracting road topology from aerial images. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 3438–3446. Citado 2 vezes nas páginas 39 e 40.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 36.
- MINGHELLI, B. Rastreio escolar: a importância na detecção precoce de posturas escolióticas em adolescentes das escolas de silves, algarve. *Rev. Port. Sau. Pub*, v. 26, n. 2, p. 61–68, 2008. Citado 2 vezes nas páginas 32 e 33.
- NEUWIRTH, M.; OSBORN, K. *The Scoliosis Sourcebook*. McGraw-Hill Education, 2001. ISBN 9780071392013. Disponível em: <<https://books.google.com.br/books?id=kxidyS7ZlgYC>>. Citado 5 vezes nas páginas 29, 30, 31, 32 e 34.
- OKTAY, O. et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. Citado 2 vezes nas páginas 46 e 47.
- ORTOPEDIA, E. Núcleo de. *Escoliose: Quais são os principais sinais clínicos?* 2026. Disponível em: <<https://nucleodeortopedia.com.br/escoliose-quais-sao-os-principais-sinais-clinicos/>>. Citado na página 31.
- PEDRYCZ, W.; CHEN, S.-M. *Deep learning: Concepts and architectures*. [S.l.]: Springer, 2020. v. 866. Citado 2 vezes nas páginas 40 e 41.
- PENG, L. et al. Cobb angle measurement based on spine segmentation using att unet 3+. *IEEE Access*, IEEE, v. 12, p. 68545–68555, 2024. Citado 5 vezes nas páginas 26, 27, 28, 74 e 75.
- QIN, X. et al. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, Elsevier, v. 106, p. 107404, 2020. Citado 3 vezes nas páginas 54, 57 e 66.
- RICHEMS, J. G.; LEE, C. M.; JOHRI, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nature communications*, Nature Publishing Group UK London, v. 11, n. 1, p. 3923, 2020. Citado na página 18.

- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241. Citado 10 vezes nas páginas 22, 25, 43, 44, 45, 54, 57, 66, 72 e 73.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 36.
- SALEHI, S. S. M.; ERDOGMUS, D.; GHOLIPOUR, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: SPRINGER. *International workshop on machine learning in medical imaging*. [S.l.], 2017. p. 379–387. Citado na página 25.
- SANTOS, V. S. d. *Coluna vertebral*. 2026. Disponível em: <<https://brasilescola.uol.com.br/biologia/coluna-vertebral.htm>>. Citado na página 30.
- SARDJONO, T. A. et al. Automatic cobb angle determination from radiographic images. *Spine*, LWW, v. 38, n. 20, p. E1256–E1262, 2013. Citado na página 18.
- SULTANA, A. et al. Leveraging artificial intelligence in neuroimaging for enhanced brain health diagnosis. *Revista de Inteligencia Artificial en Medicina*, v. 14, n. 1, p. 1217–1235, 2023. Citado na página 18.
- SURI, A. et al. Conquering the cobb angle: a deep learning algorithm for automated, hardware-invariant measurement of cobb angle on radiographs in patients with scoliosis. *Radiology: Artificial Intelligence*, Radiological Society of North America, v. 5, n. 4, p. e220158, 2023. Citado 4 vezes nas páginas 21, 24, 27 e 75.
- TORTORA, G.; NIELSEN, M. *Principles of Human Anatomy*. Wiley, 2017. ISBN 9781119444466. Disponível em: <<https://books.google.com.br/books?id=e0rcDwAAQBAJ>>. Citado na página 29.
- TU, Y. et al. Automatic measurement algorithm of scoliosis cobb angle based on deep learning. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2019. v. 1187, n. 4, p. 042100. Citado 5 vezes nas páginas 25, 27, 28, 74 e 75.
- VIALLE, L. *AOSpine Masters Series, Volume 9: Pediatric Spinal Deformities*. [S.l.]: Georg Thieme Verlag, 2017. Citado na página 35.
- WANG, J. et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 43, n. 10, p. 3349–3364, 2020. Citado na página 23.
- WANG, X. et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2097–2106. Citado na página 25.
- WANG, Z.; SIMONCELLI, E. P.; BOVIK, A. C. Multiscale structural similarity for image quality assessment. In: IEEE. *The thirty-seventh asilomar conference on signals, systems & computers, 2003*. [S.l.], 2003. v. 2, p. 1398–1402. Citado na página 39.

- WOO, S. et al. Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*. [S.l.: s.n.], 2018. p. 3–19. Citado 3 vezes nas páginas 26, 48 e 49.
- WU, H. et al. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2017. p. 127–135. Citado 5 vezes nas páginas 23, 25, 26, 27 e 64.
- YANG, J.; WANG, J.; MENG, M. Q.-H. A landmark-aware network for automated cobb angle estimation using x-ray images. *arXiv preprint arXiv:2405.19645*, 2024. Citado 4 vezes nas páginas 23, 24, 27 e 75.
- YI, J. et al. Vertebra-focused landmark detection for scoliosis assessment. In: IEEE. *2020 IEEE 17th international symposium on biomedical imaging (ISBI)*. [S.l.], 2020. p. 736–740. Citado 5 vezes nas páginas 21, 23, 24, 27 e 75.
- YUNES, R. M. *Escoliose: saiba tudo sobre ela*. 2026. Disponível em: <<https://escoliose.dryunes.com/o-que-e-escoliose.php>>. Citado 2 vezes nas páginas 33 e 34.
- ZHAO, Y. et al. Automatic cobb angle measurement method based on vertebra segmentation by deep learning. *Medical & Biological Engineering & Computing*, Springer, v. 60, n. 8, p. 2257–2269, 2022. Citado 5 vezes nas páginas 26, 27, 73, 74 e 75.
- ZHOU, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In: SPRINGER. *International workshop on deep learning in medical image analysis*. [S.l.], 2018. p. 3–11. Citado 3 vezes nas páginas 54, 57 e 66.
- ZHOU, Z.; ZHU, J.; YAO, C. Vertebral center points locating and cobb angle measurement based on deep learning. *Applied Sciences*, MDPI, v. 13, n. 6, p. 3817, 2023. Citado 4 vezes nas páginas 22, 24, 27 e 75.