



**UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE TECNOLOGIA  
ENGENHARIA DE TELECOMUNICAÇÕES**

**RODRIGO ALVES DE SOUZA**

**SEGMENTAÇÃO SEMÂNTICA EM IMAGENS CARDÍACAS DE  
RESSONÂNCIA MAGNÉTICA USANDO REDE NEURAL  
CONVOLUCIONAL E MODELO U-NET**

**LIMEIRA  
2024**

RODRIGO ALVES DE SOUZA

**Segmentação semântica em imagens cardíacas de ressonância  
magnética usando rede neural convolucional e modelo U-Net**

Monografia apresentada ao curso de Engenharia de Telecomunicações da Faculdade de Tecnologia da Universidade Estadual de Campinas como requisito para obtenção do título de bacharelado em Engenharia de Telecomunicações.  
Orientador: Prof. Dr. Rangel Arthur

LIMEIRA  
2024

## Dedicatória

Dedico esse trabalho à minha família, em especial meus pais e minha irmã, que sempre me ofereceram toda confiança e apoio incondicional para tornar isso possível.

Dedico à Beatriz, meu amor, com a qual eu tenho a sorte de partilhar a vida.

Dedico a Juliano Ignácio Pedro (*in memoriam*), colega e amigo, que me ensina todo dia a ser alguém melhor.

## Agradecimentos

Agradeço a minha família por toda força e por todo apoio, em especial meus pais Edson e Jucilei, e a minha irmã Gabriela, por estarem comigo sempre, em todos os momentos.

Agradeço a minha namorada, Beatriz, por me acompanhar e me apoiar em toda essa jornada, e por ser fortaleza em todos os momentos de dificuldade.

Agradeço ao meu professor orientador Rangel Arthur, pela oportunidade e por me guiar durante a realização desse trabalho, desde a escolha do tema até a resolução deste trabalho.

Agradeço aos meus amigos, em especial Matheus, Miriã, Marcos e Roger, o “time dos sonhos”, grandes amigos que tive a oportunidade de conhecer nessa jornada. Os desafios da graduação não seriam os mesmo sem nossos momentos de estudo e descontração.

## Resumo

A análise clínica é uma fase fundamental e crítica para definição de um diagnóstico ágil e preciso. Quando observado por áreas médicas de alto risco, como a cardiologia, a identificação de doenças é crucial, tendo em vista que as doenças cardiovasculares são as principais causas de morte em todo mundo. Dos exames cardiovasculares, os mais utilizados para avaliação do quadro médico são os de ressonância magnética. Dessa forma, a segmentação das regiões do coração é geralmente realizada por um profissional, levando considerável tempo e permitindo a possibilidade de falhas no diagnóstico. Este trabalho teve como objetivo a construção e avaliação de uma rede neural para segmentação automática das câmaras cardíacas, utilizando métodos de segmentação semântica e atingindo os valores de acurácia de 0,8567 e 0,8517, analisando os efeitos de técnicas de pré-processamento de imagem, realce de imagens e *data augmentation* no desempenho global da rede.

Palavras-chave: Segmentação de imagens, imagens médicas, aprendizado profundo, redes neurais convolucionais, U-Net

# Abstract

Clinical analysis is a fundamental and critical phase for defining an agile and precise diagnosis. When observed in high-risk medical areas, such as cardiology, disease identification is crucial, considering that cardiovascular diseases are the main causes of death worldwide. Among cardiovascular tests, the most commonly used for medical evaluation are magnetic resonance tests. Therefore, the segmentation of heart regions is generally performed by a professional, taking considerable time and allowing the possibility of diagnostic failures. This work aimed to build and evaluate a neural network for automatic segmentation of cardiac chambers, using semantic segmentation methods and achieving accuracy values of 0,8567 and 0,8517, analyzing the effects of image preprocessing techniques, image enhancement, and data augmentation on the global performance of the network.

Keywords: Image segmentation, medical images, deep learning, convolutional neural networks, U-Net

# Sumário

<b>1. Introdução</b>	10
1.1. Motivação	10
1.2. Problemática	11
1.3. Objetivos	11
<b>2. Fundamentação Teórica</b>	12
2.1. Inteligência Artificial	12
2.2. <i>Machine Learning</i>	12
2.3. Redes Neurais Artificiais	14
2.3.1. Redes Neurais Convolucionais	16
2.4. Arquitetura de CNN	18
2.4.1. U-Net	18
2.5. Segmentação de Imagem	19
2.5.1. Segmentação Semântica	19
2.6. Métricas de Avaliação de Desempenho	20
<b>3. Metodologia</b>	21
3.1. Materiais	21
3.1.1. Tipo de Linguagem	21
3.1.2. <i>Framework</i>	22
3.1.3. <i>Dataset</i>	22
3.1.4. Ambiente de Desenvolvimento	23
<b>4. Desenvolvimento e Análise de Resultados</b>	24
4.1. Tratamento de Imagem	24
4.1.1. Equalização de Histograma	24
4.1.2. <i>Data Augmentation</i>	25
4.1.3. Criação de <i>Ground-Truth</i>	26
4.2. Desenvolvimento do Modelo	26
4.3. Treinamento do Modelo	27
4.4. Análise dos Resultados	28
<b>5. Considerações Finais</b>	31
<b>6. Bibliografia</b>	32

# Lista de Figuras

1.1	Número de artigos sobre ML em <i>Smart Cities</i> .....	10
2.1	Tipos de Aprendizado .....	13
2.2	Representação de um <i>Perceptron</i> .....	14
2.3	Exemplo de uma camada de convolução .....	17
2.4	Exemplo de uma camada de <i>pooling</i> .....	17
2.5	Arquitetura da rede U-Net .....	18
2.6	Exemplo de segmentação semântica .....	29
3.1	Orientações dos planos cardíacos em relação ao coração .....	23
4.1	Histograma Imagem original da base de imagens <i>Sunnybrook</i> e seu respectivo histograma .....	25
4.2	Pré-processamento (a) Imagem original da base de imagens <i>Sunnybrook</i> (b) Imagem após realização da equalização de histograma .....	25
4.3	Definição de <i>Ground-Truth</i> : Imagem pré-processada e sua respectiva máscara .....	26
4.4	<i>Dataset</i> original - Decaimento do <i>val loss</i> por épocas no treinamento....	28
4.5	Predição da rede sem <i>Data Augmentation</i> .....	29
4.6	<i>Dataset</i> original + <i>Data Augmentation</i> - Decaimento do <i>val loss</i> por épocas no treinamento .....	29
4.7	Predição da rede com <i>Data Augmentation</i> .....	30



## Lista de Abreviaturas e Siglas

<b>IA</b>	<i>Inteligência Artificial</i>
<b>ML</b>	<i>Machine Learning</i>
<b>DL</b>	<i>Deep Learning</i>
<b>PDS</b>	<i>Processamento Digital de Sinais</i>
<b>CNN</b>	<i>Convolutional Neural Network</i>
<b>ANN</b>	<i>Artificial Neural Network</i>
<b>DICOM</b>	<i>Digital Imaging and Communications in Medicine.</i>
<b>PLN</b>	<i>Processamento de Linguagem Natural</i>
<b>RNA</b>	<i>Redes Neurais Artificiais</i>
<b>PNG</b>	<i>Portable Network Graphics</i>
<b>FCN</b>	<i>Fully Convolutional Network</i>
<b>ReLU</b>	<i>Rectified Linear Units</i>
<b>CPU</b>	<i>Central Processing Unit</i>
<b>GPU</b>	<i>Graphic Processing Unit</i>
<b>TPU</b>	<i>Tensor Processing Unit</i>

# 1. Introdução

## 1.1. Motivação

Com o advento de novos avanços nos estudos de IA, novas áreas de aplicação e desenvolvimento são exploradas de maneira a utilizar esta inovação. Aliado a chegada do 5G, os estudos na área de Cidades Inteligentes (*Smart Cities*) tiveram um aumento substancial desde 2017, utilizando do Aprendizado de Máquina para dinamização de processos, previsão e predição de cenários urbanos, como pode ser observado na Figura 1.1.

Embora a fundamentação teórica de *Smart City* esteja fundamentada na literatura desde o início da década de 90, os recentes avanços e pesquisas em métodos e algoritmos de ML, aliado a chegada do 5G, possibilitaram o resgate de suas ideias, trazendo para holofote um grande leque de possibilidades de áreas de pesquisa e inovação [1].

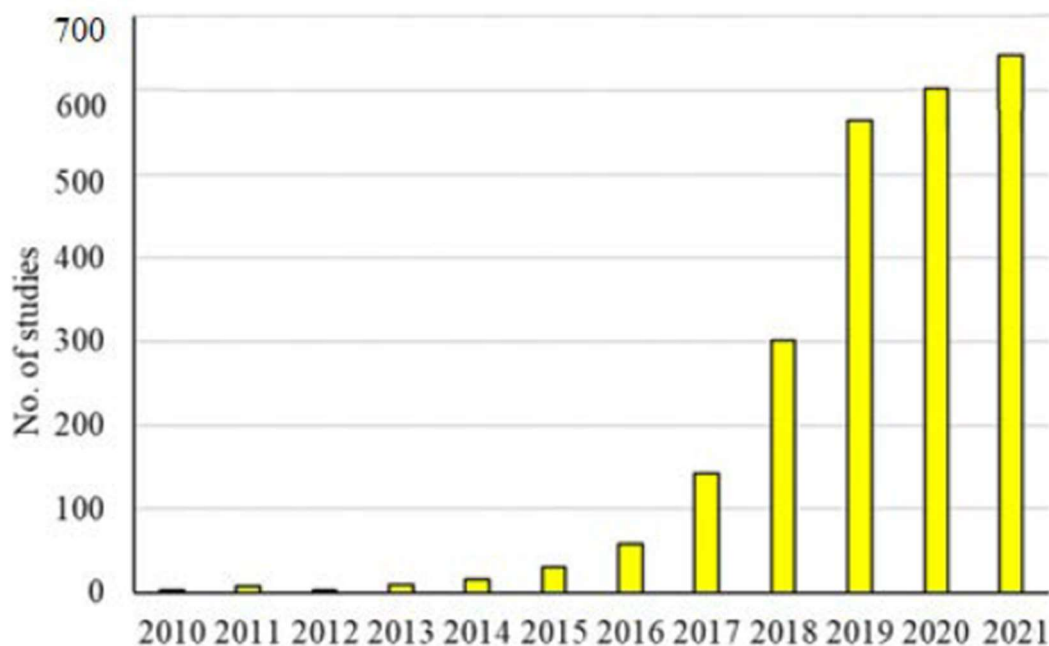


Figura 1.1: Número de artigos sobre ML em Smart Cities - a partir de 2013, houve um aumento vertiginoso no número de publicações, incentivada pelos avanços tecnológicos [2]

Amparado pelos últimos avanços tecnológicos e a busca pelo desenvolvimento sustentável nas grandes cidades, com gestão consciente de recursos, avaliação de demandas regionais e construção de novos modelos de infraestrutura urbana, o entendimento e elaboração de projetos de *Smart Cities* torna-se cruciais no

panorama de médio e longo prazo, onde buscam-se a integração dos modelos metropolitanos existentes com novas tecnologias para melhorar a qualidade de vida populacional.

## 1.2. Problemática

Das áreas exploradas em *Smart Cities*, uma das mais críticas é a Assistência Médica Inteligente (*Smart Healthcare*), pois lida diretamente com os indivíduos em situação de fragilidade, e dessa forma, a precisão e agilidade para avaliação e oferecimento de um diagnóstico médico é crucial para o ambiente hospitalar [3].

Utilizando imagens de exames médicos, fazendo uso de técnicas de PDS e DL [4], redes neurais são capazes de comparar exames médicos, como imagens pulmonares ou cardíacas, com um banco de imagens vasto, e após certo aprendizado, identificar e caracterizar alterações em exames, de maneira a auxiliar e acelerar o diagnóstico médico.

Deste modo, a análise de imagens cardíacas tem um papel fundamental para identificação e avaliação de doenças e mazelas críticas, possuindo um elevado grau de importância nas etapas de triagem médica, podendo definir em certos casos a recuperação ou regressão do tratamento do paciente [5]. Das partes constitutivas do sistema cardíaco, o coração destaca-se, não só por sua função, mas como sua ação e/ou reação a indicadores externos, como exemplo, a dilatação arterial causada pela administração de certos remédios. Isolar em imagens suas regiões de interesse, além de acelerar o processo clínico, permite a identificação de doenças, como o infarto do miocárdio e a cardiomegalia.

## 1.3. Objetivos

O presente trabalho teve como intuito o desenvolvimento de uma CNN para segmentação de imagens de tomografia cardíaca computadorizada, realizando pré-tratamento e normalização a um conjunto de aprendizado e um conjunto teste, definindo um ciclo de iteração e buscando otimizar o sistema utilizando *backpropagation* com crescimento do coeficiente *Dice*.

Neste modelo, o foco para segmentação é definido em duas regiões: o ventrículo esquerdo, localizado abaixo do átrio esquerdo e responsável por bombear o sangue para todo o corpo (exceto o pulmão) e; o miocárdio, parte da parede do coração, formado por células chamadas cardiomiócitos [6].

## 2. Fundamentação Teórica

### 2.1. Inteligência Artificial

A IA é um amplo ramo da ciência da computação, notabilizada pelo uso e/ou aplicação de tecnologias com o intuito de criar sistemas, sejam eles implementados em máquinas ou computadores, que sejam capazes de reproduzir funções cognitivas associadas à inteligência humana, como a capacidade de ler, compreender e responder linguagem falada ou escrita, analisar dados ou observar imagens.

Possuindo uma ampla gama de aplicações, a Inteligência Artificial pode ser subdividida em várias subáreas, como ML, ANNs, Visão Computacional, PLN e Sistemas Especialistas [7]. Neste trabalho, iremos nos concentrar nas três primeiras áreas.

### 2.2. Machine Learning

O Aprendizado de Máquina tem como objetivo o desenvolvimento de algoritmos e técnicas que permitam computadores, a partir da análise prévia de uma base de dados, tomar decisões sem serem explicitamente programados para cada tarefa específica. Dessa forma, algumas técnicas de aprendizado são demonstradas na Figura 2.1.

Das técnicas de aprendizado [7], podemos destacar:

- **Aprendizado supervisionado:** Aprendizado a partir de informações rotuladas ou pré-ordenadas para predição de resultados para dados imprevistos;
- **Aprendizado não supervisionado:** Aprendizado que utiliza estatísticas descritivas para examinar padrões e relações que ocorrem naturalmente entre os dados;

- **Aprendizado semi-supervisionado:** Aprendizado que tem como circunstâncias de treinamento um apanhado de informação rotulada em meio a uma grande quantidade de informação não rotulada.
- **Aprendizado por reforço:** Aprendizado que ocorre entre um aprendizado supervisionado e não supervisionado, onde o algoritmo é tratado como “agente”, permitindo aprender com o ambiente para criar rotas de ação e comportamento inteligente. Este aprendiz interage por meio de ações, e recebe “recompensas” para incentivar e criar uma política de bom comportamento.

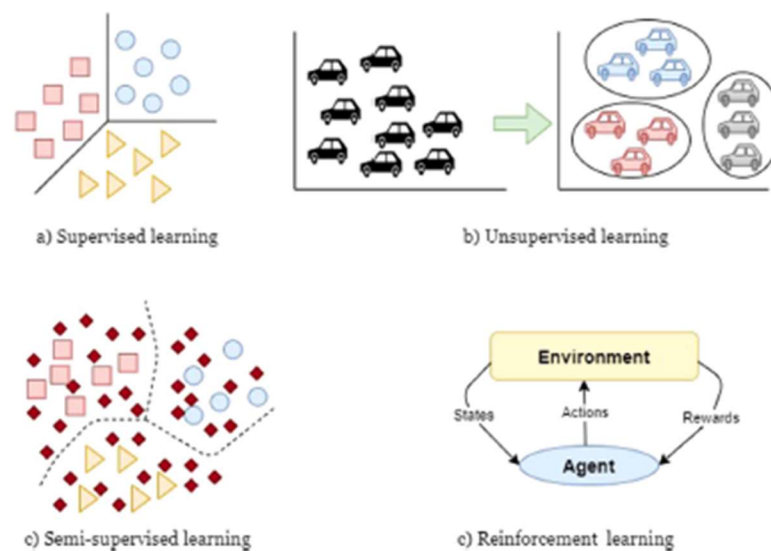


Figura 2.1: Tipos de Aprendizado - Supervisionado, Não supervisionado, Semi-supervisionado e reforço, da esquerda para direita, de cima para baixo [7]

A capacidade de aprendizado deste sistema está intrinsicamente ligada a qualidade dos dados fornecidos, tanto na etapa de aprendizado, quanto na etapa de teste. O Aprendizado de Máquina é capaz de resolver diversos problemas, como categorização, detecção de movimentos, ou segmentação.

Contudo, conforme os problemas se aproximam da realidade, a natureza das soluções deixa de tomar nuances lineares e passam a ser relacionadas às respostas associadas a funções cada vez mais complexas. Desta forma, precisamos compreender como as redes aprendem.

## 2.3. Redes Neurais Artificiais

RNAs são modelos computacionais inspirados no sistema nervoso biológico, especificamente no funcionamento do cérebro humano. Elas são compostas por unidades de processamento chamadas neurônios artificiais ou *perceptrons*, organizados em camadas interconectadas.

Um *Perceptron* ou neurônio artificial é a unidade básica de uma rede neural simples, um modelo computacional capaz de simular o funcionamento do neurônio biológico [8]. Sua representação básica é definida na Figura 2.2.

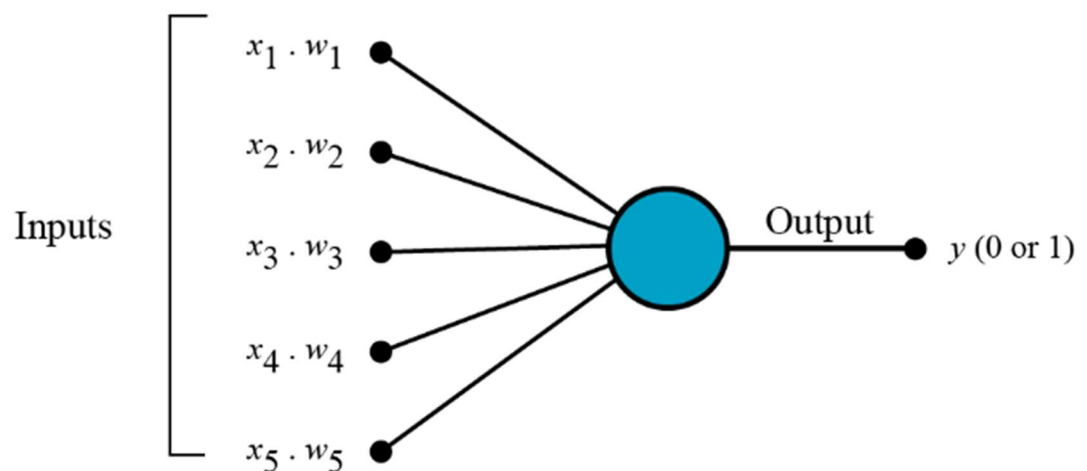


Figura 2.2: Representação de um *Perceptron* - com 5 entradas lineares  $X_N$  associadas a 5 pesos  $W_N$ . Neste caso,  $Y_N$  é o somatório das entradas ponderadas após a função de ativação [9]

Um *Perceptron* é formado por:

**Entradas Ponderadas:** Cada entrada para o *perceptron* é multiplicada por um peso correspondente. Esses pesos representam a importância relativa de cada entrada para o neurônio. Os pesos são ajustados durante o treinamento para que o *perceptron* possa aprender a fazer previsões ou classificações mais precisas.

**Soma Ponderada:** As entradas ponderadas são somadas juntas com um viés (ou bias), que é um parâmetro adicional adicionado à soma. Essa soma ponderada representa a ativação líquida do *perceptron*.

**Função de Ativação:** A soma ponderada é então passada por uma função de ativação, que determina a saída do *perceptron* com base nesse

valor. A função de ativação introduz não-linearidades na saída do *perceptron*, permitindo que ele aprenda a representar relações mais complexas nos dados.

**Saída:** A saída do *perceptron* é o resultado da aplicação da função de ativação à soma ponderada das entradas.

Devido a função de ativação introduzir não linearidades a saída de um *Perceptron*, a associação de *n-Perceptrons* forma uma rede interconectada, permitindo que *Perceptrons* de camadas posteriores sejam capazes de tomar decisões em um nível mais complexo e mais abstrato que os *Perceptrons* de primeira camada [10].

Uma rede neural simples é formada por três camadas: **Camada de Entrada**, onde são incluídos os dados a serem ponderados, **Camadas Escondidas**, onde estes dados são associados entre si de forma a extrair o máximo de características deles, e **Camadas de Saída**, onde são observados os resultados da análise a partir do objetivo definido.

Uma rede Neural aprende seguindo um processo iterativo de treinamento, definido nas seguintes etapas:

**Inicialização dos Pesos:** Os pesos das conexões entre os neurônios são inicializados com valores aleatórios ou com algum valor pré-definido.

**Propagação Direta (*Forward Propagation*):** Durante a propagação direta, os dados de entrada são passados através da rede neural, camada por camada, até que a saída seja produzida. Cada neurônio na rede passa o resultado por uma função de ativação e transmite o resultado para os neurônios na próxima camada.

**Cálculo da Função de Perda:** Após a propagação direta, a saída da rede é comparada com os rótulos verdadeiros dos dados de treinamento para calcular a função de perda.

**Retropropagação do Erro (*Backpropagation*):** Na retropropagação do erro, os gradientes da função de perda em relação aos pesos da rede são calculados usando o algoritmo de diferenciação automática. Os pesos são

então atualizados na direção oposta ao gradiente, usando um algoritmo de otimização.

**Atualização dos Pesos:** Com base nos gradientes calculados na retropropagação do erro, os pesos das conexões entre os neurônios são atualizados para reduzir a função de perda. A magnitude da atualização dos pesos é determinada pela taxa de aprendizado, que controla a velocidade com que a rede neural converge para uma solução.

**Época:** Os passos de propagação direta, cálculo da função de perda, retropropagação do erro e atualização dos pesos são repetidos iterativamente para todo o conjunto de dados de treinamento, totalizando 1 época. O número de épocas pode variar dependendo da complexidade do problema.

Redes neurais podem ser definidas em: **Rasas (ML)**, caso possuam até duas camadas escondidas e; **Profundas (DP)**, caso possuam mais que duas camadas escondidas [11]. Redes Neurais simples conseguem resolver uma ampla gama de problemas, mas encontram barreiras quando lidamos com problemas que envolvam uma grande quantidade de dados, como problemas envolvendo Visão Computacional, onde a dimensionalidade, quantidade de imagens e complexidade dessas matrizes necessitaria de um poder computacional massivo.

### 2.3.1. Redes Neurais Convolucionais

CNNs surgem como uma maneira de resolver os problemas de processamento de imagem observados em Redes Neurais Simples [12]. Caracterizadas como Redes Profundas, este tipo de rede é definido em:

**Camadas Convolucionais:** A característica central das CNNs são suas camadas convolucionais. Nessas camadas, *kernels* são aplicados à entrada usando convolução. Esses *kernels* agem como filtros, detectando padrões simples, como bordas, texturas e gradientes de cores, em diferentes regiões da imagem. Durante o treinamento, os filtros são ajustados automaticamente para aprender características relevantes da imagem, como demonstra a Figura 2.3.



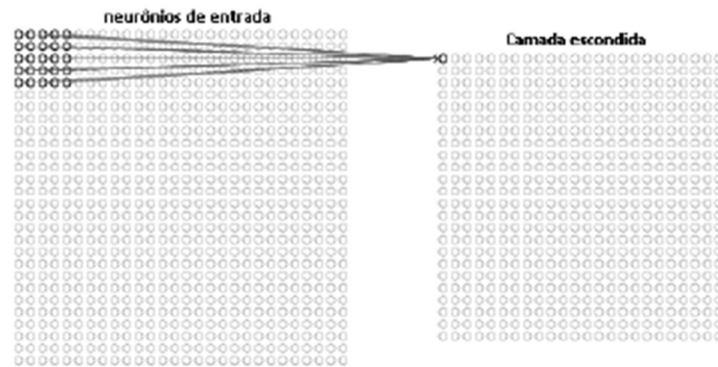


Figura 2.3: Exemplo de uma camada de convolução: Neste caso, em cada posição do filtro 5X5 aplicado a imagem original será calculado uma soma ponderada, gerando um único valor na matriz de saída (mapa de ativação). [13]

**Camadas de Pooling (ou Subamostragem):** Após as camadas convolucionais, as CNNs geralmente incluem camadas de *pooling* para reduzir a dimensionalidade dos dados e extrair características mais importantes. O *pooling* envolve a redução do tamanho da entrada, agrupando regiões adjacentes e selecionando um valor representativo, como o máximo (*max pooling*) ou a média (*average pooling*). Isso ajuda a tornar a representação dos dados mais compacta e eficiente. Um exemplo desse processo é mostrado na Figura 2.4.

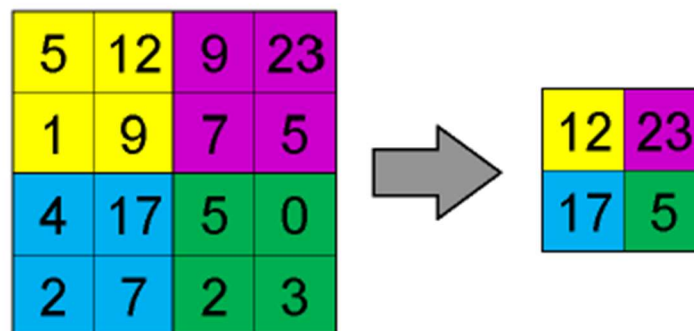


Figura 2.4: Exemplo de uma camada de *pooling* - na imagem, é representado o método de *max pooling* com deslocamento(*stride*) de 2, onde o maior valor representativo representa cada filtro 2X2 no mapa redimensionado [14]

**Camadas Totalmente Conectadas:** Após várias camadas convolucionais e de *pooling*, as CNNs geralmente têm uma ou mais camadas totalmente conectadas, semelhantes às redes neurais tradicionais. Essas camadas são usadas para combinar as características extraídas anteriormente e produzir uma saída final.

Normalmente, a saída é passada por uma função de ativação, que produz probabilidades para cada saída.

## 2.4. Arquitetura de CNN

A arquitetura da rede neural representa a forma como os neurônios artificiais de uma rede são agrupados.

### 2.4.1. U-Net

Desenvolvida por Ronneberger et al. [15] para segmentação de imagens biomédicas no *Computer Science Department and BIOS Centre for Biological Signalling Studies* da *University of Freiburg* na Alemanha, a rede foi baseada numa FCN proposta por Long et al. [16], onde as camadas totalmente conectadas são substituídas por camadas de convolução para permitir a geração de mapas de características para cada imagem.

Consiste em uma rede com dois caminhos: um de contração e outro de expansão, cuja simetria define o formato em U observado na Figura 2.5. Durante o caminho de contração, a rede se comporta como rede convolucional típica, seguida sequencialmente por convolução, uma função de ativação ReLU e um operador de *max pooling*, alternando entre camadas de convolução e *pooling*.

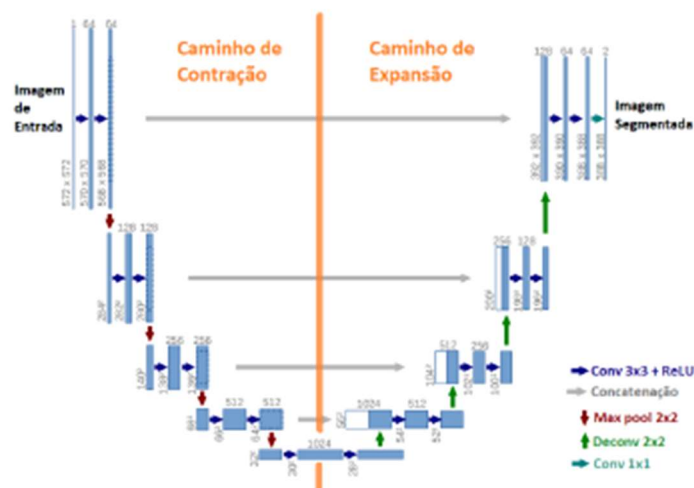


Figura 2.5: Arquitetura da rede U-Net: Composta por um caminho de contração e um de expansão, a rede ajusta os pesos dos *kernels* na contração e incrementa as *features* na fase de expansão. [15]

Reduzindo as informações espaciais, a rede incrementa as *features* durante o caminho de expansão, combinando-as para adquirir features de alta resolução, características e padrões das imagens.

## 2.5. Segmentação de Imagem

A segmentação de imagem é um processo em visão computacional cujo objetivo consiste em, por meio de diferentes técnicas [17], simplificar ou mudar a representação de uma imagem, como forma a torná-la mais significativa e ou mais fácil de se analisar.

### 2.5.1. Segmentação Semântica

A segmentação semântica é uma subárea da segmentação de imagem. Sua técnica consiste na atribuição de rótulos de classe para cada pixel de uma imagem, de modo a caracterizar os pixels pertencentes a uma determinada classe como de mesmo rótulo. Um exemplo é apresentado na Figura 2.6.



Figura 2.6: Exemplo de segmentação semântica: Regiões com características semelhantes são definidas com o mesmo rótulo [18]

Utilizando as imagens de treino, uma rede neural é capaz de identificar padrões nas imagens e realizar discretizações, de forma a separar as diferentes regiões de uma imagem. A qualidade da segmentação depende de vários fatores, como

quantidade de épocas de treinamento, tamanho do *dataset* de treinamento e principalmente a qualidade das imagens fornecidas ao treinamento.

## 2.6. Métricas de Avaliação de Desempenho

Métricas de Avaliação de Desempenho são critérios utilizados para comparar modelos de segmentação, de forma avaliar aspectos específicos do modelo. A principal métrica utilizada é a matriz de confusão, sendo as demais métricas combinações de seus valores [19].

A matriz de confusão é composta pela comparação dos *pixels* gerados pelo método de segmentação e os *pixels* do *ground-truth*. Dessa forma, pode-se obter 4 possíveis classificações:

- Verdadeiro Positivo (VP): Pixel pertencente a classe positiva foi classificada corretamente.
- Falso Positivo (FP): Pixel pertencente a classe positiva foi classificada incorretamente.
- Verdadeiro Negativo (VN): Pixel pertencente a classe negativa foi classificada corretamente.
- Falso Negativo (FN): Pixel pertencente a classe negativa foi classificada incorretamente.

Assim, obtemos as seguintes métricas:

- Acurácia (*Accuracy*): Taxa de acerto do modelo, ou seja, quantidade de segmentos classificados corretamente.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

- Precisão: Quantidade de segmentos da classe positiva (VP) que realmente pertencem a classe positiva (VP e FP).

$$Precisão = \frac{VP}{VP + FP}$$

- Revocação: Quantidade de elementos que pertencem a classe positiva (VP) que foram classificados corretamente (VP e FN).

$$Revocação = \frac{VP}{VP + FN}$$

- Coeficiente *Dice*: Métrica para medir a similaridade entre dois objetos. Ele é calculado pela média harmônica entre a precisão e a revocação, variando de 0 a 1, onde quanto mais próximo de 1 mais os valores aferidos da matriz não possuem grandes distorções:

$$Dice = 2 * \frac{Precisão * Revocação}{Precisão + Revocação}$$

- *Intersection over Union*: Métrica utilizada para avaliar a precisão de modelos de segmentação em redes neurais. O IoU é definido como a razão entre a área de interseção e área da união entre dois conjuntos (no nosso caso, a área predita pela rede e a área real). O valor do IoU varia de 0 a 1, sendo quanto mais próximo de 1 mais similar a região predita é em relação a região real.

$$IoU = \frac{Área(Intersecção)}{Área(União)} = \frac{Área predita}{Área real}$$

### 3. Metodologia

#### 3.1. Materiais

##### 3.1.1. Tipo de Linguagem

A bibliografia mais recente tem destacado um vasto número de linguagens para o desenvolvimento da interface, sendo algumas delas como *Java*, *C++* e *Python* [20]. A escolha do tipo de linguagem a ser empregado na construção na rede se baseia principalmente tanto na capacidade operacional da rede, de forma a mitigar o gasto operacional, quanto na facilidade e suporte oferecido pelos frameworks mais utilizados, a depender de sua finalidade.

CNNs são redes robustas, necessitando uma grande quantidade de dados coerentes e poder computacional para atingir resultados ótimos. De maneira a reduzir a complexidade e explorando a ampla gama de bibliotecas e funcionalidades, optou-se pela utilização da linguagem *Python*.

### 3.1.2. *Framework*

O *Framework* possui um papel importantíssimo para a construção, desenvolvimento e otimização da rede neural a ser desenvolvida. Cada framework possui pontos favoráveis e desfavoráveis cujo desenvolvedor deve optar a depender do foco do desenvolvimento [2].

Quando tratamos de processamento de imagens, nosso framework deve possuir como requisitos: um ótimo suporte para fluxo de dados, pois haverá uma alta quantidade de matrizes de bits; deve ser otimizado para o uso de expressões matemáticas, para realizar manipulações algébricas nas matrizes de bits e; permitir a utilização de CPU, GPUs e/ou TPUs, pois a quantidade massiva de dados pode exigir a necessidade de processamento paralelo, oferecido pelas GPUs/TPUs [22].

Dessa forma, optou-se pela utilização do *TensorFlow*, que além de suportar *Python*, integrando as suas bibliotecas como *NumPy* e *pandas*, oferece uma alta flexibilidade para desenvolver e implementar modelos de DL.

### 3.1.3. *Dataset*

Dentre os fatores determinantes para melhora na capacidade de aprendizado das redes neurais, devemos destacar qualidade dos dados oferecidos a rede, e quantidade de dados, que deve ser bastante elevada para obter melhores resultados. *Datasets* de imagens de ressonância magnética são bastante recorrentes, gerados em concursos e desafios com o intuito de melhorar a pesquisa e desenvolvimento de algoritmos de segmentação e detecção.

Dessa forma, foi utilizado o *dataset Sunnybrook Cardiac Data*, da *Cardiac Atlas Project* [21]. Este *dataset* é composto por imagens de 45 pacientes, divididos em 4 grupos de patologias: insuficiência cardíaca com infarto, insuficiência cardíaca sem infarto, hipertrofia cardíaca e pacientes saudáveis, totalizando um conjunto com 32000 imagens monocromáticas com dimensões distintas.

Dentre os 4 grupos, fora selecionado o grupo de pacientes saudáveis. Uma característica presente deste *dataset* foi obtenção manual das imagens, não caracterizando em sua divisão a separação das imagens coerentemente a seu ângulo de obtenção. Em outras palavras, as imagens obtidas de cada paciente

apresentam diversos ângulos da região do coração, como demonstrado na imagem abaixo. A aquisição de imagens foi realizada a partir do método de ressonância magnética cinemática, onde múltiplas imagens são obtidas ao longo do intervalo de um batimento, para cada posição de corte, como demonstra a Figura 3.1 [23].

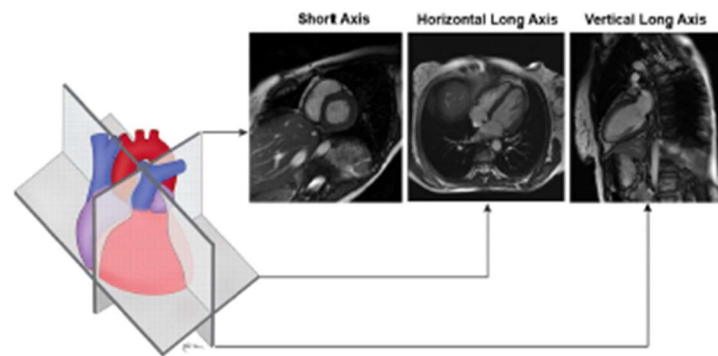


Figura 3.1: Orientações dos planos cardíacos em relação ao coração: Imagens do coração no *dataset* variam em diferentes planos cardíacos [24]

Dessa forma, dos pacientes saudáveis, foram selecionadas o total de 760 imagens, em formato DICOM, para o treinamento da rede neural. Estas imagens, além de estarem em um formato de extensão incompatível ao *TensorFlow*, possuíam tamanhos diversos (variando entre 64 X 64 bits até 256 X 256 bits) passaram por alguns processos de pré-processamento como forma a regularizar as imagens em um padrão compatível a rede (512 X 512), como também aplicar algumas técnicas para melhorar a qualidade da imagem, que serão exploradas mais profundamente na sessão de Métodos.

#### 3.1.4. Ambiente de Desenvolvimento

A rede neural convolucional, composta por todo o seu processamento, treinamento e testes foram realizados na plataforma *Google Colaboratory* [25]. Como forma de diminuir o tempo computacional de execução e melhorar o treinamento, fora utilizada a versão paga do *Colaboratory*, o *Colab Pro*.

Esta versão dá acesso a GPUs/TPUs mais avançadas, permitindo a utilização de RAM estendida para aumento de processamento. No treinamento e testes da rede, fora utilizada a GPU T4, com RAM estendida, cuja composição é definida abaixo. Além da melhora da CPU, a RAM estendida permite a realização de

maior quantidade de cálculos, e assim, nos da capacidade de aumentar a quantidade de imagens fornecidas a rede.

## 4. Desenvolvimento e Análise de Resultados

### 4.1. Tratamento de Imagem

O pré-processamento de imagens é um processamento realizado com o intuito de melhorar a qualidade das imagens antes do treinamento do modelo. Em imagens médicas, os procedimentos padrões são o realce de contraste, analisando o histograma das imagens, a normalização dos pixels das imagens e, em casos em que possuímos um número pequeno de imagens de treinamento, é utilizado o procedimento de *Data Augmentation*.

Partindo do *dataset*, são dispostas 760 imagens de pacientes saudáveis em formato DICOM (.dcm) [26]. DICOM, ou *Digital Imaging and Communications in Medicine* (Comunicação de Imagens Digitais em Medicina), é um conjunto de normas que padronizam o arquivamento e comunicação de imagens e dados obtidos em exames. Arquivos neste tipo de extensão possuem a imagem codificada e um cabeçalho sobre as informações do exame, como dados do paciente, número de fatias que o coração foi dividido, dimensões da imagem capturada, dentre outras.

Por possuir diversos tipos de enquadramento da região torácica, os tamanhos de arquivo foram normalizados para o formato 512 X 512, compatível para uso nos tensores do *TensorFlow*. Também foi realizada a conversão das imagens, do formato DICOM, para o formato PNG separando as imagens em dois diretórios: *train*, composto pelas imagens pré-processadas e suas respectivas máscaras, e *test*, composto por uma parte das imagens pré processadas para verificação do nível de acurácia da rede após treinamento. Como o *dataset* possuía definições de contorno para um mínimo número de imagens, foi necessário se criar máscaras (*ground-truth*) para o mapeamento da rede neural.

#### 4.1.1. Equalização de Histograma

Histograma é a representação gráfica da quantidade de cada pixel em intensidade de cor em uma imagem. Quando tratamos de imagens de



ressonância magnética, em preto e branco, a intensidade é definida em escala de cinza, variando de 0 a 255 em imagens de 24 bits. Podemos observar o exemplo de uma imagem e seu respectivo histograma na Figura 4.1.

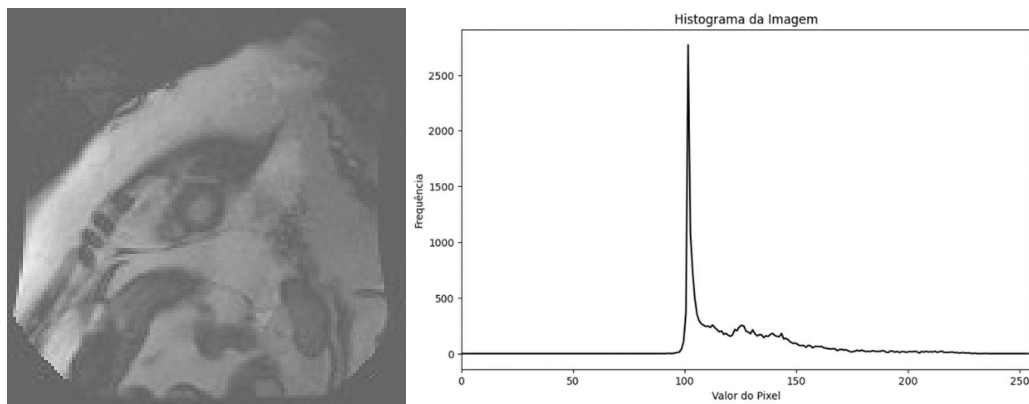


Figura 4.1: Histograma de Imagem original da base de imagens *Sunnybrook* e seu respectivo histograma: As imagens variam em opacidade/claridade dependendo do momento que a foto foi tirada durante o ciclo de um batimento

Através da distribuição da quantidade de pixels no histograma da imagem, podemos obter informações sobre brilho, qualidade e contraste médio da imagem [27]. A equalização do histograma é um procedimento que permite suavizar o a distribuição dos pixels na imagem, ajustando seu brilho, como se mostra na Figura 4.2 [28].

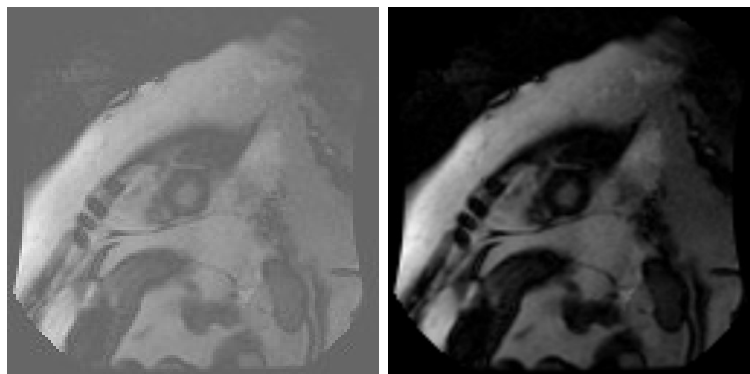


Figura 4.2: Pré-processamento (a) Imagem original da base de imagens *Sunnybrook* (b) Imagem após realização da equalização de histograma: A suavização na distribuição de pixels acentua e realça os contornos da imagem.

#### 4.1.2. *Data Augmentation*

*Data Augmentation* é uma técnica de regularização utilizada para aumentar o número de imagens em um banco de dados adicionando cópias levemente

modificadas dos dados originais [29]. Alguns tipos de modificação são no brilho, saturação, zoom, rotação ou reflexão horizontal e vertical [30].

#### 4.1.3. Criação de *Ground-Truth*

*Ground-Truth* são nossos dados da rede segmentados, ou seja, representam o resultado que desejamos que nossa rede aprenda a identificar. De maneira geral, CNNs utilizam definições de contorno criadas manualmente para, a partir do aprendizado de máquina, conseguirem identificar os padrões aprendidos e definidos por esse dimensionamento.

Imagens de ressonância magnética possuem como característica principal a centralização do órgão alvo de avaliação na chapa de imagem. Como nosso enfoque é segmentar a região do coração do restante da imagem, aplicamos além de um clareamento a imagem original (de forma a destacar os seus contornos principais) uma máscara circular, com alvo na região do coração. Dessa forma, ao realizar a etapa de treinamento, a rede focará na definição da máscara em seu aprendizado, ignorando as demais regiões. Um exemplo de definição desta máscara é apresentado na Figura 4.3.

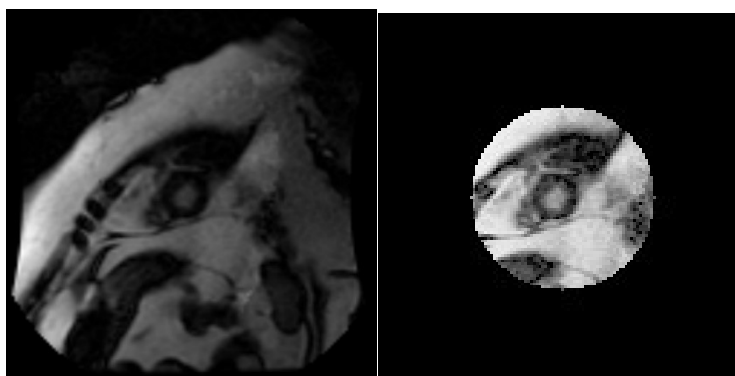


Figura 4.3: Definição de *Ground-Truth*: Imagem pré-processada e sua respectiva máscara:

Além da inclusão do alvo no centro da imagem, houve destacamento da região a ser segmentada, para auxiliar na definição do contorno.

## 4.2. Desenvolvimento do Modelo

Para avaliarmos o impacto da técnica de *Data Augmentation* em nosso modelo e na qualidade dos resultados, propusemos três cenários de treinamento:

- Sem *Data Augmentation*: A rede foi treinada com as 760 imagens pré-processadas, sem adição de novas imagens.

- *Data Augmentation* (rotação horizontal): Realizou-se *Data Augmentation* nas 760 imagens originais e suas respectivas máscaras, gerando 760 novas imagens espelhadas horizontalmente, totalizando 1520 imagens de treinamento.
- *Data Augmentation* (todas as manipulações): Realizou-se *Data Augmentation* nas 760 imagens originais e suas respectivas máscaras, gerando 5 variações de cada imagem original, com espelhamento horizontal e vertical, realizando leves distorções, saturação e rotação de imagem. Somando as imagens originais, totalizou-se 4560 imagens.

### 4.3. Treinamento do Modelo

A definição dos hiper parâmetros da rede neural é um critério definido pelo projetista de maneira a maximizar a qualidade dos resultados da rede em detrimento da capacidade operacional disponível. Dessa forma, todos os testes foram definidos com os mesmos parâmetros, para que se fossem analisados qualitativamente o impacto do aumento de dados na rede.

O hiper parâmetros definidos foram:

- Número de épocas: 100
- Tamanho de batch: 4
- Taxa de aprendizado: 0,0001 (ou  $1 \cdot 10^{-4}$ )

Os modelos foram treinados utilizando como parâmetro o decaimento do Coeficiente *Dice*, calculado ao final de cada época e comparado com o maior valor anterior encontrado. Caso o modelo de uma época mais recente obtivesse um valor de *Dice* maior, o modelo era salvo em diretório a parte. Caso não fosse, uma mensagem de aviso era notificada, e se ela se repetisse por consecutivamente por 5 épocas, o treinamento era encerrado.

Por fim, para avaliação dos resultados no teste, foram utilizadas as imagens de treinamento, criadas como espelhos das imagens teste, e comparadas com o modelo de predição, recuperando os níveis de acurácia e IoU de cada uma das imagens. Como uma avaliação métrica geral da rede, foi calculado a acurácia e o IoU médio dos modelos.

## 4.4. Análise dos Resultados

### Caso 1 - *Dataset* sem *Data Augmentation*

Neste primeiro caso, fora utilizado o dataset padrão, com 760 imagens. Dessa forma obtivemos o seguinte desenvolvimento durante o treinamento, demonstrado na Figura 4.4.

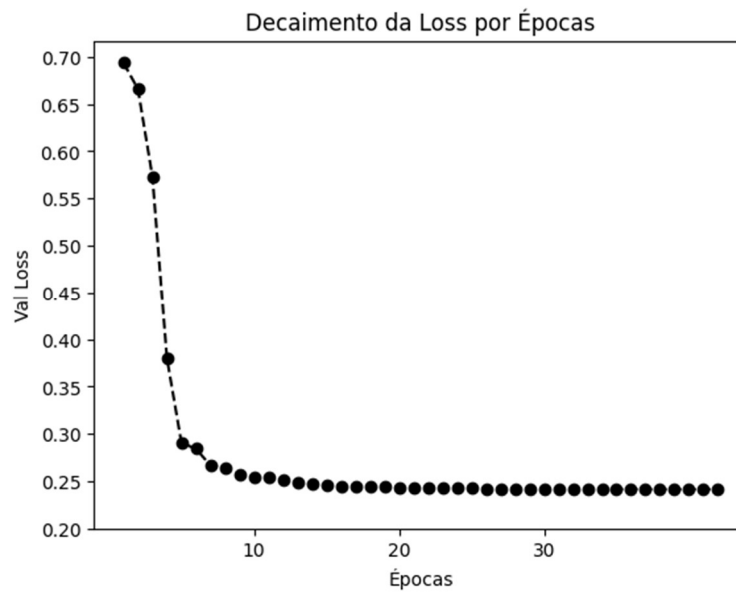


Figura 4.4: *Dataset* original - Decaimento do *val loss* por épocas no treinamento: observa-se um decaimento gradual do *val loss*, com uma estabilização a partir da vigésima época

O tempo médio de época foi em torno de 3 minutos e foram necessárias 42 épocas para a convergência. Devido ao valor de *batch*, cada época foi composta por 190 imagens.

Podemos perceber um decaimento consistente durante as primeiras épocas, com pouca turbulência, gerando uma notável convergência em poucas épocas. Neste caso, a rede respondeu de maneira correspondente, gerando um coeficiente *Dice* de 0,7586.

Durante os testes, a rede se comportou de maneira estável, gerando uma média de acurácia de 0,8567 e uma média de IoU de 0,42854. Um comparativo da imagem original, sua respectiva máscara e o modelo preditivo da rede são apresentadas na Figura 4.5.

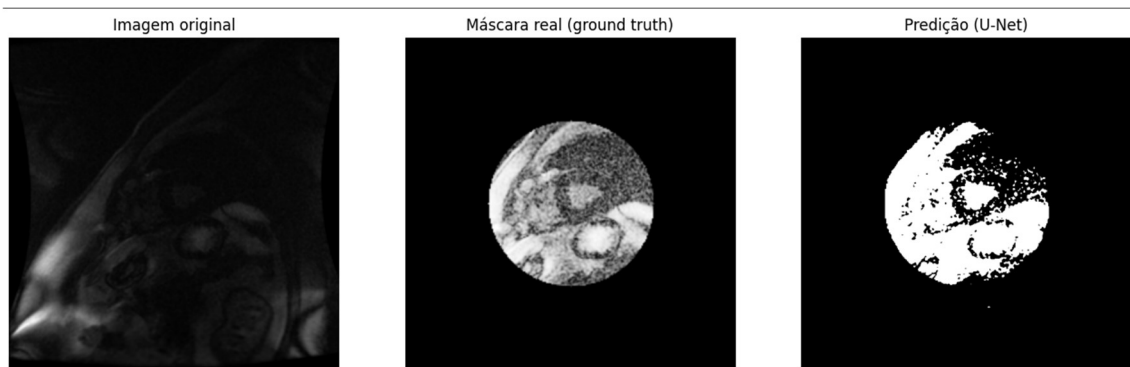


Figura 4.5: Predição da rede sem *Data Augmentation*: Podemos notar os primeiros indícios de segmentação mais fina da região do coração mesmo em imagens de difícil identificação.

## Caso 2 - *Dataset* com *Data Augmentation* (reflexão horizontal)

No segundo caso, adicionamos junto ao *dataset* original as imagens refletidas horizontalmente, totalizando 1540 imagens. Dessa forma, obtivemos o seguinte desenvolvimento durante o treinamento, demonstrado na Figura 4.6.

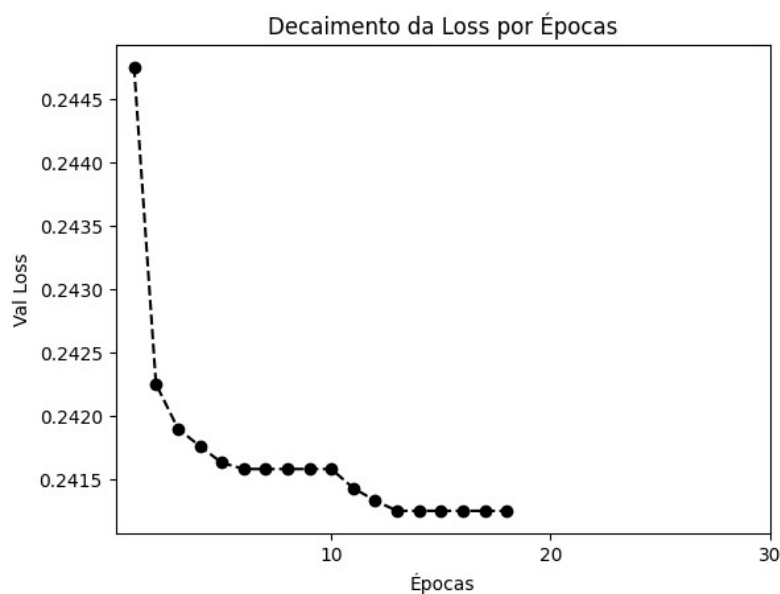


Figura 4.6: *Dataset* original + *Data Augmentation* - Decaimento do *val loss* por épocas no treinamento: O aumento no número de amostras promoveu um efeito de decaimento mais rápido

Houve um incremento de tempo significativo nas épocas, saindo de 3 min para 15 min em cada iteração. Devido ao valor de *batch size*, cada batch fora definido com um tamanho de 385 amostras por época, dobrando a quantidade com relação ao modelo anterior.

Neste segundo treinamento, note que o aumento no número de amostras por época produziu um efeito de decaimento mais abrupto que no primeiro modelo. De fato, o aumento na capacidade de aprendizado acabou por acelerar a convergência do modelo, melhorando o Coeficiente *Dice* significativamente com uma quantidade adequada de épocas.

Durante os testes, a rede se comportou de maneira estável, gerando uma média de acurácia de 0,85193 e uma média de IoU de 0,42614. Um comparativo da imagem original, sua respectiva máscara e o modelo preditivo da rede são apresentadas na Figura 4.7.

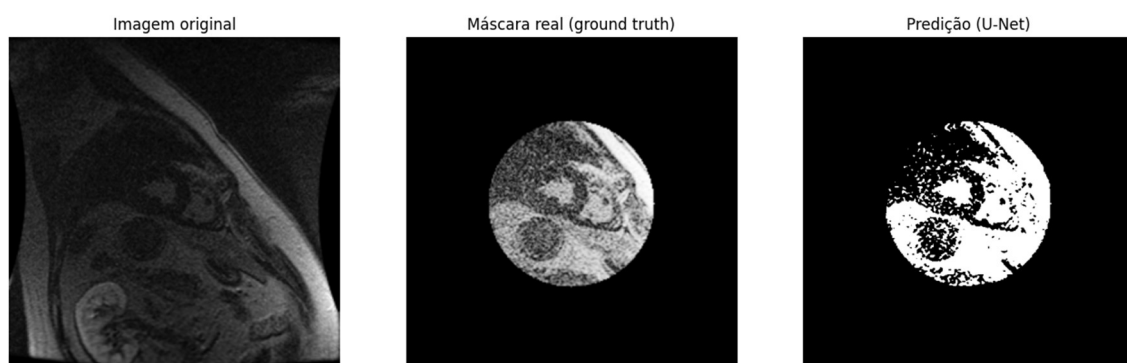


Figura 4.7: Predição da rede com *Data Augmentation*: Podemos notar a acentuação dos detalhes da região do coração, com algum certo ruído presente na imagem tomográfica

### Caso 3 – *Dataset* com *Data Augmentation* (Todas as variações)

Por fim, analisou-se o caso do *dataset* com *Data Augmentation* para todas as variações. Contabilizando 4560 imagens, com o *batch size* definido em 4, cada época deveria processar em torno de 1140 imagens por época, representando o caso mais pesado para nossa rede.

Embora a memória da GPU fosse suficiente para calcular e separar as imagens integrantes de cada batch, o tempo de processamento da rede escalou abruptamente, de 15 min por época para 60 min por época. Considerando apenas o treinamento simples da rede, que já estava utilizando uma das melhores GPUs disponíveis para o pacote do *Google Colab Pro*, seu término ocorreria, sendo mais otimista e pensando na convergência dos casos anteriores, por volta da 25ª época, ou seja, 1 dia e 1 hora.

Mesmo aguardando para podermos analisar as primeiras épocas, o modelo travou ao fim da primeira época, permanecendo em looping. Alterações foram tentadas para disponibilizar melhores resultados: aumento do *batch size*, redução da velocidade de leitura da rede, e até mesmo redução no número de épocas para 1/4 da capacidade, mas não houve convergência.

## 5. Considerações Finais

Diversos fatores podem impactar na velocidade e na efetividade do aprendizado de um modelo. Em nosso caso, a definição manual de máscaras acaba por reduzir a qualidade nos resultados obtidos em nossa rede, pois afetam diretamente o treinamento do modelo preditivo. De todo modo, os resultados encontrados são positivos, uma vez que o modelo segmenta com certa qualidade a região central das imagens tomográficas.

Embora o uso de *Data Augmentation* permite uma melhora na capacidade de generalização da rede, a aplicação de transformações aleatórias às imagens de treinamento pode introduzir ruído nos dados de treinamento, o que interfere significativamente na convergência da rede. Uma alternativa para solucionar este problema seria a aplicação dos dados em redes mais complexas, como a *Attention U-Net*, uma variante da rede U-Net.

Algumas sugestões de melhoria seriam a definição de uma base de dados mais vasta e coerente, que não exija tanto a necessidade de aplicação de variedade espacial como *Data Augmentation*; investimento em GPUs/TPUs mais robustas que permitam tempo de processamento mais longos; depuração e otimização dos métodos de pré-processamento com intuito tornar a execução e preparação dos dados mais leve e; analisar a possibilidade de um modelo de classificação/segmentação para análise de cada ângulo de imagem cardíaca.

## 6. Bibliografia

- [1] CHEN, Qi et al. **A survey on an emerging area: Deep learning for smart city data.** *IEEE Transactions on Emerging Topics in Computational Intelligence*, v. 3, n. 5, p. 392-410, 2019.
- [2] BAND, Shahab S. et al. **When smart cities get smarter via machine learning: An in-depth literature review.** *IEEE Access*, v. 10, p. 60985-61015, 2022.
- [3] MUHAMMAD, Amina N. et al. **Deep learning application in smart cities: recent development, taxonomy, challenges and research prospects.** *Neural computing and applications*, v. 33, p. 2973-3009, 2021.
- [4] Gonzalez, Rafael e Richard Woods: **Digital Image Processing.** Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2002.
- [5] Ridgway, John P: **Cardiovascular magnetic resonance physics for clinicians: part i.** *Journal of cardiovascular magnetic resonance*, 12(1):1–28, 2010.
- [6] Tortora, Gerard J e Bryan Derrickson: **Corpo Humano-: Fundamentos de Anatomia e Fisiologia.** Artmed Editora, 2016.
- [7] MUHAMMAD, Amina N. et al. **Deep learning application in smart cities: recent development, taxonomy, challenges and research prospects.** *Neural computing and applications*, v. 33, p. 2973-3009, 2021.
- [8] THAKUR, Amey; KONDE, Archit. **Fundamentals of neural networks.** *International Journal for Research in Applied Science and Engineering Technology*, v. 9, n. VIII, p. 407-426, 2021.
- [9] SHANKAR, R. Shiva et al. **Optimized model for predicting gestational diabetes using ML techniques.** In: **2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA).** IEEE, 2021. p. 1623-1629.
- [10] Data Science Academy. **Deep Learning Book.** 2022. Disponível em <<https://www.deeplearningbook.com.br>>. Acesso em 18 jun. 2024.
- [11] CASTIGLIONI, Isabella et al. **AI applications to medical images: From machine learning to deep learning.** *Physica medica*, v. 83, p. 9-24, 2021.
- [12] MUHAMMAD, Khan et al. **Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey.** *IEEE Transactions on Neural Networks and Learning Systems*, v. 32, n. 2, p. 507-522, 2020.
- [13] Alves, Gisely. **Entendendo Redes Convolucionais (CNNs).** Medium. Disponível em <<https://medium.com/neuronio-br/entendendo-redes-convolucionais-cnns-d10359f21184>>. Acesso em 18 jun. 2024.



[14] NAZARÉ, Marcelo Áxel Chiapinotto de. **Segmentação do ventrículo esquerdo em imagens de ressonância magnética usando rede neuralconvolucional e modelo Chan-Vese**. 2021.

[15] Ronneberger, Olaf, Philipp Fischer e Thomas Brox: **U-net: Convolutional networks for biomedical image segmentation**. Em International Conference on Medical image computing and computer-assisted intervention, páginas 234–241. Springer, 2015.

[16] Long, Jonathan, Evan Shelhamer e Trevor Darrell: **Fully convolutional networks for semantic segmentation**. Em Proceedings of the IEEE conference on computer vision and pattern recognition, páginas 3431–3440, 2015.

[17] CUNHA, Leonardo Cardoso da. **Redes neurais convolucionais e segmentação de imagens: uma revisão bibliográfica**. 2020. 51 f. Monografia (Graduação em Engenharia de Controle e Automação) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 2020.

[18] Junior, Geraldo Braz. **Segmentação Semântica: U-Net**. Disponível em <<https://nca.ufma.br/~geraldovcm/9.unet.pdf>>. Acesso em 18 jun. 2024.

[19] BERTELS, Jeroen et al. **Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice**. In: **Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22**. Springer International Publishing, 2019. p. 92-100.

[20] ATITALLAH, Safa Ben et al. **Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions**. *Computer Science Review*, v. 38, p. 100303, 2020.

[21] Project, Cardiac Atlas: **Sunnybrook cardiac data - cardiac atlas project**. <https://www.cardiacatlas.org/studies/sunnybrook-cardiac-data/>, acesso em 2021-08-07.

[22] Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA. **Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI**. *The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge*, <http://hdl.handle.net/10380/3070>

[23] Ishida, Masaki, Shingo Kato e Hajime Sakuma: **Cardiac mri in ischemic heart disease**. *Circulation Journal*, páginas 0908060454–0908060454, 2009.

[24] Ginat, Daniel T., Michael W. Fong, David J. Tuttle, Susan K. Hobbs e Rajashree C. Vyas: **Cardiac imaging: Part 1, mr pulse sequences, imaging planes, and basic anatomy**. *American Journal of Roentgenology*, 197(4):808–815, 2011. <https://doi.org/10.2214/AJR.10.7231>, PMID: 21940567. 4, 5, 6

[25] Google. **Colaboratory - Perguntas Frequentes**. Disponível em <<https://research.google.com/colaboratory/intl/pt->

BR/faq.html#:~:text=O%20Colab%20é%20um%20serviço,ciência%20de%20da dad%20e%20educação.&text=Sim.,financeiros%20para%20usar%20o%20Cola C.>. Acesso em 18 jun. 2024.

[26] Mildenberger, Peter, Marco Eichelberg e Eric Martin: **Introduction to the dicom standard**. European radiology, 12(4):920–927, 2002.

[27] Zimmerman, J.B., S.M. Pizer, E.V. Staab, J.R. Perry, W. McCartney e B.C. Brenton: **An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement**. IEEE Transactions on Medical Imaging, 7(4):304–312, 1988.

[28] Pisano, Etta D, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning e Stephen M Pizer: **Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms**. Journal of Digital imaging, 11(4):193, 1998. 24

[29] GARCEA, Fabio et al. **Data augmentation for medical imaging: A systematic literature review**. Computers in Biology and Medicine, v. 152, p. 106391, 2023.

[30] Eaton-Rosen, Zach, Felix Bragman, Sebastien Ourselin e M Jorge Cardoso: **Improving data augmentation for medical image segmentation**. 2018.