

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Joel Ramos Michaliszen Junior

Uma nova abordagem para segmentação de Acidente Vascular Cerebral em Imagens de Tomografia Computadorizada usando Aprendizagem de Zero-Shot

Fortaleza – CE

15 de Janeiro de 2024

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

M621n Michaliszen, Joel Ramos.

Uma nova abordagem para segmentação de Acidente Vascular Cerebral em Imagens de Tomografia Computadorizada usando Aprendizagem de Zero-Shot / Joel Ramos Michaliszen. - 2024.
86 f. : il. color.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2024.

Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho.

Coorientação: Prof. Dr. Leandro Bezerra Marinho.

1. Detecção e segmentação em TCs, aprendizagem profunda, acidente vascular cerebral hemorrágico. I.
Titulo.

CDD 004

Joel Ramos Michaliszen Junior

Uma nova abordagem para segmentação de Acidente Vascular Cerebral em Imagens de Tomografia Computadorizada usando Aprendizagem de Zero-Shot

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Processamento Digitais de Sinais e Imagens

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho

Coorientador: Leandro Bezerra Marinho

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

15 de Janeiro de 2024



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Joel Ramos Michaliszen Junior

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho
Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

Coorientador: Leandro Bezerra Marinho
Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE)

**Prof. Dr. Carlos Maurício Jaborandy de Mattos
Dourado Junior**
Instituto Federal de Educação, Ciência e Tecnologia do
Ceará (IFCE), Brasil.

Prof. Dr. Antonio Carlos da Silva Barros
Universidade da Integração Internacional da Lusofonia
Afro-Brasileira (UNILAB), Brasil.

Fortaleza – CE
15 de Janeiro de 2024

Agradecimentos

Agradeço ao meu Orientador Prof. Dr. Pedro Pedrosa Rebouças Filho, por sua paciência, incentivo e generosidade durante o período acadêmico.

Agradeço também ao meu Co-orientador Prof. Leandro Bezerra Marinho por toda sua ajuda, disponibilidade e direcionamento para que conseguisse concluir minha pesquisa.

Agradecer finalmente ao Laboratório de Processamento de Imagens, Sinais e Computação aplicada (LAPISCO) - especialmente a todos os membros, pelas trocas de experiência de vida, vida acadêmica que nos trouxe e continua a proporcionar muitas conquistas.

*“Mas graças a Deus, que sempre nos conduz vitoriosamente em Cristo e por nosso
intermédio exala em todo lugar a fragrância do seu conhecimento!”*

1 Coríntios 15:57

Resumo

O AVC é uma das principais causas de morte em todo o mundo. Apesar da alta taxa de mortalidade, as chances de recuperação são significativamente maiores quando o diagnóstico é precoce e preciso, seguido do tratamento adequado. Várias técnicas de neuroimagem são empregadas para detecção de AVC, sendo a tomografia computadorizada (TC) o método principal. Os sistemas de diagnóstico assistido por computador (CAD) podem desempenhar um papel crucial na captura de informações imperceptíveis ao olho humano. Portanto, este trabalho propõe uma nova abordagem para a segmentação e detecção de AVC hemorrágico na TC usando *Deep Learning*. Os métodos propostos apresentaram resultados promissores usando os modelos base (*Foundation Models*) para segmentação *Zero-Shot*, que demonstrou uma boa eficiência para lidar com os desafios de segmentação impostos por imagens médicas, sem a necessidade de uma grande base de dados rotulados para treinamento, atingindo uma precisão de 99,82%. Para viabilizar a segmentação *Zero-Shot* de forma automática da região lesionada pelo AVC hemorrágico, uma etapa de inicialização foi adicionada ao método. A etapa de inicialização foi desenvolvida utilizando uma CNN para detecção da região de interesse (RoI). Também foram avaliadas outras abordagens para detecção, porém a abordagem utilizando a rede Detectron2 foi a mais precisa, com 99,36% de acurácia e Dice de 92,99%. Também foi realizado um experimento, afim de avaliar o método proposto, com redes para segmentação em uma única etapa (*One-Stage*), utilizando a mesma base de dados para treinamento e validação, a rede You Only Look At Coefficients (YOLACT). Também mostrou resultados promissores com uma precisão de 99,74%.

Palavras-chave: imagens médicas, TC cerebrais, aprendizado profundo, detecção e segmentação.

Abstract

Stroke is one of the leading causes of death worldwide. Despite its high mortality rate, chances of recovery are significantly higher when the diagnosis is early and accurate, followed by appropriate treatment. Various neuroimaging techniques are employed for stroke detection, with computed tomography (CT) being the primary method. Computer-aided diagnosis (CAD) systems can play a crucial role in capturing information imperceptible to the human eye. Therefore, this work proposes a new approach for segmentation and detection of hemorrhagic stroke in CT using Deep Learning. The proposed methods showed promising results using Foundation Models for Zero-Shot segmentation, which demonstrated good efficiency in dealing with segmentation challenges posed by medical images, without the need for large labeled datasets for training, achieving an accuracy of 99.82%. To enable automatic Zero-Shot segmentation of the hemorrhagic stroke lesion region, an initialization step was added to the method. The initialization step was developed using a CNN for detection of the region of interest (RoI). Other approaches for detection were also evaluated, but the approach using the Detectron2 network was the most accurate, with 99.36% accuracy and a Dice coefficient of 92.99%. An experiment was also conducted to validate the proposed method with single-stage segmentation networks, using the same dataset for training and validation, the You Only Look At Coefficients (YOLACT) network. It also showed promising results with an accuracy of 99.74%.

Keywords: medical images, brain, deep learning, detection and segmentation

Lista de ilustrações

Figura 1 –	Imagens de diferentes fatias de TC de lesão cerebral por AVC Hemorrágico. É possível identificar as variações de textura e área da região do AVC, que se modificam de acordo com a fatia do exame, e a região de sangue coagulado na massa cerebral.	29
Figura 2 –	Ilustração dos tipos de RNAs. Fonte: DSA (2022)	36
Figura 3 –	(A)- Ilustração da ligação(conexão) entre as camadas de uma rede CNN. (B)- Representação de 3 neurônios da camada oculta interligados ao mesmo Mapa de Características. Fonte: O próprio autor.	38
Figura 4 –	Ilustração do processo de camadas convolucionais.	39
Figura 5 –	Ilustração de diferentes camadas encarregadas em um processo de convolucional.	40
Figura 6 –	Ilustração do <i>loop</i> da camada oculta de uma RNN Vs. Rede neural <i>Feed-Foward</i> . Fonte: Kalita (2022)	41
Figura 7 –	Arquitetura geral de uma rede Transformer. Fonte:(VASWANI et al., 2023)	42
Figura 8 –	<i>Scaled Dot-Product Attention</i> - Cálculo para matriz de atenção dos tensores Q, K, V. Representa a função de atenção em 3.4 <i>Multi-Head Attention</i> - Múltiplas camadas de atenção que executam em paralelo. Representa a função de atenção multi-cabeça em 3.3 Fonte:(VASWANI et al., 2023)	43
Figura 9 –	Visão geral do ViT . Fonte:(DOSOVITSKIY et al., 2021)	45
Figura 10 –	Tarefa : entradas da solicitação de segmentação de um objeto em uma imagem. Fonte:(KIRILLOV et al., 2023)	47
Figura 11 –	Visão geral do modelo SAM : Um codificador de imagem gera um <i>embedding</i> de imagem que pode então ser consultada com eficiência por uma variedade de solicitações de entrada para produzir máscaras de objetos em velocidade amortizada em tempo real. Para solicitações ambíguas correspondentes a mais de um objeto, o SAM pode gerar múltiplas máscaras válidas e pontuações de confiança associadas. Fonte:(KIRILLOV et al., 2023)	48

Figura 12 – Detalhes do Decodificador da Mascara (<i>Mask Decoder</i>). Um decodificador de duas camadas atualiza embedding de imagens e os tokens da solicitação por meio de atenção cruzada. Em seguida, a incorporação da imagem é ampliada, a partir da qual os tokens de saída atualizados são usados para prever máscaras dinamicamente. (Não ilustrado para maior clareza da figura: em cada camada de atenção, codificações posicionais são adicionadas à incorporação da imagem e todo o token de solicitação original (incluindo codificação de posição) é adicionado novamente às consultas e chaves do token.) Fonte:(KIRILLOV et al., 2023)	49
Figura 13 – Tarefa: Destilação dissociada. Fonte:(ZHANG et al., 2023)	50
Figura 14 – Arquitetura da rede YOLACT Fonte:(BOLYA et al., 2019)	52
Figura 15 – Exemplos do conjunto de dados - Imagens de diferentes fatias de TC de lesão cerebral por AVC Hemorrágico.	54
Figura 16 – Método proposto para segmentação de avc hemorrágico.	55
Figura 17 – Exemplos de resultados da segmentação do método proposto MobileSAM _{centroid} - Ilustrando a mascara gerada aplicada na imagem original junto com o centroid calculado.	65
Figura 18 – Detecção de AVC Hemorrágico com o uso de aprendizagem profunda.	70
Figura 19 – Ilustração dos resultados obtidos para detecção de AVC hemorrágico na tabela 6.	71

Lista de tabelas

Tabela 1 – Comparação da quantidade de parâmetros e velocidade do modelo original SAM (KIRILLOV et al., 2023) e o modelo MobileSAM (ZHANG et al., 2023). A verificação da velocidade se deu utilizando apenas uma GPU.	51
Tabela 2 – Comparação dos resultados obtidos na segmentação automática de AVC Hemorrágico em imagens de CT proposto neste trabalho, em comparação com os resultados de trabalhos que utilizaram o mesmo conjunto de dados.	67
Tabela 3 – Comparação dos resultados obtidos na segmentação do AVC Hemorrágico em imagens de TC propostas neste trabalho com os relatados na literatura utilizando aprendizagem profunda (ABBASI et al., 2023). . .	68
Tabela 4 – Comparação do tempo para segmentação entre os métodos propostos e os melhores resultados de precisão geral dos métodos propostos por (HAN et al., 2020) e (MARQUES et al., 2022) que utilizam o mesmo conjunto de dados.	69
Tabela 5 – Comparação da precisão geral (Acc) dos resultados obtidos na segmentação de AVC Hemorrágico em imagens de TC do mesmo conjunto de dados.	70
Tabela 6 – Comparação dos resultados obtidos na detecção de AVC Hemorrágico em imagens de CT proposto neste trabalho, em comparação com os resultados da literatura (KARTHIK et al., 2020).	72

Lista de abreviaturas e siglas

AB	Balão adaptativo (<i>Adaptive Balloon</i>)
ACM	Métodos de Contorno Ativo (<i>Active Contours Methods</i>)
ADP	Análise de Densidade Pulmonar
AHTD	Análise da Densidade do Tecido Humano (<i>Analysis Human Tissue Density</i>)
CAD	Diagnóstico Auxiliado por Computador (<i>Computer-Aided Diagnosis</i>)
CNN	Rede Neural Convolucional (<i>Convolutional Neural Network</i>)
COMEPE	Comitê de Ética em Pesquisa
CRAD	Crisp Adaptativo
DC	Coefficiente dos dados (<i>Dice's Coefficient</i>)
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DPI	Doença Pulmonar Intersticial
DPM	Densidade Pulmonar Média
DPOC	Doença Pulmonar Obstrutiva Crônica
EM	Expectativa - Maximização (<i>Expectation – Maximization</i>)
GB	Gigabyte
GHz	GIGA-HERTZ
GMMs	Modelos de Mistura Gaussiana
GPU	Unidade de Processamento Gráfico (<i>Graphics Processing Unit</i>)
GVF	Fluxo de Vetor Gradiente (<i>Gradient Vector Flow</i>)
MCA	Método de Contorno Ativo
Mask R-CNN	Redes Neurais Convolucionais baseadas em Regiões com Máscara
MLP	Rede Perceptron Multicamadas
OMS	Organização Mundial da Saúde

OPF	Floresta de Caminho Ideal (<i>Optimum-Path Forest</i>)
OPS	<i>Optimum Path Snake</i>
PDI	Processamento Digital de Imagens
PPV	Valor Preditivo Positivo
RAM	Memória de Acesso Aleatório (<i>Random Access Memory</i>)
RBF	Função de Base Inicial (<i>Radial Basis Function</i>)
R-CNN	Redes Neurais Convolucionais baseadas em Regiões
RPN	Rede de Propostas de Região
RNAs	Redes Neurais Artificiais
RM	Ressonância Magnética
Sen	Sensibilidade
SISDEP	Sistema para a Detecção e a Quantificação de Enfisema Pulmonar
Spe	Especificidade
SSE	Soma do Erro Quadrático (<i>Sum of Squared Error</i>)
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)
TC	Tomografia Computadorizada
TCAR	Tomografia Computadorizada de Alta Resolução
UFC	Universidade Federal do Ceará
UH	Unidade Hounsfield
VFC	Convolução de Campo de Vetor (<i>Vector Field Convolution</i>)

Lista de símbolos

Π	Letra grega Pi Maiúscula, denominado Produto ou Produtório, no conceito matemático refere-se a multiplicação.
Φ	Letra grega Fi, equivale ao conjunto de parâmetros definidos pela i-ésima componenteda mistura.
Θ	Letra grega Teta, na representa a medida de um determinado Vetor, sendo um conjunto de parâmetros da mistura,
θ	Letra grega teta, referência ao ângulo.
Σ	Notação Sigma, somatório ou somatória, significa a soma de termos.
π	Símbolo Pi, representa uma proporção numérica definida pela relação entre o perímetro de uma circunferência e seu diâmetro.
ϕ	Letra grega Fi, (variante). correspondete a uma função de densidade Gaussiana.
α	Letra grega Alpha, conotação angular de início ou inicialização.
β	Letra grega Beta, conotação ponderação ao ângulo.
$\int$	Símbolo Integral, no qual representa um sinal que indica a integração e o resultado de integrar uma diferencial ou derivada.
μ	Letra grega Mi (Pronúncia: Miú), refere-se ao peso específico de uma coeficiente.
∂	Símbolo "del", usado para distinguir derivadas parciais das derivadas ordinárias de uma variável.
$\cap$	Símbolo Intersecção, corresponde a Intersecção entre conjuntos.
ω	Letra grega Ômega minúsculo, indica o mínimo local.
Ω	Letra grega Ômega maiúsculo, indica a composição do mínimo com o máximo local.
ν	Letra grega Nu, indica o limiar da força balão.
∇	Símbolo grego Nabla, denomina o operador diferencial.
$\in$	Símbolo que indica a relação de pertinência.

Sumário

1	INTRODUÇÃO	25
1.1	Descrição do problema	29
1.2	Objetivo Geral e Específicos	30
1.3	Produção científica	30
1.4	Estrutura da Dissertação	32
2	TRABALHOS RELACIONADOS	33
3	FUNDAMENTAÇÃO TEÓRICA	35
3.1	Aprendizagem de Máquina	35
3.2	Aprendizado Profundo - <i>Deep Learning</i>	35
3.2.1	Redes Neurais Convolucionais - CNN	37
3.2.2	Redes Neurais Recorrentes - RNN	40
3.2.3	Redes Transformers	41
3.2.3.1	Arquitetura	41
3.2.4	Transformers para visão computacional - <i>Vision Transformers</i>	44
3.2.5	Deteção - Detectron2	45
3.2.6	Segmentação de instancia	46
3.2.6.1	SAM (<i>Segment Anything Model</i>)	46
3.2.6.2	MobileSAM (<i>Mobile Segment Anything Model</i>)	50
3.2.6.3	YOLACT (<i>You Only Look At CoefficientTs</i>)	50
4	MATERIAIS E MÉTODOS	53
4.1	Conjunto de Dados	53
4.2	Métodos propostos	54
4.3	Avaliação de Desempenho	58
4.3.1	Acurácia	58
4.3.2	Sensitividade	59
4.3.3	Especificidade	60
4.3.4	Sensitividade $\times$ Especificidade - (<i>Sen $\times$ Spe</i>)	60
4.3.5	F1-Score	61
4.3.6	Coefficiente de Correlação de Matthews	61
4.3.7	Valor Preditivo Negativo	61
4.3.8	Valor Preditivo Positivo	62
4.3.9	Coefficiente de Dice - (<i>Dice's coefficient</i>)	62
4.3.10	Precisão Geral - (<i>Accuracy</i>)	62

4.3.11	Teste Estatístico	62
5	RESULTADOS E DISCUSSÃO	65
5.1	Resultados de Segmentação.	65
5.2	Validação dos Resultados	69
5.3	Resultados da Detecção.	70
6	CONCLUSÃO	73
6.1	Proposta de Trabalhos Futuros	74
7	ANEXO A	78
	REFERÊNCIAS	79

1 Introdução

O Acidente Vascular Cerebral (AVC) é a segunda maior causa de mortes no mundo, segundo a Organização Mundial da Saúde (OMS), o AVC ocorre quando algum vaso sanguíneo responsável pelo fluxo sanguíneo no cérebro é obstruído, normalmente ocasionado por algum outro fator de risco. O AVC pode ser classificado em 2 tipos: isquêmico e hemorrágico. O AVC isquêmico ocorre quando a obstrução impede o fluxo sanguíneo em uma região do cérebro, e o hemorrágico quando o vaso sanguíneo se rompe, causando sangramento no cérebro. Dado a quantidade de fatalidades inferidas, é possível afirmar que esta doença não afeta somente o indivíduo, mas todo o sistema de saúde público e privado ([BENVENUTTI et al., 2022](#)). O AVC hemorrágico constitui 15% de todos os casos de AVC, e concentra a maior parte das fatalidades, o diagnóstico é realizado por exame de imagem, como Tomografia Computadorizada (TC) ou Ressonância Magnética (RMN), os quais permitem identificar as áreas afetadas pela hemorragia. Dentre estas duas técnicas de exame para diagnóstico a Tomografia Computadorizada (TC) é a mais utilizada, por ser mais barata, mais rápida e menos traumática para algumas pessoas. Por ser uma hemorragia, quanto mais rápido se diagnosticar e tratar, maiores são as chances de recuperação e menores as chances de causar algum dano permanente ou uma fatalidade ([KAMTCHUM-TATUENE; JICKLING, 2019](#)).

A tomografia computadorizada (TC) é uma ferramenta diagnóstica e analítica fundamental, pois pode fornecer imagens transversais detalhadas dos órgãos internos, ossos, vasos sanguíneos e tecidos moles, permitindo assim o diagnóstico de uma variedade de condições médicas. Esta técnica de imagens médicas utiliza-se de raios-x e um programa de computador capaz de organizar e calcular com precisão os dados obtidos por um detector que mede a quantidade de raios-X que atravessa o corpo, e então formar imagens tridimensionais ou bidimensionais dos órgãos e estruturas internas, que podem ser visualizadas por médicos para fazer diagnósticos precisos ([HSIEH, 2003](#)).

A detecção precoce do AVC hemorrágico é de suma importância para garantir o tratamento efetivo e evitar complicações mais graves e até mesmo o óbito. Isso destaca a necessidade de melhorias constantes nas técnicas de diagnóstico e na tecnologia para permitir a detecção mais rápida possível. Entre os exames para diagnosticar o AVC a TC é o mais viável, considerando; custo, velocidade e acessibilidade da população. Porém o diagnóstico por imagem demanda uma grande atenção do médico especialista, pois ele deve realizar uma segmentação visual manualmente, contornando as áreas de interesse que podem apresentar alguma alteração patológica. No entanto, as alterações que aparecem nas imagens de TC são sutis e às vezes indistinguíveis para uma análise visual. Considerando a urgência em identificar o tipo de lesão, os sistemas de suporte à decisão

para o tratamento clínico em questão se tornam muito úteis. Sistemas de Diagnóstico Auxiliado por Computador (CAD), combinados com conhecimento médico, têm ganhado uso mais amplo em vários tipos de diagnósticos por imagem, reduzindo o tempo de avaliação do exame.

No entanto, os Sistemas de Diagnóstico Auxiliado por Computador (CAD) oferecem uma variedade de aplicações, como por exemplo segmentação automática de imagens de TC, através de técnicas de visão computacional, aplicação de redes neurais convolucionais (CNN) e aprendizagem profunda (MARQUES et al., 2022). Porém um diagnostico automatico com precisão ainda é um desafio para a area de inteligencia computacional. Estes sistemas inteligentes vem se mostrando muito promissores para pré-diagnosticos, destacando-se como um método não invasivo.

Estes sistemas podem identificar e analisar imagens médicas através de processamento computacional, e efetivamente detectar e identificar varias doenças. Dessa forma, é notavel que o avanço nos métodos computacionais voltados para diagnostico por imagem utilizando visão computacional como técnicas de detecção e segmentação, em conjunto do aumento da capacidade de processamento computacional seja por hardware ou ferramentas de otimização, tornaram o diagnostico mais rápido e preciso.

As Redes Neurais Convolucionais (CNN) apresentam resultados muito promissores, tornando-se muito utilizada nos sistemas CAD. O que torna as CNN uma metodologia altamente eficiente é seu grande poder de processamento que pode ser usado para identificar e classificar pixels anômalos, com um custo computacional relativamente alto. Outra técnica que tem recebido muita atenção dos pesquisadores é o pós-processamento com Ajuste Fino (FT) por meio de transferência de aprendizado, que consiste em um conjunto de técnicas computacionais aplicadas ao resultado de uma segmentação, com o objetivo de refinar e ajustar as bordas obtidas para uma melhor fidelidade à região afetada real.

Os avanços recentes em técnicas de aprendizado profundo têm revolucionado a área de imagens médicas, trazendo consigo um enorme potencial para a detecção e segmentação de áreas e lesões específicas. Além de que Redes Neurais Convolucionais (CNNs) têm se mostrado altamente eficazes na tarefa de classificação em exames por imagem. Essas redes são capazes de aprender padrões complexos presentes nas imagens e, assim, distinguir precisamente a área da lesão, permitindo que o especialista tome decisões reduzindo riscos para o paciente (INKEAW et al., 2022).

Para além da classificação, métodos baseados em deep learning também têm sido amplamente explorados para a detecção automática de lesões em imagens médicas. A utilização de CNNs treinadas em grandes conjuntos de dados de imagens cerebrais permite que o sistema localize e identifique regiões suspeitas, indicando a possível presença de um AVC. Isso pode ser especialmente útil em casos onde a hemorragia é muito estriada ou de difícil identificação aos olhos humanos, possibilitando intervenções mais precisas e

melhorando as chances de sucesso no tratamento ([GUIMARÃES et al., 2020](#)).

A Aprendizagem profunda (*Deep Learning*) e Redes Neurais Totalmente Convolucionais (FCNs) aplicados em imagens médicas afim de segmentar regiões de interesse são muito importantes, pois permitem a delimitação automática muito precisa e detalhada. O que é essencial para o planejamento cirúrgico, uma vez que fornece informações valiosas sobre a extensão exata da região afetada e sua relação com estruturas circundantes. Além disso, a segmentação precisa também facilita o acompanhamento contínuo e em tempo real da evolução da lesão, permitindo que os médicos avaliem a eficácia do tratamento ao longo do tempo ([RONNEBERGER; FISCHER; BROX, 2015](#); [DOU et al., 2017](#))

No contexto de imagens médicas, muitas aplicações demandam conjuntos de dados extensos, o que acarreta em altos custos tanto computacionais quanto financeiros. Isso inclui a marcação das imagens por especialistas, anotações de dados e armazenamento. Além disso, a diversidade de problemas e a falta de foco na abordagem podem limitar a criação de modelos específicos para tarefas particulares.

Por outro lado, os avanços nas CNNs, aliados à automação e ao uso de técnicas como o fine-tuning e a transferência de aprendizado, têm permitido o desenvolvimento de modelos mais robustos. Essas técnicas permitem a adaptação dos modelos para resolver problemas específicos, combinando-se com outras abordagens de aprendizado de máquina ([MARQUES et al., 2022](#)). Isso abre caminho para estratégias mais eficazes na busca pela melhoria dos resultados. Como as redes especializadas em detecção e segmentação em uma única etapa ou *One-Stage*. Se destacam pela velocidade de processamento sem um alto custo, possibilitando realizar a segmentação em tempo-real ([BOLYA et al., 2019](#)).

Além de novas arquiteturas como *Vision Transformers*, inspirada em processamento de linguagem natural, vem sendo utilizada para visão computacional com sucesso em tarefas de segmentação e detecção. Porém, para que se consiga implementar, é necessário uma grande base de dados rotulados. Então muitos modelos bases que implementam esta arquitetura de Transformers, acabam por atingir um alto grau de generalização ([DOSOVITSKIY et al., 2021](#)). Contudo, é difícil encontrar modelos desta natureza na área de imagens médicas, justamente pela falta de uma grande base de dados rotulados, vale lembrar que nestes casos é de suma importância que o dado seja rotulado por um especialista.

Então, com um alto grau de generalização, é possível realizar experimentos *Zero-Shot*, e até mesmo aplicação de algumas técnicas de aprendizagem como a destilação, mesmo com uma base de dados relativamente pequena. Como isso pode-se extrair mascaras de segmentação em alta resolução e qualidade.

Apesar de todo o potencial dessas técnicas, é importante enfatizar que sua implementação na prática clínica deve ser cuidadosamente validada e acompanhada por

profissionais médicos. A interpretação correta dos resultados e a confiabilidade dos modelos são aspectos críticos para garantir que essas ferramentas de inteligência artificial realmente contribuam para uma melhoria significativa para diagnósticos e tratamentos.

Em conclusão, o uso de métodos de *deep learning* para detecção e segmentação de AVC's em imagens de TC representa uma revolução na medicina. Essas técnicas têm o potencial de economizar tempo dos profissionais médicos, melhorar a precisão do diagnóstico e permitir tratamentos mais eficazes e personalizados. Com a contínua evolução dessas tecnologias e a colaboração entre a comunidade médica e científica, podemos esperar avanços ainda maiores no campo da radiologia e neurologia, levando a melhores resultados de saúde para os pacientes.

Nas seções a seguir, esta descrito o desenvolvimento deste estudo e suas justificativas. Nas Seções 2 e 3 é apresentado uma revisão da literatura relevante, comparando nossos métodos e resultados com pesquisas de ponta publicadas em periódicos conceituados, como o IJCNN. A Seção 4 detalha os materiais, conjunto de dados e os métodos para detecção e segmentação, utilizados na pesquisa. Além de apresenta uma descrição detalhada da metodologia proposta, explicando seu funcionamento e potencial benefícios.

Na seção de resultados são comparados os métodos através de métricas comuns entre a metodologia desta pesquisa e outro trabalhos realizados na área. Além da discussão da eficacia dos métodos propostos em termos de precisão, sensibilidade, especificidade e outras métricas relevantes.

Este estudo propõe uma nova abordagem para auxiliar no diagnóstico de AVC hemorrágico, utilizando técnicas de Aprendizado Profundo. A proposta visa segmentar e detectar automaticamente as áreas afetadas em imagens de Tomografia Computadorizada (TC), com o objetivo de reduzir custos médicos e aumentar a eficacia e precisão no diagnostico realizado por um médico especialista. Mais especificamente, o estudo tem como objetivo:

- Apresentar uma nova abordagem para o processamento de imagens médicas de TC.
- Detectar a região afetada pelo AVC hemorrágico utilizando o método de aprendizado profundo.
- Melhorar as métricas através da aplicação de técnicas de ajuste fino ao processo de segmentação, afim de aumentar a precisão.

Para alcançar esses objetivos, seguimos uma abordagem sistemática e metodológica. Inicialmente, desenvolvemos um novo método de processamento de imagens médicas que se baseia nos avanços recentes em visão computacional, mais especificamente na detecção de objetos. Essa abordagem representa um avanço significativo em relação aos métodos de processamento de imagens tradicionais, que muitas vezes são demorados e exigem

intervenção manual de especialistas médicos.

1.1 Descrição do problema

A detecção e segmentação automática de Acidentes Vasculares Cerebral (AVC's) em imagens de Tomografia Computadorizada é uma atividade extremamente complexa e importante para área médica. A complexidade se da por conta da irregularidade das regiões afetadas, oque dificulta muito a identificação dos limites da hemorragia, tanto para os especialistas, como para técnicas matemáticas que visam distinguir a região afetada pelas características dos pixels da imagem.

A imagem da Figura 1 exibe imagens de tomografia computadorizada (TC) de AVC hemorrágico, cada uma destacando questões específicas relacionadas a cada paciente. Embora seja possível fazer uma identificação visual da região do AVC, determinar com precisão os limites da hemorragia pode ser desafiador, uma vez que, em alguns casos, essa área pode se assemelhar aos *pixels* da massa cerebral circundante. A detecção a olho nu pode ser viável em alguns cenários, mas é fundamental garantir uma localização precisa dos *pixels* que compõem a hemorragia cerebral para uma avaliação clínica precisa. A complexidade da tarefa de segmentação varia conforme a necessidade de distinguir contornos que se assemelham às áreas da massa cerebral, permitindo, assim, um acompanhamento preciso da evolução do quadro clínico.



Figura 1 – Imagens de diferentes fatias de TC de lesão cerebral por AVC Hemorrágico. É possível identificar as variações de textura e área da região do AVC, que se modificam de acordo com a fatia do exame, e a região de sangue coagulado na massa cerebral.

Avanços recentes nas técnicas de aprendizagem profunda revolucionaram o campo da imagem médica, trazendo imenso potencial para a detecção e segmentação de áreas e lesões

específicas. Redes Neurais Convolucionais (CNNs) têm se mostrado altamente eficazes na tarefa de classificação de imagens. Essas redes podem aprender padrões complexos nas imagens e distinguir com precisão a área da lesão, permitindo ao especialista tomar decisões, reduzindo assim os riscos para o paciente (INKEAW et al., 2022).

Embora grandes conjuntos de dados representem um desafio para aplicações médicas, exigindo anotações de médicos especialistas, trabalhos recentes exploraram *foundation models* baseados em Vision Transformers especificamente para tarefas de visão computacional, como segmentação e detecção. Esses trabalhos mostram a notável capacidade desses modelos de generalizar com precisão na detecção e segmentação de objetos, abrindo caminho para aplicações potenciais na área médica. Isso levanta a questão: **os recursos precisos e generalistas de detecção de objetos deste tipo de modelo podem ser aproveitados para imagens médicas?**

1.2 Objetivo Geral e Específicos

Este trabalho propõe uma nova abordagem para segmentação e detecção de AVC hemorrágico em Tomografia Computadorizada para o auxílio de especialistas utilizando aprendizagem de Zero-Shot e segmentação por instância em tempo real.

Em seu sentido mais específico, essa dissertação tem os seguintes objetivos:

- Propor uma nova abordagem segmentação de acidente vascular cerebral hemorrágico em imagens de tomografia computadorizada que aproveita a Tecnologia da Saúde (*Health of Things*);
- Avaliar duas técnicas de segmentação que se baseiam em generalização zero-shot, SAM (*Segment Anything Model*) e MobileSAM;
- Avaliar uma técnica de segmentação de instâncias em tempo real, (*You Only Look At Coefficients*);
- Analisar o uso de um método de detecção de objetos na inicialização da abordagem proposta, Detectron2;
- Examinar entre as abordagens propostas neste estudo e da literatura, a segmentação mais próxima de marcações manuais realizadas pelo especialista.

1.3 Produção científica

Durante este projeto de mestrado, os seguintes trabalhos científicos foram aceitos e publicados, a saber:

- (IJCNN 2024)[SUBMETIDO] -A New Approach to Segmentation and Detection of Stroke in CT Images: Zero-Shot and One-Stage Deep Learning Techniques - 1º Autor - Qualis A1 (Eng IV) ;

- (IJCNN 2023) -Divisible Cell-Segmentation: A New Approach for Stroke Detection and Segmentation in CT Scans Using Deep Learning and Fine-tuning - 2º Autor - Qualis A1 (Eng IV);

1.4 Estrutura da Dissertação

Nos Capítulos 2 e 3 deste estudo apresenta uma revisão literária de trabalhos relacionados a métodos de segmentação em imagens médicas e utilização de Redes Neurais Convolucionais(CNN), abrangendo métodos tradicionais a métodos mais recentes baseados em aprendizagem profunda. Em seguida são apresentados métodos que serviram como base de análise e comparação de resultados entre a base de dados em estudo. No Capítulo 3, são detalhados modelos de aprendizado de máquina usados para criação e desenvolvimento dos experimentos e métodos propostos.

No Capítulo 4 é apresentado os experimentos e uma nova abordagem de segmentação de AVC hemorrágico baseado em *deep learning* e aprendizagem *Zero-shot* e redes one-stage. No Capítulo 5 serão discutidos os resultados dos métodos e experimentos propostos em comparação com estudos publicados e com reconhecimento acadêmico. Por fim, a Conclusão e Trabalhos Futuros são apresentados no Capítulo 6.

2 Trabalhos relacionados

Nesta seção, discutimos trabalhos relacionados ao assunto desta pesquisa, bem como o uso de técnicas de classificação, detecção e segmentação em diversos tipos de imagens médicas.

Técnicas de classificação são amplamente utilizadas em imagens médicas e desempenham um papel essencial no diagnóstico de doenças. Um método amplamente utilizado para realizar a classificação no campo médico é a Rede Neural Convolutiva (CNN), e tem mostrado resultados satisfatórios, como no caso de Tripathi e Kopparapu ([TRIPATHI; KOPPARAPU, 2020](#)), que utilizaram CNN para realizar a classificação de sinais e identificar níveis de gravidade de pacientes com a doença de Parkinson.

Além disso, Raj *et al.* ([RAJ et al., 2020](#)) aplicaram Redes Neurais Artificiais (ANN) para realizar a classificação de imagens de câncer de pulmão, imagens cerebrais e doença de Alzheimer em duas classes, benignas e malignas. No entanto, a técnica utilizada nesta pesquisa requer um alto custo computacional, e os resultados obtidos poderiam ser melhorados através do uso de diferentes características e uma maior quantidade de dados de treinamento.

Métodos de detecção de imagens são utilizados para os mais diversos fins, desde detecção de eventos multimídia ([KHAN; CURRY, 2020](#)) até aplicações para auxiliar no diagnóstico médico, onde tem obtido resultados satisfatórios, como foi o caso do estudo produzido por Gaur *et al.* ([GAUR et al., 2021](#)), que utilizaram CNN e transfer learning para realizar a detecção de Covid-19 em imagens de tomografia computadorizada. No entanto, esse modelo foi treinado com uma quantidade reduzida de dados, o que pode afetar seu desempenho em diferentes conjuntos de dados.

Outro estudo que obteve bons resultados na área de detecção de imagens médicas foi o realizado por Han *et al.* ([HAN et al., 2019](#)), no qual eles utilizaram o You Only Look Once v3 (YOLOv3) para realizar a detecção de metástases cerebrais, juntamente com uma técnica de aumento de dados (DA) chamada Generative Adversarial Network (GAN), que garantiu uma maior quantidade de dados durante o treinamento. No entanto, a técnica de DA é extremamente arriscada, pois, apesar de gerar um aumento na sensibilidade do modelo, também aumenta o número de falsos positivos.

A segmentação de imagens médicas desempenha um papel importante na automação do diagnóstico médico, sendo capaz de acelerar processos que teriam que ser feitos manualmente.

Tripathi *et al.* ([TRIPATHI; VERMA; SHARMA, 2021](#)) utilizaram uma adaptação

da rede Seg-Net, na qual aplicaram normalização cruzada de canal com parâmetros modificados para melhorar o desempenho na segmentação de tumores cerebrais no mesmo conjunto de dados usado nesta pesquisa. No entanto, embora tenha obtido resultados satisfatórios, deixou algumas lacunas para melhoria, que foram exploradas durante este estudo.

Buda *et al.* (BUDA; SAHA; MAZUROWSKI, 2019) também realizaram a segmentação de tumores cerebrais, desta vez usando uma CNN com a arquitetura U-Net. No entanto, apesar de terem obtido um resultado satisfatório, o estudo considerou apenas três características, o que abre espaço para possíveis melhorias no modelo apresentado. Além disso, uma pequena amostra de pacientes foi analisada no estudo (cerca de 110 pessoas), o que faz com que a eficiência do modelo não seja garantida em uma quantidade maior de amostras.

3 Fundamentação Teórica

Nesta seção, são apresentados diferentes métodos utilizados no modelo proposto e os materiais utilizados no estudo.

3.1 Aprendizagem de Máquina

Na última década, houve avanços notáveis no campo da inteligência artificial, com especial destaque para a área conhecida como aprendizado de máquina. Essa vertente concentra-se no desenvolvimento de algoritmos capazes de automatizar a tomada de ações inteligentes a partir de dados, sem a necessidade de programação explícita para cada atividade. Em resumo, o âmago desse ramo está na capacidade de transformar informações em decisões automatizadas, representando um marco significativo no avanço da inteligência artificial (PROVOST; KOHAVI, 1998).

O processo de aprendizagem fundamenta-se em métodos probabilísticos de otimização, os quais visam minimizar os erros cometidos até que se alcance um resultado satisfatório para a tarefa desejada. A otimização dos erros tem como objetivo principal gerar previsões baseadas nos dados de entrada, possibilitando a detecção de padrões e a compreensão das relações e sensibilidades entre os dados. Esses elementos essencialmente definem o conceito de inteligência computacional.

Essa evolução culminou no surgimento do aprendizado profundo, também conhecido como deep learning, a partir da década de 2000. Originado como uma extensão do aprendizado de máquina, o aprendizado profundo foi impulsionado pela disponibilidade de vastos conjuntos de dados e pelos avanços no poder computacional, especialmente o uso de unidades de processamento gráfico (GPUs).

3.2 Aprendizado Profundo - *Deep Learning*

O cérebro humano é uma máquina biológica complexa capaz de processar informações de forma notável. Desde as tarefas mais simples, como perceber o ambiente, até as mais complexas, como tomar decisões, o cérebro é responsável por um vasto conjunto de funções. Essas atividades são possíveis graças à sua intrincada rede de neurônios, que se comunicam entre si para processar informações e gerar respostas (KOCH; TSUCHIYA, 2007).

Na tentativa de compreender e replicar o processo de processamento de informações do cérebro, surgiu o campo das Redes Neurais Artificiais (RNAs). Essas redes são modelos

computacionais inspirados no cérebro humano, capazes de tomar decisões ou realizar atividades específicas com base nos dados fornecidos (HAYKIN, 2007).

Junto à abordagem inspirada na biologia, a computação tem desempenhado um papel vital no desenvolvimento da inteligência artificial. Utilizando conceitos matemáticos, estatísticos e linguagens de programação, os cientistas foram capazes de construir modelos que simulam e, em alguns casos, otimizam as atividades de inteligência realizadas pelo cérebro humano.

O avanço tecnológico contínuo permitiu o desenvolvimento do Aprendizado Profundo, um subcampo da aprendizagem de máquinas que utiliza redes neurais profundas para modelar e processar informações complexas. Na Figura 2 é possível identificar a diferença entre uma RNA simples e uma RNA profunda ou de *Deep Learning*. O Aprendizado Profundo tem revolucionado diversas áreas, como visão computacional e processamento de linguagem natural.

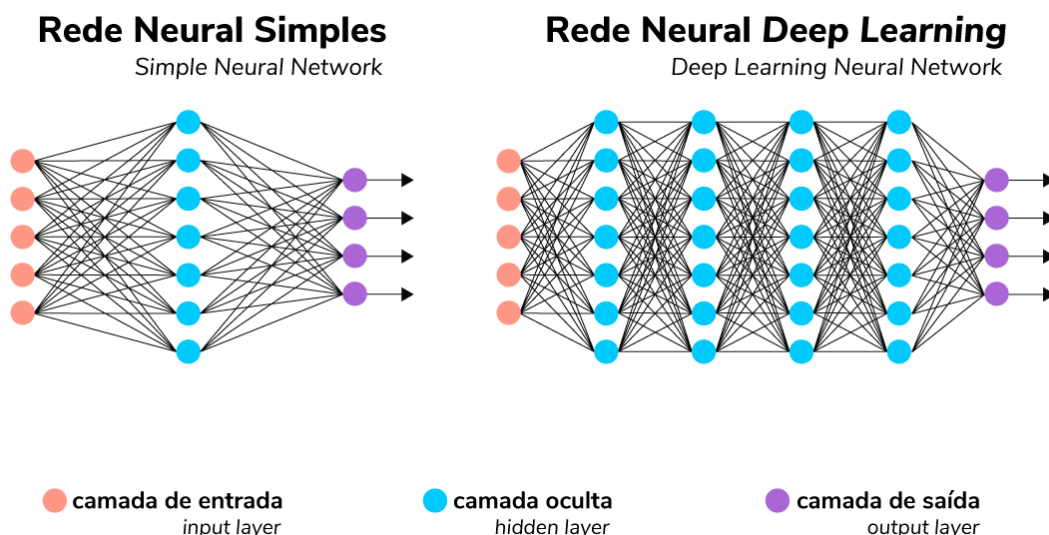


Figura 2 – Ilustração dos tipos de RNAs. Fonte: DSA (2022)

A revolução trazida pelas RNAs e, posteriormente, pelo Aprendizado Profundo, é refletida não apenas em avanços acadêmicos, mas também em aplicações práticas. Setores como saúde, finanças, transporte e entretenimento foram drasticamente transformados.

- Na área médica, por exemplo, sistemas de diagnóstico assistido por computador estão sendo implementados para identificar doenças com precisão comparável ou, em alguns casos, superior à dos especialistas.
- No setor financeiro, algoritmos de Aprendizado Profundo são empregados para detectar fraudes e para automatizar decisões de investimentos, tornando os processos mais rápidos e precisos.
- No setor de transporte, carros autônomos estão sendo desenvolvidos com o uso de Aprendizado Profundo para navegar pelas ruas com segurança.

- No setor de entretenimento, algoritmos de Aprendizado Profundo são usados para gerar imagens, vídeos e música de alta qualidade.

No entanto, a implementação e generalização do Aprendizado Profundo e das RNAs não estão isentas de desafios. A necessidade de grandes conjuntos de dados para treinamento, a explicabilidade dos modelos e a otimização de sua arquitetura são questões em aberto na pesquisa.

- A necessidade de grandes conjuntos de dados para treinamento é um desafio significativo, pois pode ser caro e demorado coletar e preparar dados de alta qualidade.
- A explicabilidade dos modelos é um desafio importante, pois é necessário entender como os modelos tomam decisões para garantir que sejam justos e confiáveis.
- A otimização da arquitetura dos modelos é um desafio técnico complexo, pois é necessário encontrar o equilíbrio certo entre desempenho e complexidade.

Assim, enquanto celebramos os progressos e benefícios destas tecnologias, é imperativo abordar e resolver os desafios inerentes para garantir seu uso eficaz.

3.2.1 Redes Neurais Convolucionais - CNN

Redes neurais convolucionais (CNNs) têm sido estudadas há pelo menos duas décadas em várias áreas de pesquisa, como segmentação, detecção, reconhecimento de objetos e regiões. No entanto, foi após 2012 que as CNNs alcançaram um grande impulso, ao vencerem a competição ImageNet, que utiliza um conjunto de dados de 1 milhão de imagens com 1.000 classes distintas. Essa vitória consolidou o sucesso das CNNs como uma tecnologia promissora para visão computacional (LECUN; BENGIO; HINTON, 2015). O avanço das CNNs também foi impulsionado pelo significativo aumento no poder de processamento nas últimas décadas, juntamente com o desenvolvimento de novas tecnologias focadas em hardware.

A abordagem da rede CNN consiste em aplicar uma variedade de filtros distintos, que podem ser utilizados em dados visuais, como imagens, mantendo a conectividade local entre os pixels vizinhos durante todo o processo de convolução. Assim, as entradas dos neurônios ocultos da camada x são compostas por um subconjunto específico da camada anterior, chamada de camada $x - 1$.

A Figura 3 (A) ilustra a conexão entre as diferentes camadas de uma rede CNN. É importante destacar que, nesse método, cada filtro f_i é replicado em todos os campos visuais da imagem. As unidades replicadas compartilham os mesmos parâmetros, como vetor de peso e polarização, formando um mapa de características, conforme ilustrado na Figura 3 (B).

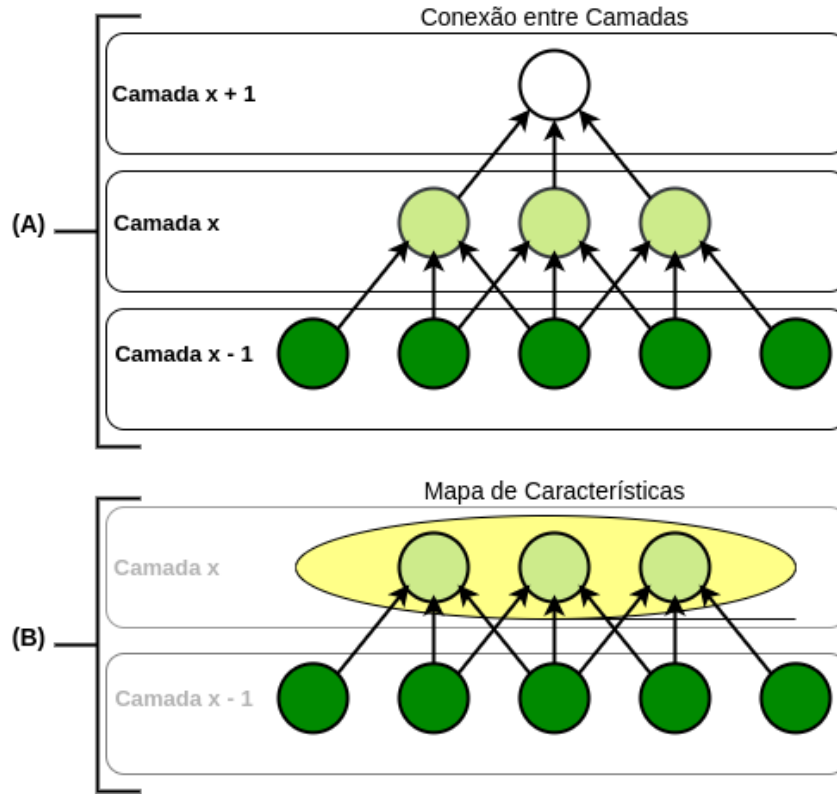


Figura 3 – (A)- Ilustração da ligação(conexão) entre as camadas de uma rede CNN. (B)- Representação de 3 neurônios da camada oculta interligados ao mesmo Mapa de Características. Fonte: O próprio autor.

A Figura 3 (B) ilustra um mapa de características formado por 3 neurônios. Cada mapa de características é formado por uma combinação de características extraídas da camada anterior. Em cada camada oculta, existem diversos mapas de características. O gradiente do peso partilhado é definido pela soma dos gradientes dos parâmetros partilhados entre si. O cálculo do k-ésimo mapa de características f é realizado pela convolução da imagem com um filtro w^k seguido da adição do bias b_k , conforme a equação 3.1.

$$f_{ij}^k = \tanh((w^k * x)_{ij} + b_k) \quad (3.1)$$

O compartilhamento de pesos nas redes CNNs torna as características facilmente detectáveis, pois permite que cada filtro seja aplicado a uma ampla área da imagem. Isso reduz o número de parâmetros a serem aprendidos, o que torna o treinamento mais eficiente.

Na Figura 4 os pesos da camada oculta da CNN ilustrada, são representados por um tensor 4D, com elementos atribuídos a cada conjunto de mapas de características: destino, fonte e posição (vertical e horizontal). As polarizações b são representadas por um vetor específico, contendo um elemento para cada mapa de características de destino.

A Figura 4 ilustra os elementos específicos pertencentes às camadas do processo

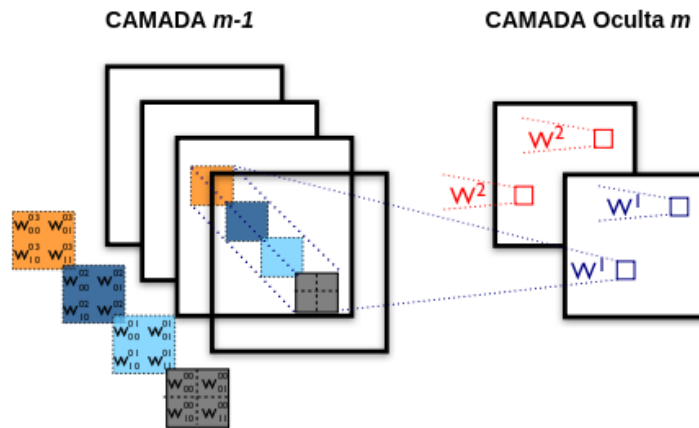


Figura 4 – Ilustração do processo de camadas convolucionais.

convolucional. A camada $m - 1$ contém 4 mapas de características, enquanto a camada Oculta m recebe 2 mapas de características denominados (f^1 e f^2). Os valores dos pixels para os neurônios (f^1 e f^2) na camada Oculta m , representados pelos quadrados azuis e vermelhos, são obtidos a partir dos pixels da camada $m - 1$, conforme ilustrado na Figura 4. Esses valores são então transferidos para as camadas de convolução.

As camadas convolucionais recebem como entrada os 4 mapas de características da camada anterior. Os pesos w^0 e w^1 dos neurônios f^1 e f^2 são representados por tensores com peso 3D. A dimensão principal identifica o mapa de características de entrada, enquanto as duas dimensões restantes representam as coordenadas dos pixels. Cada peso é responsável por conectar todos os pixels do k -ésimo mapa de características da camada m com o pixel de localização (coordenadas; i, j) do l -ésimo mapa de características da camada $m - 1$.

Vale ressaltar que *Max-Pooling* é uma operação importante para Redes Neurais Convolucionais. Ele funciona como um agrupamento máximo, dividindo a imagem de entrada em conjuntos de retângulos não sobrepostos, cada um com uma largura e altura de $p \times p$ pixels. Para cada conjunto, é gerado um valor máximo, que é o valor máximo de todos os pixels no conjunto.

Essa operação reduz o processo computacional para as camadas subsequentes, estabelecendo uma conexão em cascata entre a camada *Max-Pooling* e a camada de convolução.

A ilustração mostrada na Figura 5 ilustra o processo de convolução em uma rede neural convolucional. O quadro destacado na imagem original é a camada de entrada, que é processada por camadas convolucionais intercaladas com camadas de *max-pooling*. Esse procedimento segue a equação 3.1 para calcular as características relevantes.

As camadas superiores são totalmente conectadas, ou seja, consistem em uma camada oculta seguida por uma regressão logística. Nota-se que a entrada para a primeira

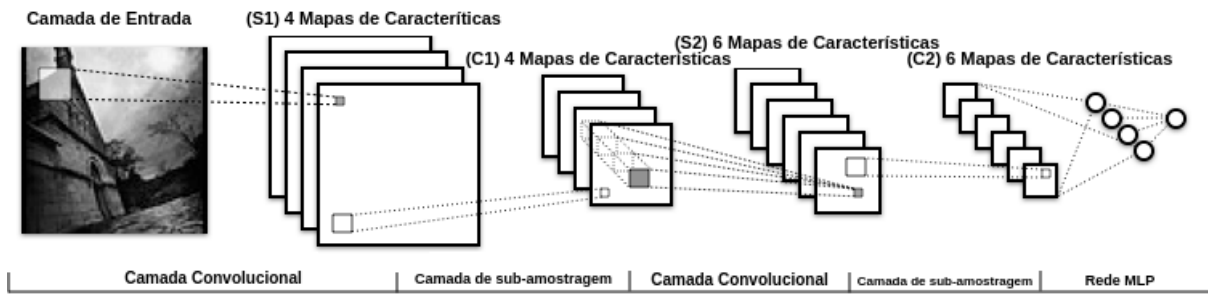


Figura 5 – Ilustração de diferentes camadas encarregadas em um processo de convolucional.

camada é o conjunto de todos os mapas de características da camada $x - 1$, conforme ilustrado na Figura 3 (B).

3.2.2 Redes Neurais Recorrentes - RNN

As Redes Neurais Recorrentes (RNNs) são uma evolução das redes neurais tradicionais que foram projetadas para lidar com dados sequenciais. As redes neurais tradicionais processam dados independentes, um de cada vez. Isso pode ser um problema para dados sequenciais, como texto ou áudio, onde cada item na sequência tem uma relação com os itens anteriores. As RNNs, por outro lado, mantêm um estado interno que armazena informações sobre o contexto de dados anteriores. Isso permite que as RNNs usem essas informações para prever o próximo item na sequência. Por exemplo, uma RNN pode ser usada para traduzir texto de um idioma para outro. A RNN pode usar o contexto da frase anterior para prever a próxima palavra na tradução.

As RNNs são uma ferramenta poderosa para lidar com dados sequenciais. Elas são usadas em uma variedade de aplicações, incluindo tradução automática, reconhecimento de fala, processamento de linguagem natural e geração de texto.

Redes neurais recorrentes (RNNs) são um tipo de rede neural que armazena informações sobre itens anteriores em uma sequência. Isso é feito por meio de um *loop* interno que permite que a rede transmita informações de uma unidade para a outra.

As RNNs são compostas por três tipos de camadas: entrada, oculta e saída. A camada de entrada recebe a sequência de dados, a camada oculta processa a sequência e mantém uma representação interna do estado da rede por meio do *loop* interno, e a camada de saída produz as previsões. A Figura 6 ilustra uma comparação entre as camadas ocultas de uma rede neural *feed-forward* e uma rede recorrente.

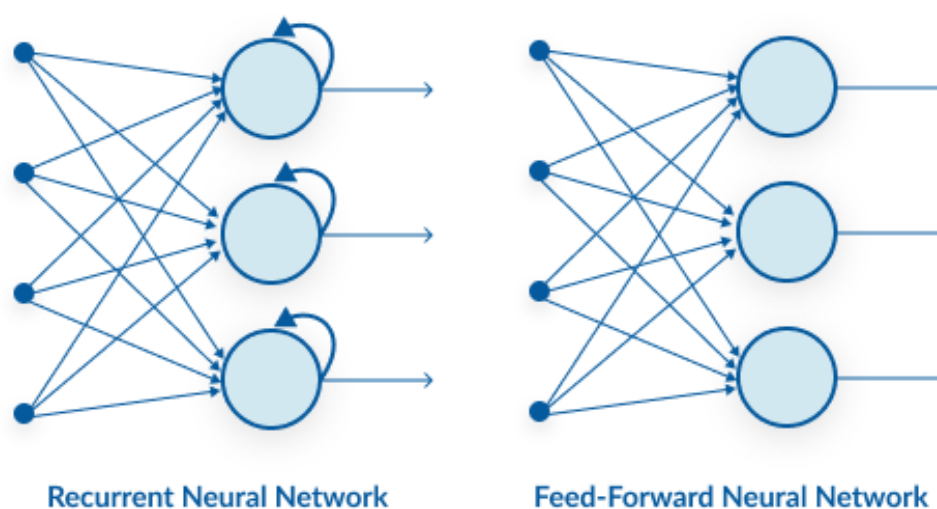


Figura 6 – Ilustração do *loop* da camada oculta de uma RNN Vs. Rede neural *Feed-Forward*.
Fonte: [Kalita \(2022\)](#)

3.2.3 Redes Transformers

As redes Transformer são uma arquitetura de rede neural que foi introduzida em 2017. Ela é capaz de processar dados sequenciais, como texto, áudio ou imagens, de forma paralela. Isso a torna uma opção eficaz para tarefas de processamento de linguagem natural, como tradução automática, análise de sentimentos, geração de texto e classificação de tópicos ([VASWANI et al., 2023](#)).

A arquitetura Transformer é uma arquitetura de rede neural que processa dados sequenciais de forma paralela, sem a necessidade de processar cada entrada da sequência sequencialmente. Isso torna a arquitetura Transformer muito eficiente para treinar modelos de redes neurais em grandes conjuntos de dados.

3.2.3.1 Arquitetura

A arquitetura Transformer original é composta por um codificador de seis camadas e um decodificador de seis camadas. O modelo utiliza mecanismos de auto atenção para capturar as dependências de longo alcance em uma sequência de entrada ([VASWANI et al., 2023](#)), como ilustrado na Figura 7.

O modelo Transformer recebe uma sequência de vetores de palavras (*embedding*) como entrada. Esses vetores são passados para uma série de camadas de codificação, que são compostas por dois subcampos: um módulo de auto atenção e um módulo de *feed-forward*. O módulo de auto atenção calcula pesos de atenção para cada palavra em relação a todas as outras palavras da sequência. Esses pesos indicam a importância relativa de cada palavra para a compreensão da sequência. O módulo de *feed-forward*, por sua vez, processa cada palavra individualmente, aplicando uma função não linear aos vetores de

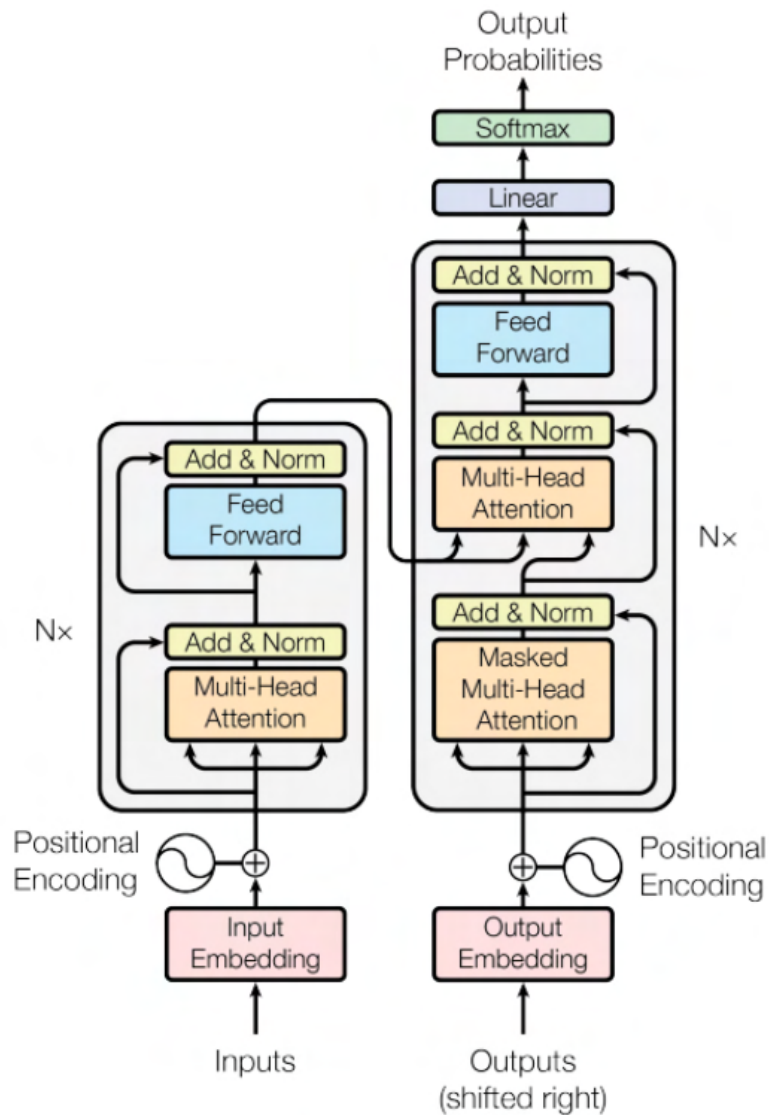


Figura 7 – Arquitetura geral de uma rede Transformer. Fonte: (VASWANI et al., 2023)

palavras. A combinação desses dois subcampos permite que o modelo Transformer capture as dependências de longo alcance entre as palavras em uma sequência. Isso é essencial para tarefas de processamento de linguagem natural, como tradução automática e análise de sentimentos.

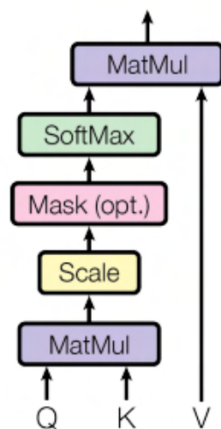
O modelo Transformer usa um mecanismo de normalização residual após cada camada de codificação. Isso ajuda a estabilizar o treinamento do modelo e a evitar que ele fique sobreajustado. O processo de codificação é realizado várias vezes, geralmente seis vezes. Isso permite que o modelo capture informações de diferentes granularidades.

Depois de serem codificadas, as sequências de entrada são passadas para o decodificador. O decodificador é semelhante ao codificador, mas também incorpora informações da saída do modelo. Isso permite que o decodificador gere uma saída que seja relevante para a tarefa em questão.

O decodificador também utiliza mecanismos de auto atenção para gerar uma saída em cada etapa do processo. Isso permite que o decodificador capture as dependências de longo alcance entre as palavras na saída.

O **mecanismo de atenção** da arquitetura Transformer é semelhante à atenção humana. Ele permite que o modelo dê mais importância a partes específicas da entrada, dependendo da tarefa em questão (VASWANI et al., 2023).

Scaled Dot-Product Attention



Multi-Head Attention

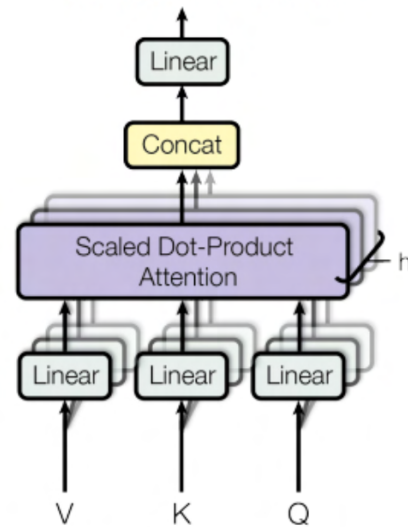


Figura 8 – *Scaled Dot-Product Attention* - Cálculo para matriz de atenção dos tensores Q , K , V . Representa a função de atenção em 3.4

Multi-Head Attention - Múltiplas camadas de atenção que executam em paralelo.

Representa a função de atenção multi-cabeça em 3.3

Fonte:(VASWANI et al., 2023)

O mecanismo de atenção da arquitetura Transformer usa três tensores para calcular a importância de cada palavra em uma sequência de entrada, como ilustrado na Figura 8. Esses tensores são chamados de consulta, chave e valor.

Primeiro, a entrada é transformada em três tensores (Q - Consulta, K - Chave, V - Valor) por meio de camadas lineares independentes. O vetor de consulta é usado para calcular a importância de cada palavra em relação a todas as outras palavras na sequência. Os vetores chave e valor são usados para codificar informações contextuais sobre cada palavra.

O modelo usa uma função de similaridade para calcular a importância de cada palavra em relação a todas as outras palavras na sequência. Essa função é chamada de função de atenção. A função de atenção calcula o *score* de atenção para cada palavra. O *score* de atenção é uma medida de quão importante é uma palavra para outra. Os scores de atenção são então normalizados por uma função Softmax. Isso significa que os scores são convertidos em uma distribuição de probabilidades. A função de atenção é representada

pela equação 3.4, onde d_k é a dimensão do vetor chave.

$$Atencao(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right) V \quad (3.2)$$

Após a normalização dos scores de atenção, o modelo usa esses scores para ponderar os tensores de valor de cada palavra. Isso significa que cada palavra da sequência é ponderada com base na importância que o modelo atribui a ela.

O resultado desse processo é uma representação contextualizada para cada palavra da sequência. Essa representação contextualizada é uma combinação das informações da palavra em si e do contexto das palavras ao seu redor.

A representação contextualizada é então passada para a próxima camada de codificação do modelo. Todo esse processo é repetido várias vezes em camadas de codificação subsequentes para permitir que o modelo capture informações de diferentes granularidades. A equação 3.3 descreve este processo de agregação dos resultados da função de atenção. Onde h é a quantidade de camadas ou cabeças de atenção

$$\begin{aligned} MultCabecas(Q, K, V) &= Concat(cabeca_1, \dots, cabeca_h)W^O \\ \text{onde } cabeca_i &= Atencao(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (3.3)$$

Onde as matrizes com os parâmetros são projetados por $W_i^Q \in \mathbb{R}^{d_{model} \times d_q}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ e $W_i^O \in \mathbb{R}^{h d_v \times d_{model}}$. Todo o processo deste mecanismo de atenção é ilustrado na Figura 8.

3.2.4 Transformers para visão computacional - *Vision Transformers*

Inspirado na arquitetura de Transfromes que revolucionou a área de **PLN** (**P**rocessamento de **L**inguagem **N**atural), foi proposto uma arquitetura de Transformer em tarefas de reconhecimento de imagem, o **ViT** (**V**ision **T**ransformer) (DOSOVITSKIY et al., 2021).

A ideia principal da arquitetura Transformer para processamento de imagens é tratar as imagens como uma sequência de *patches* - recortes retangulares da imagem original. Na primeira etapa, os *patches* são extraídos da imagem original. Em seguida, esses *patches* são passados pela arquitetura Transformer, que extrai características de cada *patch*. Na segunda etapa, as características de cada *patch* são agregadas para criar uma representação global da imagem. Essa representação global pode ser usada para várias tarefas de processamento de imagens, como classificação, detecção de objetos e segmentação de imagens. Então Essa representação global é então usada como entrada para uma camada classificadora, que atribui um rótulo à imagem, como representado pela Figura 9

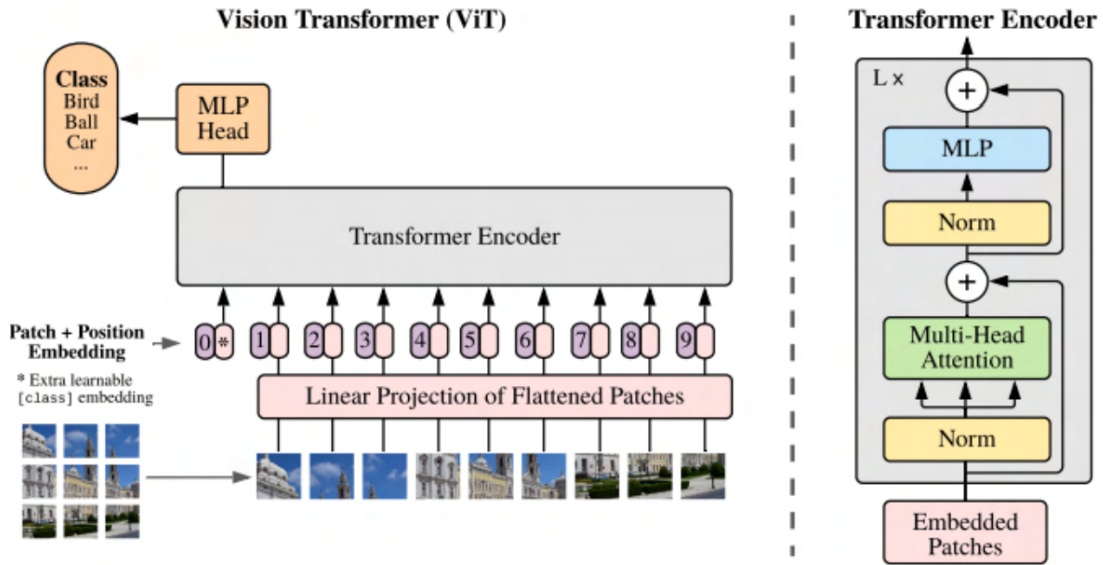


Figura 9 – Visão geral do ViT. Fonte: (DOSOVITSKIY et al., 2021)

A arquitetura Transformer aplicada a visão computacional é composta apenas por camadas codificadoras. Essas camadas são suficientes para tarefas de extração de características e classificação de imagens.

No entanto, para tarefas de geração de imagens, a adição de camadas decodificadoras é necessária. Essas camadas são responsáveis por gerar novas representações a partir de uma representação codificada. Assim, a arquitetura Transformer pode ser usada para uma variedade de tarefas de visão computacional, desde classificação e detecção de objetos até geração de imagens.

3.2.5 Detecção - Detectron2

O Detectron2 é uma poderosa CNN capaz de classificar o conjunto de pixels na imagem, delimitando a região de interesse por uma caixa delimitadora (bounding box) e uma máscara delimitadora (delimiting mask). A capacidade do Detectron2 em fazer isso se deve ao fato de, em seu processo de extração e classificação, delimitar as regiões com maior probabilidade de localizar a região de interesse através da marcação da caixa delimitadora durante a primeira etapa de detecção. Na segunda etapa de funcionamento do Detectron2, a rede classifica os pixels dentro das caixas delimitadoras, construindo uma máscara binária correspondente aos pixels que pertencem à região de interesse. A segunda etapa de detecção do Detectron2 pode ser usada na segmentação de regiões, pois a máscara binária fornecida corresponde aos resultados de várias técnicas do estado da arte (WU et al., 2019). Para o problema de detecção da hemorragia causada pelo AVC em imagens de TC, o Detectron2 pode fornecer resultados promissores, apesar do desafio de detectar e segmentar a região afetada pela lesão, mesmo em tipos de exames desfavoráveis para visualização (DIVYA; PETER, 2021).

3.2.6 Segmentação de instancia

Neste estudo, são apresentados diferentes métodos de segmentação, porém com ênfase nos experimentos de 2 tipos de modelos, *Zero-Shot* e *One-Stage*.

Modelos *One-Stage* são um tipos de modelo de detecção que tenta detectar todos os objetos em uma imagem em uma única etapa. Eles geralmente são treinados em um conjunto de dados de imagens com rótulos de objetos (BOLYA et al., 2019).

Modelos *Zero-Shot* são um tipo de modelo de aprendizado de máquina que pode realizar uma tarefa sem ter sido treinado explicitamente para essa tarefa. Eles geralmente são treinados em um conjunto de dados de imagens com rótulos de atributos, como cor, forma ou tamanho (KIRILLOV et al., 2023).

A principal diferença entre modelos *One-Stage* e *Zero-Shot* é que modelos *One-Stage* precisam ser treinados em um conjunto de dados de imagens com rótulos de objetos, enquanto modelos *Zero-Shot* podem ser treinados em um conjunto de dados de imagens com rótulos de atributos.

Modelos *One-Stage* geralmente são mais eficientes em tempo de execução do que modelos *Zero-Shot*, pois precisam realizar apenas uma etapa para detectar todos os objetos em uma imagem. No entanto, Modelos *One-Stage* geralmente são menos precisos do que modelos *Zero-Shot*, pois precisam aprender a identificar objetos com base em seus atributos, que podem ser menos discriminatórios do que rótulos de objetos.

Modelos *Zero-Shot* geralmente são mais precisos do que Modelos *One-Stage*, pois podem aprender a identificar objetos com base em seus atributos, que podem ser mais discriminatórios do que rótulos de objetos. No entanto, modelos *Zero-Shot* geralmente são menos eficientes em tempo de execução do que Modelos *One-Stage*, pois precisam realizar duas etapas para detectar todos os objetos em uma imagem: primeiro, eles precisam identificar os objetos na imagem com base em seus atributos; segundo, eles precisam classificar os objetos identificados.

Em geral, modelos *One-Stage* são uma boa opção para tarefas de detecção de objetos que precisam ser executadas de forma eficiente em tempo de execução. Modelos *Zero-Shot* são uma boa opção para tarefas de detecção de objetos que precisam ser precisas, mesmo para objetos que não foram vistos no conjunto de dados de treinamento.

3.2.6.1 SAM (**Segment Anything Model**)

O modelo **SAM** (*Segment Anything Model*) tem como objetivo ser um *foundation model* ou modelo de base para a tarefa de **segmentação**. Ainda que para que tenha este título de modelo de base, seja necessário observar como esta sendo usado pela comunidade. *Foundation models* ou modelos de base, basicamente são modelos pré-treinados que podem ser adaptados para tarefas específicas, um conceito entendido desde os primeiros estudos

de aprendizagem de maquina (THRUN, 1995). Porém, nos ultimos anos este paradigma vem sendo redefinido por conta da enfase em escala, então foi cunhado o nome "*foundation model*": *i.e.*: modelos que são "treinados em uma grande quantidade de dados em escala e são adaptáveis em uma ampla variedade de tarefas especificas"(BOMMASANI et al., 2022).

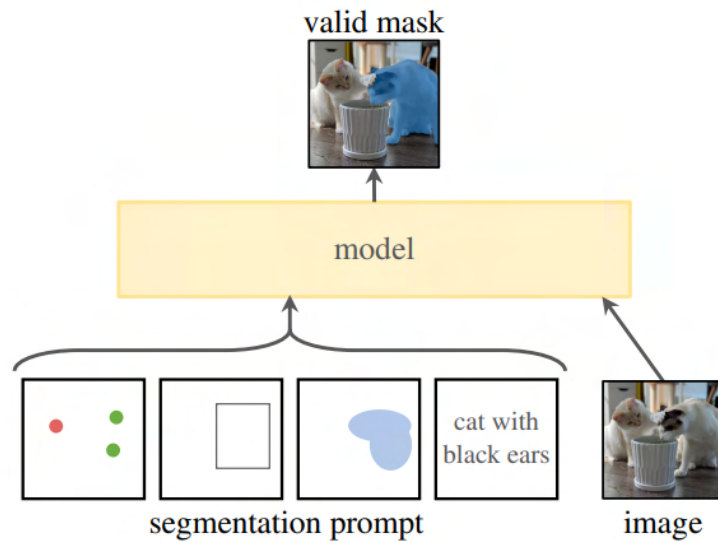


Figura 10 – **Tarefa**: entradas da solicitação de segmentação de um objeto em uma imagem. Fonte:(KIRILLOV et al., 2023)

O modelo **SAM** tem a capacidade de produzir mascaras de objetos com alta qualidade a partir de entradas ou solicitações como: um ponto, uma caixa delimitadora ou texto livre, como mostra a Figura 10. Para isso o modelo foi construído sobre *Transformer vision models*(CARION et al., 2020; DOSOVITSKIY et al., 2021; CHENG; SCHWING; KIRILLOV, 2021; LI et al., 2022) e possui três componentes, apresentados na Figura 11: um *image encoder*, um *flexible prompt encoder* e um *mask decoder*. Esta arquitetura possibilita a generalização com um excelente desempenho na segmentação em novas imagens digitais, sem ser previamente treinado para tal.

Codificador da Imagem (*Image Encoder*): Em suma pode ser qualquer rede neural computacional que retorne uma representação matemática da imagem no formato $C \times A \times L$ onde C representa os canais, A altura e L largura. Contudo, foi utilizado o ViT-H/16 um ViT (*Vision Transformer*)(DOSOVITSKIY et al., 2021) um MAE (*Masked AutoEncoder*) (HE et al., 2021) pré-treinado e preparado para imagens de alta resolução. O resultado é uma representação matemática da imagem de entrada 16× menor que a imagem de entrada original. Afim de facilitar o entendimento, define-se estas representações matemáticas como *embeddings*. Seguindo as melhores práticas de (GHIASI et al., 2021), a resolução da imagem de entrada é ajustada para 1024×1024 por meio de redimensionamento e preenchimento se for menor. E para que tamanho do

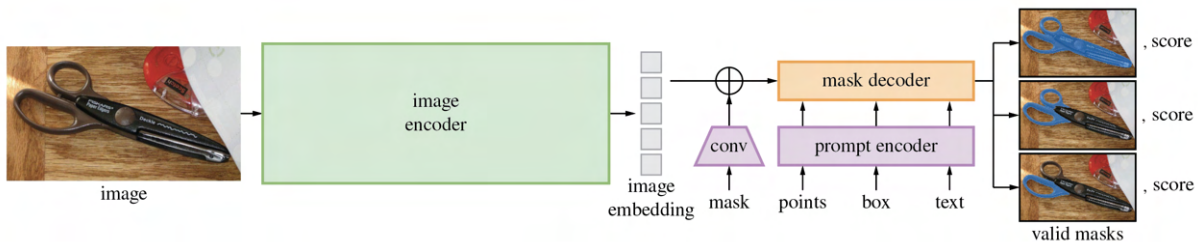


Figura 11 – Visão geral do modelo **SAM**: Um codificador de imagem gera um *embedding* de imagem que pode então ser consultada com eficiência por uma variedade de solicitações de entrada para produzir máscaras de objetos em velocidade amortizada em tempo real. Para solicitações ambíguas correspondentes a mais de um objeto, o SAM pode gerar múltiplas máscaras válidas e pontuações de confiança associadas. Fonte: (KIRILLOV et al., 2023)

resultado seja 64×64 a dimensão de cores da imagem deve ser reduzida para 256 canais, como apresentado em (LI et al., 2022), é aplicado uma convolução 1×1 seguida de outra 3×3 . Cada convolução é seguida por uma normalização de camada (BA; KIROS; HINTON, 2016).

Codificador de Solicitação Flexível (*Flexible Prompt Encoder*): As solicitações ou *prompts* de segmentação, são mapeadas para um *embedding* vetorial de 256 dimensões, os três tipos de solicitações são representados da seguinte forma: Um **ponto**, e representado como a soma de uma codificação posicional (TANCIK et al., 2020) dada a localização do ponto e uma **codificação de aprendizado** que indica se o ponto está em primeiro plano ou no fundo da imagem; Uma **caixa**, é constituído por um par de codificações - (1) a soma da codificação posicional do ponto referente ao canto superior esquerdo e a **codificação de aprendizado** e (2) seguindo a mesma lógica mas para o canto inferior direito da caixa; E por fim, para codificar solicitações do tipo **texto livre** foi utilizado o *text encoder* CLIP (RADFORD et al., 2021).

Decodificador da Mascara (*Mask Decoder*): Responsável por mapear os *embeddings* da imagem e do prompt para contruir uma mascara como resultado. Para isso, antes de aplicar o decodificador nos *embeddings*, é adicionado um identificador nos *embeddings* de solicitação chamado de *output token*, analogo ao token *[class]* definido em (DOSOVITSKIY et al., 2021). Para facilitar o entendimento, este identificador será chamado de "*token*". Cada camada do decodificador executa quatro etapas, apresentadas na Figura 12:

1. Autoatenção ou *self-attention* nos tokens: Esta etapa permite que os tokens interajam entre si e aprendam as dependências entre eles;
2. Atenção cruzada ou *cross-attention* dos tokens para os *embeddings* da imagem: permite que os tokens aprendam informações da imagem;

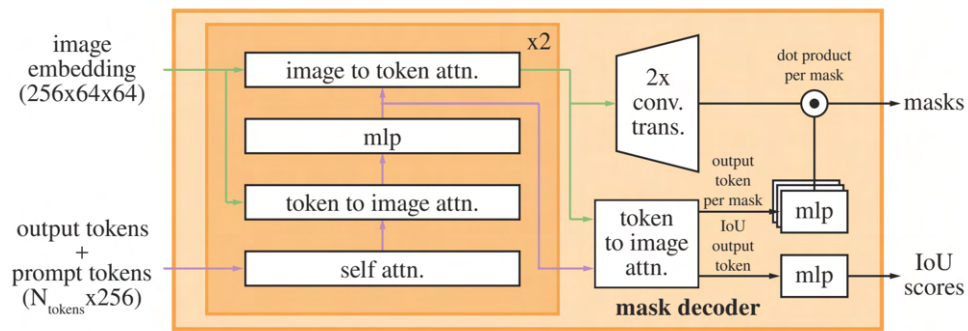


Figura 12 – Detalhes do Decodificador da Mascara (*Mask Decoder*). Um decodificador de duas camadas atualiza embedding de imagens e os tokens da solicitação por meio de atenção cruzada. Em seguida, a incorporação da imagem é ampliada, a partir da qual os tokens de saída atualizados são usados para prever máscaras dinamicamente. (Não ilustrado para maior clareza da figura: em cada camada de atenção, codificações posicionais são adicionadas à incorporação da imagem e todo o token de solicitação original (incluindo codificação de posição) é adicionado novamente às consultas e chaves do token.) Fonte: (KIRILLOV et al., 2023)

3. uma MLP atualiza cada token: Esta etapa atualiza cada token de acordo com as informações aprendidas nas etapas anteriores.
4. Atenção cruzada ou *cross-attention* do *embeddings* da imagem para os tokens: Esta etapa atualiza a *embeddings* da imagem com informações da solicitação.

Para garantir que o decodificador tenha acesso a informações geométricas críticas, as codificações posicionais são adicionadas ao *embedding* da imagem sempre que participam de uma camada de atenção. Além disso, todos os tokens da solicitação original (incluindo suas codificações posicionais) são readicionados aos tokens atualizados sempre que participam de uma camada de atenção. Isso permite uma forte dependência tanto da localização geométrica quanto do tipo de token da solicitação.

Após a execução do decodificador, o *embedding* da imagem atualizada é aumentada em 4x com duas camadas convolucionais transpostas e uma é uma MLP (*Multilayer Perceptron* - Rede Neural Perceptron Multicamada) mapeia o token de saída para um classificador linear dinâmico, que então calcula a probabilidade de primeiro plano da máscara em cada local da imagem.

Contudo, para o modelo SAM atingir um alto grau de generalização, foi pré-treinado em uma nova base de dados **SA-1B** (KIRILLOV et al., 2023), contendo 11M imagens variadas em alta resolução, licenciadas e com proteção de privacidade e mais de 1B de máscaras de segmentação.

3.2.6.2 MobileSAM (**M**obile **S**egment **A**nthing **M**odel)

O modelo MobileSAM tem como principal objetivo melhorar a performance de seu modelo base o SAM 3.2.6.1, e habilitar o seu uso em dispositivos moveis, por isso o prefixo *Mobile*. Para atingir tal objetivo, foi realizado um estudo na topologia do método afim de identificar os pontos de melhoria. Os principais pontos identificados afim de tornar o modelo SAM otimizado para dispositivos móveis são: o Codificador de Imagem (*Image Encoder*); e o Decodificador da Máscara (*Mask Decoder*) (ZHANG et al., 2023).

Através do processo de destilação dissociada, o aprendizado do *Image Encoder* do modelo original SAM 3.2.6.1, o ViT-H/16 um ViT (**V**ision **T**ransformer)(DOSOVITSKIY et al., 2021) foi destilado para um novo *Image Encoder* mais leve e mais rapido. Por conta do processo de transferencia de aprendizado utilizado ser a destilação de conhecimento, o novo *Image Encoder* é automaticamente compativel com o *Mask decoder* original, por isso os experimemntos realizados em (ZHANG et al., 2023) não realizaram nenhuma melhoria neste componente, porém deixando de maneira empirica que é possivel melhorar os resultados de precisão, realizando um processo de ajuste fino no componente *Mask decoder*. A Figura 13 ilustra o processo de destilação dissociada, indicando que o ajuste fino no *Mask Decoder* é opcional.

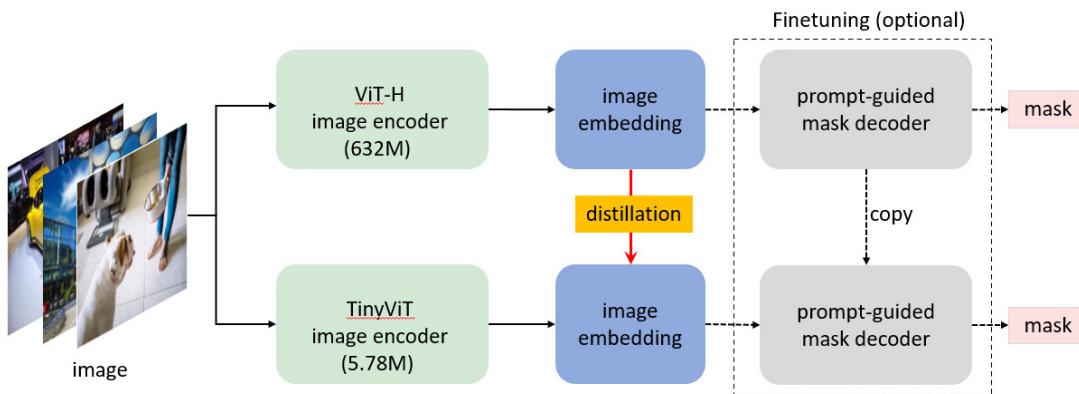


Figura 13 – **Tarefa:** Destilação dissociada. Fonte:(ZHANG et al., 2023)

O vale ressaltar que o processo de destilação do conhecimento, foi realizado o treinamento utilizado apenas 1% da base de dados SA-1B (KIRILLOV et al., 2023) por 8 épocas e com uma unica GPU. O processo seguiu o método proposto por (TOUVRON et al., 2021) para torna-lo mais rápido. Contudo, a Tabela 1 mostra uma comparação de quantidade de parâmetros e velocidade do componente *Image Encoder*, e a quantidade de parametros foi reduzido de 632M de parâmetros para apenas 5,78M, e ficou quase 60× mais rapido que o modelo original de 452ms para 8ms.

3.2.6.3 YOLACT (**Y**ou **O**nly **L**ook **A**t **C**oefficien**T**s)

A rede YOLACT é uma rede neural convolucional profunda que foi projetada para realizar segmentação de instâncias em tempo real e **One-Stage** ou em unica etapa.

Tabela 1 – Comparação da quantidade de parâmetros e velocidade do modelo original SAM (KIRILLOV et al., 2023) e o modelo MobileSAM (ZHANG et al., 2023). A verificação da velocidade se deu utilizando apenas uma GPU.

Método	Parâmetros	Tempo
SAM	632M	452ms
MobileSAM	5,78M	8ms

O processo de Segmentação em única etapa, utiliza técnicas como mapas sensíveis à posição e *pooling* sensível à posição (DAI et al., 2016; LI et al., 2017), onde os **mapas sensíveis a posição** são: representações intermediárias de imagens que codificam a localização espacial de características relevantes, como bordas ou regiões de objetos; e ***pooling* sensível à posição** é uma técnica de processamento de imagens que combina informações de mapas sensíveis à posição para gerar máscaras de objetos mais precisas. Além de combinações de: **logits de segmentação semântica** - valores numéricos que representam a probabilidade de cada pixel de uma imagem pertencer a uma determinada classe de objeto; e **Logits de previsão de direção** - representam a probabilidade de cada pixel de uma imagem estar na borda de um objeto (CHEN et al., 2017). Para gerar máscaras de objetos precisas.

Em contraste, a rede Mask-RCNN (HE et al., 2018) é um exemplo segmentação em duas etapas, onde a primeira etapa gera candidatos a **RoI** (*Region of Interests*) ou regiões de interesse e então os classifica e segmenta na segunda etapa. Estas 2 etapas necessitam que seja aplicado métodos de *re-pooling* nas características em cada região de interesse e processa-los novamente, o que acaba onerando muito o tempo de computação criando uma forte resistência para que se obtenha uma velocidade compatível com tempo real.

Dada a diferença entre os processos de segmentação de etapa única ou duas etapas, ainda que conceitualmente a etapa única seja mais rápida, é necessário aplicar *re-pooling* ou algum outro processamento como um cálculo por votos, para definir a máscara mais assertiva para o problema. Isto posto, este processamento afeta a velocidade, impondo limites para que se atinja uma performance considerável para segmentação em tempo real.

Para habilitar a segmentação em etapa única em tempo real a rede YOLACT propõem executar a tarefa de segmentação junto com a detecção em uma tarefa única, adicionando um codificador de máscara no cálculo. Para isso, além de juntar as duas tarefas, dividiu-se a tarefa de segmentação em 2 sub-tarefas executadas em paralelo: (1) gera os protótipos das máscaras; e (2) predição das máscaras por instância com base no coeficiente de máscara, como ilustrado na Figura 14. E por serem processos independentes, os resultados são facilmente combinados linearmente para produzir as máscaras de cada objeto (BOLYA et al., 2019).

Contudo, para gerar as máscaras com base nos resultados dos processos em paralelo,

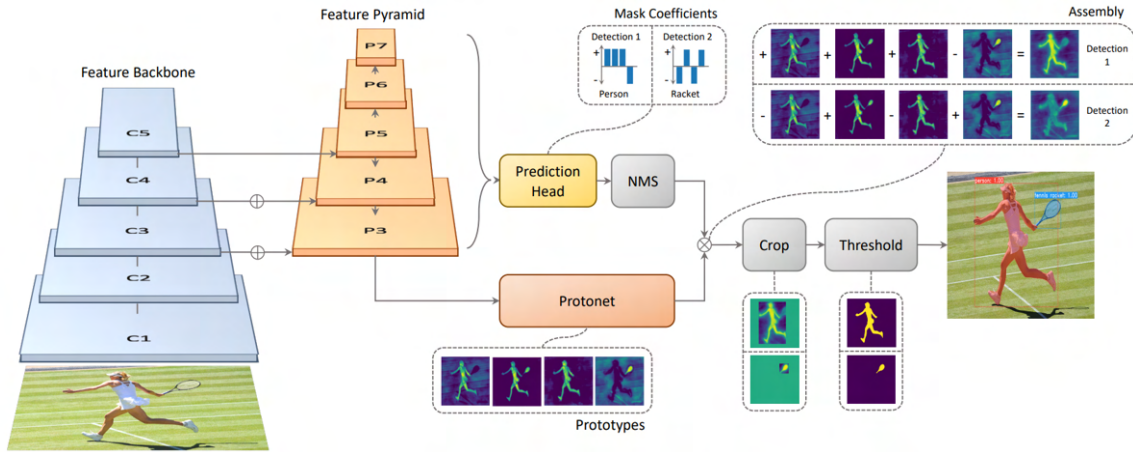


Figura 14 – Arquitetura da rede YOLACT

Fonte: (BOLYA et al., 2019)

a combinação é feita de forma linear, para ponderar esses resultados. Após a combinação linear, é aplicada uma função não linear chamada sigmoide para refinar os resultados e gerar as máscaras finais. Esta operação é implementada de forma eficiente usando uma simples multiplicação de matrizes e uma sigmoid, como apresentado em ??:

$$M = \sigma(PC^T) \quad (3.4)$$

onde P é uma matriz $a \times l \times k$ dos protótipos de máscaras e C é uma matriz $n \times k$ dos coeficientes de máscaras para as n instancias que sobraram após aplicação do **NMS** (*Non-Maximum Suppression* - Supressão de Sobreposições) e do limiar de coeficiente. Outras combinações são possíveis, porém os autores decidiram manter simples e rápido (BOLYA et al., 2019).

No entanto, ainda que a YOLACT consiga produzir máscaras de alta qualidade, ainda está muito atrás do estado da arte em segmentação, ainda que seja muito mais rápido.

4 Materiais e Métodos

A segmentação precisa da região afetada por AVC hemorrágico em imagens de TC cranianas é crucial para o diagnóstico e tratamento adequados. Este capítulo apresenta o método proposto para realizar essa tarefa de forma precisa e eficiente. Abordamos em detalhes o conjunto de dados de imagens de TC utilizado, a arquitetura da rede neural profunda utilizada para segmentação, as métricas de avaliação de performance, a etapa de detecção de objetos utilizada para inicializar o processo de segmentação e a forma como as tarefas de detecção e segmentação são combinadas para delinear com precisão a região do AVC hemorrágico. Ao fornecer uma descrição completa do método e da combinação de tarefas, este capítulo contribui para a compreensão e replicação do estudo, além de fomentar pesquisas futuras na área de segmentação de AVC hemorrágico em imagens de TC.

4.1 Conjunto de Dados

O banco de dados foi composto por 25 exames de Tomografia Computadorizada (TC) de pacientes diagnosticados com acidente vascular cerebral isquêmico, acidente vascular cerebral hemorrágico e de pacientes saudáveis. Para o banco de dados de teste, foram selecionadas 100 imagens de TC de pacientes com acidente vascular cerebral hemorrágico, a Figura 15 apresenta alguns exemplo. Estas imagens foram gentilmente fornecidas pelo Hospital do Coração de Fortaleza, Ceará, Brasil, e já foram utilizadas com sucesso em estudos anteriores.

As imagens de TC foram adquiridas utilizando o modelo tomográfico General Medical System HiSpeed (GEMSH). Elas foram obtidas no plano axial, com uma espessura de fatia de 0,7 mm e um campo de visão de 230 mm. As condições de aquisição incluíram uma tensão do tubo de 120 kV, uma corrente do tubo de 80 mA e dimensões de 512 x 512 pixels. Além disso, a janela de voxel foi de 0,585 x 0,585 x 1,5 mm, com profundidade de 16 bits, uma faixa de 40 HU e um nível igual a 80 HU.

As imagens foram armazenadas no formato Digital Imaging and Communications in Medicine (DICOM). O banco de dados original, fornecido pelos autores, já estava pré-processado, contendo segmentação da região de acidente vascular cerebral e mapas de densidade radiológica. Para manter a privacidade dos pacientes, o banco de dados não inclui qualquer identificação pessoal.

Aplicamos a técnica de Data Augmentation no conjunto de dados original, afim de melhorar a qualidade do treinamento dos modelos utilizando técnicas como rotação,



Figura 15 – Exemplos do conjunto de dados - Imagens de diferentes fatias de TC de lesão cerebral por AVC Hemorrágico.

espelhamento e translação. Isso resultou na criação de 300 amostras sintética adicionais, expandindo, assim, o número de amostras disponíveis para o treinamento.

4.2 Métodos propostos

Esta seção apresenta a metodologia do estudo proposto para detecção e segmentação de AVC hemorrágico em imagens de TCs. O experimento é dividido em duas fases: Na primeira fase é apresentada o método de detecção da região do AVC hemorrágico por meio de rede de aprendizagem profunda. Na segunda fase, é apresentado o fluxo do método proposto para segmentação por instancia da região afetada pelo AVC hemorrágico se aproveitando da generalização do aprendizado das redes com a segmentação *Zero-Shot*. E finalmente a apresentação de um experimento para segmentação utilizando uma rede do tipo *One-Stage* afim de comparação.

A Figura 16 representa as diferentes etapas deste estudo. O modelo proposto por este estudo é dividido em 2 etapas: Fase 1 - Detecção de AVC Hemorrágico e Fase 2 - Segmentação, seguidas pela etapa de resultados.

Etapas 1 - Detecção de AVC Hemorrágico: Nesta primeira fase, é utilizada a rede convolucional Detectron2, uma rede de aprendizagem profunda capaz de identificar regiões em imagens, previamente treinada. A Figura 16 apresenta a Fase 1 na imagem que corresponde o processo de treinamento do modelo com a rede Detectron 2. Na Etapa (A), é realizado o treinamento do modelo com as imagens especificadas na Subseção 4.1. Após o processo de treinamento é gerado um modelo capaz de identificar AVC hemorrágico.

Após esse processo de treinamento, Na Etapa (B) é ilustrado o processo de inicia-

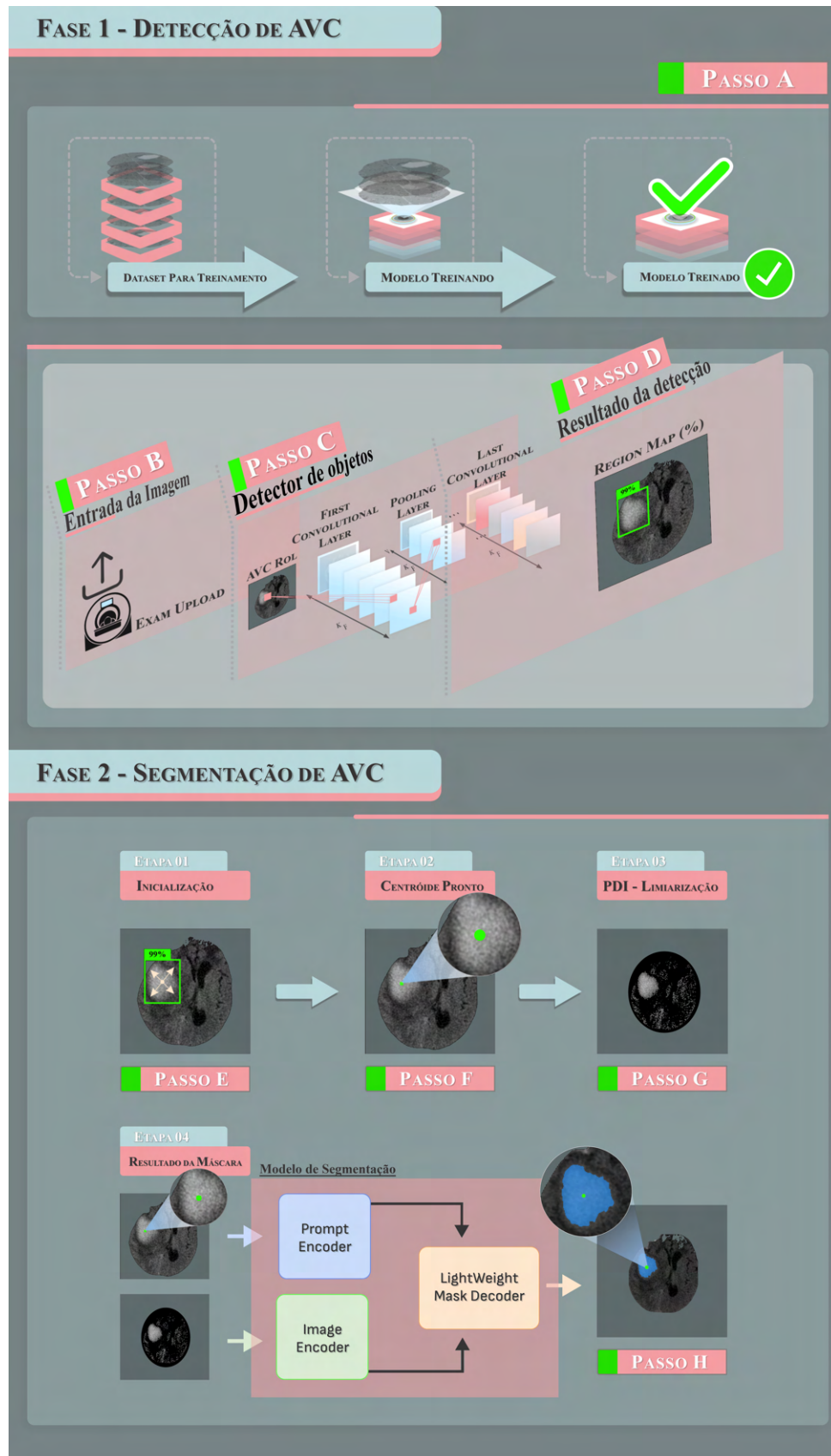


Figura 16 – Método proposto para segmentação de AVC hemorrágico.

lização da detecção de AVC hemorrágico dentro desta primeira fase, (Fase 1), onde (B)

representa a imagem de entrada, (C) representa o modelo da rede Detectron2 treinado previamente para detecção da região de interesse (AVC hemorrágico), como mostra a Figura 16 Módulo (C).

E por fim, na Etapa (D) na Figura 16, mostra o resultado da detecção por meio de *bounding box*, ou seja, a delineação enquadrando o AVC hemorrágico com uso do treinamento da rede, e o respectivo valor de acurácia para detecção do objeto de interesse, como mostra a imagem em (D) Fase 1. Assim, o resultado da saída da rede Detectron2, por meio do bounding box, é a base da inicialização do método proposto para segmentação totalmente automático.

Etapa 2 - Segmentação: Nesta Etapa, vários modelos computacionais são utilizados para delinear com precisão a imagem de saída da Etapa 1, que contém a detecção da região com o AVC. Então são aplicados o métodos propostos neste estudo para comparação com outros modelos da literatura que utilizam a mesma base de dados e apurar sua eficácia para segmentação de AVC hemorrágico em imagens de TCs. A Etapa 2 representa a segmentação do AVC hemorrágico, ou seja, a delineação das bordas do AVC, identificando a região de barreira final da hemorragia cerebral. Métodos desenvolvidos neste estudo:

A) Métodos Propostos $SAM_{centroid}$ e $MobileSAM_{centroid}$: A aplicação deste método é dividida em (Estágios) com as respectivas etapas do processo, para melhor entendimento da segmentação. Estágios:

1. **Etapas 1 e 2 - Inicialização** A Etapa 1 de Segmentação, representa a inicialização da segmentação, onde a imagem de saída da detecção, é insumo para calcular a entrada para o método de segmentação. Em (E), a imagem de entrada pós identificação da região do AVC é recebida e calculada o centroide do AVC com base no *bounding box* capturado na primeira etapa da metodologia (Fase 1). O cálculo do centroide pode ser representado pela equação 4.1, onde trás o processo de inicialização da segmentação do AVC hemorrágico.

$$\begin{aligned} R_D &= I_i + P_{CNN} \\ I_{Mg(p)} &= R_D + B_b \\ I \leftarrow C &= \left(\frac{B_b(x)}{2}, \frac{B_b(y)}{2} \right) \end{aligned} \quad (4.1)$$

Onde R_D representa o resultado do Processamento da Detectron2 com o bounding box, que é chamado de I_i que representa a Imagem de entrada com seu respectivo índice, mais o processo da CNN Detectron2. Logo em $I_{Mg(p)}$ é o resultado da imagem processada, onde pode ser representada pela soma de $R_D + B_b$. Assim, entende-se que B_b é a própria caixa delimitador.

Então, a Inicialização pode ser representada por I , onde I simbolizada a Inicialização do processo de segmentação, que recebe $C = \left(\frac{B_b(x)}{2}, \frac{B_b(y)}{2} \right)$, na qual C é o próprio

centroide na imagem de detecção do bounding box, que é igual aos eixo x, y de cada bounding box dividido por 2. Ou seja, a metade do eixo x e a metade do y dentro do quadrante de detecção detectado pela rede Detectron2 previamente na primeira fase da metodologia. Logo, nesta equação temos como localização de forma automática, o centroide referente ao centro do AVC hemorrágico, tendo como o resultado o ponto em (F).

2. **Etapa 3 - PDI** O pré-processamento das imagens de entrada para segmentação, afim de melhorar e facilitar o reconhecimento das bordas. Em (G) a imagem passa por um processo de limiarização, o *Multi-Otsu Thresholding* (LIAO TSE-SHENG CHEN, 2001), que determina o limiar do pixel de forma automática, afim de normalizar os tons de cinza da imagem, dado que a base de dados conta diferentes fatias da TC cerebral, em certos níveis a região da hemorragia pode ser mais clara ou mais escura que outra, oque pode dificultar na delimitação das bordas.
3. **Etapa 4 - Segmentar** Finalmente a aplicação dos modelos *Zero-Shot* de cada método proposto; $SAM_{centroid}$ e $MobileSAM_{centroid}$, os quais utilizam dos resultados dos estágios anteriores. Logo, a entrada para os modelos é a imagem pré processada e as coordenadas (x, y) do centroide da região detectada com AVC, e retorna como resultado a mascarará com a delineação da região com a hemorragia, como apresentado na Imagem 16 (H).

B) Experimento com Modelo YOLACT (You Only Look At CoefficientTs):

Nesta Etapa, é realizado o treinamento do modelo YOLACT para segmentar a região da hemorragia do AVC em imagens de TC. O conjunto de treinamento atua como a verdade básica, pois utiliza apenas imagens já segmentadas por um especialista como base de seu treino, denominadas de Padrão Ouro (*Ground Truth - GT*), assim descrito na seção 4.1. O conhecimento gerado pelo modelo é armazenado no final do treinamento. O processo de treinamento se deu seguindo os seguintes parâmetros:

- Tamanhos máximo das imagens 550 x 550 pixels;
- Arquitetura utilizada foi a **Resnet50**;
- Taxa de aprendizado de 0,0001;
- A rede foi treinada por 800.000 iterações ou épocas, com uma redução da taxa de aprendizado em 4 etapas: 280.000 iterações, 600.000 iterações, 700.000 iterações e 750.000 iterações; e
- lote de 8 imagens.

A rede foi treinada em um conjunto de dados de 100 imagens de objetos de uma única classe. O conjunto de dados foi dividido em 80% para treinamento e 20% para validação. Os parâmetros de treinamento da rede YOLACT foram ajustados para melhorar a precisão da detecção de objetos. O tamanho da imagem de entrada foi aumentado para

550 x 550 pixels para ajudar a capturar mais detalhes dos objetos. A taxa de aprendizado inicial foi reduzida para 0,0001 para evitar o *overfitting*. O número de épocas foi aumentado para 800.000 iterações para melhorar a precisão da detecção e da classificação de objetos e da segmentação.

Os resultados mostram que os ajustes nos parâmetros de treinamento melhoraram significativamente a precisão da detecção e segmentação das regiões afetadas pelo AVC hemorrágico.

4.3 Avaliação de Desempenho

Esta seção apresenta as métricas adotadas na avaliação dos resultados obtidos nos testes da abordagem proposta neste trabalho.

Foi empregado um conjunto de métricas para avaliar as tarefas de detecção e segmentação nos experimentos realizados e validar o método proposto, estabelecendo comparações com estudos relacionados. As características específicas dessas métricas são abordadas detalhadamente nesta subseção.

A avaliação dos resultados gerados por este estudo baseia-se no emprego de métricas de sobreposição, incluindo Acurácia (ACC), Precisão (PRE) e Sensibilidade (SEN). Além disso, as métricas de sobreposição compreendem Especificidade (SPE), F1-Score, Coeficiente de Correlação de Matthews (MCC), Valor Preditivo Negativo (NPV) e Valor Preditivo Positivo (PPV) (TAHA; HANBURY, 2015).

Em linhas gerais, as métricas exploradas neste estudo derivam de cálculos envolvendo as quantidades de Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN). Essas variáveis são, por sua vez, calculadas mediante uma matriz de confusão, resultante da comparação entre os resultados obtidos pelo método desenvolvido e o padrão ouro de referência, representado pelo valor de gabarito (TAHA; HANBURY, 2015).

4.3.1 Acurácia

A Acurácia é uma métrica amplamente empregada em várias categorias de problemas, incluindo classificação, detecção e segmentação. No contexto de segmentação, a acurácia está diretamente relacionada ao número de pixels segmentados corretamente em relação ao número total de pixels correspondentes à região de interesse em uma imagem (ALJABAR et al., 2009). Sua fórmula é expressa da seguinte maneira:

$$ACC = \frac{VP + VP}{VP + VP + FP + FN} \quad (4.2)$$

A avaliação da performance geral de algoritmos de segmentação é fundamental, e a acurácia desempenha um papel essencial nesse contexto, oferecendo uma visão abrangente da capacidade do modelo em segmentar adequadamente os pixels de interesse. Entretanto, é crucial observar que em situações de desequilíbrio acentuado entre as classes (por exemplo, quando uma classe de interesse possui significativamente mais pixels do que a outra), a acurácia por si só pode não ser suficiente para uma avaliação completa da eficácia do modelo. Em tais casos, torna-se pertinente recorrer a outras métricas, como a F1-Score ou a Sensitividade, proporcionando uma análise mais abrangente do desempenho da segmentação (ALJABAR et al., 2009).

No contexto da tarefa de detecção de objetos, a acurácia mede a proporção de objetos que foram detectados corretamente na imagem. Um modelo com alta acurácia terá detectado a maioria dos objetos presentes na cena e terá um número baixo de falsos positivos (objetos detectados que não existem) e falsos negativos (objetos reais que não foram detectados) (PADILLA; NETTO; SILVA, 2020).

A **Precisão**, também conhecida como Valor Preditivo Positivo (PPV), representa uma métrica de classificação que avalia a proporção de pixels corretamente classificados como positivos em relação ao total de pixels identificados como positivos na região de interesse pelo modelo de segmentação. Esta métrica torna-se particularmente relevante quando é crucial minimizar os falsos positivos, ou seja, quando é essencial evitar que pixels sejam erroneamente considerados como pertencentes à região de interesse. Em contextos como aplicações médicas, por exemplo, na detecção de tumores em imagens de ressonância magnética, a minimização de diagnósticos incorretos é vital (TAHA; HANBURY, 2015). Para detecção de objetos, a precisão é a capacidade de um modelo identificar apenas objetos relevantes. É a porcentagem de previsões positivas corretas (PADILLA; NETTO; SILVA, 2020). A fórmula para calcular a precisão é a seguinte:

$$\text{Precisão (PPV)} = \frac{VP}{VP + FP} \quad (4.3)$$

4.3.2 Sensitividade

A Sensitividade é uma métrica de classificação que avalia a proporção de pixels corretamente segmentados como pertencentes à região de interesse em relação ao total de pixels que realmente pertencem a esse grupo (BHADAURIA; DEWAL, 2014). Em resumo, a Sensitividade é uma medida que destaca a capacidade do modelo em detectar de maneira precisa os pixels da região de interesse, evitando falsos negativos, ou seja, pixels relevantes erroneamente classificados como não pertencentes à região de interesse (TAHA; HANBURY, 2015). Para a tarefa de detecção de objetos, ela representa a habilidade do modelo em identificar todos os casos relevantes, ou seja, todas as caixas delimitadoras de referência presentes no padrão ouro. A Sensitividade é expressa como a porcentagem

de previsões positivas corretas em relação a todas as referências fornecidas ([PADILLA; NETTO; SILVA, 2020](#)). Sua fórmula é apresentada da seguinte maneira:

$$Sen = \frac{VP}{VP + FN} \quad (4.4)$$

4.3.3 Especificidade

A Especificidade, métrica de classificação, destaca o índice de pixels corretamente detectados ou segmentados como parte da região de fundo entre todos os pixels que verdadeiramente compõem o fundo, ou seja, aqueles que não fazem parte da região de interesse ([BHADAURIA; DEWAL, 2014](#)). A fórmula que a representa é a seguinte:

$$Spe = \frac{VN}{VN + FP} \quad (4.5)$$

A Especificidade desempenha um papel crucial na avaliação abrangente do desempenho de algoritmos de detecção e segmentação, oferecendo uma perspectiva complementar à Sensitividade. Enquanto a Sensitividade se concentra na detecção precisa dos pixels da região de interesse, a Especificidade destaca a precisão na identificação dos pixels que não pertencem a essa região. Em conjunto, essas métricas proporcionam uma avaliação completa da qualidade da segmentação, permitindo ajustes e aprimoramentos no modelo para uma variedade de aplicações ([TAHA; HANBURY, 2015](#)).

4.3.4 Sensitividade $\times$ Especificidade - ($Sen \times Spe$)

Sensibilidade, também conhecida como *recall*, avalia o número de positivos reais corretamente detectados. Refere-se à habilidade do sistema em identificar de maneira precisa as amostras que verdadeiramente pertencem à classe desejada, representando, neste caso, a capacidade do método em identificar corretamente as regiões que fazem parte da região de hemorragia do AVC. Por outro lado, a Especificidade (negativo real) calcula o número de negativos reais que foram corretamente identificados ([FAWCETT, 2006](#)).

$$Sen = 1 - \frac{VP}{VP + FN} = \frac{VP}{PixelsDaRegiãoDoAVC}. \quad (4.6)$$

$$Spe = \frac{VN}{VN + FP} = \frac{VN}{PixelsDaRegiãoDoAVC}. \quad (4.7)$$

onde VP significa Verdadeiro Positivo, FN significa Falso Negativo, VN significa Verdadeiro Negativo e FP são os pixels de Falso Positivo.

4.3.5 F1-Score

O F1-Score é uma métrica que unifica a precisão e a sensibilidade (recall) em uma única medida para a avaliação de detecção e segmentação de imagens. Esta métrica se destaca em situações de desequilíbrio entre as classes nas imagens, como diferentes tipos de objetos ou regiões. Ao calcular o F1-Score, são consideradas tanto a capacidade do modelo em fornecer detecções e segmentações corretas dos pixels pertencentes às regiões de interesse (verdadeiros positivos) quanto a habilidade de capturar a maioria dos pixels corretos que pertencem às regiões de interesse existentes na imagem (evitando falsos negativos). Essa abordagem torna o F1-Score uma métrica valiosa para avaliar o desempenho global de um modelo de segmentação em cenários nos quais as classes (regiões de interesse) possuem diferentes tamanhos ou formas, assegurando uma avaliação equilibrada e abrangente (BHADURIA; DEWAL, 2014). O F1-Score pode ser calculado utilizando a seguinte equação:

$$F1 = \frac{2 \times \text{PREC} \times \text{SEN}}{\text{PREC} + \text{SEN}} \quad (4.8)$$

4.3.6 Coeficiente de Correlação de Matthews

O Coeficiente de Correlação de Matthews (MCC) representa uma métrica altamente informativa que vai além da consideração de verdadeiros positivos e verdadeiros negativos, incorporando também falsos positivos e falsos negativos na avaliação de algoritmos de detecção e segmentação de imagens. Essa métrica destaca-se, especialmente, em cenários nos quais há desequilíbrio entre as classes de pixels ou objetos detectados, ou quando há um interesse específico em avaliar o desempenho do modelo em todas as quatro possíveis classificações: Falsos Positivos (FP), Verdadeiros Positivos (VP), Verdadeiros negativos (VN) e Falsos Negativos (FN) (BHADURIA; DEWAL, 2014). A fórmula para o cálculo do MCC é a seguinte:

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VP + FP)(VN + FN)}} \quad (4.9)$$

4.3.7 Valor Preditivo Negativo

O Valor Preditivo Negativo (NPV) é uma medida que avalia a proporção de pixels corretamente identificados como pertencentes ao fundo (região que não é de interesse) em relação ao número total de pixels classificados como fundo pelo modelo de segmentação (BHADURIA; DEWAL, 2014). A fórmula para o NPV é a seguinte:

$$NPV = \frac{VP}{VP + FN} \quad (4.10)$$

4.3.8 Valor Preditivo Positivo

O Valor Preditivo Positivo (PPV) avalia a capacidade do modelo em identificar corretamente os pixels pertencentes à região de interesse, evitando falsos positivos, ou seja, pixels que seriam erroneamente classificados como pertencentes a essa região (BHADARIA; DEWAL, 2014). A fórmula para o PPV é a seguinte:

$$PPV = \frac{VP}{VP + FP} \quad (4.11)$$

4.3.9 Coeficiente de Dice - (*Dice's coefficient*)

O coeficiente de Dice (Dice) é empregado para avaliar a semelhança entre as regiões detectadas ou segmentadas automaticamente e aquelas marcadas manualmente. Essa métrica compara a sobreposição entre a detecção ou segmentação gerada pelo algoritmo e o padrão ouro estabelecido por um especialista.

$$DC = 2 \times \frac{|S_m \cap S_t|}{|A_k| + |A_l|}, \quad (4.12)$$

Conforme evidenciado na Equação 4.12, onde S_m denota a região definida automaticamente e S_t representa o padrão de referência. Quando o valor de sobreposição dos intervalos é 0, não há sobreposição entre as duas regiões, enquanto que quando é 1, a sobreposição é total (DICE, 1945). O coeficiente de Dice aumenta à medida que a semelhança entre a região marcada e o padrão ouro aumenta (QIAN et al., 2013).

4.3.10 Precisão Geral - (*Accuracy*)

Accuracy (Precisão), é o número de ocorrências positivas em P dividido pelo número total de amostras N (FAWCETT, 2006). Ou seja, é a relação entre as respostas do algoritmo e o conjunto das possibilidades de erros. Essa medida avalia os acertos do algoritmo em relação ao total de amostras apresentadas, indicando desta forma, a proximidade dos resultados obtidos daqueles desejado. Podendo ser compreendida pela equação:

$$Accuracy = \frac{P}{P + N}. \quad (4.13)$$

4.3.11 Teste Estatístico

O teste estatístico de Kolmogorov-Smirnov desempenha um papel fundamental e versátil na análise de dados ao comparar distribuições de amostras contínuas. Ao contrário dos testes paramétricos, que demandam suposições sobre a distribuição dos dados, o teste K-S é não paramétrico e é apropriado para situações em que a distribuição subjacente é

desconhecida ou não segue um padrão específico, como a distribuição normal ([ASLAM, 2019](#)).

Ao aplicar o teste de Kolmogorov-Smirnov, são formuladas duas hipóteses: a hipótese nula (H_0), que postula que as duas amostras compartilham a mesma distribuição, e a hipótese alternativa (H_1), que sustenta que as amostras possuem distribuições distintas. Em seguida, são obtidas duas amostras de dados, representando diferentes grupos ou populações que se pretende comparar. Essas amostras devem ser contínuas e ter o mesmo tamanho ([BAUMGARTNER; KOLASSA, 2023](#)).

Para calcular a estatística de teste, procede-se ordenando os valores de cada amostra em ordem crescente e, em seguida, calculando a Função de Distribuição Cumulativa Empírica (ECDF) para ambas as amostras. A ECDF é uma estimativa da função de distribuição cumulativa real, fundamentada nos dados coletados, e atribui a probabilidade acumulada de encontrar um valor menor ou igual a cada valor da amostra ([OTSU; TANIGUCHI, 2020](#)).

A estatística de teste de Kolmogorov-Smirnov é definida como a maior discrepância absoluta entre as duas Funções de Distribuição Cumulativa Empírica (ECDF), representando a máxima diferença vertical entre elas. Posteriormente, é calculado o P-valor, que corresponde à probabilidade de obter uma estatística de teste tão extrema quanto a observada, sob a suposição de que a hipótese nula é verdadeira ([WARE et al., 2019](#)).

A interpretação do valor P é crucial na tomada de decisões. Quando o P-valor é superior ao coeficiente de confiança escolhido, representado por α (como 5% ou 3%), não há evidências suficientes para rejeitar a hipótese nula. Nesse caso, conclui-se que as duas amostras compartilham a mesma distribuição (igualdade estatística). No entanto, se o valor P for inferior a α , rejeita-se a hipótese nula em favor da hipótese alternativa, indicando que as amostras possuem distribuições distintas (diferença estatística) ([WARE et al., 2019](#)).

5 Resultados e Discussão

Nesta seção, são apresentados os resultados dos experimentos em três etapas: detecção, segmentação e comparação com os resultados obtidos neste estudo e da literatura. *Etapa 1:* Segmentação de AVC hemorrágico em imagens de CT usando os métodos propostos e experimento com a rede de segmentação em única etapa, YOLACT; *Etapa 2:* Validação dos experimentos por comparação com estudos relevantes na literatura; e *Etapa 3:* Detecção de AVC hemorrágico para inicialização da segmentação automática.

Para avaliar os resultados na tarefa de *segmentação*, foram empregadas as seguintes métricas: Acurácia (ACC), Sensibilidade (SEN), Especificidade (SPE), Coeficiente de Dice (Dice), Valor Preditivo Negativo (NPV), Valor Preditivo Positivo (PPV) e Falsos Positivos (FPR) (TAHA; HANBURY, 2015). No processo de escolha do método de detecção utilizado pelo método proposto para segmentação, foram consideradas as métricas de Acurácia (ACC), Sensibilidade (SEN), Especificidade (SPE) e Coeficiente de Dice (Dice). É relevante destacar que essas métricas são suficientes, dada a finalidade da avaliação do processo de detecção.

5.1 Resultados de Segmentação.

Nesta seção são apresentados os resultados dos experimentos de segmentação realizados neste estudo em comparação com outros modelos da literatura, na Figura 17 é apresentado um exemplo de mascaras de segmentação geradas pelo métodos proposto.

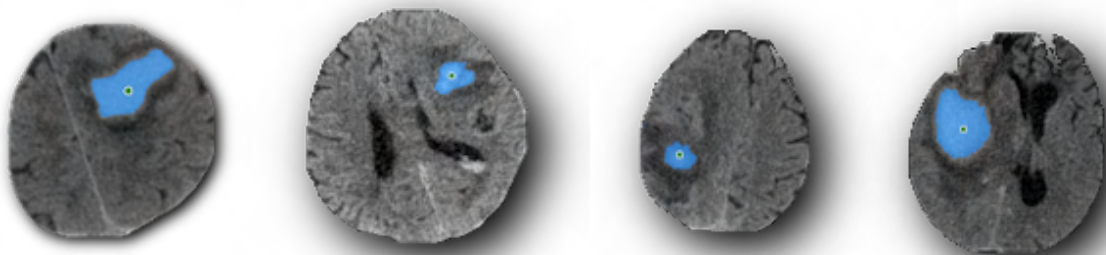


Figura 17 – Exemplos de resultados da segmentação do método proposto MobileSAM_{centroid} - Ilustrando a mascara gerada aplicada na imagem original junto com o centroid calculado.

Para validar os resultados obtidos neste estudo, esta subseção apresenta comparações com trabalhos encontrados no estado da arte. Esses trabalhos têm como premissa a delimitação/segmentação da área afetada pelo AVC hemorrágico com o uso de CNN e o

uso de ajuste fino. Além disso, todos os trabalhos utilizaram o mesmo banco de dados para os experimentos realizados.

A Tabela 2 apresenta diferentes modelos que utilizam detecção de rede convolucional e segmentação com modelos de ajuste fino, aplicados na mesma base dados. A Tabela 2 também apresenta diferentes modelos de aprendizado profundo em comparação com os métodos propostos, nela é apresentado o melhor e o pior resultado de cada método proposto por estudo comparado. Os autores de (HAN et al., 2020) e (MARQUES et al., 2022) também utilizam a rede Mask R-CNN para detecção de AVC. O estudo de (MARQUES et al., 2022) utiliza diferentes combinações de técnicas de processamento de imagens digitais (PDI) para segmentação de hemorragia cerebral. Vale ressaltar, que os estudos de (HAN et al., 2020) e (MARQUES et al., 2022) são métodos totalmente automáticos.

Os resultados obtidos na Tabela 2 mostram diferentes modelos do estudo de Han et al. (2020). Vale ressaltar que o autor não se ateve apenas a a tarefa de segmentação mas propôs um método de classificação e segmentação. Para realizar a tarefa de segmentação o trabalho utilizou-se de transferência de aprendizado e técnicas clássicas de ajuste fino, tendo como base a rede de aprendizagem profunda *Detectron2* (WU et al., 2019).

Os métodos desenvolvidos neste estudo também obtiveram acurácia (Acc) semelhante ao melhor resultado obtido por Han et al. (2020) respectivamente; 99,82% e 99,88%. As métricas de Dice (**91,08***% Vs. 94,09%), sensibilidade (Sen) (**94,97***% Vs. 90,89%) e especificidade (Spe) (**99,89***% Vs. 99,97%) também apresentaram valores semelhantes. No entanto, o estudo de Han et al. (2020) não abordou a possibilidade de melhorar os resultados com a combinação dos modelos, deixando o melhor modelo, o *Detectron2*, sem as combinações para o modelo de segmentação.

Ainda na Tabela 2 mostra que os métodos propostos por este estudo, alcançaram resultados satisfatórios em termos de acurácia (Acc) e precisão (Npv), em comparação aos melhores resultados dos métodos apresentados pelo estudo de Marques et al. (2022) - Acc: **99,82%** Vs. 99,84% e NPV: **99,82%** Vs. 99,78. Neste caso é importante verificar que o desvio padrão da Acc do método proposto por Marques et al. (2022) é de 3,43 contra 0,11 do método proposto e 0,05 de Han et al. (2020), o que indica uma certa instabilidade.

Embora os métodos propostos por Marques et al. (2022) não tenham aplicado validação cruzada, isso pode ter afetado os resultados obtidos. A validação cruzada é um processo de revalidação do teste realizado por diferentes partes do conjunto de dados. Isso torna as imagens de teste diferentes em cada ciclo do teste de validação cruzada, o que pode melhorar os resultados obtidos.

Além da comparação com métodos de aprendizagem profunda para segmentação da região afetada pelo AVC hemorrágico, a Tabela 2 também apresenta uma comparação

Tabela 2 – Comparação dos resultados obtidos na segmentação automática de AVC Hemorrágico em imagens de CT proposto neste trabalho, em comparação com os resultados de trabalhos que utilizaram o mesmo conjunto de dados.

Trabalho	Categoria	Método	Acc	Resultados
Métodos Propostos 1	(Zero-Shot) CNN + MAE + ViT-H	$SAM_{centroid}$	$99,82 \pm 0,11$	Dice: 91,08 Sen: 94,97 Spe: 99,89 PPV: 89,00 NPV: 99,92
Métodos Propostos 2	(Zero-Shot) CNN + MAE + TinyViT	$MobileSAM_{centroid}$	$99,82 \pm 0,10$	Dice: 91,02 Sen: 94,84 Spe: 99,90 PPV: 89,05 NPV: 99,92
Experimento	(One-Stage) CNN + FPN	YOLOACT	$99,74 \pm 0,29$	Dice: 83,02 Sen: 78,66 Spe: 99,96 PPV: 91,17 NPV: 99,77
Han et al. (2020)	RNN + Fine-Tuning	Detectron-f λ	$99,83 \pm 0,06$	Dice: 91,37 Sen: 85,21 Spe: 99,99
Han et al. (2020)	RNN + Fine-Tuning	Detectron-f δ	$99,88 \pm 0,05$	Dice: 94,09 Sen: 90,89 Spe: 99,97
Marques et al. (2022)	RNN + Fine-Tuning	$ft_{mask}Parzen$	$98,08 \pm 4,76$	Sen: 98,25 PPV: 81,35 NPV: 99,82
Marques et al. (2022)	RNN + Fine-Tuning	$ft_{mask}Clustering_{parzen}$	$98,84 \pm 3,43$	Sen: 99,05 PPV: 87,06 NPV: 99,78
Rebouças et al. (2017)	Algoritmo LS	LSBRD	$99,83 \pm 0,10$	F-1 Score: 99,05
Rebouças et al. (2017)	Algoritmo LS	LSCPM	$99,72 \pm 0,23$	F-1 Score: 88,88
Peixoto et al. (2017)	Algoritmo	Optimum Path Snakes (OPS)	$99,82 \pm 0,12$	MCC: 89,83% Dice: 89,46% HD: 3,11 HM: 97,85%

com métodos que aplicaram algoritmos de processamento de imagem afim de delinear a região afetada, os trabalhos de [Rebouças et al. \(2017\)](#) e [Peixoto et al. \(2017\)](#). E como apresentado, os valores de acurácia são bem próximos do método proposto que utiliza técnicas mais recentes como *Vision Transformers* e RNN's por exemplo.

Quando comparado com o método proposto podemos observar uma pequena diferença de 0,01% entre os 99,82% de acurácia pelo método proposto e 99,83% proposto por [Rebouças et al. \(2017\)](#), além de se mostrar um método consideravelmente estável dado o desvio padrão de 0,10, junto com o método de [Peixoto et al. \(2017\)](#) com uma acurácia 99,82% e um desvio padrão de 0,12. Além disso o método de [Rebouças et al. \(2017\)](#) possui uma certa semelhança para sua aplicação, tendo uma etapa de inicialização para encontrar o ponto inicial (*seed*) da região a ser segmentada.

A Tabela 3, apresenta uma comparação de forma geral com métodos encontrados na literatura ([ABBASI et al., 2023](#)) e os métodos e experimentos propostos neste estudo para segmentação de AVC hemorrágico. Também é possível verificar que vários estudos utilizam métricas diferentes, porém como já argumentado na seção 4 as métricas utilizadas

Tabela 3 – Comparação dos resultados obtidos na segmentação do AVC Hemorrágico em imagens de TC propostas neste trabalho com os relatados na literatura utilizando aprendizagem profunda (ABBASI et al., 2023).

Trabalho	Método	Resultado
Método proposto 1	RNN + MAE + ViT-H	Acc: 99,82; Dice: 91,08; Sen: 94,97; Spe: 99,89; PPV: 89,00; NPV: 99,92;
Método proposto 2	RNN + MAE + TinyViT	Acc: 99,82; Dice: 91,02; Sen: 94,84; Spe: 99,90; PPV: 89,05; NPV: 99,92;
Experimento	CNN + FPN	Acc: 99,74; Dice: 83,02; Sen: 78,66; Spe: 99,96; PPV: 91,17; NPV: 99,77;
Alis et al. (2021)	Residual ConvLSTM U-Net	DSC de A: 0,858
Clèrigues et al. (2020)	U-Net based CNN	DSC para SISS: $0,59 \pm 0,31$; DSC para SPES: $0,84 \pm 0,10$
Kumar et al. (2020)	CSNet	DSC para SISS: 0,83
Nazari-Farsani et al. (2020)	Crawford-Howell	Acc: 73,00
Nazari-Farsani et al. (2023)	AG-DCNN	DSC: 0,50; AUC: 91,00; Sen: 60,00; Spe: 97,00
Cui et al. (2021)	DeepSym-3D-CNN	Acc: 85,00
Praveen et al. (2018)	SSA+SVM	Acc: 90,00; DSC: 0,94

neste estudo são suficientemente satisfatórias para apurar a performance para o problema.

A Tabela 4 mostra a comparação entre a média de tempo de resposta para segmentação de AVC hemorrágico em imagens de TC em segundos. Vale ressaltar que o ambiente de execução dos modelos dos trabalhos apresentados são diferentes. No entanto como já esperado, dada a natureza do tipo de rede *Zero-Shot*, o método proposto que possui um alto grau de generalização é o mais demorado com 79.338s desde o carregamento da imagem no modelo até a resposta com o contorno da região da hemorragia. Porém, com uma grande diferença o método proposto que utiliza uma variação do modelo *Zero-Shot* apresentou um tempo interessante de 3,2s, ainda mantendo uma precisão geral equivalente ao de 99,82%. E o experimento que utiliza a rede com foco em segmentação em uma única

etapa a YOLACT, foi o mais rápido, com 1,845s e estável se comparado com o método proposto por [Rebouças et al. \(2017\)](#) com 1,760s e um desvio padrão de 0,29 contra 0,13 do experimento deste estudo. Inclusive em comparação com os métodos com maior precisão geral dos estudos, Detectron- $\mathcal{f}\delta$ de [Han et al. \(2020\)](#) com 2,186s e 99,88% de acurácia, e $ft_{mask}Parzen_{GrowingRegion}$ de [Marques et al. \(2022\)](#) com 6,492s e 99,72% de acurácia.

Tabela 4 – Comparação do tempo para segmentação entre os métodos propostos e os melhores resultados de precisão geral dos métodos propostos por ([HAN et al., 2020](#)) e ([MARQUES et al., 2022](#)) que utilizam o mesmo conjunto de dados.

Método	Tempo (s)
(Métodos propostos)	
SAM _{centroid}	79,338 $\pm$ 3,632
MobileSAM _{centroid}	3,200 $\pm$ 0,481
YOLACT	1,845 $\pm$ 0,133
Han et al. (2020)	2,186 $\pm$ 1,890
Marques et al. (2022)	6,492 $\pm$ 0,503
Rebouças et al. (2017)	1,760 $\pm$ 0,290

O processo de treinamento da rede YOLACT para o experimento, foi realizado utilizando o sistema operacional Linux, distribuição Ubuntu 22.04, usando 12,7 GB de RAM, Processador Intel Xeon 2.00GHz e Placa de Vídeo Tesla T4 - 15GB VRAM CUDA version 12.2. E para execução dos modelos *Zero-Shot* propostos e do experimento de tempo de resposta para segmentação para o experimento com a rede YOLACT, foi utilizando o sistema operacional Windows 10 Home, usando 32 GB de RAM, Processador Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz, 2901 Mhz, 2 Núcleos, 4 Processadores Lógicos e Placa de Vídeo NVIDIA GeForce 940MX - 20GB VRAM.

5.2 Validação dos Resultados

Para validar os resultados obtidos neste estudo, foram realizadas comparações com estudos encontrados na literatura. A Tabela 5 apresenta os métodos com as melhores métricas de precisão (Acc), coeficiente de Dice (Dice), sensibilidade (Sen) e especificidade (Spe), todos eles com o objetivo de segmentar AVC hemorrágico utilizando o mesmo conjunto de dados utilizado neste estudo.

Os resultados da Tabela 5 mostram que os métodos propostos neste estudo obtiveram resultados bastante satisfatórios. Os métodos *Zero-Shot* baseados no *Foundation Model* (modelo base) de segmentação SAM, em particular, atingiram uma precisão de 99,82% e um desvio padrão entre 0,10 e 0,11, sem a necessidade de treinamento com o conjunto de dados. Esse resultado é bastante promissor, pois indica que o método é confiável e pode ser usado na prática clínica.

Tabela 5 – Comparação da precisão geral (Acc) dos resultados obtidos na segmentação de AVC Hemorrágico em imagens de TC do mesmo conjunto de dados.

Métodos	Acc
(Métodos Propostos)	
$SAM_{centroid}$	$99,82 \pm 0,11$
$MobileSAM_{centroid}$	$99,82 \pm 0,10$
YOLACT	$99,74 \pm 0,29$
Marques et al. (2022)	$99,72 \pm 0,24$
Han et al. (2020)	$99,88 \pm 0,05$
Rebouças et al. (2017)	$99,83 \pm 0,10$
Peixoto et al. (2017)	$99,78 \pm 0,12$

5.3 Resultados da Detecção.

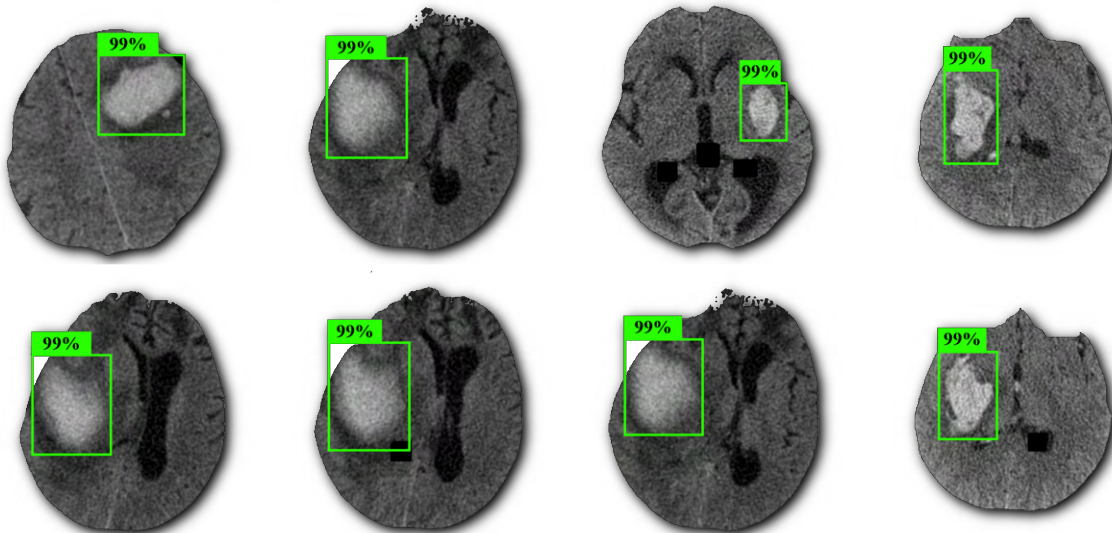


Figura 18 – Detecção de AVC Hemorrágico com o uso de aprendizagem profunda.

Nesta subseção são apresentados os resultados referente a detecção do AVC hemorrágico com uso de deep learning. O método tem como objetivo identificação da região de interesse na imagem de TC, ilustrado na Figura 18. A rede neural convolucional Detectron2 foi treinada para identificar regiões lesionadas no cérebro, como apresentado na Seção 4. Esta fase é executada exclusivamente para inicialização da segmentação automática com métodos propostos $SAM_{centroid}$ e $MobileSAM_{centroid}$.

Desta forma, foi gerado um modelo para detecção inteligente em imagens de hemorragia cerebral em TCs, com uso de deep learning. O processo de avaliação se deu pela utilização das métricas de avaliação (Acurácia- Acc, Sencitividade-Sen, Especificidade-Spe e Coefficient Dice - Dice), e cada uma tem sua importância para a atividade de detecção de AVC hemorrágico. **Acc** - Mede a precisão geral da detecção; **Sen** - Mede a capacidade do modelo de detectar todos os objetos de interesse; **Spe** - Mede a capacidade do modelo de evitar detectar objetos não de interesse; e **Dice** - Mede a semelhança entre

as regiões de interesse e detecção. Além de serem métricas comuns em diferentes trabalhos encontrados no estado da arte.

Todo o processo de aprendizagem profunda para a detecção, foi utilizando o sistema operacional Linux, distribuição Ubuntu 22.04, usando 16 GB de RAM, Processador Intel® Core™ i7-7700K CPU @ 4.20GHz e Placa de Vídeo NVIDIA GeForce GTX 1050 Ti (com 4 GB de memória). A implementação da detecção foi feita através da linguagem Python, versão 3.11.1, com uso das bibliotecas OpenCV v4.1.0, scikit-learn v0.20.2, v.1.14.0, Keras v.2.2.4.

A Tabela 6 apresenta diferentes resultados com uso das CNNs Mask R-CNN, Detectron2 e outros estudos presentes na literatura. Na tabela estão os resultados do experimento deste estudo para detecção da região do AVC hemorrágico com a base de dados especificada na subsecção 4.1. Seguindo da comparação com o trabalho de (HAN et al., 2020), que apresenta uma solução para o mesmo problema de detecção em imagens de AVC hemorrágico, o autor usa a mesma base de dados deste estudo, com a abordagem de uso da rede Mask R-CNN. O gráfico presente na Figura 19 ilustra os resultados entre as redes Mask R-CNN e Detectron2.

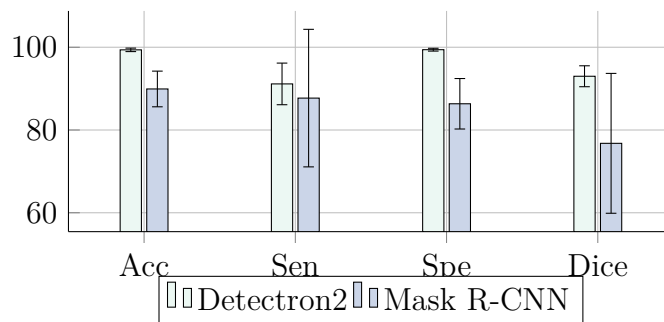


Figura 19 – Ilustração dos resultados obtidos para detecção de AVC hemorrágico na tabela 6.

Em análise dos valores obtidos, é possível identificar a diferença da rede Detectron2 em diferentes valores métricos, a rede obteve melhor performance para identificação da região do AVC em todas as métricas apresentadas, comparada a rede Mask R-CNN. Na métrica de acurácia foi obtida uma diferença de cerca de 10% em relação a rede Mask R-CNN (99,37% contra 89,92%), com desvio padrão com quase 4% de diferença. Isso permite concluir que neste processo de detecção da região de interesse, a Detectron2 demonstrou melhor desempenho em identificar os pixels pertencentes a região da hemorragia cerebral, e o que de fato não era a região pertencente ao AVC. Isso pode ser validado e consolidado mediante as métricas; (Sensitividade-Sen, Especificidade e Coeficiente de Dice), ambos os valores obtidos foram superiores ao modelo Mask R-CNN, mostrando que a rede identificou erroneamente em diferentes regiões da imagem do AVC.

Vale ressaltar, que o autor (HAN et al., 2020) não apresentou resultados baseados em cross valitation, o que ainda pode ser cabível de novos testes para uma análise mais

Tabela 6 – Comparação dos resultados obtidos na detecção de AVC Hemorrágico em imagens de CT proposto neste trabalho, em comparação com os resultados da literatura (KARTHIK et al., 2020).

Trabalho	Método	Biblioteca	Resultado
Método proposto	CNN	Detectron2	Acc: 99,37; Sen: 91,14; Spe: 99,41; Dice: 92,99
Han et al. (2020)	CNN	Maks R-CNN	Acc: 89,92; Sen: 87,71; Spe: 86,34; Dice: 76,78
Islam et al. (2019)	CNN	Maks R-CNN	Dice: 87,60
Patel et al. (2019b)	CNN + FCN	Theano e Lasagne	Dice: 91,00
Cho et al. (2019)	CNN + FCN	Caffe	Acc: 98,28; Dice: 84,00
Patel et al. (2019a)	CNN + RNN	Keras + Theano	Acc: 87,00; AUC: 0,96
Kuo et al. (2019)	FCN	PyTorch	AUC: 0,991
Arbabshirani et al. (2018)	CNN	Caffe	Acc: 84,00
Phong et al. (2017)	CNN	TensorFlow	Acc: 99,70

criteriosa, com valores que representem testes concisos na base de dados. Porém como para o fim deste estudo a fase de detecção será utilizada para aplicação de um dos métodos de segmentação, o qual utilizado para a identificação do ponto centro da *bounding box*, isso para que este ponto seja utilizado como entrada para os métodos propostos $SAM_{centroid}$ e $MobileSAM_{centroid}$

6 Conclusão

O estudo proposto por este trabalho aborda a aplicação de modelos *zero-shot* e *one-stage* em sistemas do tipo *Health of Things* (CDS) para classificação, detecção e segmentação de AVC hemorrágico em imagens de tomografia computadorizada (CT) cranianas. O estudo teve como foco a tarefa de segmentação.

O estudo abordou alternativas, que visam acelerar o desenvolvimento de ferramentas que auxiliem o diagnóstico médico em imagens de TC cranianas, de forma que se mantenha uma precisão aceitável e garanta agilidade na identificação da área da hemorragia causada pelo AVC hemorrágico.

Considerando os resultados apresentados de 99,82% de acurácia (Acc) com um desvio padrão de 0,11 e 0,10 pelos métodos propostos **SAM**_{centroid} e **MobileSAM**_{centroid} ue utilizam redes do tipo *Zero-Shot*, é possível afirmar que sera muito promissor para o desenvolvimento de ferramentas para assistência na tomada de decisões médicas, pois não necessitam de treinamento prévio em uma base de dados rotulados. Contudo é importante ter uma amostragem rotulada para conseguir mensurar a eficacia na execução da atividade.

Ainda, o método **MobileSAM**_{centroid} apresentou um tempo de resposta muito melhor que a sua variação utilizando rede base SAM, sendo em média 3,2s com **MobileSAM** contra 79,34s com SAM, uma diferença de 76,14s é extremamente considerável dependendo da atividade. Por exemplo, em uma aplicação em tempo real este tempo de 79,34s não seria satisfatório, ainda que com uma ótima precisão.

Após uma análise mais profunda no tempo de resposta da rede SAM, foi identificado o momento de execução da rede que mais onera o tempo de resposta, sendo a etapa de *encoding* da imagem, que utiliza um MAE (*M*asked *A*uto*E*ncoder) ViT (*V*ision *T*ransformer) pré-treinado, adaptado para imagens de alta resolução (LI et al., 2022). Sendo esta exatamente a camada ajustada pela rede **MobileSAM** afim de melhorar o tempo de resposta e consumo de processamento, permitindo uma execução mais suave inclusive em CPU invés de GPU.

A rede YOLACT do tipo *One-Stage* apresentou o melhor tempo de resposta dentre todos os estudos e experimentos comparados neste trabalho, e uma precisão satisfatória de 99,74%, mas, vale ressaltar que ela demanda do processo de treinamento com uma base de dados rotulados no formato COCO. Ou seja, é necessário uma etapa massante e trabalhosa de rotulação do objeto qual deseja-se detectar e segmentar, neste caso foi a lesão causada pelo AVC hemorrágico. Ainda após a base de dados pronta para treinamento, é necessário uma serie de experimentos e estudos para que se possa garantir a efetividade do treinamento.

Contudo, na comparação de tempo não foi considerada a etapa de inicialização automática da segmentação para os modelos de segmentação *Zero-Shot*, a qual determina a entrada ou *prompt* para os métodos que utilizam os modelos bases de segmentação, dado que o foco do estudo na atividade de segmentação. Porém ao implementar em um sistemas do tipo *Health of Things* (CDS), é possível solicitar esta entrada do usuário, como por exemplo, apenas selecionar o ponto central da região que se deseja delinear. Sendo possível até aumentar a performance de um processo de rotulação para preparação da base de dados para treinamento, afim de resolver outros problemas de segmentação.

Porém, inicialmente foi realizado um estudo superficial para inicializar a entrada ou *prompt* para o modelo base (*Foundation Model*), e então aplicar a segmentação *Zero-Shot* dos métodos propostos: **SAM_{centroid}** e **MobileSAM_{centroid}**. Levando em consideração estudos realizados utilizando a mesma base de imagens de CT. Dentre eles foram descritas e analisadas métodos que utilizam CNN's para detecção das regiões de interesse (RoI), ou seja as lesões causadas pelo AVC hemorrágico. Os melhores métodos utilizam as redes *Mask R-CNN* e *Detectron2*. E dada a precisão de 99,37% na detecção de região afetada pela hemorragia do AVC, foi escolhido o método que utiliza rede *Detectron2*. Então foi realizada a detecção da região, e por meio das coordenadas da caixa delimitadora é calculado o seu centroide, e então utilizar as coordenadas do centroid como entrada para inicio da segmentação.

A aplicação de modelos base (*Foundation Models*) especializados em segmentação junto com uma técnica para definir automaticamente a entrada do modelo, demonstrou uma ótima eficiência para o problema de segmentação em imagens médicas, quando comparado com trabalhos do estado da arte. Contudo, sem a necessidade de uma grande base de dados rotulados, possibilitando a segmentação *Zero-Shot*.

No entanto, é importante ressaltar que o estudo também abordou um experimento que utiliza uma rede de segmentação em uma única etapa (*One-Stage*). E as diferenças entre os dois tipos de modelos abordados atendem a requisitos específicos. Por exemplo, os modelos que possibilitam a segmentação *Zero-Shot* são mais preciso por conta do seu alto grau de generalização, porém mais lento. Por outro lado, os modelos de segmentação em uma única etapa (*One-Stage*), por outro lado, são mais rápidos, no entanto menos preciso.

6.1 Proposta de Trabalhos Futuros

Os resultados obtidos nesta dissertação abrem novas possibilidades de pesquisa. Como trabalhos futuros, sugerem-se:

- Avaliação do método proposto em outros contextos de imagens médicas de TC e RMI;

- Utilização da caixa delimitadora da detecção como entrada dos métodos baseados no *Foundation Model* SAM;
- Estudo da topologia da rede SAM e aplicar processos de ajustes finos para especialização em imagens médicas de CT;
- Aplicação de redes Zero-Shot em tarefas de *Self-Supervised Learning* para classificação em imagens de CT de AVC's hemorrágicos;
- Segmentação 3D por Método recursivo com base no centro de massa das mascaras das fatias de TC craniana.

7 Anexo A



Universidade Federal do Ceará
Comitê de Ética em Pesquisa

Of. Nº 107/06

Fortaleza, 31 de março de 2006

Protocolo COMEPE nº 36/06

Pesquisador responsável: Simone Castelo Branco Fortaleza

Deptº./Serviço: Hospital Universitário Walter Cantídio

Título do Projeto: "Efeitos da administração da CPAP sobre o parênquima pulmonar nas posições prona e supina em pacientes com doença pulmonar obstrutiva crônica: alterações na tomografia computadorizada de alta resolução e correlação com a dispnéia"

Levamos ao conhecimento de V.Sª. que o Comitê de Ética em Pesquisa da Universidade Federal do Ceará – COMEPE, dentro das normas que regulamentam a pesquisa em seres humanos, do Conselho Nacional de Saúde – Ministério da Saúde, Resolução nº196 de 10 de outubro de 1996 e complementares, aprovou o projeto supracitado na reunião do dia 30 de março de 2006.

Outrossim, informamos, que o pesquisador deverá se comprometer a enviar o relatório parcial e final do referido projeto.

Atenciosamente,

Dr. Fernando A. Frede Bezerra
Coordenador do Comitê
de Ética em Pesquisa
COMEPE/UFC

Referências

- ABBASI, H. et al. Automatic brain ischemic stroke segmentation with deep learning: A review. *Neuroscience Informatics*, v. 3, n. 4, p. 100145, 2023. ISSN 2772-5286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2772528623000304>>. Citado 3 vezes nas páginas 17, 67 e 68.
- ALIS, D. et al. Inter-vendor performance of deep learning in segmenting acute ischemic lesions on diffusion-weighted imaging: a multicenter study. *Scientific reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 12434, 2021. Citado na página 68.
- ALJABAR, P. et al. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, Elsevier, v. 46, n. 3, p. 726–738, 2009. Citado 2 vezes nas páginas 58 e 59.
- ARBABSHIRANI, M. R. et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ digital medicine*, Nature Publishing Group UK London, v. 1, n. 1, p. 9, 2018. Citado na página 72.
- ASLAM, M. Introducing kolmogorov–smirnov tests under uncertainty: an application to radioactive data. *ACS omega*, ACS Publications, v. 5, n. 1, p. 914–917, 2019. Citado na página 63.
- BA, J. L.; KIROS, J. R.; HINTON, G. E. *Layer Normalization*. 2016. Citado na página 48.
- BAUMGARTNER, D.; KOLASSA, J. Power considerations for kolmogorov–smirnov and anderson–darling two-sample tests. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 52, n. 7, p. 3137–3145, 2023. Citado na página 63.
- BENVENUTTI, L. et al. Poliglobulia ou trombocitose em contexto de primeiro AVC isquêmico: Mortalidade, recorrência e frequência de encaminhamento à hematologia. *Hematology, Transfusion and Cell Therapy*, Elsevier, v. 44, p. S619–S620, 2022. Citado na página 25.
- BHADAURIA, H.; DEWAL, M. Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain ct imaging. *Signal, Image and Video Processing*, Springer, v. 8, n. 2, p. 357–364, 2014. Citado 4 vezes nas páginas 59, 60, 61 e 62.
- BOLYA, D. et al. Yolact: Real-time instance segmentation. In: *ICCV*. [S.l.: s.n.], 2019. Citado 5 vezes nas páginas 16, 27, 46, 51 e 52.
- BOMMASANI, R. et al. *On the Opportunities and Risks of Foundation Models*. 2022. Citado na página 47.
- BUDA, M.; SAHA, A.; MAZUROWSKI, M. A. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, Elsevier, v. 109, p. 218–225, 2019. Citado na página 34.

- CARION, N. et al. *End-to-End Object Detection with Transformers*. 2020. Citado na página 47.
- CHEN, L.-C. et al. *MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features*. 2017. Citado na página 51.
- CHENG, B.; SCHWING, A. G.; KIRILLOV, A. *Per-Pixel Classification is Not All You Need for Semantic Segmentation*. 2021. Citado na página 47.
- CHO, J. et al. Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models. *Journal of digital imaging*, Springer, v. 32, p. 450–461, 2019. Citado na página 72.
- CLÈRIGUES, A. et al. Acute and sub-acute stroke lesion segmentation from multimodal mri. *Computer methods and programs in biomedicine*, Elsevier, v. 194, p. 105521, 2020. Citado na página 68.
- CUI, L. et al. Deep symmetric three-dimensional convolutional neural networks for identifying acute ischemic stroke via diffusion-weighted images. *Journal of X-Ray Science and Technology*, IOS Press, v. 29, n. 4, p. 551–566, 2021. Citado na página 68.
- DAI, J. et al. *Instance-sensitive Fully Convolutional Networks*. 2016. Citado na página 51.
- DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, Wiley Online Library, v. 26, n. 3, p. 297–302, 1945. Citado na página 62.
- DIVYA, R.; PETER, J. D. Smart healthcare system-a brain-like computing approach for analyzing the performance of detectron2 and posenet models for anomalous action detection in aged people with movement impairments. *Complex & Intelligent Systems*, Springer, p. 1–20, 2021. Citado na página 45.
- DOSOVITSKIY, A. et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. Citado 7 vezes nas páginas 15, 27, 44, 45, 47, 48 e 50.
- DOU, Q. et al. 3d deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*, Elsevier, 2017. Citado na página 27.
- DSA, E. *Capítulo 3 - O que são redes Neurais Artificiais Profundas ou deep learning?* 2022. Disponível em: <<https://www.deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>>. Citado 2 vezes nas páginas 15 e 36.
- FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado 2 vezes nas páginas 60 e 62.
- GAUR, L. et al. Medical image-based detection of covid-19 using deep convolution neural networks. *Multimedia systems*, Springer, p. 1–10, 2021. Citado na página 33.
- GHIASI, G. et al. *Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation*. 2021. Citado na página 47.
- GUIMARÃES, M. T. et al. An optimized approach to huntington’s disease detecting via audio signals processing with dimensionality reduction. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2020. p. 1–8. Citado na página 27.

- HAN, C. et al. Learning more with less: conditional ppgan-based data augmentation for brain metastases detection using highly-rough annotation on mr images. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. [S.l.: s.n.], 2019. p. 119–127. Citado na página 33.
- HAN, T. et al. Internet of medical things—based on deep learning techniques for segmentation of lung and stroke regions in ct scans. *IEEE Access*, IEEE, v. 8, p. 71117–71135, 2020. Citado 7 vezes nas páginas 17, 66, 67, 69, 70, 71 e 72.
- HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007. Citado na página 36.
- HE, K. et al. *Masked Autoencoders Are Scalable Vision Learners*. 2021. Citado na página 47.
- HE, K. et al. *Mask R-CNN*. 2018. Citado na página 51.
- HSIEH, J. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. [S.l.]: SPIE Press, 2003. ISBN 978-0819447131. Citado na página 25.
- INKEAW, P. et al. Automatic hemorrhage segmentation on head ct scan for traumatic brain injury using 3d deep learning model. *Computers in Biology and Medicine*, v. 146, p. 105530, 2022. ISSN 0010-4825. Citado 2 vezes nas páginas 26 e 30.
- ISLAM, M. et al. Ichnet: intracerebral hemorrhage (ich) segmentation using deep learning. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*. [S.l.], 2019. p. 456–463. Citado na página 72.
- KALITA, D. *A brief overview of recurrent neural networks (RNN)*. 2022. Disponível em: <<https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neural-networks-rnn/>>. Citado 2 vezes nas páginas 15 e 41.
- KAMTCHUM-TATUENE, J.; JICKLING, G. C. Blood biomarkers for stroke diagnosis and management. *Neuromolecular medicine*, Springer, v. 21, n. 4, p. 344–368, 2019. Citado na página 25.
- KARTHIK, R. et al. Neuroimaging and deep learning for brain stroke detection - a review of recent advancements and future prospects. *Computer Methods and Programs in Biomedicine*, v. 197, p. 105728, 2020. ISSN 0169-2607. Citado 2 vezes nas páginas 17 e 72.
- KHAN, M. J.; CURRY, E. Neuro-symbolic visual reasoning for multimedia event processing: Overview, prospects and challenges. In: *CIKM (Workshops)*. [S.l.: s.n.], 2020. Citado na página 33.
- KIRILLOV, A. et al. Segment anything. *arXiv:2304.02643*, 2023. Citado 9 vezes nas páginas 15, 16, 17, 46, 47, 48, 49, 50 e 51.
- KOCH, C.; TSUCHIYA, N. Attention and consciousness: two distinct brain processes. *Trends in cognitive sciences*, Elsevier, v. 11, n. 1, p. 16–22, 2007. Citado na página 35.

- KUMAR, A. et al. Csnets: A new deepnet framework for ischemic stroke lesion segmentation. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 193, p. 105524, 2020. Citado na página 68.
- KUO, W. et al. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 116, n. 45, p. 22737–22745, 2019. Citado na página 72.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 37.
- LI, Y. et al. *Exploring Plain Vision Transformer Backbones for Object Detection*. 2022. Citado 3 vezes nas páginas 47, 48 e 73.
- LI, Y. et al. *Fully Convolutional Instance-aware Semantic Segmentation*. 2017. Citado na página 51.
- LIAO TSE-SHENG CHEN, P.-C. C. P.-S. A fast algorithm for multilevel thresholding. *Journal of Information Science and Engineering*, JISE, v. 17, p. 713–727, 2001. Citado na página 57.
- MARQUES, A. G. et al. Automatic segmentation of hemorrhagic stroke on brain ct images using convolutional neural networks through fine-tuning. In: IEEE. *2022 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2022. p. 01–08. Citado 7 vezes nas páginas 17, 26, 27, 66, 67, 69 e 70.
- NAZARI-FARSANI, S. et al. Automated segmentation of acute stroke lesions using a data-driven anomaly detection on diffusion weighted mri. *Journal of neuroscience methods*, Elsevier, v. 333, p. 108575, 2020. Citado na página 68.
- NAZARI-FARSANI, S. et al. Predicting final ischemic stroke lesions from initial diffusion-weighted images using a deep neural network. *NeuroImage: Clinical*, Elsevier, v. 37, p. 103278, 2023. Citado na página 68.
- OTSU, T.; TANIGUCHI, G. Kolmogorov–smirnov type test for generated variables. *Economics Letters*, Elsevier, v. 195, p. 109401, 2020. Citado na página 63.
- PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. da. A survey on performance metrics for object-detection algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2020. p. 237–242. Citado 2 vezes nas páginas 59 e 60.
- PATEL, A. et al. Image level training and prediction: intracranial hemorrhage identification in 3d non-contrast ct. *Ieee Access*, IEEE, v. 7, p. 92355–92364, 2019. Citado na página 72.
- PATEL, A. et al. Intracerebral haemorrhage segmentation in non-contrast ct. *Scientific reports*, Nature Publishing Group UK London, v. 9, n. 1, p. 17858, 2019. Citado na página 72.
- PEIXOTO, S. A. et al. A new approach to segment hemorrhagic stroke in computed tomography via optimum path snakes. In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.: s.n.], 2017. p. 402–407. Citado 2 vezes nas páginas 67 e 70.

- PHONG, T. D. et al. Brain hemorrhage diagnosis by using deep learning. In: *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*. [S.l.: s.n.], 2017. p. 34–39. Citado na página [72](#).
- PRAVEEN, G. et al. Ischemic stroke lesion segmentation using stacked sparse autoencoder. *Computers in biology and medicine*, Elsevier, v. 99, p. 38–52, 2018. Citado na página [68](#).
- PROVOST, F.; KOHAVI, R. Glossary of terms. *Journal of Machine Learning*, v. 30, n. 2-3, p. 271–274, 1998. Citado na página [35](#).
- QIAN, W. et al. Mapping and candidate-gene screening of the novel turnip mosaic virus resistance gene retri02 in chinese cabbage (brassica rapa l.). *Theoretical and applied genetics*, Springer, v. 126, p. 179–188, 2013. Citado na página [62](#).
- RADFORD, A. et al. *Learning Transferable Visual Models From Natural Language Supervision*. 2021. Citado na página [48](#).
- RAJ, R. J. S. et al. Optimal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access*, IEEE, v. 8, p. 58006–58017, 2020. Citado na página [33](#).
- REBOUÇAS, E. d. S. et al. Level set based on brain radiological densities for stroke segmentation in ct images. In: *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.: s.n.], 2017. p. 391–396. Citado 3 vezes nas páginas [67](#), [69](#) e [70](#).
- RONNEBERGER, O.; FISCHER, P.; BROX, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. Citado na página [27](#).
- TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 1–28, 2015. Citado 4 vezes nas páginas [58](#), [59](#), [60](#) e [65](#).
- TANCIK, M. et al. *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains*. 2020. Citado na página [48](#).
- THRUN, S. . *Is learning the n-th thing any easier than learning the first?* 1995. Citado na página [47](#).
- TOUVRON, H. et al. *Training data-efficient image transformers distillation through attention*. 2021. Citado na página [50](#).
- TRIPATHI, A.; KOPPARAPU, S. K. Cnn based parkinson’s disease assessment using empirical mode decomposition. In: *CIKM (Workshops)*. [S.l.: s.n.], 2020. Citado na página [33](#).
- TRIPATHI, S.; VERMA, A.; SHARMA, N. Automatic segmentation of brain tumour in mr images using an enhanced deep learning approach. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, v. 9, n. 2, p. 121–130, 2021. Citado na página [33](#).
- VASWANI, A. et al. *Attention Is All You Need*. 2023. Citado 4 vezes nas páginas [15](#), [41](#), [42](#) e [43](#).

WARE, J. H. et al. P values. In: *Medical uses of statistics*. [S.l.]: CRC Press, 2019. p. 181–200. Citado na página 63.

WU, Y. et al. *Detectron2*. 2019. <<https://github.com/facebookresearch/detectron2>>. Citado 2 vezes nas páginas 45 e 66.

ZHANG, C. et al. Faster segment anything: Towards lightweight sam for mobile applications. *arXiv preprint arXiv:2306.14289*, 2023. Citado 4 vezes nas páginas 16, 17, 50 e 51.