

ErrorSim: a deep learning error simulator in left ventricle segmentations

Bruno F. Raquel¹, Matheus A. O. Ribeiro¹, Fátima L. S. Nunes¹

¹Laboratório de Aplicações de Informática em Saúde,
Escola de Artes, Ciências e Humanidades, Universidade de São Paulo
Rua Arlindo Bettio, 1000, São Paulo – SP, Brazil, 03828-000

{brunofraquel, matheus.alberto.oliveira, fatima.nunes}@usp.br

Abstract. *Left ventricle segmentation in Cardiac Magnetic Resonance exams is important for medical diagnosis. Deep learning methods have excelled in obtaining segmentations similar to those of experts. However, one of the current limitations is the arbitrary production of anatomical errors that can compromise the diagnosis. Given this problem, this work proposes a framework for categorization and controlled automatic simulation of different anatomical errors. The framework favors the development of methods aimed at detecting and correcting these errors. Results indicate that the proposed framework is capable of generating errors of several categories and replicating the same errors produced by networks ($Dice > 0.8$).*

Resumo. *A segmentação do ventrículo esquerdo em exames de Ressonância Magnética Cardíaca é importante para o diagnóstico médico. Métodos de aprendizado profundo têm se destacado ao obter segmentações semelhantes à de especialistas. Entretanto, uma das limitações atuais é a produção arbitrária de erros anatômicos que podem comprometer o diagnóstico. Dado esse problema, esse trabalho propõe um arcabouço para categorização e simulação controlada e automática de diferentes erros anatômicos. O arcabouço favorece o desenvolvimento de métodos voltados à detecção e à correção desses erros. Resultados indicam que o arcabouço proposto é capaz de gerar erros de diversas categorias e replicar os mesmos erros produzidos por redes ($Dice > 0.8$).*

1. Introdução

A correta segmentação do ventrículo esquerdo (VE) em imagens médicas viabiliza a extração de biomarcadores fundamentais para o diagnóstico e acompanhamento de doenças cardíacas. Dentre as diferentes modalidades de exames de imagem disponíveis para a análise do coração e obtenção de biomarcadores, destaca-se a Ressonância Magnética Cardíaca (RMC), reconhecida como o “padrão-ouro” para a obtenção de métricas cardíacas, com vantagens que incluem a precisão dos dados obtidos e a consistência desses dados quando o exame é repetido [Lohr et al. 2024].

A segmentação manual, realizada por especialistas, é um processo exaustivo, demorado e subjetivo, podendo apresentar variabilidade interobservador e intraobservador [Waite et al. 2017]. Para superar essas limitações, diversos métodos computacionais têm

sido desenvolvidos com o objetivo de automatizar a segmentação do VE, propiciando maior eficiência e velocidade na extração de biomarcadores clínicos [Wang et al. 2021].

Apesar de métodos baseados em redes de aprendizado profundo (AP) alcançarem, geralmente, resultados próximos aos especialistas [Bernard et al. 2018], os métodos de segmentação usualmente produzem erros anatômicos arbitrários que especialistas não cometeriam, sendo comuns e observados em diversas redes de AP [Bernard et al. 2018, Painchaud et al. 2020]. Como esses erros são arbitrários e os modelos de AP tendem a ser pouco interpretáveis, há uma dificuldade significativa em aprovar seu uso na prática clínica [Guo et al. 2020].

Embora os erros anatômicos em segmentações automáticas sejam comuns, há uma escassez de estudos dedicados à categorização e à simulação automática dessas falhas. Como os erros são arbitrários e há poucos estudos, há uma ausência de bases anotadas com erros anatômicos, o que dificulta o treinamento de métodos de AP especializados na caracterização e detecção. A capacidade de categorizar e reproduzir erros anatômicos de maneira controlada estimula o desenvolvimento de abordagens mais eficazes na detecção e correção dessas imprecisões, bem como a avaliação da robustez de algoritmos de segmentação.

Diante desse cenário, o presente trabalho tem como objetivo principal o desenvolvimento de um arcabouço para a simulação e análise sistemática de erros na segmentação do VE. A motivação central reside no fato de que, ao compreender e reproduzir artificialmente diferentes tipos de falhas, é possível criar um banco de casos controlados que favoreça o estudo, de forma isolada, o impacto de cada erro sobre métricas e análises subsequentes. Essa capacidade contribui para o desenvolvimento e a validação de métodos automáticos de detecção e correção, pois fornece exemplos representativos e calibráveis dos problemas que esses métodos precisam resolver. Para isso, propõe-se a criação de uma “fábrica de erros”, um ambiente controlado no qual diferentes tipos de falhas anatômicas em segmentações podem ser gerados, manipulados e estudados de forma calibrável e replicável. As principais contribuições desta pesquisa são: (i) a proposição de definições para os principais tipos de erros observados em segmentações médicas, com a descrição de características específicas para cada um; e (ii) o desenvolvimento de um sistema capaz de reproduzir erros comuns em segmentações automáticas a partir de segmentações consideradas corretas, permitindo controle sobre qual erro será aplicado e sua intensidade.

2. Trabalhos correlatos

A segmentação de estruturas cardíacas em imagens médicas tem sido amplamente explorada na literatura, com abordagens baseadas em atlas, modelos deformáveis, inteligência artificial, dentre outros [Ribeiro and Nunes 2022]. Tais trabalhos, em geral, se dedicam ao desenvolvimento de abordagens capazes de produzir segmentações análogas às de especialistas. No entanto, poucos trabalhos se dedicam à detecção, análise e replicação de erros anatômicos, que podem ocorrer arbitrariamente e muitas vezes são indetectáveis por métricas de avaliação comumente empregadas, exigindo inspeção manual dos resultados [Bernard et al. 2018]. Nesse contexto, a criação de erros controláveis em segmentações pode ser útil para, por exemplo, avaliar a sensibilidade de métricas de segmentação.

Como exemplo de abordagem que utiliza a replicação de erros para aumentar a eficiência do método, é possível citar o “ErrorNet” [Tajbakhsh et al. 2020], um arcabouço

de segmentação projetado para reduzir erros causados por variações entre domínios de treinamento e teste ou conjuntos de dados limitados, desenvolvido para o contexto de vasos da retina em imagens de fundo de olho. O método combina três redes de AP: uma rede de segmentação, uma rede de injeção de erro e uma rede de predição de erro. Durante o treinamento, a rede de injeção degrada a segmentação inicial e a rede de predição de erro estima um mapa de erro entre essa segmentação degradada e a segmentação real. Na fase de teste, a rede de predição recebe a segmentação e corrige os erros adicionando o mapa de erro estimado. Apesar de obter bons resultados, a utilização de AP para injeção de erros implica em uma solução menos calibrável e explicável, além de ser mais complexa e exigir dados para treinamento.

Para a sistematização de erros de segmentação gerados por modelos de aprendizado de máquina, o “MISS-tool” (*Medical Image Segmentation Synthesis Tool*) [Guan et al. 2023] emula erros em segmentações médicas, possibilitando a geração de segmentações sintéticas a partir de segmentações verdadeiras, categorizando diferentes tipos de erro parametrizáveis. Esta ferramenta tem foco em lesões pulmonares e outras estruturas que possuem apenas um contorno e uma máscara. Além disso, a geração de seus erros segue apenas a forma, não levando em conta informações da imagem como, por exemplo, a média das intensidades dos *pixels* da região de interesse.

Com o objetivo de superar o desafio de obter exemplares representativos da heterogeneidade dos dados do mundo real para melhorar o desempenho de redes de AP em imagens médicas [Graves et al. 2020], várias técnicas de aumento de dados têm sido propostas. Métodos como rotações, translações, deformações e adição de ruídos são frequentemente aplicados para expandir conjuntos de treinamento, reduzindo a dependência de grandes volumes de imagens anotadas manualmente [Cong and Zhang 2018, Khened et al. 2019]. No entanto, muitas vezes tais técnicas não são suficientes para gerar segmentações que contenham erros realistas. Diferentemente dessas técnicas, o arcabouço proposto é capaz de replicar artificialmente erros anatômicos em segmentações, como será apresentado na Seção 3.2, gerando novas segmentações que simulam falhas tipicamente observadas em métodos automáticos. A abordagem enriquece o conjunto de anotações com exemplos que abrangem diversas categorias, favorecendo o treinamento de modelos especializados na detecção e correção dessas inconsistências.

3. Materiais e métodos

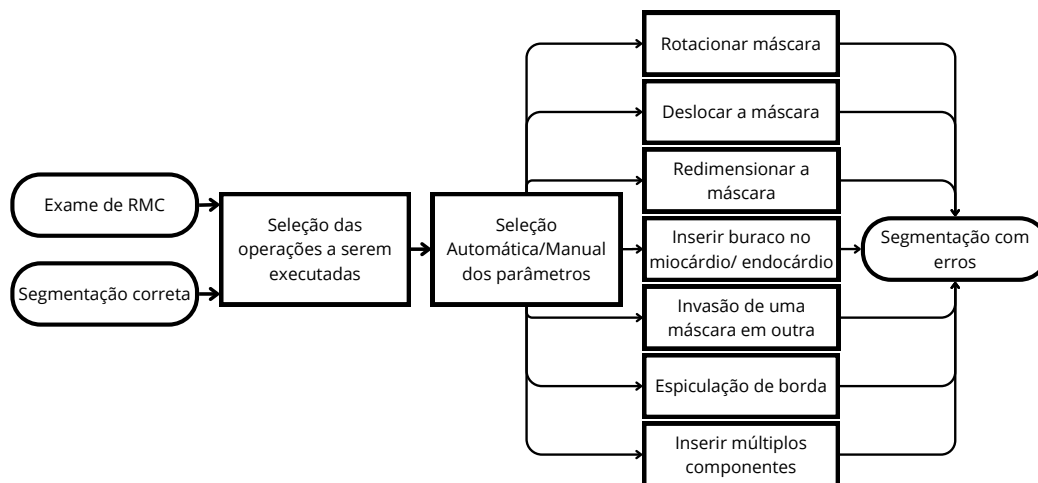
A segmentação do VE tem como objetivo identificar duas estruturas principais: o miocárdio e a câmara cardíaca. O miocárdio corresponde ao músculo cardíaco, enquanto a câmara cardíaca indica a região interna do VE, preenchida de sangue. Dessa forma, a segmentação é representada por duas máscaras que delimitam essas regiões.

O método proposto para simulação de erros assume como entrada uma segmentação correta usada como referência e a imagem original do exame de RMC. Em seguida, aplica operações selecionadas de forma aleatória ou manual para reproduzir falhas observadas em segmentações automáticas. A Figura 1 apresenta o processo para obtenção de erros e as diferentes operações implementadas.

3.1. Dataset e visão geral do método

Os exames de RMC foram obtidos a partir de três bases disponibilizadas publicamente: ACDC [Bernard et al. 2018], Sata13 [Landman and Warfield 2013] e Sunny

Figura 1. Fluxograma do método de geração de erros.



[Radau et al. 2009]. Essas bases são amplamente utilizadas pela literatura e possuem segmentações produzidas por especialistas que são utilizadas como segmentações corretas [Ribeiro and Nunes 2022, Cong and Zhang 2018, Graves et al. 2020].

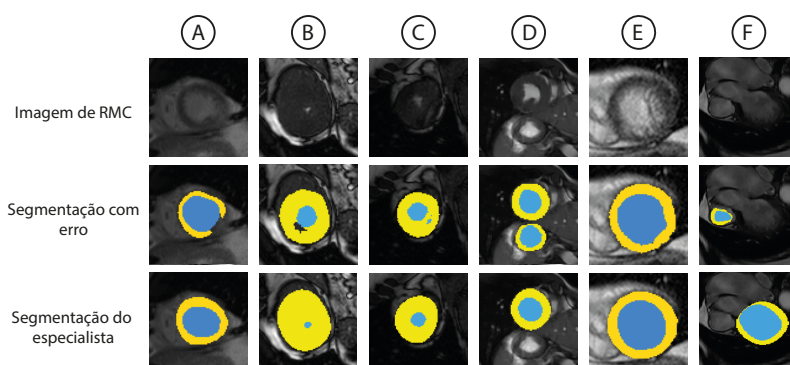
O método de segmentação é composto por três etapas principais: (1) a obtenção automática da região de interesse (ROI), que corresponde a uma subimagem centralizada no ventrículo esquerdo (VE); (2) o pré-processamento da ROI, que inclui equalização de histograma e redimensionamento para 256×256 pixels; e (3) a aplicação da rede de segmentação U-Net, responsável por classificar os pixels entre câmara cardíaca, miocárdio e fundo. O método completo é apresentado em [Ribeiro and Nunes 2021]. A partir das segmentações geradas por esse método, foram selecionados 390 casos contendo erros anatômicos diversos, que serviram de base para a catalogação e o desenvolvimento do arcabouço metodológico. A seleção desses casos foi realizada por meio da comparação visual entre as segmentações produzidas pelo método automático e aquelas feitas por um especialista médico.

3.2. Simulação de segmentações com erros

A partir da análise manual das segmentações contendo erros anatômicos produzidos pelo método descrito na Seção 3.1, seis categorias de erros foram definidas com base em características anatômicas do VE, conforme exemplificadas na Figura 2. A identificação de erros considera os seguintes critérios anatômicos comumente considerados em trabalhos correlatos [Painchaud et al. 2020]: i) VE possui formato arredondado; ii) miocárdio envolve completamente a câmara cardíaca; iii) não existência de buracos dentro das máscaras de segmentação; iv) regiões de uma mesma máscara sejam conexas. Destaca-se que, a existência de erros não é mutuamente exclusiva, isto é, é possível que um ou mais erros existam simultaneamente em uma mesma segmentação, com diferentes intensidades. O processo para simular segmentações para cada uma das categorias mencionadas é

detalhado a seguir.

Figura 2. Visão geral dos tipos de erros anatômicos identificados. Para cada erro, a primeira imagem não possui segmentação, a segunda imagem é a segmentação com erro anatômico e a terceira imagem é a segmentação do especialista. As regiões destacadas em azul e amarelo indicam respectivamente, a máscara da câmara cardíaca e do miocárdio: (A) Buraco no miocárdio; (B) Câmara cardíaca com tamanho maior do que o correto; (C) Máscara da câmara cardíaca invadindo o miocárdio; (D) Dois componentes desconexos segmentados; (E) Borda do miocárdio espiculada; (F) Segmentação está com seu centro deslocado para a esquerda.



3.2.1. Buraco no miocárdio e/ou câmara cardíaca

As áreas correspondentes ao miocárdio e à câmara cardíaca devem aparecer como regiões únicas e contínuas, sem interrupções internas e, portanto, não devem apresentar buracos ou lacunas. A presença de buracos nessas regiões ou em suas bordas caracteriza um erro de segmentação, que pode ser causado por artefatos de imagem ou por falhas na diferenciação de tecidos adjacentes. Esse tipo de erro é especialmente problemático, pois pode levar à obtenção incorreta de biomarcadores importantes como a massa do VE ou volume de sangue bombeado.

Para simular esse erro, o processo inicia definindo um ponto de partida para a criação do buraco. Caso o usuário não especifique esse ponto, o algoritmo analisa a região da estrutura indicada pela máscara na qual será inserido o buraco e seleciona, dentre os *pixels* presentes, aquele cujo valor difere mais significativamente da média das intensidades dos *pixels* da estrutura de interesse, medido a partir de um mapa de calor associada à região.

Uma vez definido o ponto inicial, o algoritmo expande o buraco de maneira iterativa. Em cada iteração, a expansão é orientada pela análise dos *pixels* adjacentes à borda atual do buraco. Essa vizinhança é determinada por uma janela de tamanho parametrizável, com valor padrão de 5×5 *pixels*, na qual calcula-se a média das intensidades na vizinhança. Os *pixels* com maiores médias – até um número máximo definido pelo parâmetro *top-K* – são então considerados candidatos a serem removidos, isto é, para serem incorporados ao buraco.

A seleção dos *pixels* com as maiores médias prioriza aqueles que se destacam da continuidade da região segmentada, promovendo uma remoção que simula a formação de

uma lacuna consistente. A escolha, dentre os candidatos, é realizada de forma aleatória, utilizando uma semente para garantir que o mesmo erro possa ser replicado com os mesmos parâmetros. Enquanto o número de iterações e o ponto inicial definem a intensidade do erro, o tamanho da janela e o limite de candidatos definem o quão disperso os *pixels* serão em relação ao centro do erro.

3.2.2. Tamanho dilatado ou reduzido do miocárdio ou câmara cardíaca

O erro de dimensionamento em segmentações médicas ocorre quando uma estrutura é identificada como maior ou menor do que deveria ser (Figura 2B). No caso do miocárdio, uma segmentação dilatada pode sugerir hipertrofia ventricular [Sayin and Oto 2022], por exemplo. Da mesma forma, uma câmara cardíaca superdimensionada leva à superestimação do volume de sangue, enquanto um subdimensionada implica na subestimação, podendo acarretar no diagnóstico incorreto de cardiomiopatias.

Esse erro é simulado por meio de uma transformação geométrica baseada em uma matriz de escala afim. Inicialmente, os parâmetros de redimensionamento no eixo vertical R_x e horizontal R_y são definidos para determinar o quanto o miocárdio, câmara cardíaca ou ambas as máscaras serão expandidas ou reduzidas ao longo dos eixos. A transformação é então aplicada utilizando uma matriz afim da forma definida na Equação (1), em que w e h representam a largura e a altura da máscara original, respectivamente.

$$T = \begin{bmatrix} R_x & 0 & (1 - R_x) \cdot w/2 \\ 0 & R_y & (1 - R_y) \cdot h/2 \end{bmatrix} \quad (1)$$

Por exemplo, ao definir $R_x = 1.2$ e $R_y = 1.0$, o contorno será dilatado horizontalmente em 20%, mantendo a altura inalterada. Embora a matriz seja 2x3, sua aplicação ocorre usando-se coordenadas homogêneas (i.e., $[x, y, 1]^T$), o que torna a transformação adequada para alterar a escala da máscara. Dessa forma, essa matriz ajusta a posição dos *pixels* para que a estrutura seja redimensionada proporcionalmente ao redor de seu centro.

3.2.3. Invasão de uma máscara em outra

Esse erro ocorre quando uma região anatômica invade indevidamente outra, alterando as fronteiras naturais do VE (Figura 2C). Ele pode ocorrer de forma interna, quando o erro de invasão começa pelo interior da outra máscara levando a múltiplos componentes segmentados; ou pela borda, quando a invasão começa pela borda da câmara cardíaca ou miocárdio.

Para simular um erro de invasão de forma interna, o procedimento define primeiramente qual máscara atuará como fonte e qual atuará como alvo; se o ponto inicial não for especificado, ele é escolhido aleatoriamente a partir dos *pixels* com maior distância da média de intensidade da região de interesse. Em seguida, utilizando parâmetros como o tamanho da janela para análise local (*kernel*) e o número de iterações, o método itera sobre a região, criando uma máscara temporária que identifica os *pixels* adjacentes à área alvo. Para cada *pixel* candidato, é calculada a média de intensidade em sua vizinhança, e

o *pixel* com a maior média é selecionado para ser transferido da região alvo para a fonte, expandindo progressivamente a invasão.

A invasão de uma máscara em outra pela borda pode ser realizada por meio de parâmetros que governam a alteração do contorno, incluindo o ângulo em que o erro será inserido (θ), a magnitude da modulação radial (h) e a extensão lateral da modificação (w). A partir da máscara em que deseja-se inserir o erro, extrai-se o exterior da estrutura, organizado de modo a possibilitar a determinação do centro de massa por meio da média das coordenadas dos pontos que compõem o contorno. Esse contorno é então convertido para um sistema de coordenadas polares, no qual os ângulos e as distâncias dos pontos em relação ao centro estabelecem uma descrição da forma original.

No estágio seguinte, a modulação é realizada ao alterar os valores radiais dos pontos do contorno com base em uma função que integra o ângulo de cada ponto, um ângulo de referência, a magnitude (altura) e a extensão lateral (largura). Essa transformação gera um novo conjunto de coordenadas que, reconvertidas para o sistema cartesiano, definem uma fronteira modificada e, conseqüentemente, uma nova máscara.

3.2.4. Presença de múltiplos componentes

O erro de múltiplos componentes ocorre quando uma mesma estrutura anatômica é representada por mais de uma região segmentada desconexa (Figura 2D). Para simular essa falha, primeiramente é realizada uma análise das áreas de interesse por meio da identificação de regiões candidatas a partir de suas intensidades serem próximas às intensidades da câmara cardíaca. A partir dessas regiões, são extraídos os seus contornos com um método baseado em uma versão do algoritmo de Suzuki-Abe [Suzuki and Abe 1985], que podem ser entendidos como possíveis bordas da câmara cardíaca do novo componente que será inserido. Em seguida, a operação seleciona, dentre todos os contornos, o mais próximo ao ponto de referência especificado pelo usuário; caso o ponto não seja fornecido, o método seleciona aleatoriamente um ponto pertencente ao VE.

Após a escolha do contorno, este é redesenhado sobre uma nova máscara. Após a definição da câmara cardíaca, é adicionado o miocárdio deste novo componente, com espessura igual em todas as direções. Além disso, outras operações podem ser aplicadas para modificar a forma e a localização desse novo componente, favorecendo a inserção de falhas adicionais. Dessa maneira, o processo possibilita a criação de múltiplos componentes desconexos.

3.2.5. Espiculação de borda

Segmentações adequadas do VE devem apresentar bordas suaves, pois a morfologia do ventrículo é geralmente arredondada. Um erro de borda espiculada ocorre quando a fronteira da segmentação apresenta uma aparência irregular, com artefatos espiculados (Figura 2E). Esse erro é simulado por meio do seguinte procedimento: primeiramente, extrai-se o contorno exterior do miocárdio. Em seguida, para cada iteração, é gerado um ruído binário aleatório que é aplicado exclusivamente aos *pixels* dessa borda, em que alguns *pixels* são removidos aleatoriamente. Essa operação ocasiona na formação gradativa de “de-

graus” ou interrupções na suavidade da linha de contorno, resultando em uma aparência espiculada.

3.2.6. Segmentação em local errado

Esse erro ocorre quando a segmentação é realizada em uma região fora dos limites esperados para o VE (Figura 2F). Isso pode acontecer devido a semelhanças de intensidade com estruturas vizinhas, como o ventrículo direito e a cárdia.

A primeira operação aplica uma transformação de rotação à máscara segmentada. Essa rotação pode simular erros em que a segmentação ocorre em uma orientação incorreta. Inicialmente, o método calcula o centro de massa da região de interesse a partir dos *pixels* pertencentes à máscara de interesse que a rotação será aplicada que pode ser o miocárdio, câmara cardíaca ou ambos. Após isso, utiliza-se o centro de massa como ponto central para definir uma matriz afim de rotação com um ângulo especificado (ou escolhido aleatoriamente). A matriz de rotação utilizada é definida na Equação (2), em que θ representa o ângulo de rotação, definido em graus e convertido para radianos; x_c, y_c corresponde ao centro de massa da máscara, calculado a partir dos *pixels* pertencentes à região de interesse; os termos adicionais na última coluna da matriz garantem que a rotação ocorra em torno do centro de massa, mantendo sua posição relativa dentro da imagem.

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & (1 - \cos(\theta))x_c + \sin(\theta)y_c \\ \sin(\theta) & \cos(\theta) & (1 - \cos(\theta))y_c - \sin(\theta)x_c \end{bmatrix} \quad (2)$$

A segunda operação, por sua vez, emprega uma transformação de translação. Os *pixels* pertencentes à região de interesse são identificados e, com base em valores de deslocamento definidos para as direções horizontal (D_x) e vertical (D_y), são movidos para novas posições. O método verifica se os novos índices permanecem válidos dentro dos limites da imagem.

3.3. Experimento de validação

Para demonstrar a capacidade do método proposto em replicar erros reais observados nas segmentações automáticas analisadas (Seção 3.1), foi conduzido um experimento comparando as segmentações geradas artificialmente com aquelas produzidas por redes de AP. Inicialmente, foram selecionados exemplos de erros reais pertencentes a cada um dos tipos de falhas previamente definidas (Figura 2). Em seguida, o arcabouço foi utilizado para degradar a segmentação correta do especialista com o intuito de replicar a mesma falha observada.

A similaridade das segmentações contendo erros sintéticos e erros reais foi quantificada por meio do coeficiente de *Dice*, uma métrica amplamente utilizada para avaliar a sobreposição entre duas máscaras segmentadas. O coeficiente de *Dice* é definido na Equação (3), em que A e B representam duas máscaras binárias, $|A \cap B|$ é o número de *pixels* na interseção entre elas, e $|A|$ e $|B|$ são os números de *pixels* em cada uma, respectivamente.

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

4. Resultados e discussões

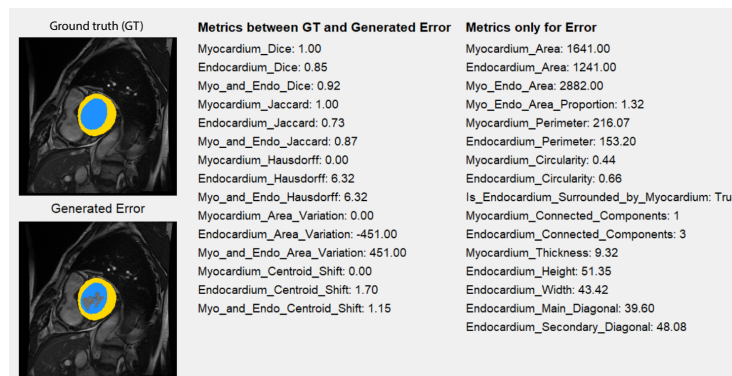
Nesta seção, a ferramenta ErrorSim será apresentada. A Seção 4.1 descreve a interface gráfica da ferramenta implementada. Já a Seção 4.2 discute a replicação de erros reais, validando a capacidade do método proposto em gerar falhas sintéticas semelhantes a erros observados em modelos de AP.

4.1. Ferramenta implementada

A ferramenta permite a inserção de erros tanto de forma automática quanto manual. Sua interface gráfica facilita tanto a configuração dos parâmetros quanto a visualização dos erros produzidos por cada método de modificação. A flexibilidade de combinar múltiplos métodos geradores de erros, cada qual com diferentes valores de parâmetro, o que confere ao arcabouço uma alta capacidade de geração de erros diversificados.

Como pode ser visualizado na Figura 3, a ferramenta dispõe de uma funcionalidade de monitoramento e visualização de métricas de qualidade da segmentação, facilitando a quantificação do impacto de cada erro. Entre as métricas disponíveis, destacam-se aquelas relacionadas à similaridade das máscaras (*Dice*), distâncias geométricas (*Hausdorff*), variação de área, deslocamento de centroide, conectividade dos componentes, circularidade e espessura do miocárdio, dentre outras. A análise conjunta dos indicadores possibilita compreender de forma mais aprofundada como cada tipo de falha afeta a confiabilidade das métricas cardíacas. Isso favorece, por exemplo, verificar se o endocárdio permanece completamente envolto pelo miocárdio ou se surgem múltiplos componentes desconexos. Essa abordagem pode revelar padrões específicos de degradação associados a determinados erros de segmentação, oferecendo subsídios para a proposição de uma métrica integrada capaz de avaliar, de forma mais abrangente, o desempenho das segmentações nesse domínio.

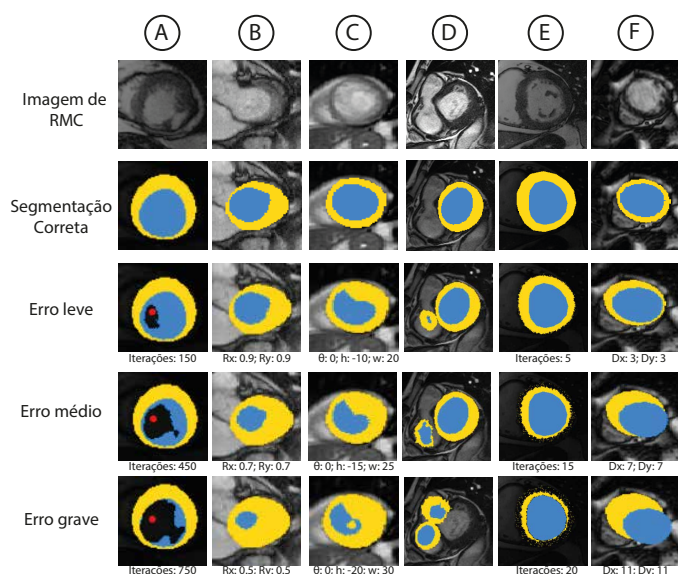
Figura 3. Interface para visualização de métricas. Neste caso, o erro gerado foi um buraco na câmara cardíaca.



A Figura 4 mostra como é possível gerar diferentes tipos e intensidades de erros em segmentações, a partir da variação de um parâmetro de controle. Em cada cenário,

o parâmetro ajustado provoca alterações específicas na máscara de segmentação, criando desde falhas leves até distorções graves na delimitação do miocárdio e da câmara cardíaca.

Figura 4. Diferentes tipos de erro introduzidos na segmentação, com variação de parâmetros que gera diferentes intensidades de erro: (A-F) mesma ordem de erros que a Figura 2; para o erro de múltiplos componentes (D), a espessura do miocárdio é de 4 *pixels*, mas outros erros são também aplicados. Em todas as representações, o ponto vermelho marca a posição definida pelos parâmetros *x* e *y* que orientam a origem do erro, para a seleção dos *pixels* de maior intensidade, selecionando sempre as 10 maiores intensidades ($K = 10$).

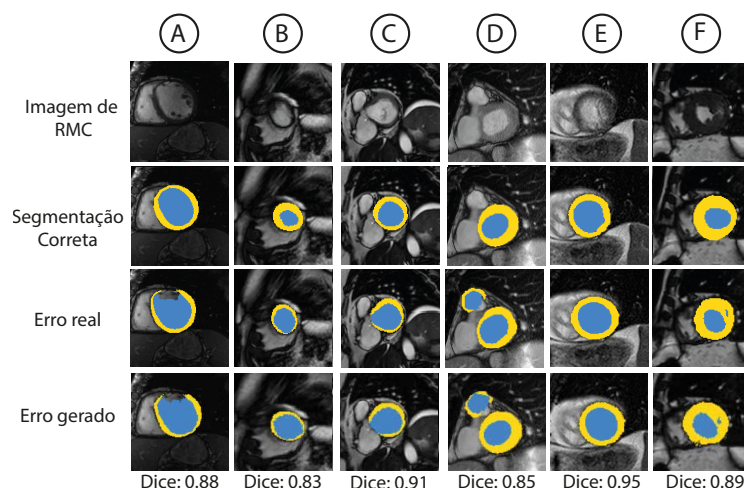


4.2. Replicação de erros reais

A Figura 5 apresenta um conjunto de exemplos nos quais se compara segmentações com erros reais produzidos por uma rede de AP e erros sinteticamente gerados pela ferramenta proposta. Cada linha, de A a F, ilustra um caso distinto de erro. O valor de *Dice* exibido em cada exemplo reflete o grau de sobreposição entre as segmentações, sendo considerado alto (variando de 0,83 a 0,95). Esses resultados indicam que o arcabouço é capaz de simular falhas anatômicas observáveis em segmentações produzidas por redes de AP.

Apesar da utilidade da geração de dados sintéticos, o arcabouço proposto apresenta algumas limitações. Os erros gerados artificialmente podem não capturar a complexidade das falhas reais, que podem surgir devido a fatores desconsiderados ou existência de ruídos específicos presentes nos dados de entrada. Além disso, alguns erros podem ser resultados de padrões que só emergem em condições específicas e imprevisíveis da rede de AP, o que torna a simulação dessas falhas um desafio adicional. Uma forma de superar essas limitações é aprimorar os métodos de geração de erros sintéticos por meio da incorporação de novos conhecimentos especializados obtidos a partir de análises de padrões em falhas produzidas por redes de AP e outros métodos de segmentação. Adicionalmente, o presente trabalho se concentra em apenas algumas possíveis categorias de falhas anatômicas, selecionadas por sua relevância prática e frequência observada em trabalhos correlatos [Painchaud et al. 2020]. A identificação e replicação de novas categorias de erros será investigada em trabalhos futuros.

Figura 5. Reprodução de erros reais a partir dos métodos descritos.(A-F) mesma ordem de erros que a Figura 2. Os valores de *Dice* informados se referem à média das regiões do miocárdio e câmara cardíaca.



5. Conclusão

Nesse trabalho, um arcabouço voltado à simulação de erros anatômicos comuns em segmentações automáticas produzidas por redes de AP foi proposto. Diferentes tipos de erros foram categorizados e métodos para simulação de cada tipo foram propostos. A ferramenta proposta fornece um ambiente estruturado para a criação de erros sintéticos calibráveis, análise de falhas e extração de métricas. Os resultados obtidos sugerem que o arcabouço é capaz de replicar erros reais observados em segmentações do VE produzidas por redes de AP. A capacidade de gerar erros representativos de casos reais favorece o desenvolvimento de novas estratégias voltadas à redução desses erros.

Como proposta para trabalhos futuros, pretende-se utilizar os exemplos gerados como técnica de aumento de dados na criação de métodos especializados na detecção e correção automáticas de falhas em segmentações, o que pode contribuir para a melhoria da precisão dos métodos existentes. Novas categorias de erros serão investigadas e incorporadas à ferramenta para simular erros ainda mais variados e realistas.

6. Agradecimentos

Este trabalho foi financiado pelo CNPq [número de concessão 122803/2024-8].

Referências

- Bernard, O. et al. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525.
- Cong, C. and Zhang, H. (2018). Invert-u-net dnn segmentation model for MRI cardiac left ventricle segmentation. *The Journal of Engineering*, 2018(16):1463–1467.
- Graves, C. V., Moreno, R. A., Rebelo, M. S., Nomura, C. H., and Gutierrez, M. A. (2020). Improving the generalization of deep learning methods to segment the left ventricle in short axis MR images. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, page 1203–1206. IEEE.

- Guan, S., Samala, R. K., Arab, A., and Chen, W. (2023). Miss-tool: medical image segmentation synthesis tool to emulate segmentation errors. In Iftekharuddin, K. M. and Chen, W., editors, Medical Imaging 2023: Computer-Aided Diagnosis, page 41. SPIE.
- Guo, F., Ng, M., Goubran, M., Petersen, S. E., Piechnik, S. K., Neubauer, S., and Wright, G. (2020). Improving cardiac mri convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach. Medical Image Analysis, 61:101636.
- Khened, M., Kollerathu, V. A., and Krishnamurthi, G. (2019). Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis, 51:21–45.
- Landman, B. and Warfield, S. (2013). 2013 cardiac atlas project standard challenge - participant project.
- Lohr, D. et al. (2024). Precision imaging of cardiac function and scar size in acute and chronic porcine myocardial infarction using ultrahigh-field mri. Communications Medicine, 4(1).
- Painchaud, N., Skandarani, Y., Judge, T., Bernard, O., Lalande, A., and Jodoin, P.-M. (2020). Cardiac segmentation with strong anatomical guarantees. IEEE Transactions on Medical Imaging, 39(11):3703–3713.
- Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A. J., and Wright, G. A. (2009). Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal.
- Ribeiro, M. A. O. and Nunes, F. L. S. (2021). Evaluating the pre-processing impact on the generalization of deep learning networks for left ventricle segmentation. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), page 3505–3512. IEEE.
- Ribeiro, M. A. O. and Nunes, F. L. S. (2022). Left ventricle segmentation in cardiac MR: A systematic mapping of the past decade. ACM Comput. Surv., 54(11s).
- Sayin, B. Y. and Oto, A. (2022). Left ventricular hypertrophy: Etiology-based therapeutic options. Cardiology and Therapy, 11(2):203–230.
- Suzuki, S. and Abe, K. (1985). Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46.
- Tajbakhsh, N., Lai, B., Ananth, S. P., and Ding, X. (2020). Errornet: Learning error representations from limited data to improve vascular segmentation. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), page 1364–1368. IEEE.
- Waite, S., Kolla, S., Jeudy, J., Legasto, A., Macknik, S. L., Martinez-Conde, S., Krupinski, E. A., and Reede, D. L. (2017). Tired in the reading room: The influence of fatigue in radiology. Journal of the American College of Radiology, 14(2):191–197.
- Wang, Y. et al. (2021). Deep learning based fully automatic segmentation of the left ventricular endocardium and epicardium from cardiac cine mri. Quantitative Imaging in Medicine and Surgery, 11(4):1600–1612.