

none



Bruno Yusuke Kitabayashi

**Estratégias para otimizar processos de
anotação e geração de datasets de
segmentação semântica em imagens de
mamografia**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática, do Departamento de Informática da PUC-Rio .

Orientador: Prof. Alberto Barbosa Raposo

Rio de Janeiro
Outubro de 2022



Bruno Yusuke Kitabayashi

**Estratégias para otimizar processos de
anotação e geração de datasets de
segmentação semântica em imagens de
mamografia**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio . Aprovada pela Comissão Examinadora abaixo:

Prof. Alberto Barbosa Raposo

Orientador

Departamento de Informática – PUC-Rio

Dr. Marcelo Blois Ribeiro

Centro de Pesquisa Global da GE

Prof. Sérgio Colcher

PUC-Rio

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Bruno Yusuke Kitabayashi

Graduado em Engenharia da Computação pela Faculdade Estácio IESAM de Belém (Pará, Brasil) em 2017.

Ficha Catalográfica

Bruno,Kitabayashi

Estratégias para otimizar processos de anotação e geração de datasets de segmentação semântica em imagens de mamografia / Bruno Yusuke Kitabayashi; orientador: Alberto Barbosa Raposo. – 2022.

107 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2022.

Inclui bibliografia

1. Informática – Teses. 2. Active Learning. 3. Anotações Assistidas. 4. Aprendizado Profundo. 5. Segmentação Semântica. 6. Imagens de Mamografia. I. Raposo, Alberto. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Ao meu avô Shozo e à minha avó Chiyoko.

Agradecimentos

Ao meu professor e orientador Alberto Barbosa Raposo por sempre me apoiar e orientar durante esta longa jornada no mestrado.

Ao César A. Sierra Franco pelos ensinamentos e orientações mediante à pesquisa.

Aos meus colegas de equipe do instituto Tecgraf que nunca mediram esforços para me ajudar e apoiar.

À minha namorada Valquíria Reis pela paciência, compreensão e apoio durante etapas difíceis que tive que confrontar durante o mestrado.

Ao Pr. Cláudio Lísias e Pra. Walnice Reis, por sempre rezarem por mim e serem meus mentores religiosos durante esta jornada.

À minha psicóloga Dra. Ana pela disponibilidade e compromisso em escutar meus problemas e ajudar a confrontar meus desafios.

Aos amigos que tive oportunidade de conhecer na PUC-Rio e me ajudaram ao longo do caminho.

Aos meus amigos de infância: Abner, Fabrício, Esther, Arthur e Fabiana, que sempre pude contar com eles desde o início.

Aos meus amigos de longa data: Jhonathas, Thiago, Renan, Emanuel e Matheus por tornarem meus dias mais alegres.

À minha família, por entenderem que esta etapa da minha vida é de suma importância.

À CAPES, Instituto Tecgraf e PUC-Rio, por me concederem condições e auxílio para realização deste trabalho.

Não menos importante, à Deus, por escutar minhas angustias e traçar o caminho em que está em suas mãos na minha vida.

Sou eternamente grato, muito obrigado!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

Bruno,Kitabayashi; Raposo, Alberto. **Estratégias para otimizar processos de anotação e geração de datasets de segmentação semântica em imagens de mamografia**. Rio de Janeiro, 2022. 107p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com o avanço recente do uso de aprendizagem profunda supervisionada (supervised deep learning) em aplicações no ramo da visão computacional, a indústria e a comunidade acadêmica vêm evidenciando que uma das principais dificuldades para o sucesso destas aplicações é a falta de datasets com a suficiente quantidade de dados anotados. Nesse sentido aponta-se a necessidade de alavancar grandes quantidades de dados rotulados para que estes modelos inteligentes possam solucionar problemas pertinentes ao seu contexto para atingir os resultados desejados. O uso de técnicas para gerar dados anotados de maneira mais eficiente está sendo cada vez mais explorado, juntamente com técnicas para o apoio à geração dos datasets que servem de insumos para o treinamento dos modelos de inteligência artificial. Este trabalho tem como propósito propor estratégias para otimizar processos de anotação e geração de datasets de segmentação semântica. Dentre as abordagens utilizadas neste trabalho destacamos o Interactive Segmentation e Active Learning. A primeira, tenta melhorar o processo de anotação de dados, tornando-o mais eficiente e eficaz do ponto de vista do anotador ou especialista responsável pela rotulagem dos dados com uso de um modelo de segmentação semântica que tenta imitar as anotações feitas pelo anotador. A segunda, consiste em uma abordagem que permite consolidar um modelo deep learning utilizando um critério inteligente, visando a seleção de dados não anotados mais informativos para o treinamento do modelo a partir de uma função de aquisição que se baseia na estimação de incerteza da rede para realizar a filtragem desses dados. Para aplicar e validar os resultados de ambas as técnicas, o trabalho os incorpora em um caso de uso relacionado em imagens de mamografia para segmentação de estruturas anatômicas.

Palavras-chave

Active Learning; Anotações Assistidas; Aprendizado Profundo; Segmentação Semântica; Imagens de Mamografia.

Abstract

Bruno,Kitabayashi; Raposo, Alberto (Advisor). **Strategies to optimize annotation processes and generation of semantic segmentation datasets in mammography images**. Rio de Janeiro, 2022. 107p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

With the recent advancement of the use of supervised deep learning in applications in the field of computer vision, the industry and the academic community have been showing that one of the main difficulties for the success of these applications is the lack of datasets with a sufficient amount of annotated data. In this sense, there is a need to leverage large amounts of labeled data so that these intelligent models can solve problems relevant to their context to achieve the desired results. The use of techniques to generate annotated data more efficiently is being increasingly explored, together with techniques to support the generation of datasets that serve as inputs for the training of artificial intelligence models. This work aims to propose strategies to optimize annotation processes and generation of semantic segmentation datasets. Among the approaches used in this work, we highlight Interactive Segmentation and Active Learning. The first one tries to improve the data annotation process, making it more efficient and effective from the point of view of the annotator or specialist responsible for labeling the data using a semantic segmentation model that tries to imitate the annotations made by the annotator. The second consists of an approach that allows consolidating a deep learning model using an intelligent criterion, aiming at the selection of more informative unannotated data for training the model from an acquisition function that is based on the uncertainty estimation of the network to filter these data. To apply and validate the results of both techniques, the work incorporates them in a use case in mammography images for segmentation of anatomical structures.

Keywords

Active Learning; Assisted Labelling; Deep Learning; Semantic Segmentation; Mammography Images.

Sumário

| | | |
|----------|--|------------|
| 1 | Introdução | 16 |
| 1.1 | Problema | 18 |
| 1.2 | Objetivo | 23 |
| 1.3 | Estrutura da dissertação | 23 |
| 2 | Conceitos Básicos | 25 |
| 2.1 | Mamografia | 25 |
| 2.2 | Segmentação Semântica | 26 |
| 2.3 | Segmentação Interativa | 28 |
| 2.4 | Active Learning | 29 |
| 2.5 | Medidas e Análise de Formas Geométricas | 34 |
| 3 | Trabalhos Relacionados | 37 |
| 3.1 | Segmentação Interativa | 37 |
| 3.2 | Active learning em aprendizagem profunda | 39 |
| 3.3 | Oportunidades na literatura | 41 |
| 4 | Anotações assistidas para dobras de pele MLO | 44 |
| 4.1 | Experimento Manual | 44 |
| 4.2 | Experimento Interativo | 46 |
| 4.3 | Resultados sobre o modelo de segmentação de dobras | 47 |
| 4.4 | Resultados e análise dos tempos de anotação entre experimentos | 48 |
| 5 | Anotações assistidas para estruturas mamárias CC e MLO | 55 |
| 5.1 | Experimento Manual e Interativo | 55 |
| 5.2 | Resultados | 57 |
| 5.3 | Análise dos resultados | 62 |
| 6 | Active Learning | 67 |
| 6.1 | Etapas de active learning | 67 |
| 6.2 | TTA: Test-Time Augmentation | 68 |
| 6.3 | Estimação da incerteza | 71 |
| 6.4 | Scores de incerteza | 76 |
| 6.5 | Resultados | 90 |
| 6.6 | Combinação dos resultados | 97 |
| 6.7 | Análise dos resultados | 99 |
| 7 | Conclusão e Trabalhos Futuros | 102 |
| 8 | Referências bibliográficas | 105 |

Lista de figuras

| | | |
|------------|--|----|
| Figura 1.1 | Mapa mental sobre o processo de human-in-the-loop, adaptado e retirado de (MONARCH, 2021) | 18 |
| Figura 1.2 | Exemplo de duas imagens de mamografia anotadas manualmente. | 21 |
| Figura 1.3 | Exemplos de dobras de pele na incidência MLO. | 22 |
| Figura 2.1 | Figura à esquerda, incidência CC. Figura à direita, incidência MLO. | 25 |
| Figura 2.2 | Estruturas anatômicas representadas em um exame de mamografia. | 26 |
| Figura 2.3 | Exemplo de um exame de mamografia e sua respectiva máscara de segmentação indicando as estruturas anatômicas da mama em classes distintas. | 27 |
| Figura 2.4 | Um dos aportes principais de active learning. | 30 |
| Figura 2.5 | Etapas iniciais do processo de active learning. | 31 |
| Figura 2.6 | Fluxograma sobre o funcionamento da abordagem e suas etapas | 32 |
| Figura 2.7 | Esquematização da técnica, test-time augmentation. | 33 |
| Figura 4.1 | Segmentações geradas pela rede na ferramenta. | 47 |
| Figura 4.2 | Distribuição dos tempos de anotação: população e amostragem do processo manual | 51 |
| Figura 4.3 | Distribuição dos tempos de anotação: população e amostragem do processo interativo | 51 |
| Figura 4.4 | Comparativo sobre tempo de anotação entre as abordagens | 52 |
| Figura 4.5 | Distribuição dos tempos de anotação a partir das amostragens | 53 |
| Figura 4.6 | Contraste entre os tempos de anotação de ambas abordagens em detalhes. | 54 |
| Figura 5.1 | Interface das etiquetas criadas e seus mecanismos. | 56 |
| Figura 5.2 | Tempo médio de anotação por estrutura mamária entre os participantes em ambos experimentos . | 59 |
| Figura 5.3 | Soma dos tempos médios das estruturas entre cada participante em ambos experimentos. | 60 |
| Figura 5.4 | Distribuição dos tempos de anotações por estrutura mamária em ambos experimentos entre os participantes | 61 |
| Figura 5.5 | Tempos de duração dos experimentos e número de interações obtidos pelo OGAMA em ambos experimentos entre os participantes. | 62 |
| Figura 5.6 | Exemplos de segmentações de papila durante os experimentos interativos. | 64 |
| Figura 5.7 | Caso raro de segmentação incorreta de dobra de pele | 65 |
| Figura 6.1 | Ciclo active learning para segmentação de imagens de mamografia. | 68 |
| Figura 6.2 | Amostras geradas a partir de TTA e as inferências do modelo sobre as imagens transformadas. | 70 |
| Figura 6.3 | Amostras geradas a partir de TTA com perturbações excessivas. | 71 |

| | | |
|-------------|--|----|
| Figura 6.4 | Esquematização do processo para obtenção dos mapas de incerteza. | 72 |
| Figura 6.5 | Imagem com baixa incerteza sobre as inferências geradas de TTA das estruturas mamárias. | 73 |
| Figura 6.6 | Imagem com alta incerteza sobre as inferências geradas de TTA das estruturas mamárias. | 74 |
| Figura 6.7 | Alguns exemplos de mapas de incerteza por estrutura mamária. | 75 |
| Figura 6.8 | Exemplos de segmentações boas e mapas de incertezas bem contidas em relação a predição das imagens de mamografia | 78 |
| Figura 6.9 | Ilustração para obtenção do comprimento de uma segmentação ideal baseado em P_{mp} | 79 |
| Figura 6.10 | Ilustração para obtenção do comprimento das áreas mais quentes em U_{mp} | 80 |
| Figura 6.11 | Exemplo onde S_1 tende à 1, pelo fato do comprimento $C(U_{mp})$ se distanciar de $C(P_{mp})$. | 80 |
| Figura 6.12 | Exemplo onde a análise topológica aborda duas segmentações de peitoral | 82 |
| Figura 6.13 | Exemplo do processo de cálculo da medida de incerteza baseada na solidez. | 82 |
| Figura 6.14 | Distância de Hausdorff representada pela reta em verde, indicando a maior distância entre os pontos correspondentes em P_x e P_y | 83 |
| Figura 6.15 | Caso onde a segmentação P_{mp} possui uma excentricidade baixa. | 84 |
| Figura 6.16 | Exemplos de segmentações boas e mapas de incertezas bem contidas em relação às suas respectivas imagens | 86 |
| Figura 6.17 | Exemplo onde S_1 tende a 1, por conta das regiões mais quentes obtidas em U_p se localizarem distantes da região de segmentação em P_p | 87 |
| Figura 6.18 | Exemplo onde observa-se duas componentes conexas em P_p . | 88 |
| Figura 6.19 | Um possível caso onde a solidez calculada é baixa, aumentando a medida de incerteza S_3 . | 89 |
| Figura 6.20 | O IoU médio de todas as estruturas mamárias por ciclo. | 92 |
| Figura 6.21 | O IoU exclusivo da estrutura mamária papila por ciclo. | 92 |
| Figura 6.22 | O IoU exclusivo da estrutura mamária músculo peitoral por ciclo. | 93 |
| Figura 6.23 | O IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via S_{papila} . | 93 |
| Figura 6.24 | O IoU exclusivo da estrutura mamária papila por ciclo, selecionando as imagens mais incertas via S_{papila} . | 94 |
| Figura 6.25 | IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$. | 95 |
| Figura 6.26 | O IoU exclusivo da estrutura mamária peitoral por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$. | 95 |
| Figura 6.27 | IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} . | 96 |
| Figura 6.28 | IoU exclusivo da estrutura mamária papila por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} . | 96 |
| Figura 6.29 | IoU exclusivo da estrutura mamária músculo peitoral por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} . | 97 |

| | | |
|-------------|--|-----|
| Figura 6.30 | Comparação do IoU da estrutura papila entre os experimentos AL-Papila e AL-Random. | 97 |
| Figura 6.31 | Comparação do IoU da estrutura músculo peitoral entre os experimentos AL-Pectoral e AL-Random. | 98 |
| Figura 6.32 | Comparação da IoU da estrutura papila entre os experimentos AL-Pec-Npl e AL-Random. | 98 |
| Figura 6.33 | Comparação da IoU da estrutura músculo peitoral entre os experimentos AL-Pec-Npl e AL-Random. | 99 |
| Figura 6.34 | IoU da estrutura peitoral ao selecionar imagens não informativas. | 100 |

Lista de tabelas

| | | |
|------------|--|----|
| Tabela 4.1 | Resultados e métricas dos processos de anotação, medidos em segundos. | 48 |
| Tabela 4.2 | Comparação entre medidas de dispersão entre amostragem e população da abordagem manual | 49 |
| Tabela 4.3 | Comparação entre medidas de dispersão entre amostragem e população da abordagem interativa | 50 |
| Tabela 5.1 | Comparação de medidas de tempo em ambos os experimentos entre todos os anotadores participantes. | 58 |

Lista de algoritmos

Lista de Abreviaturas

AL – *Active Learning*

IoU – *Intersection Over Union*

mIoU – *Mean Intersection Over Union*

ILS – *Image Labelling Software*

AID – *Annotation by Iterative Deep learning*

OGAMA – *Open Gaze And Mouse Analyzer*

TTA – *Test-time Augmentation*

DICOM – *Digital Imaging and Communications in Medicine*

Só sei que nada sei

Sócrates, .

1

Introdução

A visão computacional abrange décadas de esforços para que algoritmos permitam a melhoria da percepção de estímulos visuais em computadores de maneira considerável em aplicações do mundo real. Estes estímulos podem ser exemplificados em tarefas de baixo nível computacional, como identificação de bordas em imagens, até tarefas de alto nível como a compreensão de cenas completas. Um dos avanços que recebe destaque neste cenário consiste na atuação do aprendizado profundo (deep learning), um tipo de aprendizado de máquina ponta a ponta de funções complexas que se torna possível a partir do uso de dados. Tal técnica é bastante presente em aplicações médicas para auxiliar na solução de problemas pertinentes à análise de imagens médicas, diagnóstico com assistência de modelos computacionais e medicina em geral (ESTEVA et al., 2021).

No entanto, o algoritmo desta categoria, em geral, consiste em aprendizado supervisionado, que conta com dados anotados em grande escala para que o modelo de inteligência tenha acurácia suficientemente refinada. Estes dados, muitas das vezes, são anotados manualmente por conta da ausência de métodos que permitam a anotação mais eficiente.

Neste mesmo tema, o processo de anotação manual em imagens é uma tarefa repetitiva e exaustiva que conta com apoio de pessoas físicas para ser realizada (PAGARE; SHINDE, 2012). Este procedimento consiste em realizar anotações/segmentações em imagens onde, na maioria das vezes, a quantidade de material não anotado é grande e necessita de esforço manual para que ela seja feita.

Estas imagens anotadas servem de insumos para a criação de datasets, conjunto de dados anotados utilizados por modelos de aprendizado profundo de máquina, capazes de solucionar problemas de classificação supervisionada, detecção e segmentação em imagens. Recentemente, redes de aprendizado profundo têm obtido um progresso extraordinário como modelos de segmentação semântica na área médica, quando comparados ao uso outras técnicas de aprendizado de máquina (TAJBAKHSI et al., 2020; JANG; CHO, 2019; BUDD; ROBINSON; KAINZ, 2021).

O esforço humano no processo de anotação manual e sua melhoria têm sido amplamente estudado nos últimos anos (SAGER; JANIESCH; ZSCHECH, 2021). Pois, com o aumento de aplicações em análises de imagens médicas que fazem uso de aprendizado profundo, surgem problemas em ala-

vancar dados anotados para realizar o treinamento de modelos de segmentação semântica que auxiliam nestes sistemas (TAJBAKHSI et al., 2020).

Uma das fases mais caras e demoradas de qualquer projeto que envolva aprendizado de máquina supervisionada é a rotulagem dos dados. O processo de anotação manual se torna ainda mais difícil caso a forma para criar anotações seja utilizando máscaras de segmentação, pelo fato de exigir uma quantidade maior de interações para ser realizada. No entanto, este método possui uma precisão maior ao representar os pixels da região de interesse em relação a bounding box, por exemplo (RAPSON et al., 2018).

Estudos recentes têm pontuado sobre o uso de técnicas como active learning, que proporcionam seleção e filtragem dos dados mais informativos de serem anotados para o treinamento de modelos de aprendizado profundo (GAWLIKOWSKI et al., 2021). Procura-se entender como estes algoritmos podem estar tendo dificuldades a partir das inferências feitas sobre os dados, na qual esta informação pode ser aproveitada para direcionar melhor quais dados podem proporcionar um aprendizado mais refinado em aplicações de dados de alta dimensionalidade (GAL; ISLAM; GHAFRANI, 2017).

Para este trabalho, existe uma preocupação com o processo de anotação para gerar imagens etiquetadas para datasets de segmentação semântica. Isso porque aplicações da área fazem uso destes insumos para que haja o treinamento de modelos de segmentação semântica que auxiliam em análises médicas. São redes que fazem uso de imagens segmentadas por máscaras de segmentação para que seja possível o treinamento desses modelos. Com isto, explora-se o uso de active learning que possibilita a filtragem dos dados não anotados mais pertinentes a serem encaminhados para os anotadores peritos, sendo assim, será possível obter uma estimativa de quantos dados anotados são necessários para obter um dataset que proporcione um treinamento razoável ao modelo. E um mecanismo de anotação interativo que conta com apoio de modelo de segmentação semântica que envolve imagens de mamografia.

Observando os parágrafos anteriores, o trabalho consiste em propor mecanismos para melhoria nos processos de anotação manual e geração de datasets de segmentação semântica em imagens de mamografia computadorizada.

Ambos mecanismos mencionados acima fazem parte de um tema mais abrangente, conhecido como human-in-the-loop machine learning, o qual consiste em um paradigma onde a combinação do ser humano com inteligência computacional garantem o surgimento de estratégias que visam melhorias em aplicações ou sistemas que usufruem de inteligência artificial (MONARCH, 2021). A ideia seria a inserção desses mecanismos para a melhoria em diferentes etapas no treino de um modelo inteligente, focadas nas fases de anotação

dos dados e construção de datasets conforme a figura 1.1.

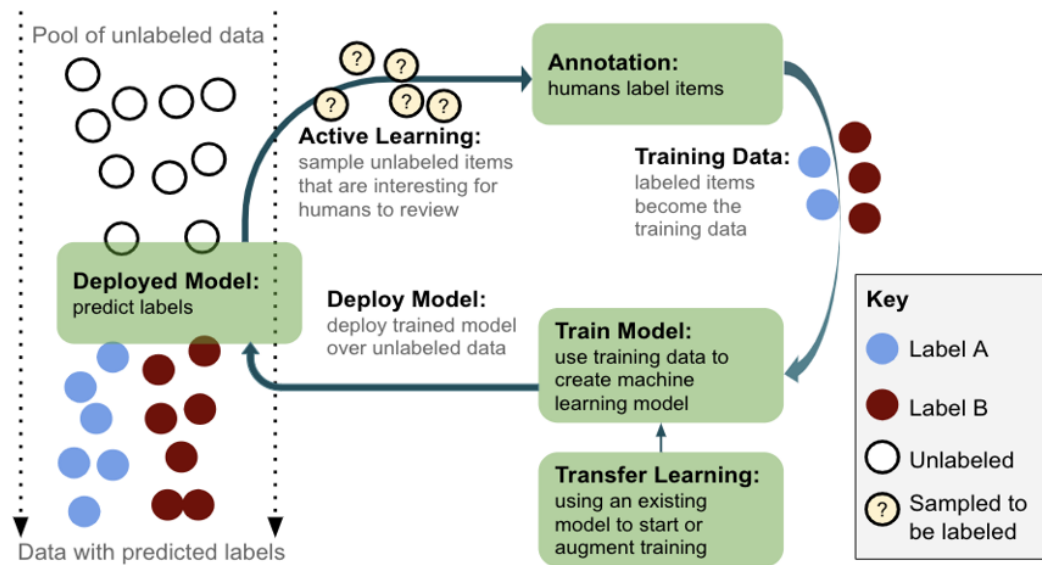


Figura 1.1: Mapa mental sobre o processo de human-in-the-loop, adaptado e retirado de (MONARCH, 2021)

A seguir, será elucidado sobre o problema e o caso de uso utilizado para validar o estudo que será desenvolvido para a melhoria dos processos. Diante disso, surge um problema pertinente ao processo de anotação manual e geração de datasets de imagens de mamografia, que será descrito na seção seguinte.

1.1 Problema

Nesta seção, contextualiza-se o problema, apresentando-o e porque ele é importante de ser resolvido.

Segmentar imagens é um processo manual e custoso, pois é necessário desenhar o contorno na região de interesse para destacar visualmente na imagem, manualmente. Na maioria das vezes, a atividade é realizada com o auxílio de ferramentas para anotar imagens gerando as máscaras de segmentação, ou seja, desenhar o contorno da região que interessa na imagem à mão. Porém, jornadas extensivas de anotação dessas imagens podem ocasionar na geração de dados ruidosos, cansaço por parte do anotador ou especialista responsável pela etiquetagem e diminuição da precisão dos modelos que utilizam estes dados gerados a partir dos pontos citados anteriormente.

É um problema importante que deve ser resolvido, devido ao tempo gasto para realizar estas anotações e o esforço físico e mental necessário para gerar imagens anotadas. O cansaço ocasionado pela extensa atividade e o curto prazo para etiquetar os exames podem apresentar riscos à qualidade das anotações

em imagens, prejudicando o aprendizado de um modelo que utiliza os dados gerados.

Além disto, um outro ponto consiste no processo de gerar datasets de segmentação semântica de imagens de mamografia computadorizada. A forma como os dados são selecionados para realizar as campanhas de anotação pelos anotadores peritos não possui uma ordem específica. Em curtas palavras, o arranjo de imagens a ser anotado pelos anotadores peritos é totalmente aleatório, não possuindo um conhecimento ou informação prévia sobre os conjuntos de dados não anotados que tragam algum norteamento de quais dados seriam mais pertinentes para o treinamento de modelos de segmentação de aprendizado profundo. Como mencionado anteriormente, para modelos atingirem uma acurácia boa, são necessárias muitas imagens anotadas. Logo, caso seja possível identificar as imagens que mais contribuiriam para o treinamento desses modelos, a quantidade de dados a serem anotados diminuiria, proporcionando uma redução nos custos e dados a serem anotados.

Por fim, a indicação das imagens de mamografia a serem etiquetadas pelos anotadores poderia contribuir para construção de datasets de segmentação semântica com dados seletos, traria benefícios na redução da quantidade de imagens a serem anotadas, treinamento de modelos com quantidades de dados reduzida e acurácia perto de um modelo completamente treinado ou, até mesmo, a superação deste limite. De certa forma, a abordagem contribui também para o processo de anotação, pelo fato de atenuar a quantidade de dados a serem anotados, beneficiando redução de custos de anotação.

Com base nos parágrafos anteriores, o problema a ser abordado neste trabalho consiste em otimizar os processos de anotação e geração de datasets de segmentação semântica que envolvem imagens de mamografia.

A seguir, para realizar estudos e aportes relacionados à pesquisa, será abordado um caso de uso pertinente ao processo de anotação e geração de datasets de segmentação semântica de imagens de mamografia.

1.1.1

Caso de Uso

O caso de uso consiste na segmentação de estruturas de interesse em exames de mamografia computadorizada. Para que haja a detecção das estruturas anatômicas, são treinados modelos de segmentação semântica que indicam as regiões prováveis de tais estruturas. No entanto, para habilitar os modelos de fazerem predições, são necessários grandes quantidades de imagens de mamografia que são manualmente anotadas de forma precisa, tal como ilustrado na figura 1.2.

No que tange a este caso de uso, são feitas segmentações por modelos de aprendizado profundo nas estruturas mamárias: papila, tecido fibroglandular, músculo peitoral e gordura mamária em incidências CC (Crânio Caudal) e MLO (Médio Lateral Oblíquo). Terminologias atreladas ao contexto de mamografia serão esclarecidas mais adiante no capítulo 2.

O objetivo neste caso de uso consiste em assessorar qualidade de exames de mamografia com as segmentações feitas por modelos inteligentes. Atua-se em assessorar casos em que a mama foi mal posicionada durante os exames ou se a mamografia atende a certos requisitos de qualidade. Este caso de uso não é focado no diagnóstico de exames de mamografia ou detecção de estruturas patológicas em imagens da categoria, e sim delimitado à assessoria de exames de mamografia via segmentação de estruturas anatômicas.

Para alavancar insumos para estes modelos, é necessário recorrer à atividade de anotação das estruturas sobre estes exames de mamografia. Servindo de insumos para o treinamento destes modelos.

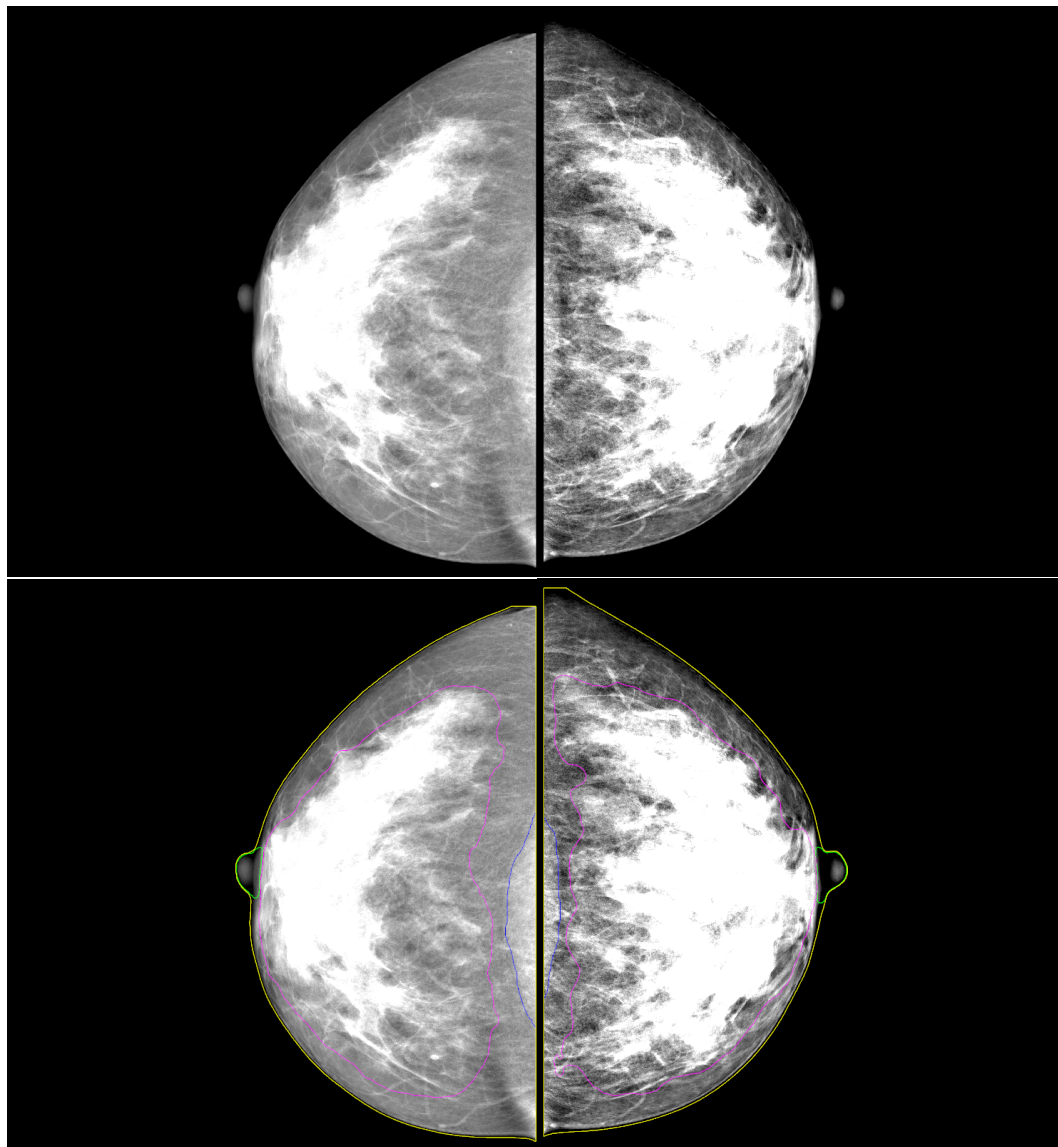


Figura 1.2: Exemplo de duas imagens de mamografia anotadas manualmente.

No entanto a atividade tem custos, como tempo e esforço, pois a anotação dos exames é feita manualmente, não possuindo meios que possam facilitar a anotação que deve ser registrada na imagem, além da ferramenta para delinear as estruturas.

Levantar insights sobre tempos de anotações destes exames poderiam contribuir para melhores noções de *deadlines* pertinentes ao tempo requerido para produção de datasets que permitam o treinamento de modelos de aprendizado de máquina para tarefas de segmentação de forma satisfatória.

Além das estruturas anatômicas da mamografia, um artefato de interesse para análise desse tipo de imagem é a detecção de dobras de peles. Uma dobra é uma sobreposição de pele, uma anomalia que pode atrapalhar a visualização do ângulo inframamário, região que é de interesse para análise médica em exames de incidência médio lateral oblíquo conforme a figura 1.3. A

anotação feita sobre esta anomalia é manual, não possuindo meios que possam tornar mais eficiente este processo de anotação. A ideia aqui é permitir que a atividade realizada possa ser melhorada e avaliada mediante métricas de tempo, relacionadas a custos de anotação.

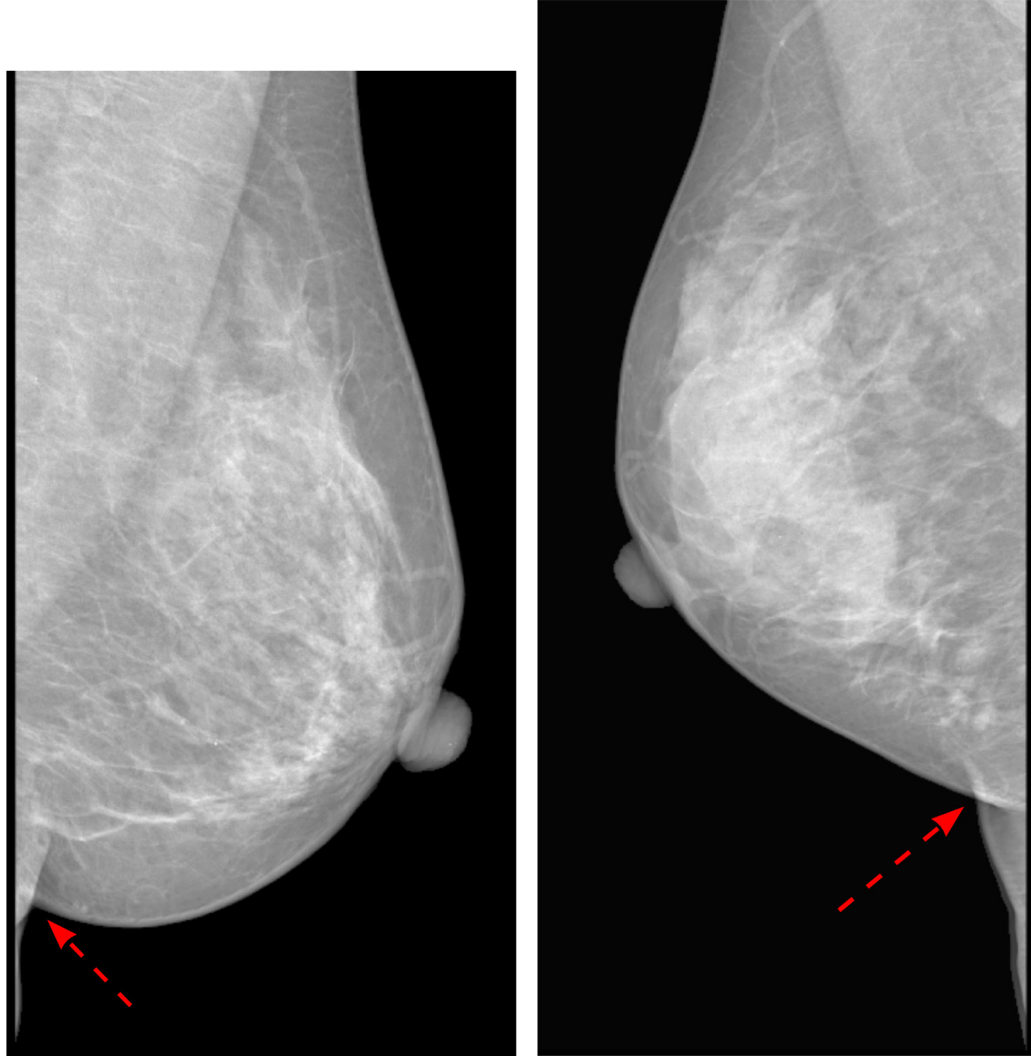


Figura 1.3: Exemplos de dobras de pele na incidência MLO.

Logo, os experimentos feitos para obter estimativas de tempo para dobras de pele poderiam ser estendidos para aquisição de mais resultados que poderiam melhorar contrastes pertinentes a noções de custo para anotações envolvendo outras estruturas mamárias.

Portanto, o procedimento de anotação manual continua sendo bastante trabalhoso, pois isso deve ser aplicado para quantidade de dados enormes que serão utilizados por aplicações de modelos de segmentação semântica de aprendizado profundo. São algoritmos que consomem estes insumos gerados para realizar tarefas de segmentação que darão apoio a análises feitas em imagens de mamografia.

Além do que foi dito, o processo de anotação, não possui uma ordem específica dos exames a serem etiquetados, pois, ao adquirir as imagens de mamografia que precisam ser anotadas, são geradas partições das mesmas em uma ordem randômica, não existindo nenhum critério a ser seguido para anotação ou alguma informação que indicasse as imagens que melhor poderiam contribuir para o treinamento de um modelo de segmentação semântica. A solução para este problema poderia contribuir para direcionar as imagens mais importantes de serem anotadas, com intuito de obter um dataset de mamografia mais compacto e com dados mais informativos para o treinamento de modelos que necessitam destes insumos. Desta forma, entende-se que custos seriam economizados perante a construção de um dataset com as características mencionadas anteriormente, poupando esforços de anotação, pelo fato de ser necessário etiquetar menos exames de mamografia.

Em resumo, a tarefa de anotar estes dados é custosa, pois a quantidade de dados não rotulados que precisam ser anotados são grandes, podendo levar a rodadas de anotações exaustivas (TAJBAKHSI et al., 2021). Além disso, o tempo gasto para realizar anotações em exames é elevado (SAKINIS et al., 2019) e, não deixando de mencionar, a fadiga ocasionada pela extensa jornada de etiquetar exames pode ocasionar na diminuição da atenção e precisão do anotador sujeitando-o ao erro.

1.2 Objetivo

O objetivo deste trabalho consiste em propor mecanismos para melhorar a eficiência nos processos de anotação e geração dos datasets de imagens de mamografia utilizados como insumos para treinar modelos de segmentação semântica de aprendizado profundo em mamografia.

Mecanismos que possam direcionar quais exames poderiam ser anotados com prioridade, com a finalidade de aprimorar a construção de datasets de mamografia com dados mais seletos e tamanho reduzido, proporcionariam uma diminuição nos custos de anotação.

1.3 Estrutura da dissertação

Este trabalho está dividido em: Capítulo 2, onde os conceitos básicos para o entendimento das técnicas e abordagens estão descritos.

Capítulo 3, no qual encontram-se os trabalhos relacionados aos temas de segmentação interativa e active learning.

No Capítulo 4, apresentam-se os experimentos e resultados pertinentes à melhoria do processo de anotação em dobras de pele na incidência MLO.

No Capítulo 5, são demonstrados os experimentos e resultados em relação à melhoria de anotações assistidas para imagens de mamografia em incidências CC e MLO.

O Capítulo 6 elucida aspectos da técnica aplicada ao estudo, experimentos e resultados sobre active learning para imagens de mamografia na visão CC.

Por fim, o Capítulo 7 apresenta algumas conclusões e frentes que possam servir como trabalhos futuros referentes a esta pesquisa.

2

Conceitos Básicos

Neste capítulo encontram-se conceitos importantes para o entendimento deste trabalho.

2.1

Mamografia

O exame de mamografia consiste na obtenção de imagens em alta resolução de mamas por meio da radiografia. Sua finalidade pode envolver diversos aspectos por meio de análise médica especializada como: inspeção de lesões em tecidos mamários, nódulos, assimetria entre estruturas da mama, dobras de pele, diagnóstico precoce de câncer de mama, entre outros.

A imagem gerada a partir do exame de mamografia possui diferentes incidências. Para o entendimento deste trabalho, será necessário conceituar duas dessas visualizações: Crânio Caudal (CC) e Médio Lateral Oblíquo (MLO).

A incidência CC permite a visualização das estruturas mamárias do paciente por uma visão aérea, como se o ponto de vista dessa visualização fosse os olhos do paciente olhando em direção as suas próprias mamas.

A incidência MLO permite a visualização das estruturas mamárias com uma visão em perfil, lateral em relação ao paciente (JESINGER, 2014).

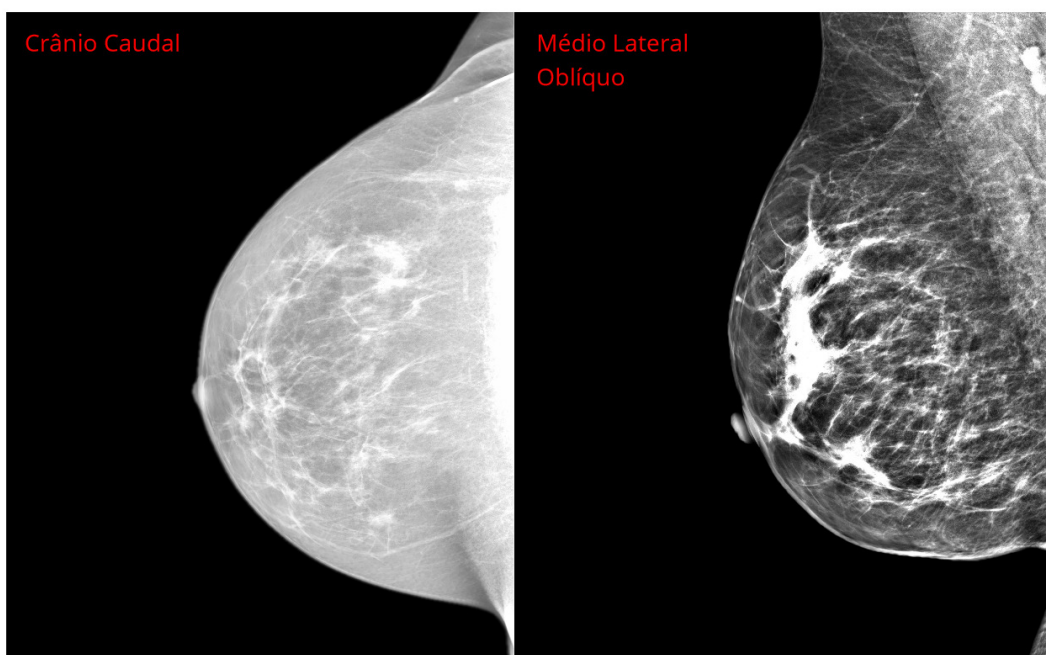


Figura 2.1: Figura à esquerda, incidência CC. Figura à direita, incidência MLO.

A figura 2.1 ilustra ambas visualizações de mamas com lateralidade à direita. A lateralidade, mama esquerda ou direita, pode ser deduzida pela direção inversa em que ela está apontando na imagem.

As mamas possuem diversas estruturas anatômicas, no entanto, este trabalho limita-se em citar as principais: Mamilo ou Papila, Tecido Fibroglandular, Músculo Peitoral e Tecido Mamário (JESINGER, 2014). Para facilitar a identificação das estruturas, a figura 2.2 elucida quanto ao seu formato e como elas se localizam em exames de mamografia em diferentes incidências.

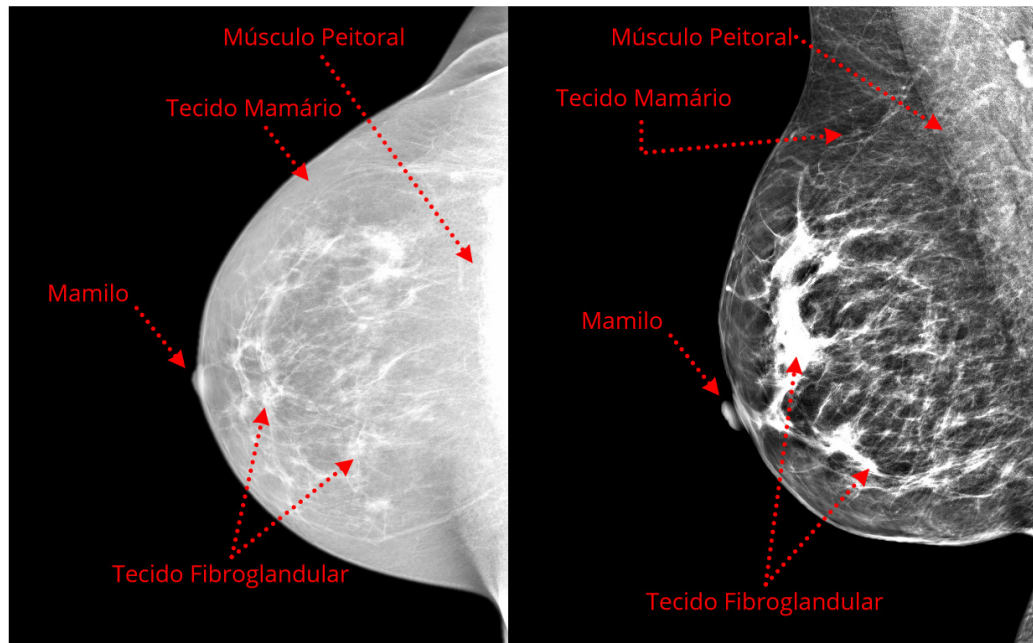


Figura 2.2: Estruturas anatômicas representadas em um exame de mamografia.

2.2

Segmentação Semântica

Segmentação é um processo no qual uma imagem é dividida em várias sub-regiões com base em informações específicas para selecionar uma região de interesse (MASOOD et al., 2015). Há diferentes formas para segmentar regiões na imagem, por exemplo, bounding boxes, polígonos, máscaras de segmentação, etc.

A tarefa de segmentação semântica de imagens, em visão computacional, consiste em classificar cada pixel de uma imagem correspondente a uma classe. Esta tarefa faz parte do conceito de compreensão da cena ou de explicar melhor o contexto global de uma imagem. No domínio da análise de imagens médicas, a segmentação de imagens pode ser usada para análise guiada por imagem, radioterapia ou diagnósticos radiológicos aprimorados (TAGHANAKI et al., 2021).

De uma maneira bem resumida, segmentação semântica é a técnica responsável por agrupar pixels de uma imagem que pertencem a uma mesma classe. Divide-se regiões da imagem que pertencem a uma classe específica, sendo visualmente representada como máscaras de segmentação. A região delimitada por pixels, ou seja, a máscara, pode ser representada por cores para se diferenciar de outras classes existentes na imagem conforme a figura 2.3 em problemas de segmentação multi classe.

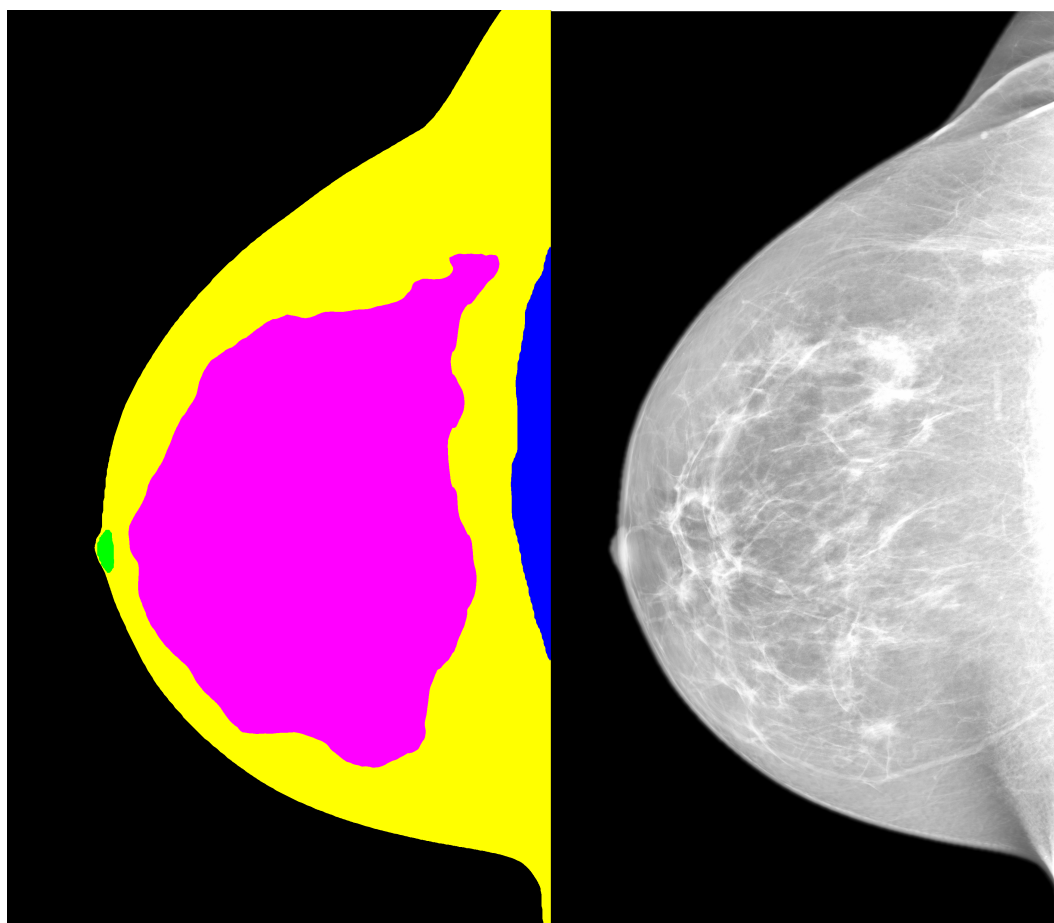


Figura 2.3: Exemplo de um exame de mamografia e sua respectiva máscara de segmentação indicando as estruturas anatômicas da mama em classes distintas.

Como resultado de uma segmentação semântica, temos uma representação visual acerca dos pixels que foram agrupados de acordo com a técnica, conforme a figura 2.3. A imagem à esquerda indica as regiões distintas segmentadas de estruturas anatômicas de um exame de mamografia. As cores preto, verde, roxo, azul e amarelo correspondem, respectivamente, ao fundo da imagem (background), o mamilo, tecido fibroglandular, músculo peitoral e tecido mamário como as classes a serem segmentadas pelo modelo de segmentação.

Trabalhos mais recentes da literatura apresentam como o estado da arte a utilização de modelos inteligentes de aprendizado profundo de segmenta-

ção semântica para solucionar problemas da computação pertinentes a visão computacional. (MO et al., 2022; GARCIA-GARCIA et al., 2017).

Sua aplicação em trabalhos recentes no domínio de imagens médicas também apresentam bons resultados ao utilizar modelos de segmentação semântica de aprendizado profundo para resolver problemas presentes no domínio de imagens médicas (TAJBAKSHI et al., 2020; JANG; CHO, 2019; BUDD; ROBINSON; KAINZ, 2021).

2.2.1

Métricas de avaliação de modelos de segmentação

Tomando como referência as pesquisas de Taghanaki e colegas (TAGHANAKI et al., 2021) e Varduhi e Irina (YEGHIAZARYAN; VOICULESCU, 2015), é possível avaliar a acurácia de modelos de segmentação semântica a partir da métrica IoU (Intersection over Union).

A métrica IoU consiste em quantificar uma pontuação, em escala de 0 até 1, o quão próximo a predição da rede está em relação ao ground truth. É estimada a partir da razão entre a intersecção e união dos pixels da máscara gerada pela rede com ground truth. Quanto mais próximo de 1, a segmentação feita pelo modelo é mais apurada.

2.3

Segmentação Interativa

Segmentação Interativa (interactive segmentation) tem como objetivo melhorar o processo de anotação, permitindo o usuário a corrigir e facilitar a marcação feita pelo modelo de segmentação interativamente, ou seja, as anotações (máscaras de segmentação) produzidas pelo modelo terão assistência de anotadores humanos que serão responsáveis pela manutenção da anotação gerada pelo algoritmo, flexibilizando o processo de etiquetagem com o auxílio de um modelo de segmentação e a inspeção de um especialista para corrigir amostras de máscaras fora de padrão (TAJBAKSHI et al., 2020; SAGER; JANIESCH; ZSCHECH, 2021).

A abordagem consiste no usuário anotar uma quantidade inicial de dados para treinar um modelo, em seguida, usar o modelo treinado para realizar as predições e, continuamente, consultar o anotador humano por correções sobre as anotações geradas pela rede (SAGER; JANIESCH; ZSCHECH, 2021).

2.4

Active Learning

Esta técnica consiste em selecionar os dados mais úteis a partir de um conjunto de dados não anotados e encaminhá-los a um anotador ou especialista para realizar a etiquetagem dessas amostras mais informativas para proporcionar uma melhoria na acurácia de modelos de aprendizado de máquina (REN et al., 2020). Segundo informações no livro de Monarch (MONARCH, 2021), a abordagem consiste em, inicialmente, treinar um modelo rudimentar que fará previsões sobre o conjunto de dados não anotados para identificar os dados em que o modelo apresentou mais "dúvida" em classificá-los. Para cada dado em que será feita a classificação, uma pontuação (score) é calculada de acordo com os valores de predição de cada observação, que quantificará a dúvida ou indecisão que a rede apresentou ao classificar a observação que será anotada em algum momento. Em seguida, será feita a separação dos dados mais informativos com base nos scores (na maioria das vezes, os scores de maior valor representam os dados mais duvidosos) e algum critério de seleção para adicioná-los ao dataset, expandindo-o com dados seletos que possam impulsionar sua acurácia de maneira mais eficiente ao retreinar o modelo. Por fim, este procedimento é repetido até o esgotamento do conjunto de dados que precisam ser anotados ou o atingimento de uma condição de parada do ciclo.

O objetivo de filtrar os dados não anotados mais informativos para o treinamento do modelo contribui para a acurácia do modelo com uma quantidade seleta e reduzida de dados para seu treinamento. Os dados mais proveitosos são direcionados a serem anotados para um treinamento de modelo mais eficiente. Para mais, a redução de dados a serem anotados diminui os custos de anotação por parte dos anotadores humanos visto que, ao invés de anotar todos dados disponíveis, o subconjunto seleta destes dados pode garantir a redução do esforço e tempo necessário para as campanhas de anotação. Logo, em comparação com um processo de anotação em que são feitas etiquetagens sem uma ordem específica, essa técnica tenta, com uma quantidade menor de dados anotados, atingir uma performance de classificação do modelo próximo ao esperado de um modelo completamente treinado, isto é, uma rede treinada com o conjunto de dados totalmente anotado caso disponível.

Em resumo, com menos dados anotados porém seletos, active learning permitirá que o modelo consiga obter uma performance próxima ou equivalente a de um modelo em que seu treinamento foi executado com todos os dados disponíveis, conforme ilustrado na figura 2.4. Pelo fato de anotar menos dados, contribui para a redução de custos de anotação.

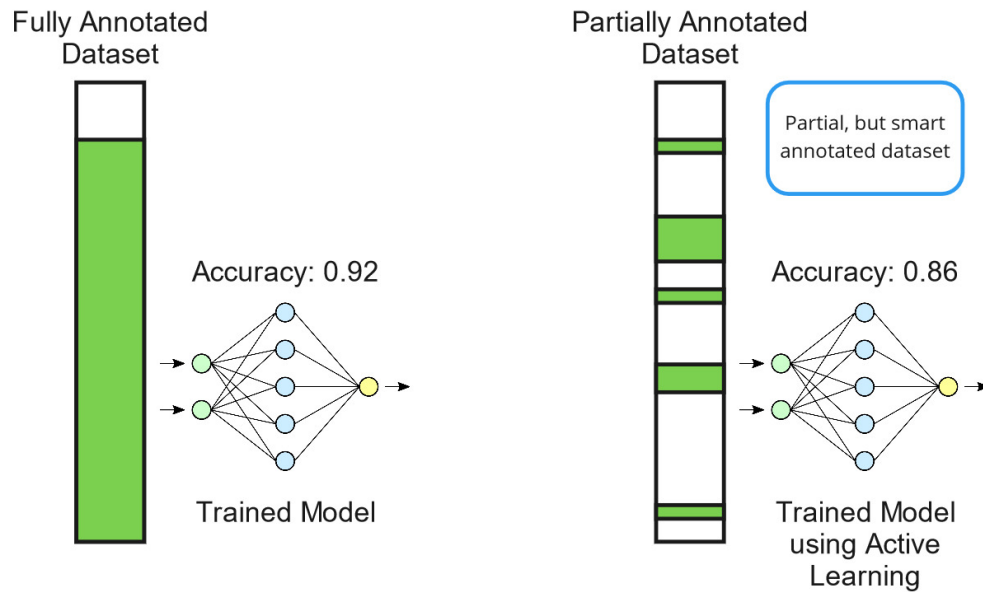


Figura 2.4: Um dos aportes principais de active learning.

Para sistemas de aprendizado supervisionado, alguns requisitos iniciais devem ser atendidos para utilização da técnica:

1. Aquisição de dados.
2. Rotular parte dos dados adquiridos e treinar um modelo inicial.

Depois de adquirir dados e anotar aleatoriamente uma parte deles (não o conjunto inteiro de dados), o próximo passo é treinar o modelo com esses insumos para preparar uma rede rudimentar para fazer algumas previsões. Os primeiros passos de active learning são importantes, pois a ausência de dados devidamente anotados, torna inviável treinar um modelo para fazer previsões para obter informações sobre as inferências feitas pela rede.

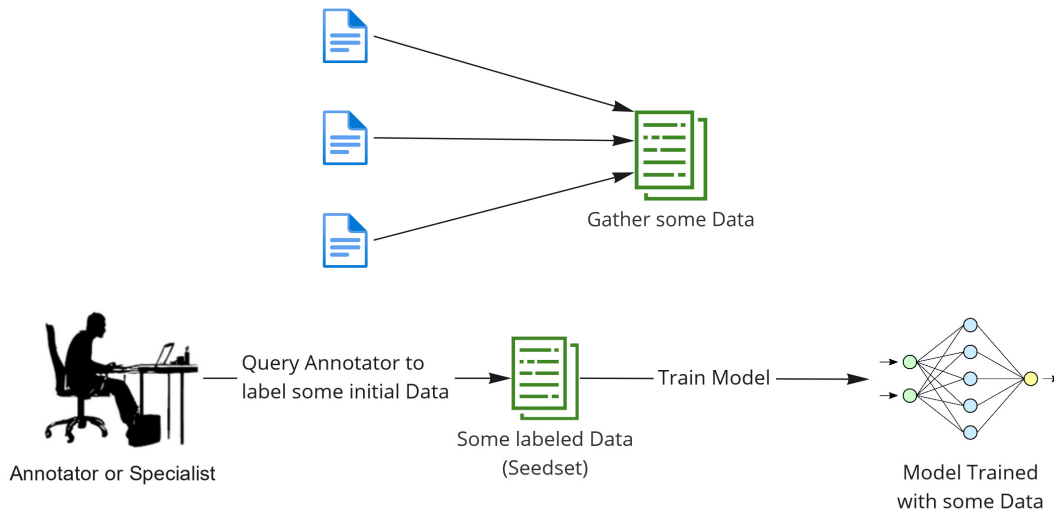


Figura 2.5: Etapas iniciais do processo de active learning.

A figura 2.5 ilustra duas imagens que representam em etapas, a obtenção dos requisitos iniciais de active learning.

Com os requisitos iniciais alcançados, a rede está pronta para realizar previsões para capturar informatividade com base em cada previsão realizada sobre o conjunto de dados não etiquetados, chamado de Pool.

A informatividade pode ser obtida por diferentes meios relacionados à uncertainty sampling, métodos que consistem em calcular a incerteza dos dados não anotados mais pertinentes para o treinamento do modelo com base nas amostras de maior incerteza obtidas pela previsão da rede. Esta filtragem é possível a partir de funções de aquisições, equações que quantificam a incerteza e ordenam para cada dado não anotado em função das previsões feitas pela rede (SCANDALEA et al., 2019). Sendo assim, é possível estimar mediante essas funções de aquisições, os dados mais pertinentes a serem anotados para obter um modelo que generaliza melhor com menos dados etiquetados.

Para quantizar incerteza sobre os dados existem alguns métodos conhecidos na abordagem clássica de active learning, como:

- Least confidence sampling.
- Margin of confidence sampling.
- Ratio of confidence.
- Entropy-based sampling.

Os métodos acima são utilizados como medidas que avaliam informações referentes à previsão do modelo para estimar uma pontuação de incerteza atrelada a uma amostra não anotada que pode ser relevante, dependendo da pontuação, para ser encaminhada ao anotador para etiquetar e adicionar ao

dataset de treinamento do modelo. Amostras com maior incerteza são dados com maior relevância de serem anotadas pois, a princípio, melhor contribuirão para o treinamento do modelo a cada ciclo de active learning.

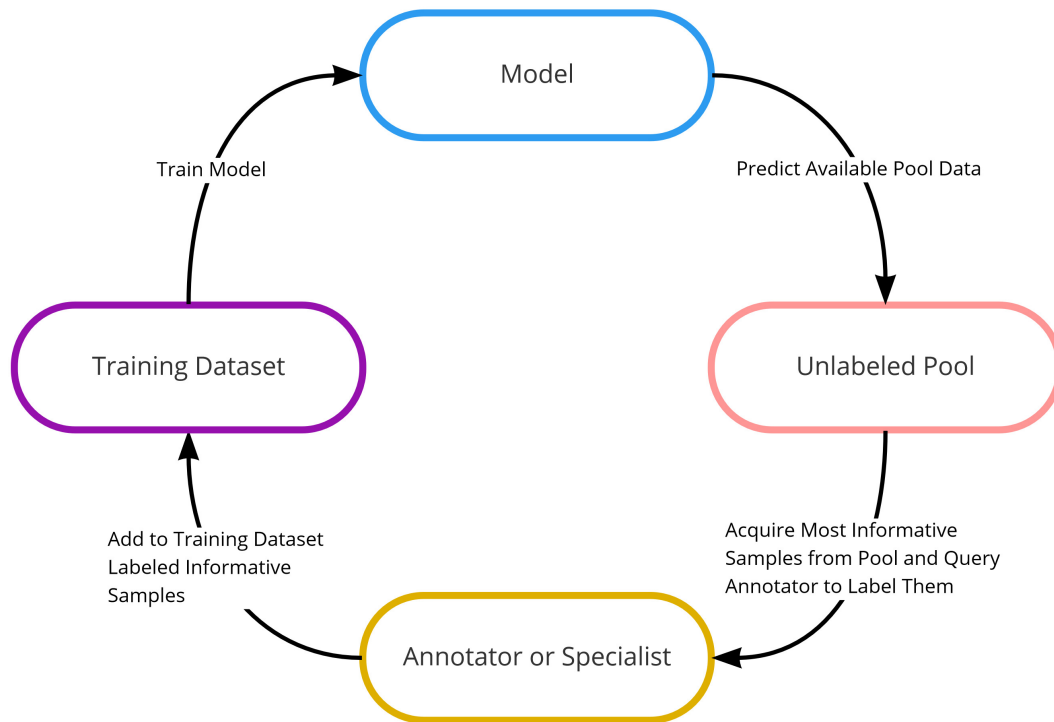


Figura 2.6: Fluxograma sobre o funcionamento da abordagem e suas etapas

A figura 2.6 mostra de maneira pictorial as etapas conceituadas acima sobre active learning e como o fluxo das informações pode ser representado. Ela se baseia em ciclos, onde cada finalização garante a melhoria de um modelo cada vez mais aperfeiçoado. A condição de parada da técnica possui diferentes possibilidades, como o modelo atingir um plateau em relação a suas métricas de acurácia, o esgotamento de todos dados contidos na pool, entre outros cenários. A condição varia de acordo com o objetivo e contexto de sua aplicação.

No entanto, active learning, ao ser aplicado no contexto de aprendizado profundo de redes de segmentação semântica, certas questões atreladas à técnica mudam por fatores vinculados a estes modelos. Essas questões serão elucidadas na próxima subseção.

2.4.1

Active learning em modelos de aprendizagem profunda de segmentação semântica

Os modelos de aprendizado profundo de segmentação semântica mais recentes retêm uma grande quantidade de parâmetros, possuem uma predição mais "sobreconfiante" e a classificação ocorre em nível de pixel (GAL; ISLAM;

GHAHRAMANI, 2017). Estimar medidas para filtrar os dados não anotados mais pertinentes a partir da abordagem clássica de active learning se torna uma tarefa mais complexa ao considerar os pontos acima. Estudos mais recentes, como (GAWLIKOWSKI et al., 2021), destacam a possibilidade de estimar a incerteza a partir de técnicas que focam em gerar diversas previsões a partir do modelo de aprendizado profundo. Tais técnicas possibilitam estender o active learning para o cenário atual de redes neurais profundas, na qual a estimação de incerteza pode ser calculada em função das diversas previsões geradas pelo modelo baseadas nas n -ésimas inferências feitas na rede.

Em resumo, aplicar a abordagem de active learning em modelos de aprendizado profundo requer técnicas de estimação de incerteza que produzam diferentes previsões a partir da rede para capturar ambiguidades nas inferências. A partir dessa variação entre os resultados do algoritmo pode-se estimar uma pontuação para filtrar os dados de maior incerteza.

2.4.1.1

Test-time Augmentation (TTA)

Para que seja possível gerar variações nas inferências feitas pela rede, test-time augmentation é uma das técnicas que permite gerar tais diferentes previsões para que seja possível estimar a incerteza (WANG et al., 2019). O método se baseia no princípio de data augmentation, uma outra técnica cuja aplicação consiste em gerar transformações em imagens durante o treino de uma rede neural para tentar atenuar overfitting (SHORTEN; KHOSHGOFTAAR, 2019). Logo, TTA explora o uso de data augmentation para gerar ruídos nos dados de entrada do modelo em tempo de predição para obter diversas inferências sob uma imagem.

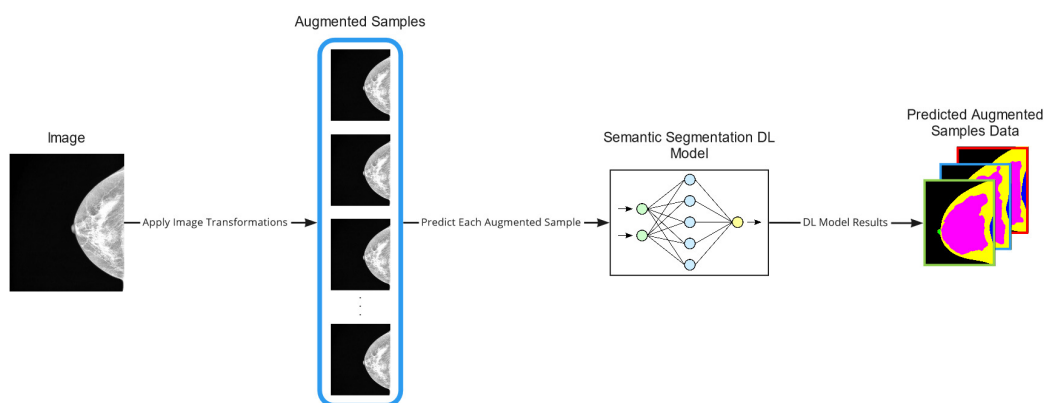


Figura 2.7: Esquematização da técnica, test-time augmentation.

A abordagem (figura 2.7) consiste em causar transformações ou perturbações na imagem que será feita a predição pela rede. Assim é possível adquirir

diferentes predições para cada perturbação aplicada na imagem, possibilitando, com base nessas diferentes inferências, a estimação de incerteza da predição que o modelo apresenta em função das ambiguidades nas segmentações geradas.

2.5

Medidas e Análise de Formas Geométricas

Nesta seção, consolidam-se breves conceitos atrelados à análise de figuras geométricas a partir de cálculos e métricas para quantificar alguns aspectos da geometria computacional. Estes princípios serão importantes para a solidificação das medidas que estimam os scores de incertezas propostos no capítulo 6.

Todas as medidas abaixo são conceituadas na dimensionalidade 2D.

2.5.1

Fecho convexo

O fecho convexo pode ser entendido como o menor polígono S que pode ser formado a partir de um conjunto de pontos $P = \{p_1, p_2, \dots, p_n\}$ (GRAHAM; YAO, 1983).

2.5.2

Distância de Hausdorff

Considere a distância euclidiana D entre dois pontos x e y que pode ser obtida via:

$$D(x, y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2-1)$$

A distância de Hausdorff, de uma maneira bem resumida, consiste na maior a distância D entre um ponto $x \in P_x$ e o ponto correspondente $y \in P_y$ entre dois os conjuntos de pontos P_x e P_y . Ou seja, a maior reta possível entre os pontos x e y (TAHA; HANBURY, 2015). Logo, a distância de Hausdorff pode ser definida como:

$$D_h(P_x, P_y) = \max\{D(x, y), D(y, x)\} \quad (2-2)$$

Por fim, ela pode ser utilizada como uma medida de maior distância entre dois pontos correspondentes $D_h(x, y)$ entre um polígono S e um fecho convexo F_c , onde cada ponto p pertence a um conjunto $P_i = \{p_1, p_2, \dots, p_i\}$ para cada estrutura acima.

2.5.3

Solidez

A solidez é uma medida que consiste em avaliar o quanto um polígono S se assemelha a um fecho convexo F_c , pela razão entre a área de S com F_c (RITI et al., 2016), representado como:

$$S_{ol} = \frac{A(S)}{A(F_c)} \quad (2-3)$$

A medida, em uma escala de 0 a 1 por exemplo, indicaria um polígono com irregularidades, caso a escala seja próximo a 0. E um polígono com superfície convexa, aproximaria a medida a 1.

2.5.4

Excentricidade

Excentricidade consiste em medir a "elipticidade" de um determinado polígono, indicando o quão próximo se assemelha à figura de uma elipse a partir do comprimento de duas retas perpendiculares $D(E_a)$ e $D(E_d)$, nomeadas, respectivamente, eixo menor e eixo maior (ADHIKARY; KUMA, 2011). Utilizando a distância euclidiana 2-1, podemos obter a medida da seguinte forma:

$$E_{cc} = \frac{D(E_d)}{D(E_a)} \quad (2-4)$$

Onde E_{cc} é calculado pela razão do comprimento da distância focal $D(E_d)$ com o comprimento do eixo maior $D(E_a)$ da elipse.

Essa medida pode ser usada para estimar a proporção de um polígono, indicando se o objeto é pequeno ou largo, se assemelhando à figura de uma elipse. Caso ambos os eixos possuam largura similar, a medida tenderia à 0. Caso contrário, considerando que os eixos possuam larguras bem diferentes, a medida se aproximaria à 1, indicando alta elipticidade.

2.5.5

Número de Euler

O número de Euler N_e pode ser usado como uma medida para análise topológica de formas geométricas. Consiste em um valor inteiro, indicando a diferença entre o número de componentes conexas C e a quantidade de buracos H (ADHIKARY; KUMA, 2011).

$$N_e = C - H \quad (2-5)$$

A medida se mostra interessante, pois a mesma é agnóstica quanto a superfície e invariante a aspectos de rotação, translação e escala do polígono

examinado.

No capítulo seguinte serão evidenciados trabalhos pertinentes ao tema da pesquisa, aberturas encontradas na literatura e artigos que darão sustentação para a solução do problema de pesquisa.

3

Trabalhos Relacionados

A seguir serão mostrados trabalhos mais recentes da literatura pertinentes aos temas e algumas oportunidades encontradas.

3.1

Segmentação Interativa

O review de Nima Tajbakhsh e colegas (TAJBAKHSH et al., 2020) cita trabalhos pertinentes à segmentação de imagens médicas com aprendizado profundo. Um dos tópicos do artigo mostra soluções com a ideia de gerar anotações em imagens médicas visando o custo benefício, melhorando a eficiência e eficácia na produção de dados anotados.

Sager et al. (SAGER; JANIESCH; ZSCHECH, 2021) demonstram um survey com guias esquematizados, revelando aspectos, orientações relacionadas na criação de ILS (Image Labelling Software) e abordagens de segmentação que servem de conhecimento para melhorias nos mecanismos de anotação.

Sambaturu e colegas (SAMBATURU et al., 2021) desenvolvem um framework para corrigir predições feitas a partir de um modelo de segmentação semântica em imagens médicas 2D e 3D com auxílio de um perito médico, baseando-se em princípios da segmentação interativa. Este trabalho, em particular, permite a múltipla correção de segmentações feitas pelo modelo; contribui com um estudo quantificando o tempo e número de interações feitas pelo usuário durante etapas de segmentação manual e segmentação interativa com auxílio da rede em vários datasets mencionados.

O trabalho de Philbrick e colegas (PHILBRICK et al., 2019) consiste na criação de uma ferramenta de anotação de imagens médicas por meio de um processo batizado de AID (Annotation by Iterative Deep learning), que acelera a tarefa de etiquetar imagens utilizando um modelo de segmentação semântica de aprendizado profundo. O método se define em criar anotações utilizando métodos tradicionais de segmentação a partir de uma base de dados não rotulados, para gerar um subconjunto de dados anotados que será encaminhado para o treinamento do modelo, que fará as próximas anotações sendo supervisionada pelo anotador. Esse processo é repetido iterativamente até que todo o conjunto de dados seja anotado ou até que o modelo atinja precisão suficiente para que não seja mais necessária uma iteração adicional.

Sun e coautores (SUN et al., 2018) propõem um método de segmentação interativo para segmentar limites com ruídos, em que o usuário insere um

ponto aproximadamente no centro do objeto e, em seguida, o modelo realiza o delineamento do objeto para a estrutura especificada pelo usuário.

Wang et al. (WANG et al., 2018) criaram um framework que consiste no uso de duas redes neurais, onde a primeira gera uma máscara de segmentação base, enquanto a segunda refina a máscara base de acordo com as sugestões fornecidas pelo anotador.

O trabalho de Sakinis et al. (SAKINIS et al., 2019) demonstra um método de segmentação de imagem semiautomático que permite uma segmentação de alta qualidade com cliques de mouse do usuário, indicando para o modelo ressegmentar a imagem com base no pixel onde foi indicado pelo usuário, possibilitando uma segmentação mais direcionada à região de interesse.

Segundo o trabalho de Yang e colaboradores (YANG et al., 2017), segmentação deep learning e active learning trazem contribuições para redução de esforço em tarefas de anotação manual em imagens biomédicas. Eles obtêm performance no estado da arte em relação ao modelo de segmentação utilizando 50% de dados disponíveis para anotação e uma abordagem de sugestão de anotações que direciona as regiões da imagem mais importantes de serem anotadas.

Entende-se que incorporar alguns conceitos de segmentação interativa, como utilizar um modelo de segmentação treinado para auxiliar nas anotações e permitir interatividade do usuário para corrigir anotações da rede, pode ser vantajoso se aplicados como mecanismos de anotação semi-automática de imagens, pelo fato do processo de anotação ser manual, abrindo oportunidades de automatizar o processo e reduzir custos de anotação, podendo contribuir na geração de anotações de uma maneira mais eficiente.

Apesar de ser mencionada outra abordagem em (SAGER; JANIESCH; ZSCHECH, 2021), como processo de anotação completamente automatizado, a mesma não permite a interação do anotador ou especialista para que ele possa realizar alguma correção na segmentação, caso necessário. Isto pode causar problemas em situações em que o modelo produza segmentações incorretas, não havendo espaço para fazer a correção na anotação na imagem.

Uma outra questão apresentada na literatura que apresenta benefícios ao tema se chama Active Learning, na qual o próximo lote de amostras para anotação é determinado pelo algoritmo, com intuito de selecionar os dados que mais contribuirão para o treinamento do modelo (TAJBAKHSI et al., 2020; BUDD; ROBINSON; KAINZ, 2021). A seguir, trabalhos relacionados a esta técnica serão abordados.

3.2

Active learning em aprendizagem profunda

O trabalho de Scandalea et al (SCANDALEA et al., 2019) implementa um framework de Deep Active Learning que se baseia em estimar a incerteza do modelo sobre imagens biomédicas não anotadas. A obtenção da medida é possível a partir do uso de modelos de aprendizado profundo modificados com a inserção de camadas de Dropout na estrutura da rede. Sendo assim, a predição do modelo varia a cada execução, pois seus parâmetros internos são modificados devido a inserção das camadas mencionadas anteriormente, que tem como função desligar algumas ligações entre neurônios da rede, fazendo com que toda vez que o modelo realize uma predição, seu resultado será não determinístico. Aproveitando-se dessa técnica, é possível gerar várias predições para explorar e quantizar a incerteza do modelo em relação às diferentes predições geradas pela rede, comumente nomeada na literatura como amostras Monte Carlo. Estes diferentes resultados servem para estimar a incerteza de um modelo de aprendizado profundo a partir das amostras obtidas, que serve de insumo para gerar mapas de incertezas sobre as ambiguidades entre esses diferentes resultados da rede. Isso permite calcular uma pontuação de incerteza, com base nos mapas obtidos, para poder identificar as amostras não anotadas que possivelmente o modelo teve maior dificuldade para realizar as predições. Em resumo, o trabalho consiste em explorar uma medida de incerteza do modelo em relação às inferências, aplicando transformações euclidianas nos mapas de incerteza para regularizar o tamanho dos axônios para calcular uma pontuação que permita ser utilizada para filtrar as imagens que o modelo teve mais dúvida ao segmentar. O mecanismo que se resume em priorizar a anotação de dados com pontuação de incerteza mais altas melhora o treinamento do modelo experimental, alcançando melhoria nas medidas de acurácia de segmentação com menos dados, em contraste com anotações feitas sem uma ordem específica.

A dissertação de Xiong (XIONG, 2019), de maneira similar, consiste em capturar a incerteza do modelo e construir um pipeline de Active Learning para a melhoria de performance do modelo de aprendizado profundo de segmentação semântica, investigando o funcionamento dos métodos de Deep Ensembles e modelos bayesianos para consolidar funções de aquisições responsáveis por filtrar os dados mais informativos para a melhoria do treinamento da rede, por fim, otimizando a performance do modelo. É de suma importância observar que ambos os métodos investigados sobressaem em relação ao selecionamento aleatório de amostras não etiquetadas para o treinamento do modelo. O autor consegue afirmar, mediante os resultados, que as técnicas experimentadas

conseguem filtrar melhor que um processo de anotação que não possui uma ordem específica para anotar dados.

Um marco para active learning em modelos de aprendizado profundo consiste no trabalho de Gal et al (GAL; ISLAM; GHAHRAMANI, 2017), pois os autores explicam os contrastes com uso da abordagem clássica de active learning em contextos de aprendizado profundo. O estudo tenta informar as dificuldades encontradas ao se aplicar a técnica com modelos que lidam com dados de alta dimensionalidade, como as redes neurais convolucionais, nas quais a representação de incerteza do modelo não é facilmente computada da maneira como é abordada na visão clássica de active learning. Então, os autores recorrem ao uso de modelos bayesianos, como também mencionados nos trabalhos de Xiong (XIONG, 2019) e Scandalea et al (SCANDALEA et al., 2019), para quantizar incertezas e consolidá-las para identificar quais dados a serem anotados seriam mais informativos para o treinamento de modelos.

O survey de Gawlikowski et al (GAWLIKOWSKI et al., 2021) revela aspectos relacionados à estimação de incerteza em modelos de aprendizado profundo. Ele demonstra trabalhos mais recentes deste tópico e as técnicas utilizadas para que seja possível estimar a incerteza do modelo para diferentes propósitos. Com isto, verificou-se que métricas podem ser calculadas a partir da estimação de incerteza para Active Learning, podendo nortear as amostras mais informativas para o treinamento de um modelo. Métricas como a entropia são bastante exploradas nos trabalhos mencionados neste survey, capazes de mensurar a incerteza em função da entropia de várias predições sobre a entrada.

O artigo de Wang et al (WANG et al., 2019) explora o uso da técnica Data Augmentation para gerar diferentes predições a partir de uma entrada, aplicada no domínio de reconhecimento de imagem com aprendizado profundo de imagens médicas nas categorias de ressonância magnética 3D e tumores cerebrais em exames 2D. O método proposto pelo autor consiste em gerar perturbações leves a partir de ruídos e transformações na imagem, para gerar diferentes predições em função da entrada de um modelo de segmentação para estimar incerteza e avaliar quais métodos em pauta descreveram a incerteza de maneira mais fidedigna. Em conclusão, entende-se que TTA obteve o melhor desempenho na estimação da incerteza de predição e descreve a equação para a obtenção da mesma baseada no cálculo de valores de entropia por pixel para problemas de segmentação semântica.

O trabalho de Roy et al (ROY et al., 2019) tenta encontrar uma correlação, a partir de três métodos propostos, a acurácia de segmentação com estimação de incerteza por estrutura anatômica cerebral. Um dos métodos tenta mapear incerteza por estrutura via visualização para que possa ser

utilizado para dois propósitos: avaliar visualmente a partir de mapas de calor os pixels onde o modelo possui menos confiança de segmentação e quantizar a incerteza a partir dos mapas de calor com ground truths, para que possam ser utilizadas para medir a incerteza na predição para cada estrutura cerebral. Observa-se um ponto interessante, de que métricas baseadas na estimação de incerteza também podem ser direcionadas para obter incertezas em nível de classe. Em outras palavras, obter uma pontuação que possa filtrar os dados não anotados mais pertinentes a serem anotados priorizando uma estrutura específica, como uma tentativa de alavancar a precisão da segmentação para aquela região anatômica específica.

Em relação aos trabalhos citados nesta seção, nota-se o uso medidas de incertezas baseadas na entropia (SHANNON, 1948). Essa grandeza vem sendo observada e utilizada desde abordagens clássicas de active learning, mostrando-se prevalente até em períodos mais recentes como métrica para estimação de incerteza em aprendizado profundo. Por fim, métricas baseadas em entropia para estimar incerteza apresentam resultados bons em trabalhos como (WANG et al., 2019; GAL; ISLAM; GHAHRAMANI, 2017) e segundo a revisão de (GAWLIKOWSKI et al., 2021) são mais consolidadas em tarefas de segmentação e classificação. Em linhas gerais, o uso de técnicas para estimação de incerteza de modelos de aprendizagem profunda em active learning, podem ser confirmados com resultados satisfatórios mediante os trabalhos dos autores (GAL; ISLAM; GHAHRAMANI, 2017; SCANDALEA et al., 2019; XIONG, 2019)

3.3

Oportunidades na literatura

3.3.1

Segmentação Interativa

Apesar das contribuições dos trabalhos mencionados (SUN et al., 2018; WANG et al., 2018; SAKINIS et al., 2019; SAMBATURU et al., 2021; PHILBRICK et al., 2019), eles carecem de medição de tempo de anotação, não possuindo detalhes sobre os ganhos pertinentes ao tempo decorrido para realizar a tarefa de segmentação com abordagem interativa. Isto abre margem para a pesquisa quantificar o tempo decorrido para realizar as tarefas de segmentação, revelando estatísticas e ganhos relacionados ao tempo de anotação manual e com uso do modelo como recurso assistente durante as campanhas de anotação.

Entende-se que a partir dos trabalhos de (SAMBATURU et al., 2021;

SAKINIS et al., 2019), a quantidade de interações e o tempo de anotação consumido para realizar as etiquetas dos dados têm relação com custos de anotação, em que a redução dessas grandezas proporcionaria melhorias no sentido de tornar o processo de anotação mais eficiente.

O trabalho de Sun et al. (SUN et al., 2018) só permite que os usuários especifiquem o ponto no início da segmentação, ou seja, as máscaras de segmentação resultantes não respondem à interação do anotador subsequentemente. Não é possível realizar ajustes no contorno da máscara resultante. A técnica usada por Sakinis et al. (SAKINIS et al., 2019) compartilha esta mesma questão, no entanto, a correção é feita pelo próprio modelo, indicando em tempo real através de cliques de mouse em uma interface de usuário, como o modelo deve ajustar a segmentação. Logo, os mecanismos propostos pelos autores não possibilita o delineamento manual para corrigir a máscara de segmentação com precisão humana. A correção manual é um fator importante para produzir dados etiquetados com maior controle.

3.3.2

Active learning em aprendizado profundo

Observa-se, a partir dos trabalhos citados (SCANDALEA et al., 2019; XI-ONG, 2019; GAL; ISLAM; GHAHRAMANI, 2017) e o survey de Gawlikowski (GAWLIKOWSKI et al., 2021), que a maior parte dos trabalhos relacionados a active learning para segmentação semântica em aprendizado profundo usam técnicas de estimação de incerteza por modelos bayesianos e outros meios. Percebe-se que existe uma escassez do uso da técnica de TTA para estimar pontuações de incertezas em Active Learning. Abre-se, portanto, a possibilidade de avaliar quais aportes poderiam ser revelados com uso desta técnica com base nos resultados da pesquisa.

Afirma-se, baseado na revisão de Gawlikowski e colegas (GAWLIKOWSKI et al., 2021), que TTA é uma técnica vantajosa de se utilizar para captar incertezas, pelo fato de não necessitar de modificações na arquitetura do modelo de segmentação e de ser computacionalmente mais econômica em relação aos outros métodos citados na revisão. Outro ponto importante reafirmado pelos autores consiste em advertir que, a partir de uma única predição de um modelo de aprendizado profundo, não é possível estimar a incerteza de predição. Logo, estima-se que o método de TTA possa auxiliar nesta questão gerando predições com sutis diferenças nas segmentações e calculando a incerteza em função dessas predições.

O trabalho de (ROY et al., 2019) tenta explorar o cálculo de incerteza no nível de estrutura. No entanto, apesar das contribuições do autor, a técnica

é aplicada em imagens de ressonância magnética cerebral, o que pode não se adaptar bem em imagens de mamografia, pelo fato das imagens ocuparem proporções diferentes de pixels. As imagens de ressonância utilizadas no trabalho do autor indicam que elas não variam muito em seu tamanho, em contraste com mamografia, que têm seu tamanho atrelado à anatomia da paciente, podendo variar de pessoa para pessoa. Muitas dessas métricas não levam em consideração o tamanho da estrutura para estimar uma pontuação de incerteza a ser utilizada em active learning. A possibilidade de usar medidas que quantificam a incerteza de predição do modelo que sejam agnósticas ao tamanho da mama pode abrir oportunidades para a filtragem dos dados não anotados com interesse em priorizar anotações de uma estrutura anatômica em específico.

Uma informação importante do trabalho de Scandalea et al (SCANDALEA et al., 2019) pode ser observada: o objetivo final do estudo consiste em limitar esforços de anotação com active learning. Os autores concluem que anotar o dataset de estudo inteiro levaria cerca de 120 horas, desconsiderando certos fatores. Logo, os resultados demonstram que utilizar a abordagem selecionando amostras altamente informativas para serem encaminhadas aos anotadores, diminuiria a quantidade de imagens necessárias para serem anotadas. Assim, entende-se que anotar menos imagens, reduziria o trabalho necessário durante as etapas de anotação.

No capítulo seguinte, pretende-se apresentar o mecanismo baseado em segmentação interativa para apoiar no processo de anotação em dobras de pele. Além de estimar, mediante experimentos, os tempos de anotação deste artefato em imagens de mamografia na visão MLO com o recurso de anotações assistidas com modelo de segmentação semântica.

4

Anotações assistidas para dobras de pele MLO

Este capítulo descreve o mecanismo que melhora o processo de anotação em dobras de pele em imagens de incidência MLO. Apresentam-se as melhorias implementadas e os resultados obtidos pelos experimentos.

A técnica implementada neste capítulo baseia-se em conceitos da segmentação interativa, como um instrumento de abordagem semi-automática para assistência de anotações de dobras de pele, em que a presença de um especialista ou anotador perito é necessária para a inspeção das segmentações geradas pela rede quando consultadas, para que haja o controle de qualidade sobre anotações produzidas pelo algoritmo.

Neste capítulo, o trabalho consiste em avaliar o tempo de anotação em processos de anotação manual e interativo, o qual, neste capítulo se limita no contexto de dobras de pele em imagens de mamografia. Para mais, foi necessário treinar um modelo rudimentar que imite as segmentações feitas pelo anotador, para auxiliar no delineamento dessas estruturas em exames MLO. Os dados etiquetados foram adquiridos durante o experimento manual que avalia o tempo de anotação obtido ao anotar dobras de pele sem mecanismo algum, ou seja, manualmente.

Em seguida, o modelo de segmentação semântica de dobras de pele foi treinado com os dados adquiridos do processo anterior, possibilitando a incorporação do modelo à ferramenta de anotação e permitindo que ele possa ser consultado para gerar a segmentação das dobras identificadas nos exames. Logo, avalia-se mediante um segundo experimento, o tempo de anotação com auxílio do modelo obtido anteriormente, sendo consultado pelo anotador humano mediante confirmação visual da estrutura.

Por fim, são avaliados os tempos de anotação em comparação com ambos os experimentos e apresentados esclarecimentos sobre os resultados obtidos.

4.1

Experimento Manual

O experimento consiste em delinear de dobras de pele manualmente, identificadas nas imagens de mamografia, na qual, servirão para treinar o modelo de segmentação semântica para auxiliar no processo de etiquetagem.

Como os exames possuem duas incidências, MLO e CC, foram escolhidas apenas as visualizações Médio Laterais Oblíquo, pelo fato das dobras estarem presentes nesta incidência. Em seguida, cuidadosamente, foram manualmente

segmentadas as dobras de pele. Exemplos em que foi identificado a não presença de dobras também foram armazenadas para indicar ao modelo situações onde não há dobras.

Logo, para anotar as dobras de pele na região do ângulo inframamário utiliza-se uma ferramenta de anotação. Esta ferramenta foi desenvolvida internamente pelo instituto para facilitar anotações em imagens em projetos de aprendizado de máquina.

Para cada imagem de incidência MLO, é contabilizado o tempo de anotação para delinear o artefato. Para cronometrar o tempo de cada anotação, um mecanismo foi implementado na ferramenta para controlar o início e fim da contagem de tempo para cada anotação de dobra de pele. Este recurso é importante para delimitar apenas o tempo decorrido durante uma anotação de estrutura mamária. Logo, a obtenção do tempo de anotação para cada imagem, consiste nos seguintes passos:

1. Visualizar a imagem de mamografia MLO.
2. Criar etiqueta de dobra de pele.
3. Acionar o cronômetro para iniciar a contagem de tempo.
4. Identificar visualmente a dobra de pele.
5. Delinear a dobra de pele, manualmente.
6. Acionar o cronômetro novamente, ao terminar a anotação do artefato.
7. Visualizar a próxima imagem MLO e repetir o passo 1.

Caso a imagem visualmente não possua dobra de pele, apenas a sua etiqueta é criada, para indicar que não há dobras e a próxima imagem de incidência MLO é analisada.

Após terminar as anotações manuais, 1372 imagens na incidência MLO foram anotadas utilizando a ferramenta para criar as máscaras de segmentação. Nesta etapa de segmentação manual foi possível contabilizar o tempo de anotação total de 26,961.715s (aproximadamente 7 hrs 29 min 21 s).

4.2

Experimento Interativo

Neste experimento foi utilizado o modelo de segmentação treinado com dados obtidos pelo processo de anotação manual, atuando como um mecanismo que irá auxiliar a etiquetagem das dobras de pele nesta etapa.

Este experimento mantém todas as regras descritas no experimento manual, com alguns ajustes nos passos que devem ser seguidos para obter o tempo de anotação com auxílio do modelo de segmentação. Logo, os passos consistem em:

1. Visualizar a imagem de mamografia MLO.
2. Criar etiqueta de dobra de pele.
3. Acionar o cronômetro para iniciar a contagem de tempo.
4. Identificar visualmente a dobra de pele.
5. Logo em seguida, acionar o modelo de segmentação para delinear a dobra de pele.
6. Caso o delineamento feito pela rede não estiver satisfatório, corrigir a segmentação, manualmente.
7. Acionar o cronômetro novamente, ao terminar a anotação do artefato.
8. Visualizar a próxima imagem MLO e repetir o passo 1.

O acionamento do modelo de segmentação é possível pressionando um mecanismo na interface da ferramenta localizada adjacente ao cronômetro. Ao confirmar visualmente a dobra de pele neste experimento, o anotador humano pressiona este mecanismo para que o modelo faça a segmentação do artefato, e caso o delineamento não estiver satisfatório, o anotador interativamente aperfeiçoa os contornos da segmentação, manualmente. Reforça-se que o acionamento do modelo para sugerir a segmentação deste artefato possui um tempo de processamento que está sendo levado em consideração na etapa 4 nos passos a serem seguidos durante o experimento interativo.

No segundo experimento foram anotadas as dobras de pele em exames MLO com o apoio do modelo de segmentação semântica treinado com os dados do experimento anterior. Totalizou-se 508 imagens anotadas, tempo de anotação total de 4,581.703s (aproximadamente 1 hr 16 min 21 s).

4.3

Resultados sobre o modelo de segmentação de dobras

Nesta seção serão apresentadas algumas questões pertinentes ao modelo de segmentação de dobras de pele.

Após levantadas estatísticas de tempo de anotação manual, foi criado um dataset de dobras de pele com as anotações obtidas para treinar o modelo e verificar a capacidade de realizar predições de maneira consistente. Até então, mostrou-se possível o uso do modelo de segmentação semântica para realizar a predição de dobras de pele, conforme alguns exemplos observados e gerados pela própria rede. Apesar do modelo conseguir segmentar dobras de pele, a rede ainda não se encontra em um nível de segmentação satisfatório. Ocorre overfitting durante o treinamento e a validação não generaliza bem.

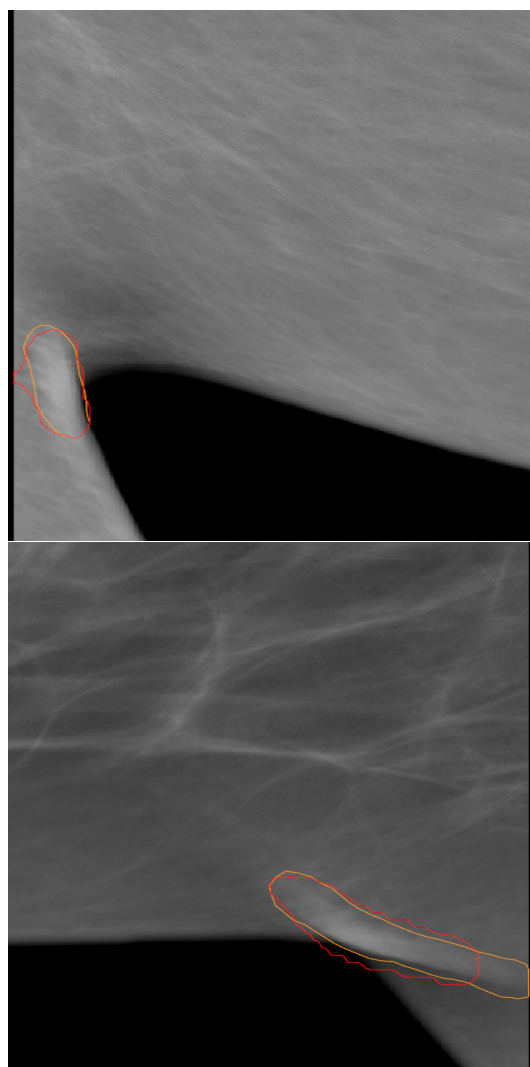


Figura 4.1: Segmentações geradas pela rede na ferramenta.

O modelo de segmentação de aprendizado profundo será utilizado na ferramenta de anotação para sugerir a presença de dobras de pele e segmentar

a mesma nas imagens. Isso facilitaria a questão de realizar a segmentação do zero, uma vez que, ao interagir com um mecanismo na ferramenta, o modelo já criaria a marcação no local de interesse e, caso a segmentação não esteja desejável, basta o usuário fazer ajustes nos contornos do desenho.

Na figura 4.1 estão alguns exemplos de segmentações geradas pelo modelo na ferramenta. O delineado vermelho representa a marcação feita pela rede, o laranja é a marcação corrigida pelo anotador. Ao testar o modelo em outros exemplos, percebe-se que o modelo possui uma tendência em segmentar a região mais destacada da dobra, não delineando em alguns casos as partes mais finas da dobra, como observado no exame de baixo na figura 4.1.

Pelo fato do modelo ser rudimentar, ele gera máscaras adicionais ou falsos positivos em alguns caso. Portanto, é indicada a consulta deste recurso para atuar na tarefa quando visivelmente há presença de dobras, segmentando a maioria dos casos de maneira satisfatória. Em adição, é compreendido que não existe necessidade de consultar o modelo para segmentar imagens que visivelmente não possuem dobras de pele, isso apenas adicionaria mais trabalho para o processo.

4.4

Resultados e análise dos tempos de anotação entre experimentos

Para uma visão melhor dos resultados, a tabela 4.1 esclarece estatísticas levantadas sobre os tempos de anotação referentes à população de ambos os processos de segmentação. Os resultados estão na escala em segundos.

| Manual | | Interactive | |
|--------------------|-----------|--------------------|----------|
| Annotated_images | 1372 | Annotated_images | 508 |
| Highest_time | 187.756 | Highest_time | 36.336 |
| Lowest_time | 0.590 | Lowest_time | 1.860 |
| Mean | 19.651 | Mean | 9.019 |
| Median | 14.488 | Median | 8.292 |
| Standard_deviation | 18.821 | Standard_deviation | 5.349 |
| Variance | 353.982 | Variance | 28.551 |
| Amplitude | 187.166 | Amplitude | 34.476 |
| Total_time | 26961.715 | Total_time | 4581.703 |

Tabela 4.1: Resultados e métricas dos processos de anotação, medidos em segundos.

Como o tamanho das populações de ambas abordagens são diferentes, a abordagem interativa (508 anotações), por possuir uma quantidade de

amostras muito inferior à da abordagem manual (1372 anotações), isso não seria uma comparação justa. Pois, como a grandeza que está sendo analisada é o tempo, pode haver um viés relacionado ao conjunto de menor população, neste caso, os resultados do experimento interativo possuir um tempo total de anotação menor.

Para resolver esta questão, foram sorteadas usando random sampling 300 amostras de cada abordagem, comparando a variância, mediana e média de população com suas amostragens respectivas, conforme as tabelas 4.2 e 4.3.

A amostragem de tamanho 300 foi definida através do cálculo da fórmula de tamanho de amostragem em relação à população do processo manual. Pode ser obtida também uma estimativa utilizando como referência valores recomendados do estudo de Taherdoost (TAHERDOOST, 2016), que elucida as estratégias de seleção de amostragens e valores indicativos sobre tamanho de amostragem. Os parâmetros usados para estimar a amostragem foram: grau de confiança, 95%; margem de erro, 5%; e variância de população, 50% (valor recomendado).

| Manual | |
|---------------------|----------|
| Annotated_images | 300 |
| Highest_time | 143.810 |
| Lowest_time | 0.590 |
| Mean | 19.098 |
| Median | 13.475 |
| Standard_deviation | 18.747 |
| Variance | 351.465 |
| Amplitude | 143.220 |
| Total_time | 5729.377 |
| Population Mean | 19.651 |
| Population Variance | 353.982 |
| Population Median | 14.488 |

Tabela 4.2: Comparação entre medidas de dispersão entre amostragem e população da abordagem manual

| Interactive | |
|---------------------|----------|
| Annotated_images | 300 |
| Highest_time | 33.104 |
| Lowest_time | 1.860 |
| Mean | 8.801 |
| Median | 7.690 |
| Standard_deviation | 5.304 |
| Variance | 28.136 |
| Amplitude | 31.243 |
| Total_time | 2640.331 |
| Population Mean | 9.019 |
| Population Variance | 28.607 |
| Population Median | 8.292 |
| Time Saved | 53.9% |

Tabela 4.3: Comparação entre medidas de dispersão entre amostragem e população da abordagem interativa

Observa-se que as medidas citadas acima possuem um certo grau de aproximação nos valores das amostragens comparadas aos valores de sua população, isto indica que a amostragem escolhida por ambos processos de anotação possuem representatividade em relação a sua população. Ou seja, 300 amostras de tempo do experimento interativo e do experimento manual, possuem uma certa aproximação em relação a média, mediana, variância e desvio padrão de suas devidas populações, servindo como indicativos de representatividade quanto os seus dados.

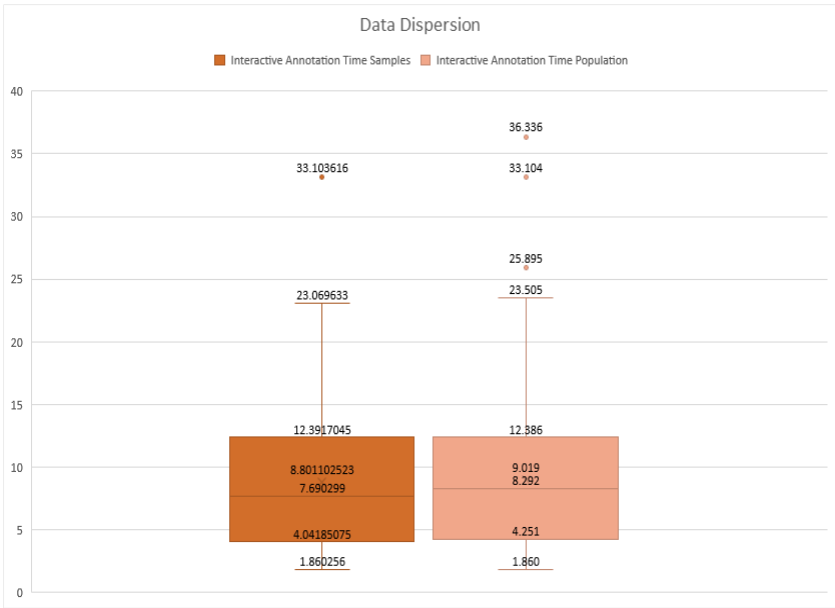


Figura 4.2: Distribuição dos tempos de anotação: população e amostragem do processo manual

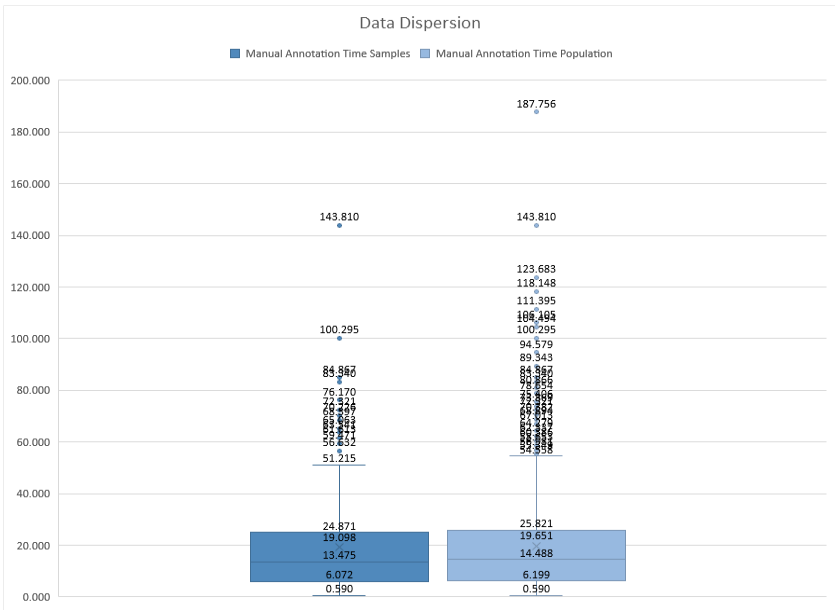


Figura 4.3: Distribuição dos tempos de anotação: população e amostragem do processo interativo

As figuras 4.2 e 4.3 mostram que as amostragens selecionadas conseguem retratar os dados a partir de sua distribuição que aparenta um comportamento homogêneo em relação a sua população. Com isto, é possível analisar ganhos entre as abordagens manual e interativa.

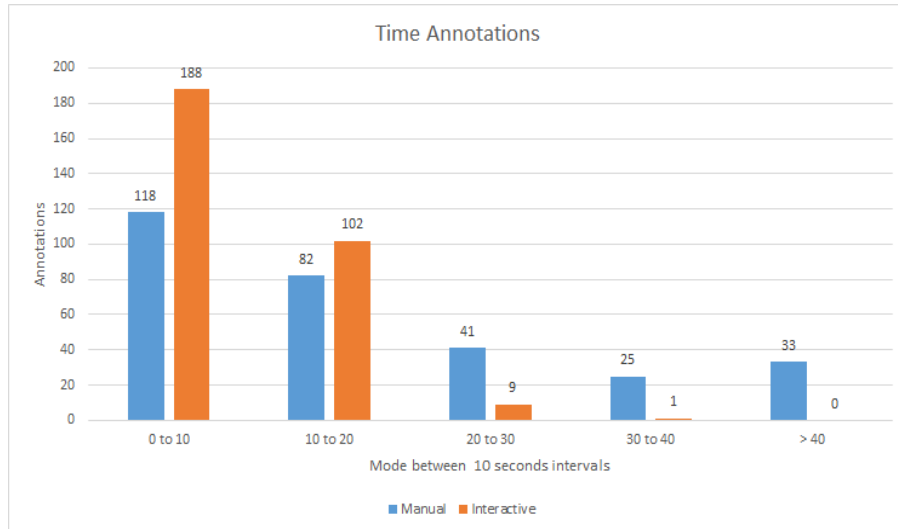


Figura 4.4: Comparativo sobre tempo de anotação entre as abordagens

A partir da figura 4.4 é possível observar que os tempos de anotação da abordagem interativa se concentram em tempos menores que 20 segundos, indicando que realização da tarefa de segmentação teve um ganho com a diminuição de tempo para realizar as etiquetagens. Em contraste com a abordagem manual, uma porção dos tempos de anotação estão distribuídas em intervalos maiores que 20 segundos, o que indica que existe uma tendência de algumas segmentações serem realizadas em durações mais elevadas se comparadas ao processo interativo. Além disso, o tempo economizado com uso do método interativo em relação ao manual foi de 53.9% como visto na tabela 4.3, obtido via equação 4-1.

$$|ts| = \left(\frac{ti - tm}{tm} \right) * 100 \quad (4-1)$$

Onde ts é o tempo economizado; ti é o tempo total no processo interativo; tm é o tempo total no processo manual.

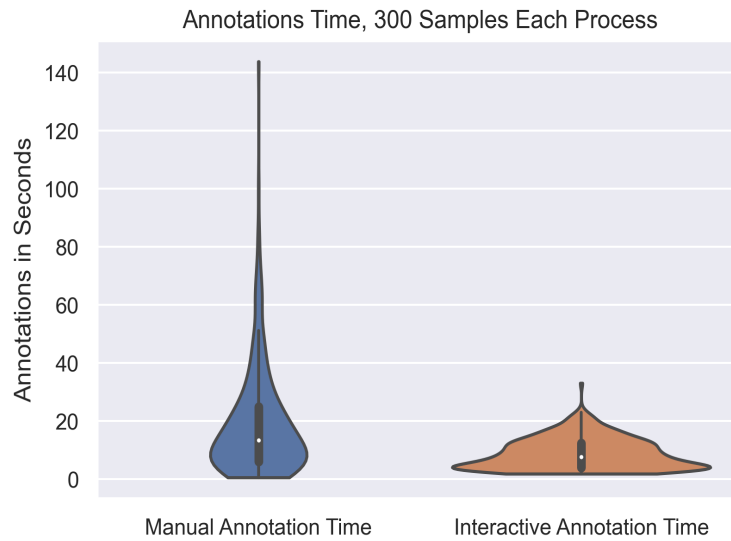


Figura 4.5: Distribuição dos tempos de anotação a partir das amostragens

A figura 4.5, representa a distribuição dos tempos de anotação das abordagens e os quartis. É possível observar que os quartis dos tempos de anotação na anotação interativa encontram-se no intervalo de 0 a 20 segundos em contraste com os tempos de anotação manual, na qual os quartis internos ressaltam uma amplitude maior que os tempo obtidos comparados com a abordagem interativa. Além disto, a mediana do conjunto de tempos de anotação com anotações assistidas, encontra-se abaixo da mediana das anotações manuais.

A figura 4.6 mostra em detalhes a distribuição total dos tempos de cada amostragem, sendo possível entender melhor o contraste entre a segmentação feita manualmente e a segmentação feita com a consulta do modelo.

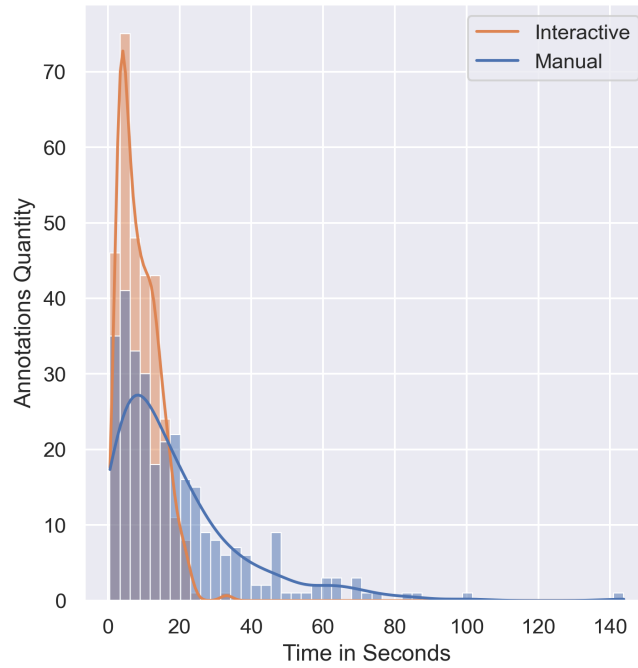


Figura 4.6: Contraste entre os tempos de anotação de ambas abordagens em detalhes.

Observa-se que os passos necessários para fazer os experimentos poderiam ser melhorados, pois a etapa de confirmação visual da dobra de pele na imagem, poderia ser feita antes de acionar o cronômetro, pelo fato deste passo em particular consistir de uma análise visual a procura do artefato na imagem, não possuindo relação com tempo de anotação em si.

Entende-se que, a partir dos resultados, a segmentação assistida em dobras de pele nos exames de incidência MLO atua como um mecanismo que consegue dar apoio à eficiência em anotações destes artefatos mediante sua confirmação visual em imagens nesta categoria, reduzindo o custo de anotação relacionado ao tempo.

No capítulo seguinte, será apresentada uma extensão dos experimentos apresentados nesta seção, focados não apenas em coletar tempos de anotações em dobras de pele. Mas sim, para outras estruturas mamárias com novas medidas pertinentes à interatividade durante as anotações de exames de mamografia.

5

Anotações assistidas para estruturas mamárias CC e MLO

Para este capítulo, propõe-se aprimorar os experimentos manual e interativo que foram feitos em anotações de dobras de pele em imagens de mamografia de incidência MLO. Os experimentos serão estendidos para obter indicativos sobre tempo de anotação e interatividade dos participantes com a ferramenta de anotação durante os experimentos, a fim de obter um panorama sobre tempo de anotação em outras estruturas mamárias e número de interações para anotar uma determinada quantidade de exames em incidências CC e MLO.

Tempo de anotação e quantidade de interações neste estudo, são recursos atrelados a custos de anotação durante os processos de etiquetagem. Sua diminuição seria indicativo de que as anotações estariam sendo produzidas de maneira mais eficiente.

O estudo apresentado neste capítulo limita-se a anotação de imagens e uso de modelos de segmentação para visualizações CC e MLO em exames de mamografia. A seguir serão elucidadas informações sobre os experimentos e os resultados.

5.1

Experimento Manual e Interativo

Os experimentos foram executados com a participação de três anotadores peritos em imagens de mamografia computadorizada. Cada participante realizou dois experimentos. Manual, sem auxílio do modelo para delinear as estruturas e Interativo, que consiste em fazer anotações assistidas com o modelo de segmentação.

Cada experimento consiste em 10 exames de mamografia, totalizando 40 imagens, pelo fato de 1 exame possuir, neste contexto, 2 imagens CC e 2 imagens MLO.

A seguir, serão esclarecidos os passos definidos nos experimentos manual e interativo para realizar as devidas anotações. Elas são empregadas da mesma maneira para as incidências CC e MLO.

Para o experimento manual, as seguintes etapas foram definidas:

1. Confirmar visualmente as estruturas mamárias existentes na imagem examinada.
2. Criar etiquetas para cada estrutura mamária confirmada na imagem.

3. Acionar o cronômetro da etiqueta correspondente à estrutura que o participante irá delinear.
4. Delinear a estrutura mamária manualmente.
5. Acionar o cronômetro novamente da etiqueta, ao terminar a anotação da estrutura.
6. Repetir o passo 3 até que todas as etiquetas criadas sejam anotadas.
7. Visualizar a próxima imagem e repetir o passo 1.

Para o experimento interativo, as seguintes etapas foram definidas:

1. Confirmar visualmente as estruturas mamárias existentes na imagem examinada.
2. Criar etiquetas para cada estrutura mamária confirmada na imagem.
3. Acionar o cronômetro da etiqueta correspondente à estrutura que o participante irá delinear.
4. Logo em seguida, acionar o modelo de segmentação para delinear a estrutura em específico.
5. Caso o delineamento feito pela rede não estiver satisfatório, corrigir a segmentação, manualmente.
6. Acionar o cronômetro novamente da etiqueta, ao terminar a anotação da estrutura.
7. Repetir o passo 3 até que todas as etiquetas criadas sejam anotadas.
8. Visualizar a próxima imagem e repetir o passo 1.

A figura 5.1 destaca em vermelho dois botões da ferramenta de anotação que controlam a cronometragem do tempo de anotação e o modelo para gerar as segmentações. O botão destacado à esquerda é responsável pela cronometragem do tempo de anotação da estrutura em específico. O botão à direita, é o mecanismo que aciona o modelo para segmentar a estrutura mamária a ser anotada.

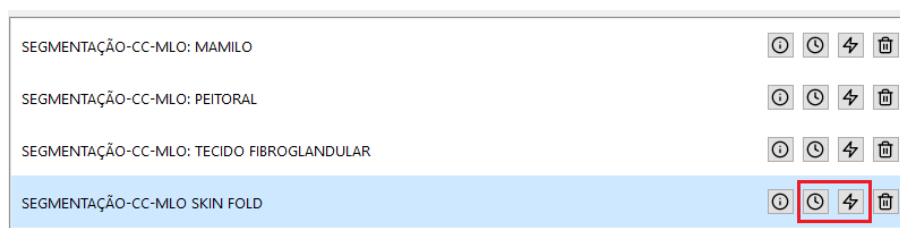


Figura 5.1: Interface das etiquetas criadas e seus mecanismos.

As estruturas mamárias a serem anotadas para ambos os experimentos são: papila, músculo peitoral, tecido fibroglandular e contorno da mama em incidências CC. Para MLO, são todas as estruturas acima com adição de dobra de pele. No entanto, não será coletado tempo de anotações pertinente ao contorno da mama, pois a maneira como é feita a anotação desta estrutura em particular consiste em usar um algoritmo de anotação automática que detecta as bordas da mama.

Os modelos de segmentação semântica usados para auxiliar nas anotações no experimento interativo são pré treinados com dados de outros equipamentos de mamografia, diferentes dos que geraram as imagens nas quais serão feitas as campanhas de anotação nos experimentos.

Em paralelo, durante os experimentos, um software será utilizado para coletar medidas de quantidade de interações de mouse e teclado, para obter dados que permitam entender uma margem de interações para anotar as determinadas quantidades de imagens de mamografia por experimento.

A ferramenta OGAMA (VOSSKÜHLER et al., 2008) é utilizada em segundo plano para mensurar interatividade do anotador humano, proporcionando medidas de interação durante os experimentos de anotação nos exames.

Os acessórios utilizados pelos anotadores peritos para fazer o delineamento das estruturas foram apenas teclado e mouse. O pressionamento dos mecanismos destes objetos serão contabilizados durante os experimentos, pois entende-se que a quantidade de interações para atividades que envolvam etiquetagem de dados, têm uma relação com custos de anotação.

Esta ferramenta é configurada e acionada no início de cada experimento, a qual habilita a coleta dos dados de interação durante a execução do experimento. Ao terminar a anotação de todos os 10 exames, o anotador ativa um mecanismo na ferramenta OGAMA para encerrar o experimento e salvar os dados coletados de interatividade. Este procedimento é seguido para ambos os experimentos manual e interativo.

5.2

Resultados

Nesta seção, serão apresentadas os resultados dos experimentos feitos pelos participantes. Cada subseção a seguir corresponde aos resultados obtidos individualmente por participante sobre quantidade de interações, tempo de experimento e gráficos que demonstram a dispersão dos tempos de anotação sobre cada estrutura mamária, divididos por experimento manual e interativo. A escala de tempo entre os resultados foram todas normalizadas para segundos.

A tabela 5.1 demonstra os resultados das medidas de tempo de anotação dos participantes sobre todas as estruturas anatômicas relacionadas.

| Annotation Time Results per Mammary Structure in Seconds | | | | | |
|--|-------------|---------------------|-----------------------|--------------------------|-----------------------|
| Manual Experiment | | | | | |
| Annotator | Total Time | Nipple | Pectoral | Fibrotissue | Skinfold |
| A | 2724.03 | 926.40 | 585.06 | 1032.43 | 180.13 |
| B | 2043.69 | 692.83 | 428.00 | 794.20 | 128.64 |
| C | 3316.71 | 1325.33 | 796.67 | 984.75 | 209.94 |
| Annotator | Sum Average | Average Time Nipple | Average Time Pectoral | Average Time Fibrotissue | Average Time Skinfold |
| A | 87.15 | 23.16 | 20.17 | 25.81 | 18.01 |
| B | 72.71 | 17.32 | 19.45 | 19.85 | 16.08 |
| C | 117.19 | 33.13 | 33.19 | 24.61 | 26.24 |
| Interactive Experiment | | | | | |
| Annotator | Total Time | Nipple | Pectoral | Fibrotissue | Skinfold |
| A | 2229.68 | 974.81 | 401.53 | 725.70 | 127.63 |
| B | 2017.52 | 1027.43 | 281.55 | 643.98 | 64.54 |
| C | 2853.06 | 1516.89 | 517.93 | 662.91 | 155.32 |
| Annotator | Sum Average | Average Time Nipple | Average Time Pectoral | Average Time Fibrotissue | Average Time Skinfold |
| A | 72.13 | 24.37 | 15.44 | 18.14 | 14.18 |
| B | 72.74 | 25.68 | 14.81 | 16.09 | 16.13 |
| C | 96.42 | 37.92 | 22.51 | 16.57 | 19.41 |

Tabela 5.1: Comparação de medidas de tempo em ambos os experimentos entre todos os anotadores participantes.

As figuras 5.2 e 5.3, apresentam respectivamente, os tempos médios de anotação por estrutura mamária divididos por cada participante e a soma dos tempos médios de anotação das estruturas entre cada participante. Os valores específicos das figuras 5.2 e 5.3 estão de acordo com a tabela 5.1. A dispersão e informação sobre os quartis dos tempos de anotação de cada estrutura mamária entre os participantes, são elucidadas na figura 5.4, onde estão agregadas várias figuras divididas por participante e qual o experimento específico.

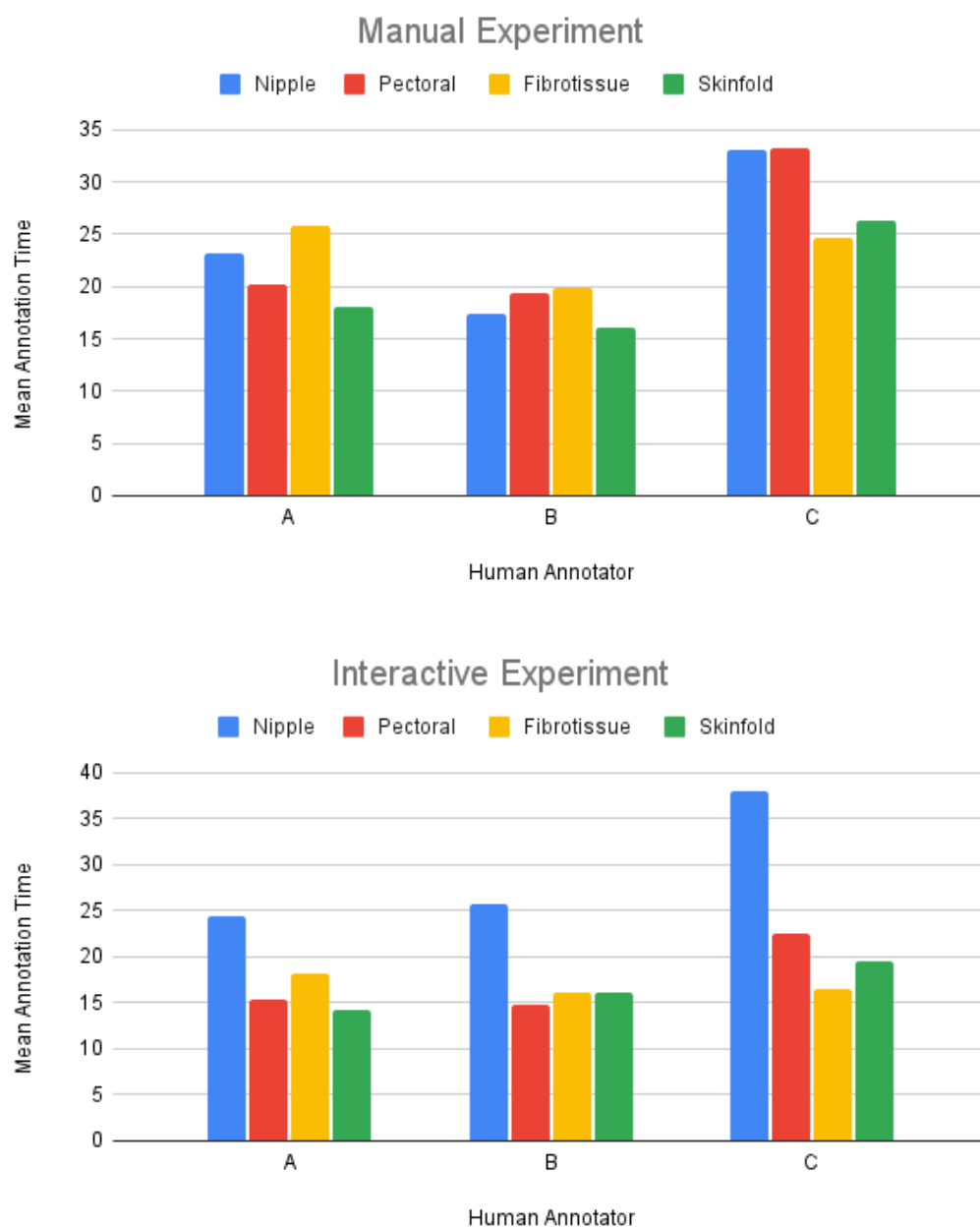


Figura 5.2: Tempo médio de anotação por estrutura mamária entre os participantes em ambos experimentos .

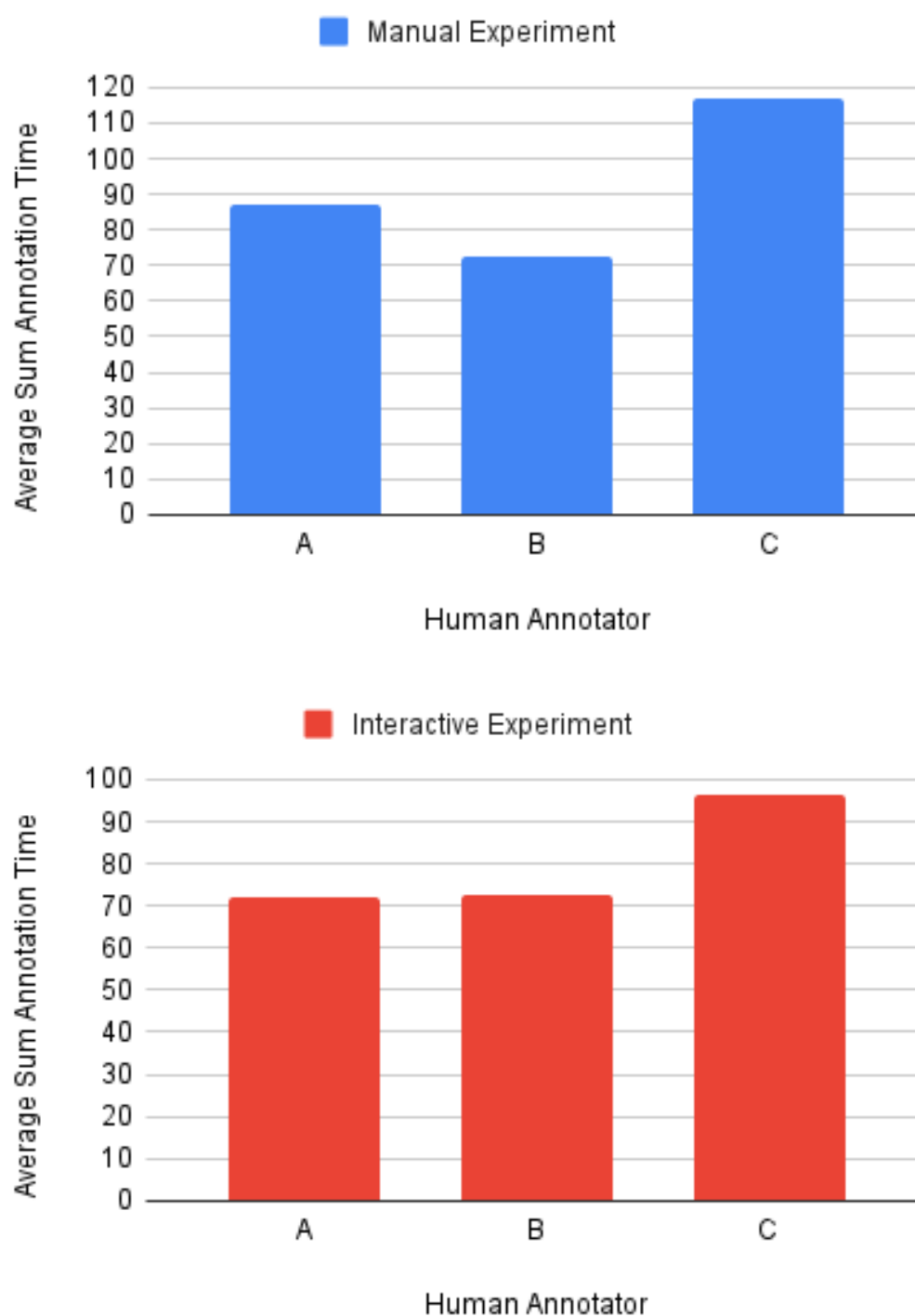


Figura 5.3: Soma dos tempos médios das estruturas entre cada participante em ambos experimentos.

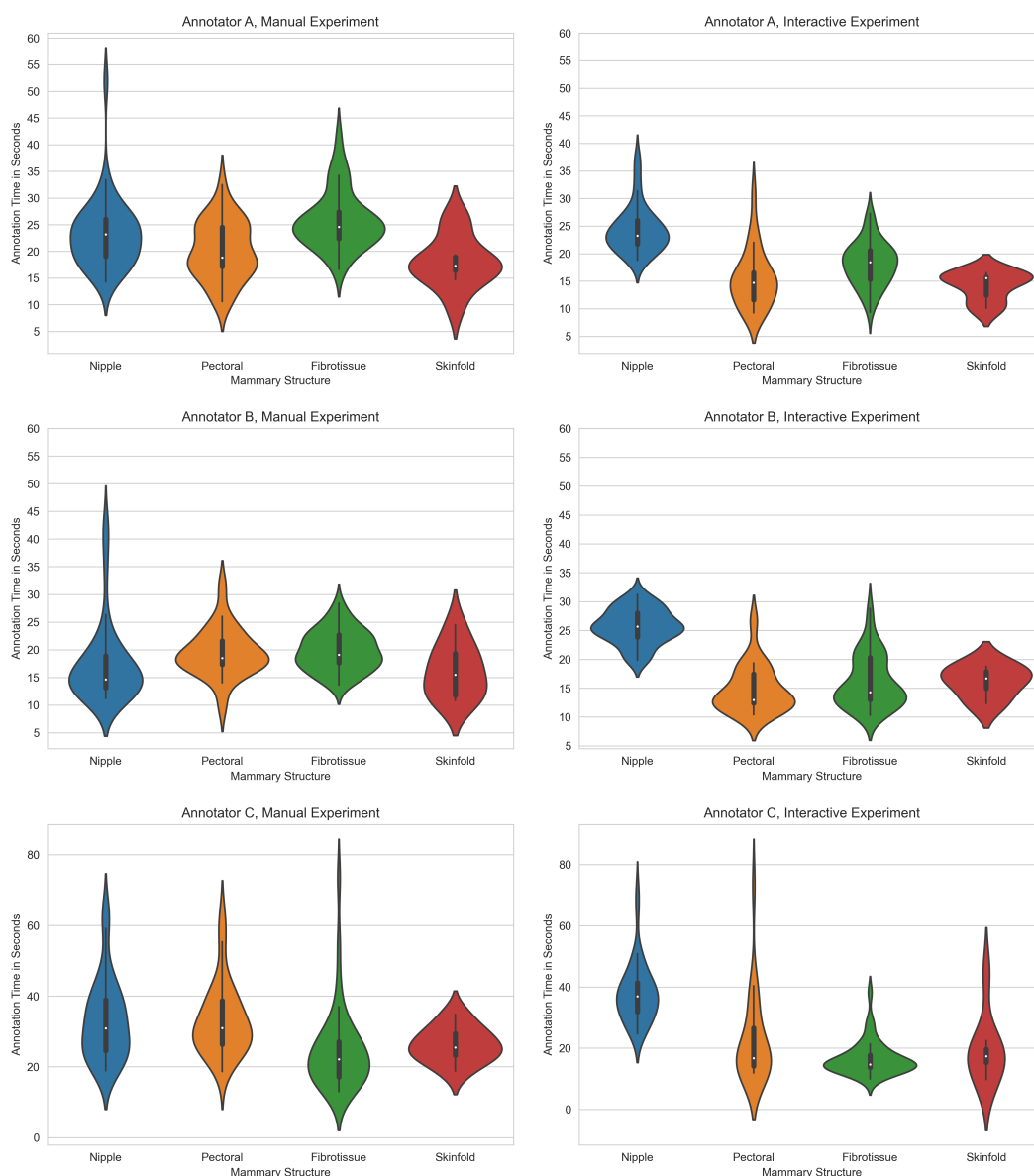
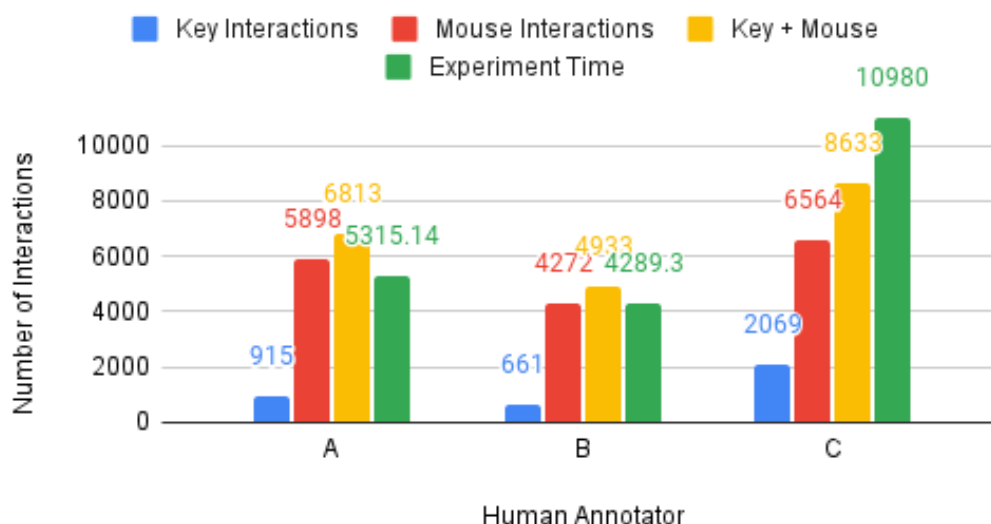


Figura 5.4: Distribuição dos tempos de anotações por estrutura mamária em ambos experimentos entre os participantes

Na figura 5.5 são ilustradas as quantidades de interações com teclado, mouse e a soma de ambos entre os participantes durante cada experimento.

Manual Experiment



Interactive Experiment

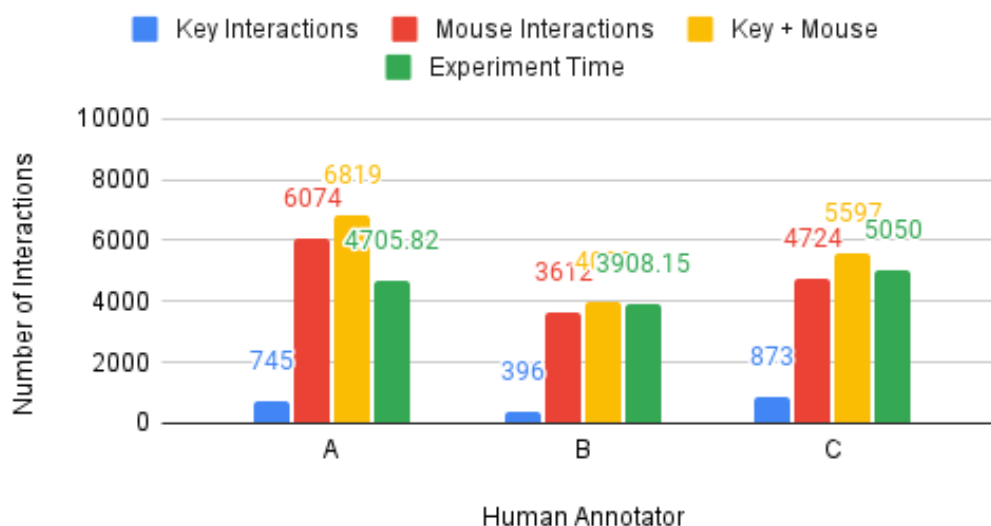


Figura 5.5: Tempos de duração dos experimentos e número de interações obtidos pelo OGAMA em ambos experimentos entre os participantes.

5.3

Análise dos resultados

Avaliando os resultados entre os participantes, com relação ao tempo de anotação total e a soma média dos tempos de anotação, conforme as figuras 5.2 e 5.3, observa-se uma redução no tempo de anotação total e médio entre as estruturas mamárias como peitoral, tecido fibroglandular e dobras de pele. No entanto, em relação à papila, verifica-se que o uso do modelo de segmentação para auxiliar no delineamento da estrutura está acrescentando

tempo de anotação, aumentando o tempo total e tempo médio de maneira geral entre os anotadores peritos. Em adição, na figura 5.4, os quartis em relação ao tempo de anotação da papila dos experimentos interativos entre os participantes encontram-se em parcelas de tempo mais elevadas, em contraste com os quartis dos experimentos manuais de um modo geral.

Observou-se, durante os experimentos interativos, que a segmentação feita pelo modelo é insatisfatória do ponto de vista dos anotadores peritos. Apesar de todas as estruturas mamárias estarem sujeitas a correções, a papila em particular, na maioria dos casos, requer uma correção total da sua máscara de segmentação, conforme exemplos ilustrados na figura 5.6, onde os delineamentos em vermelho e verde representam, respectivamente, o delineamento feito pelo modelo e o delineamento corrigido pelo anotador. Nota-se que o acionamento do mecanismo para o modelo sugerir a segmentação da estrutura mamária específica possui um tempo de processamento, ou seja, o modelo, ao ser consultado para sugerir a segmentação, consome um tempo adicional para a segmentação ser desenhada no exame, sendo um fator que estaria adicionando tempo anotação para as demais estruturas, em específico a papila e dobras de pele. Podendo nestes casos, o acionamento do modelo não ser lucrativo em relação ao tempo de anotação.

Além disso, notaram-se casos raros em que o modelo de dobras de pele, previamente treinado no capítulo anterior, segmentou indevidamente a dobra de pele em alguns exames de incidência MLO durante os experimentos interativos conforme a figura 5.7, onde a segmentação estaria sendo feita em etiquetas presentes nas imagens MLO ao invés da dobra de pele em si. Entende-se que o modelo possa confundir com localização da dobra de pele, dependendo dos níveis de contraste e brilho presentes na imagem e também pelo fato dos dados utilizados para o treinamento da rede serem de imagens de equipamentos de mamografia diferentes.

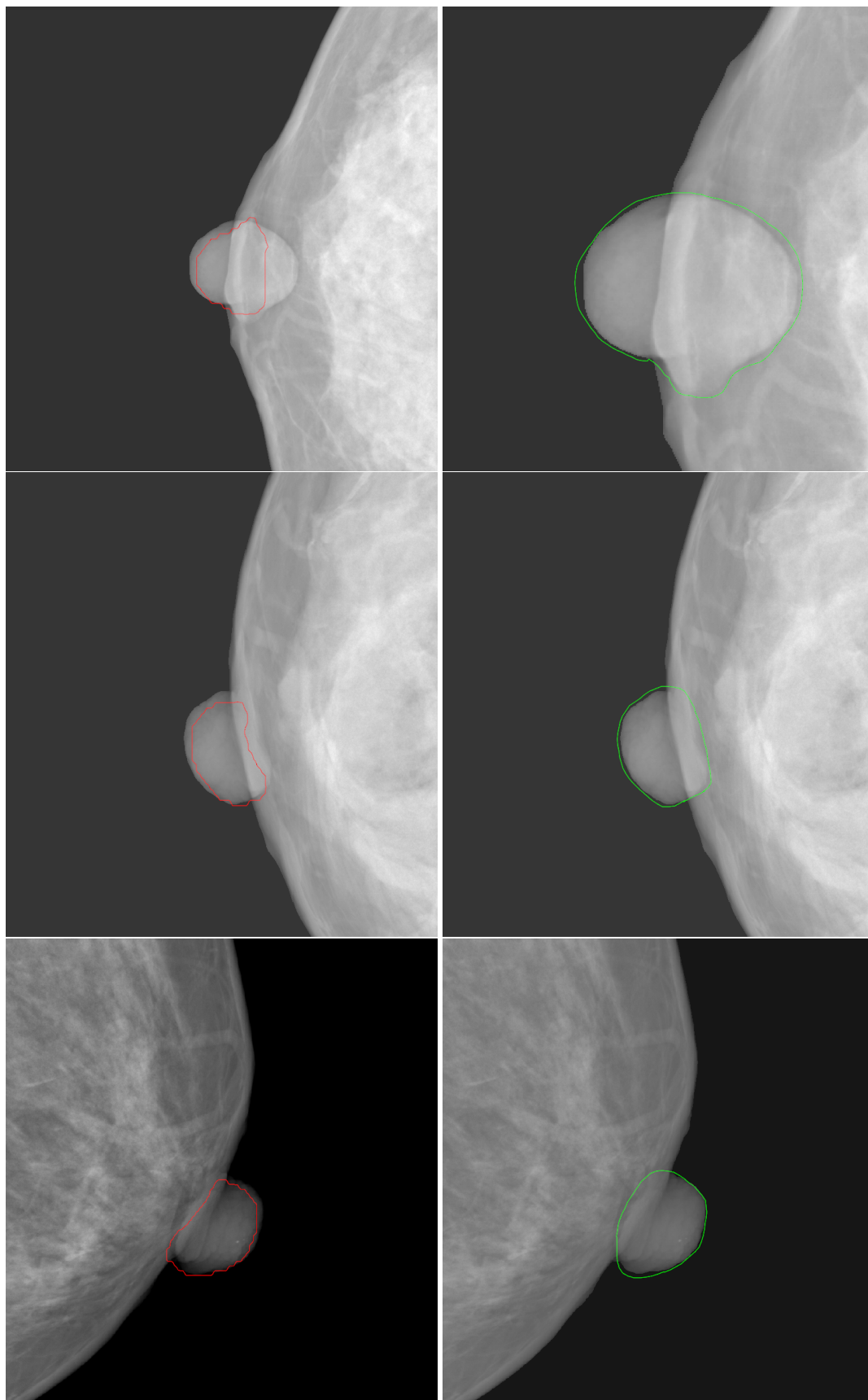


Figura 5.6: Exemplos de segmentações de papila durante os experimentos interativos.

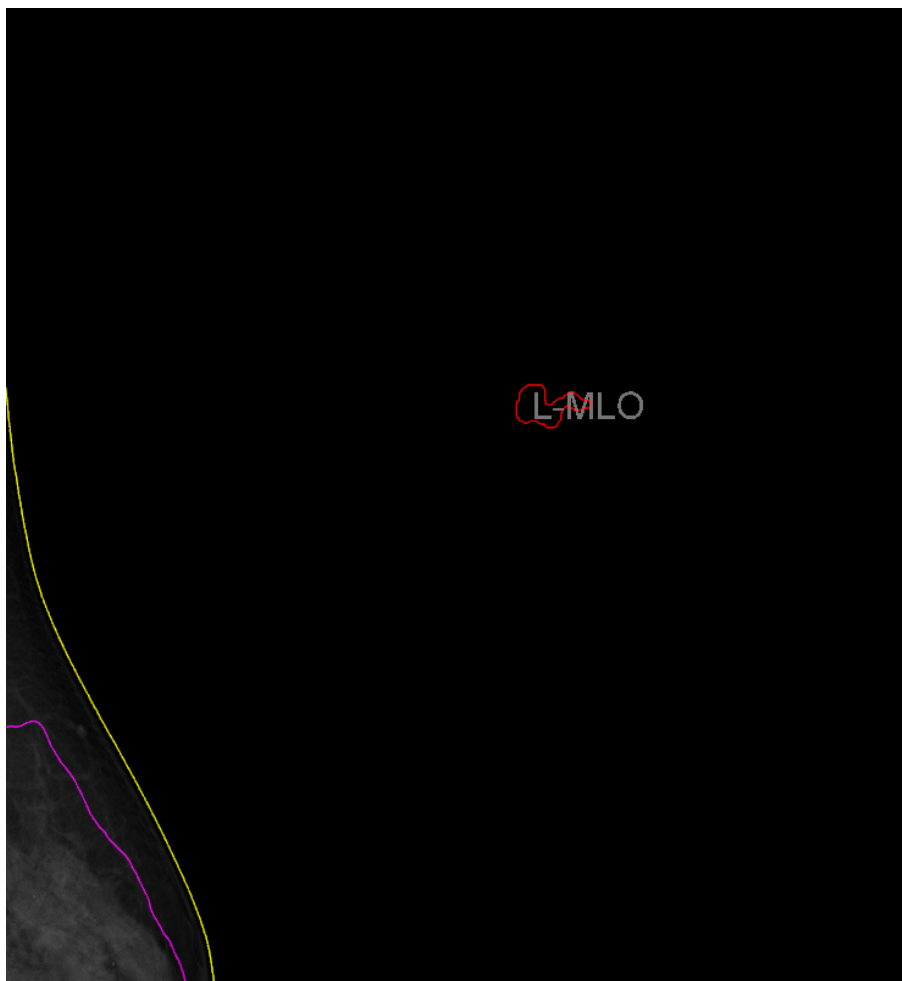


Figura 5.7: Caso raro de segmentação incorreta de dobra de pele

Sobre a quantidade de interações realizadas para completar os experimentos, houve uma diminuição por parte de interações com o teclado e mouse de um modo geral. Isso pode indicar que a consulta do modelo de segmentação para auxiliar em anotações de estruturas mamárias, em geral, ajuda a interagir menos com a ferramenta durante as campanhas de anotação de exames de mamografia e houve uma diminuição nos tempos de experimento entre os participantes, conforme observado na figura 5.5.

Em resumo, o mecanismo de anotação assistida para estruturas mamárias consegue proporcionar certa eficiência de anotação para estruturas como músculo peitoral, tecido fibroglandular e dobras de pele entre dois peritos com base nos exames anotados neste capítulo. A papila demonstrou o contrário, indicando que seriam necessárias outras melhorias para auxiliar no tempo de anotação desta estrutura mamária em particular. Sobre as medidas de interação obtidas nos experimentos interativos, houve uma redução no pressionamento de teclas entre todos os anotadores e mouse entre 2 participantes, podendo indicar, de um modo geral, uma redução na quantidade de interações ao utili-

zar modelos de segmentação para sugerir os delineamentos para anotar exames de mamografia. O acionamento do modelo para sugestão de segmentações de estruturas como mamilo neste cenário não se demonstra proveitoso, pelo fato do uso do mecanismo incrementar tempo de anotação, demorando mais para completar a anotação para esta estrutura específica por conta do tempo de processamento para a rede gerar o delineamento. O que pode ter impactado neste cenário, entende-se pelo fato dos modelos utilizados para sugestão de segmentações, de serem treinados com imagens de mamografia de domínios diferentes, ou seja, as redes foram treinadas com dados de outros equipamentos de mamografia, imagens que diferem do cenário em que os experimentos foram executados.

6

Active Learning

Para este capítulo, propõe-se implementar a abordagem de active learning para reduzir e selecionar os dados mais informativos a serem anotados, a partir da aplicação de TTA e um conjunto de técnicas para estimar incertezas sobre a predição, utilizando uma conjuntura de medidas que retornem pontuações de incerteza das imagens de mamografia mais pertinentes para anotação. Para avaliar a abordagem, serão apresentados experimentos para demonstrar que as pontuações de incerteza atreladas às imagens conseguem servir de parâmetros para melhorar a acurácia de segmentação do modelo a cada ciclo de active learning e direcionar as imagens de maior relevância para a anotação.

O tipo de dado utilizado consiste em imagens de mamografia DICOMs. São dados de alta-dimensionalidade que comportam em meta dados informações de exames de mamografia, inclusive a imagem e sua respectiva incidência CC ou MLO. Para este trabalho, a aplicação da técnica limita-se ao uso apenas para visualizações de mamas CC.

6.1

Etapas de active learning

O objetivo, neste contexto, é obter um modelo rudimentar para fazer predições e separar os dados mais informativos.

O primeiro passo consiste na aquisição de imagens de mamografia, na anotação de uma parte dos dados adquiridos para gerar um dataset inicial e no treinamento do modelo para habilitá-lo a fazer as inferências sobre o conjunto dados não anotados (pool). A seguir, para otimizar o tempo de pesquisa, as etapas de anotações dentro de active learning são simuladas, ou seja, todos os dados da pool já possuem anotações de antemão que são "ligadas" no momento em que o modelo indica que a respectiva imagem deve ser anotada.

Ao atingir os requisitos iniciais descritos acima, o ciclo de active learning proposto consiste nos passos abaixo:

1. Usar o modelo de segmentação semântica para fazer predição sobre todas as imagens da pool com a técnica TTA.
2. Calcular scores de incerteza sobre as predições das imagens da etapa anterior.
3. Encaminhar aos anotadores as imagens com score de incerteza mais elevados ou separados por algum outro critério.

4. Adicionar ao dataset da rede os dados etiquetados pelos anotadores¹.
5. Treinar um novo modelo com o dataset aperfeiçoado e gerar as métricas de avaliação deste novo modelo.
6. Repetir o passo 1 até que a condição de parada definida seja atingida.

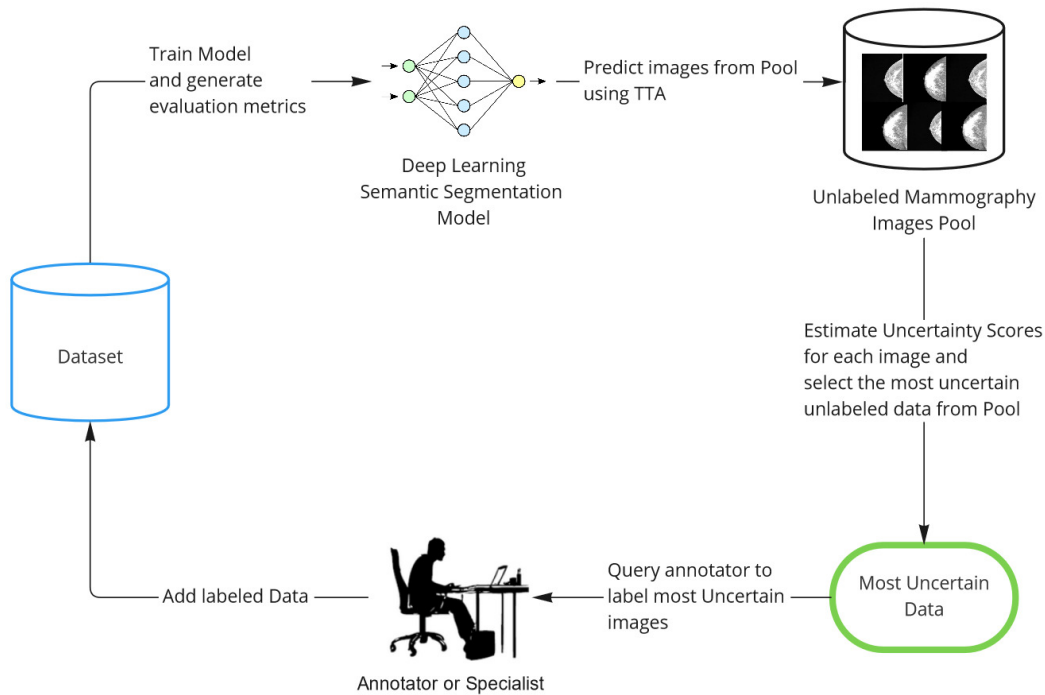


Figura 6.1: Ciclo active learning para segmentação de imagens de mamografia.

Para ter um melhor entendimento da abordagem, a figura 6.1 demonstra um panorama sobre as etapas necessárias da técnica apresentada.

Após treinado o modelo, o passo seguinte será realizar inferências sobre todas as imagens de mamografia da pool utilizando o método de TTA.

6.2

TTA: Test-Time Augmentation

Para que seja possível estimar a incerteza a partir das inferências do modelo de segmentação semântica, é necessário gerar diferentes predições feitas pelo modelo sobre a imagem para capturar ambiguidades que permitam calcular a incerteza da rede sobre as várias predições efetuadas acima. Esta técnica é crucial para obter informações que serão utilizadas nos scores de incerteza, que permitirão pontuar as imagens de mamografia da pool.

¹Lembrando que não usamos anotadores humanos. Nesta etapa, simplesmente passamos a usar as anotações já existentes na imagem, que não estavam sendo usadas até então.

Baseando-se na técnica proposta por Wang et al. (WANG et al., 2019), TTA irá permitir estimar a incerteza do modelo a partir de suas predições, aplicando perturbações sutis e válidas sobre a imagem que será processada pela rede. Para cada transformação aplicada sobre a imagem, será feita a predição sobre a mesma. Vale ressaltar que as transformações, filtrações ou perturbações aplicadas nas imagens devem ser sutis e válidas, isto é, as manipulações acima não devem alterar severamente o semblante da imagem ou dificultar muito a predição feita pelo modelo de segmentação. Caso contrário, a imagem pode se tornar muito diferente da original, prejudicando os resultados obtidos pela rede por conta de transformações excessivas a imagem. Por exemplo, a figura 6.3 ilustra casos onde a imagem de entrada é perturbada intensamente, modificando a identidade das estruturas mamárias e comprometendo inferências feitas pelo modelo de segmentação semântica.

Em resumo, as perturbações devem ser suficientes para gerar predições em que os resultados mostrem diferenças entre as inferências feitas sobre as amostras alteradas em relação à imagem original, conforme a figura 6.2, onde são causadas perturbações em escalas variadas sob os níveis de gama na imagem, com limites contidos, para que as amostras modificadas não sejam muito diferentes entre si, garantindo segmentações com resultados aproximados. Gama, em contexto de processamento de imagem, consiste nos níveis de contraste e brilho em que sua alteração proporciona um acréscimo ou decréscimo não linear em função do contraste e brilho na imagem. Quando o valor de gama for > 1 , a imagem se torna mais clara e caso o valor seja < 1 , a mesma se torna mais escura (SINGNOO; FINLAYSON, 2010). Esta técnica se destaca em relação a outros tipos de transformações pelo fato de se adaptar bem ao incorporar, com limites entre 0.6 e 1.4, em imagens de mamografia durante o processo de TTA, não gerando amostras muito diferentes entre si e a segmentação feita sobre as imagens transformadas não são prejudicadas.

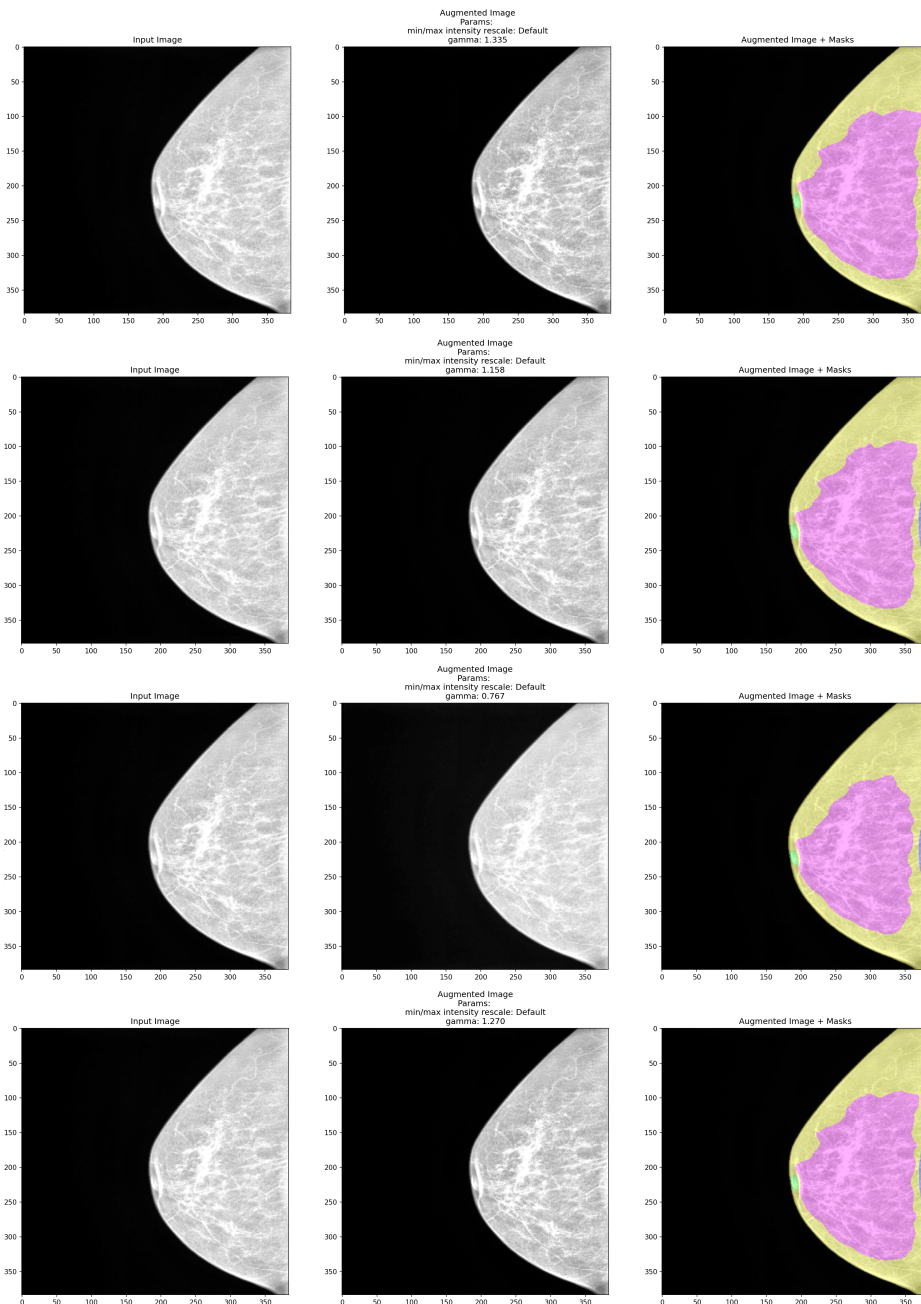


Figura 6.2: Amostras geradas a partir de TTA e as inferências do modelo sobre as imagens transformadas.

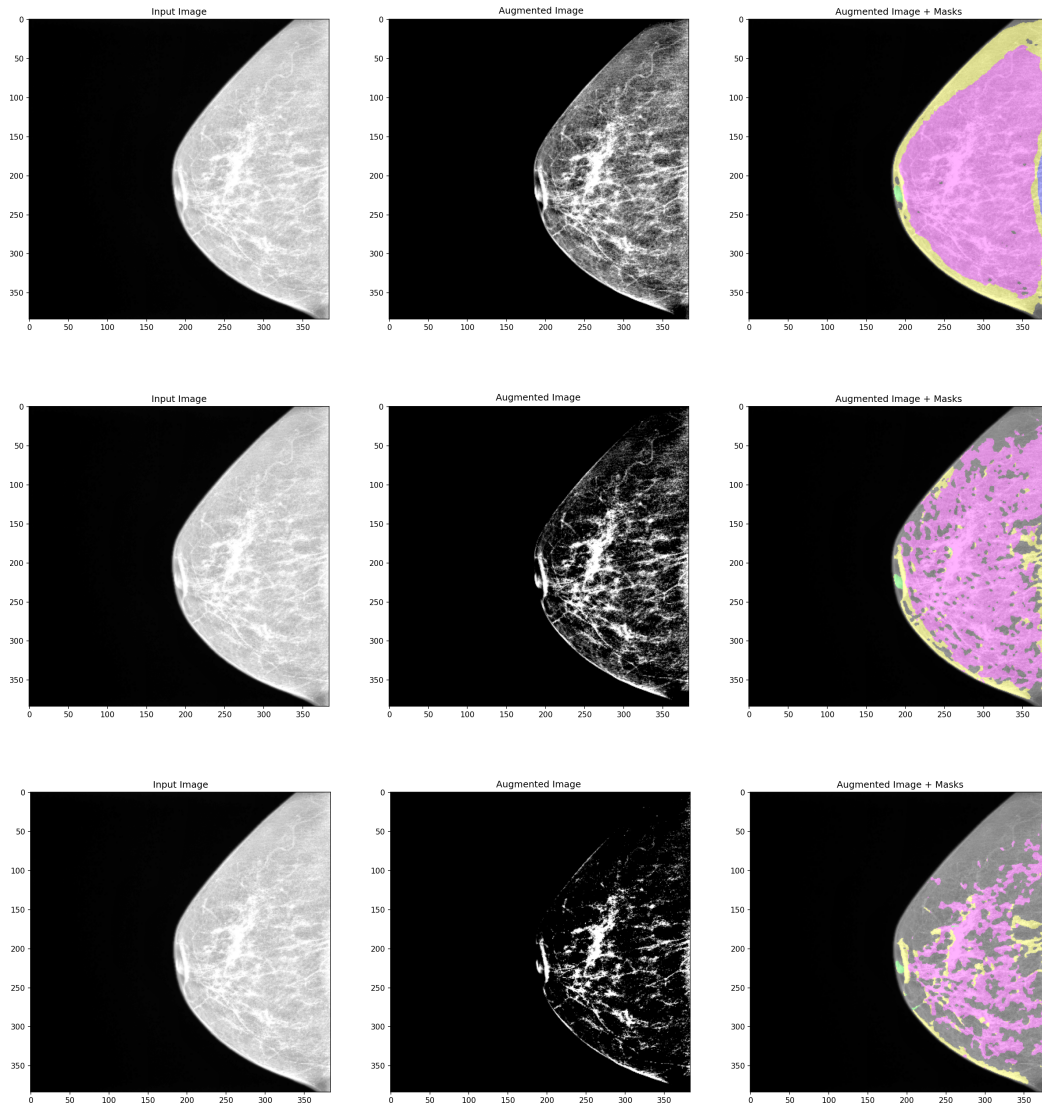


Figura 6.3: Amostras geradas a partir de TTA com perturbações excessivas.

6.3 Estimação da incerteza

O processo para estimar incerteza de uma maneira global e por estrutura mamária pode ser obtido baseando-se na figura 6.4, que será elucidado ao longo desta subseção.

Esta etapa é importante para obter os insumos que serão utilizados para quantizar, a partir da estimação da incerteza, pontuações que serão computadas e serão atribuídas aos dados da pool para poder selecionar as imagens em que o modelo teve maior dificuldade em segmentar. Além disso, essa etapa proporciona mapas de calor de incerteza sobre os pixels classificados pelo modelo de segmentação semântica que servem como um horizonte sobre

onde possivelmente o modelo estaria confuso em segmentar na imagem.

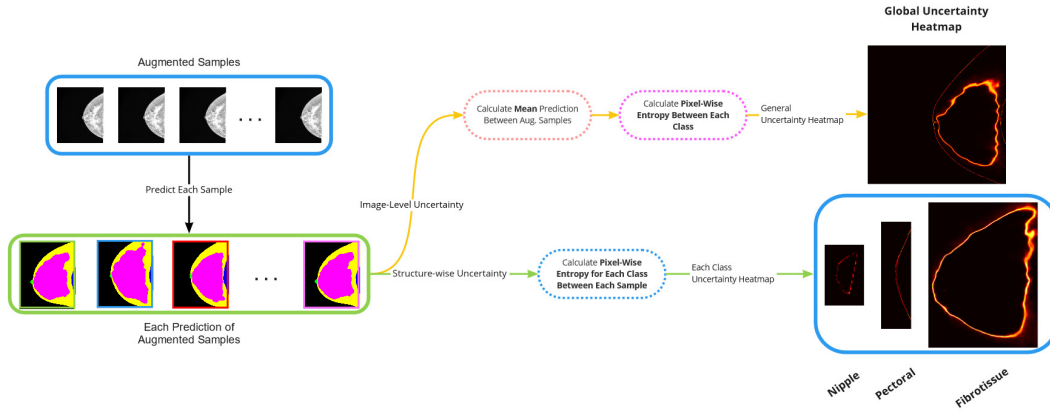


Figura 6.4: Esquematização do processo para obtenção dos mapas de incerteza.

Sendo assim, podemos estimar a incerteza baseando se na entropia de Shannon (SHANNON, 1948):

$$H(X) = - \sum_{i=1}^N P(x_i) * \log_2 P(x_i) \quad (6-1)$$

Na qual $P(x_i)$ é a probabilidade do i -ésima ocorrência de x .

Segundo o livro de Monarch (MONARCH, 2021), a entropia pode se entender como uma métrica de incerteza ao medir probabilidades em ocorrências de eventos. Quanto mais alta a entropia H , maior a incerteza.

A maneira como entropia é computada deve ser ajustada, pois, em problemas que envolvem segmentação semântica, a classificação é feita no nível de pixel, diferente de problemas de classificação supervisionada.

Como TTA gera diferentes inferências, o primeiro passo para computar a incerteza consiste em tirar a média das n -ésimas previsões feitas pela rede sobre as imagens transformadas geradas por TTA, conforme a equação:

$$P_i^k = \frac{1}{S} \sum_s (p_i^k | p = k) \quad (6-2)$$

Para a qual P_i^k é o valor de predição média ao i -ésimo pixel pertencente à estrutura mamária k (classe). S é o número de previsões de TTA e $(p_i^k | p = k)$ é a probabilidade do pixel p_i^k pertencer à classe k . Para resumir, define-se apenas P como a média das previsões das imagens transformadas por TTA.

Obtendo uma predição P que representa a média das previsões geradas pelas S transformações aplicadas sobre a imagem, é possível adaptar a equação 6-1 para estimar a incerteza a nível de pixel a partir de P da seguinte forma:

$$E_i = - \sum_i^I \sum_k^K P_i^k * \log_2(P_i^k) \quad (6-3)$$

Em que E_i corresponde à entropia no nível de pixel de P_i^k . Ao computar E_i para cada i -ésimo pixel, obtém-se o mapa de incerteza que serve como um panorama sobre a visualização das regiões onde possivelmente o modelo indica maior incerteza, conforme as figuras 6.5 e 6.6.

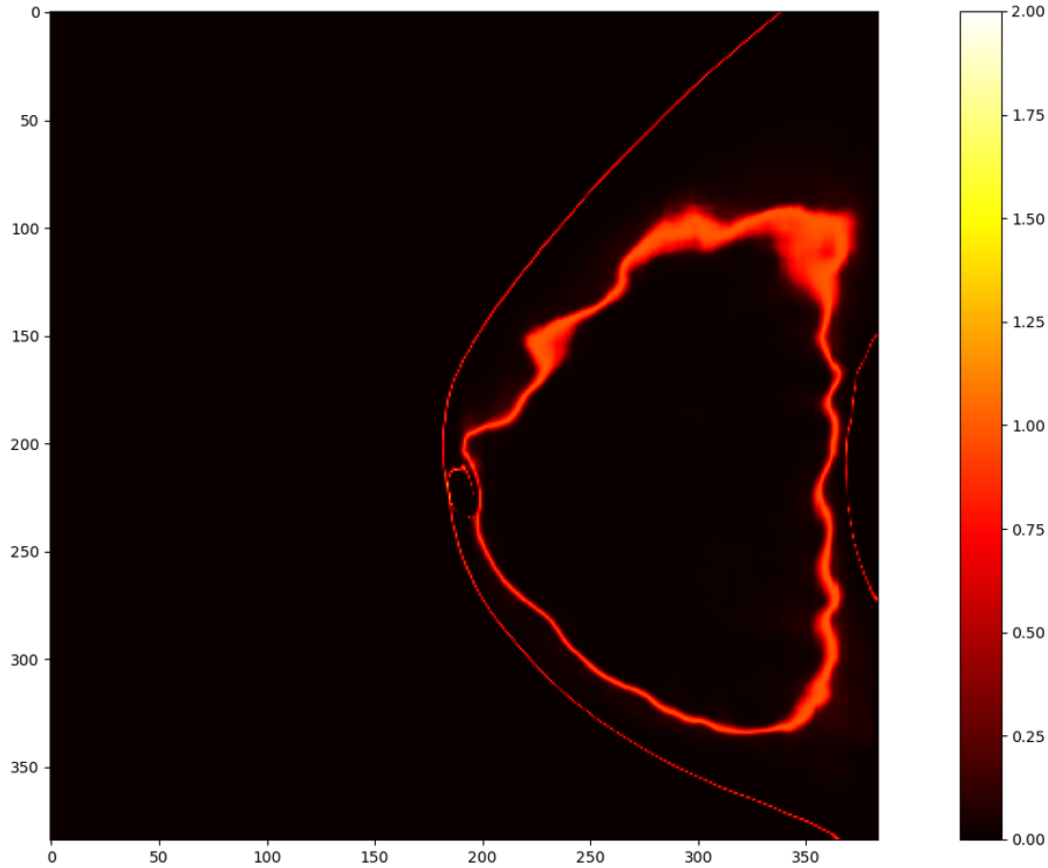


Figura 6.5: Imagem com baixa incerteza sobre as inferências geradas de TTA das estruturas mamárias.

Uma predição com incerteza baixa é indicada por alguns fatores visuais que podem ser identificados na figura 6.5, como: localização da incerteza predominantemente em bordas finas nas regiões das estruturas mamárias; poucos ruídos e oscilações nos limites das estruturas; região da incerteza bem delimitada, sem ramificações internas e temperatura baixa em torno de cada estrutura.

Já uma predição com incerteza alta consiste em aspectos visuais encontrados na figura 6.6, como: incerteza presente fora da mama; temperaturas elevadas sobre as estruturas mamárias; muita oscilação, ruídos e ramificações da incerteza por toda imagem.

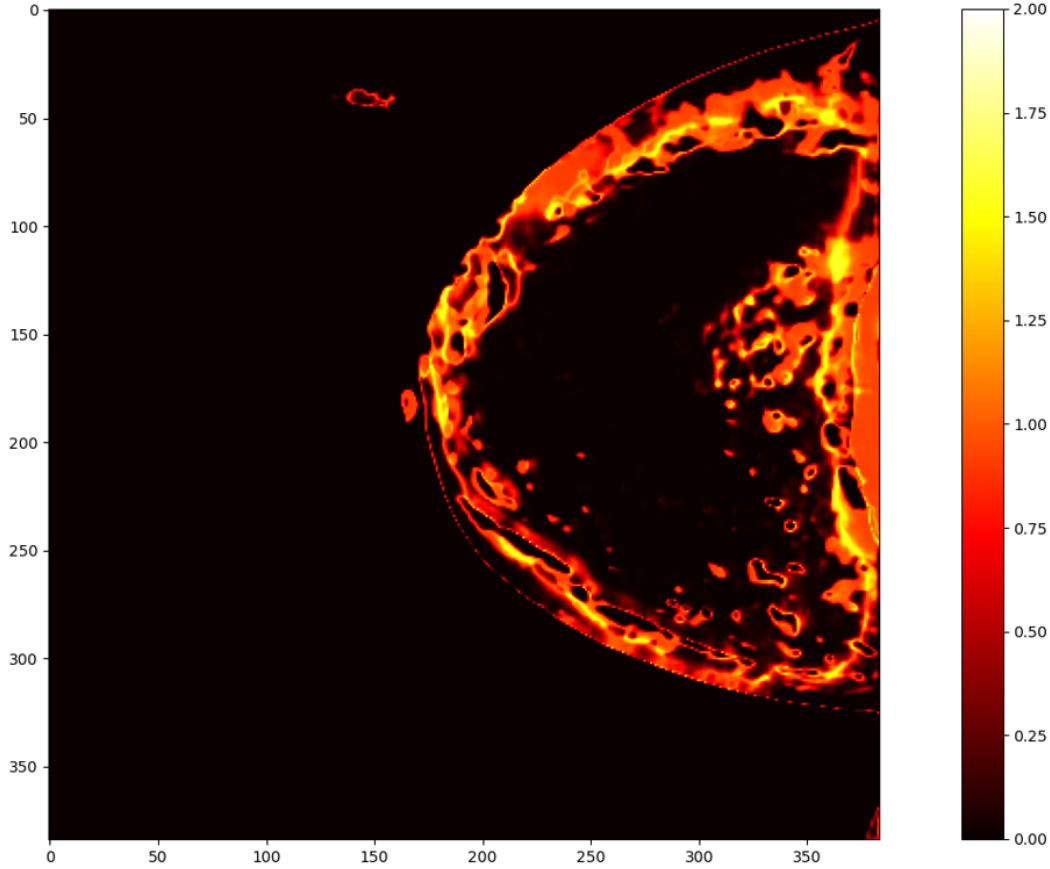


Figura 6.6: Imagem com alta incerteza sobre as inferências geradas de TTA das estruturas mamárias.

A visualização dos mapas de incertezas acima ajuda a entender, de uma maneira geral, como as inferências feitas na imagem pelo modelo exibem uma certa dúvida sobre os pixels. Além disso, um panorama visual da incerteza do modelo ao fazer previsões a partir da técnica TTA.

No entanto, uma análise mais ponderada sobre a incerteza das estruturas mamárias e uma métrica para filtrar os dados com maior incerteza para active learning, somente será possível mediante mapas de incerteza e medidas de incerteza focadas por estrutura mamária.

Para gerar mapas de incertezas por estrutura mamária, é necessário adaptar a equação 6-3 para, ao invés de calcular a entropia no nível de pixel E_i em função da média das previsões P , computar a entropia em função das previsões das imagens transformadas s para uma estrutura específica k . Define-se E_k como entropia no nível de estrutura anatômica, que corresponde ao mapa de incerteza para estrutura k .

A figura 6.4 demonstra o esquema de obtenção dos mapas de incerteza para diferentes propósitos de maneira resumida. Na qual, ao obter E_k para

cada estrutura anatômica k , é possível visualizar a incerteza da estrutura em específico.

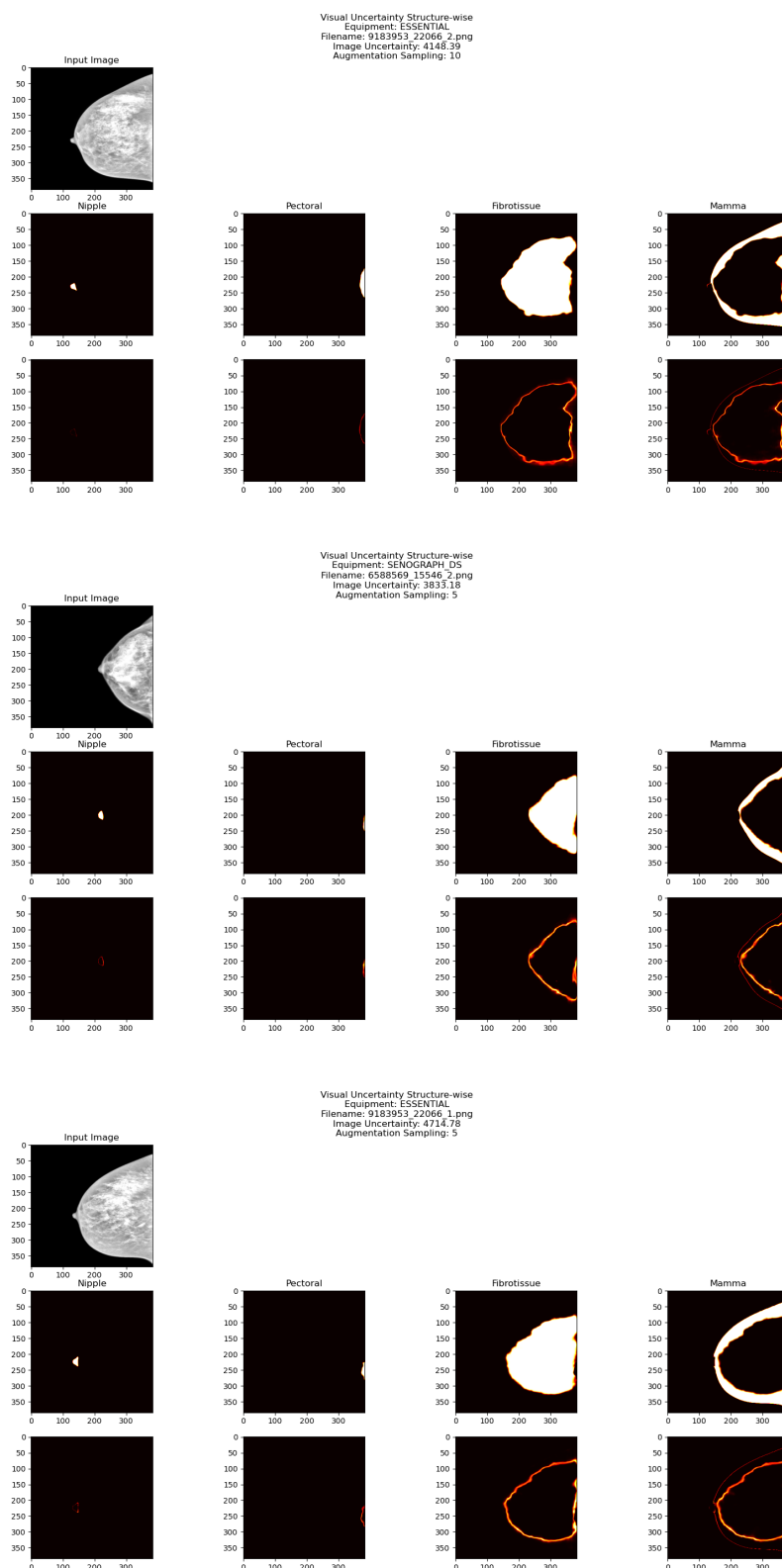


Figura 6.7: Alguns exemplos de mapas de incerteza por estrutura mamária.

A figura 6.7 ilustra a imagem de mamografia a ser segmentada, as predições por cada estrutura anatômica e a incerteza no nível de pixel por cada estrutura mamária, respectivamente, de cima para baixo.

Agora que é possível gerar os mapas de incertezas, estes materiais visuais irão ajudar na construção de pontuações que indicam a incerteza do modelo em relação a uma estrutura mamária específica, para filtrar as imagens mais pertinentes para serem anotadas que o modelo demonstrou mais dúvida ao segmentar.

6.4

Scores de incerteza

Para atribuir e separar as imagens por pontuações de incerteza, deve-se utilizar como insumo as visualizações produzidas na subseção anterior e um conjunto de técnicas para definir os scores de incerteza.

Estes scores serão responsáveis por atribuir uma pontuação de incerteza baseada nas segmentações feitas pelo modelo para cada imagem da pool. Assim é possível separar as imagens que a rede não obteve muito êxito ou demonstrou dúvida no momento da segmentação, indicado pela pontuação. As imagens desafiadoras segmentadas pelo modelo conforme a pontuação e critério estabelecido podem então ser encaminhadas aos anotadores peritos.

O primeiro score de incerteza define-se como o somatório dos valores de entropia no nível de pixel para uma imagem Y , pode ser descrito como:

$$E(Y) = \sum_i^I E_i \quad (6-4)$$

$E(Y)$ define-se como a incerteza no nível de imagem, cuja soma das entropias E_i por pixel resultará em uma pontuação na qual, quanto maior $E(Y)$, maior seria a incerteza do modelo ao segmentar a imagem Y . Logo, as imagens com maior pontuação $E(Y)$ seriam as imagens de mamografia segmentadas com maior incerteza de modo global. No entanto, esta medida teria dificuldades em quantizar a incerteza para mamas de tamanhos diferentes, pois o fato de elas ocuparem mais pixels na imagem poderia majorar a pontuação de incerteza, ou vice-versa. Isso tornaria a métrica menos fidedigna, caso o tamanho das mamas variem muito.

Para contornar tal problema, serão propostos scores de incerteza a seguir que tentarão ser agnósticos ao tamanho das mamas, especializados em quantizar scores pertinentes a estruturas críticas como papila e músculo peitoral, em função das predições e mapas de incerteza adquiridos.

A dimensionalidade dos mapas de segmentação e incerteza das estruturas são em 2D.

Os scores de incerteza específicos de uma estrutura mamária são consolidados por uma conjuntura de medidas de incerteza que permitirá assessorar diferentes casos que divergem de uma segmentação ideal da estrutura mamária específica. Com isto, a partir das medidas obtidas analisando os mapas de incertezas e segmentações feitas pelo modelo, torna-se factível quantizar um score limitado na escala de 0.0 até 1.0. Isso serve como uma quantização da "dúvida" que o modelo teve ao segmentar o exame a ser anotado, onde, quanto mais próximo de 0.0, menor a incerteza e vice-versa.

A quantização dos scores e das medidas são controladas dentro da escala a partir de uma função gaussiana que recebe três parâmetros a seguir:

X , o valor obtido a partir de uma medida específica (e.g., medida de solidez);

$mean$, o valor meio da função gaussiana, a variável que se deseja usar como referência para comparar X com $mean$, e dependendo da distância entre X e $mean$, a medida de incerteza tenderá a 0 ou 1;

δ , parâmetro de ajuste de "penalização" da medida, pode ser relativo à largura de alguma estrutura ou fixo para poder ajustar o score.

Logo, a equação gaussiana base para controlar os scores de incerteza a partir das medidas obtidas pode ser definida como:

$$S(Y) = 1 - \exp\left(-0.5 * \frac{|mean - X|^2}{\delta^2}\right) \quad (6-5)$$

Portanto, a equação demonstrativa 6-5, computa um score de incerteza S contido entre 0.0 a 1.0. Neste caso, quanto mais distante X de $mean$, S tenderá a 1.0, indicando o percentual de incerteza que o modelo obteve ao segmentar um exame Y .

Nas subseções a seguir, esta função será adaptada dependendo da medida e dos critérios de análise definidos para cada caso em particular.

6.4.1

Score de incerteza sobre o músculo peitoral

Para estimar a pontuação de incerteza para o músculo peitoral, utiliza-se o cálculo de cinco medidas baseadas nos mapas de incerteza U_{mp} e a predição do peitoral P_{mp} . Elas podem ser resumidas nos itens abaixo:

1. Análise do comprimento do fecho convexo de P_{mp} com U_{mp} .
2. Análise topológica via número de Euler e componentes conexas sobre P_{mp} .
3. Solidez de P_{mp} .

4. Distância de Hausdorff entre os pontos correspondentes do fecho convexo de P_{mp} e P_{mp} .
5. Excentricidade de P_{mp} .

É considerada uma boa segmentação do peitoral P_{mp} , uma estrutura convexa, contínua, sem ruídos nas bordas, similar à metade de uma elipse e com tamanho razoável, não sendo uma segmentação muito pequena. Já um mapa de incerteza U_{mp} considerado bom, possui características bem similares aos aspectos citados acima, em adição com as temperaturas mais altas situadas exatamente nos contornos aproximados de P_{mp} . Alguns exemplos serão mostrados na figura 6.8, onde a imagem à esquerda, meio e direita representam, respectivamente, a imagem a ser segmentada pelo modelo, o mapa de incerteza do peitoral U_{mp} e a segmentação do peitoral P_{mp} .

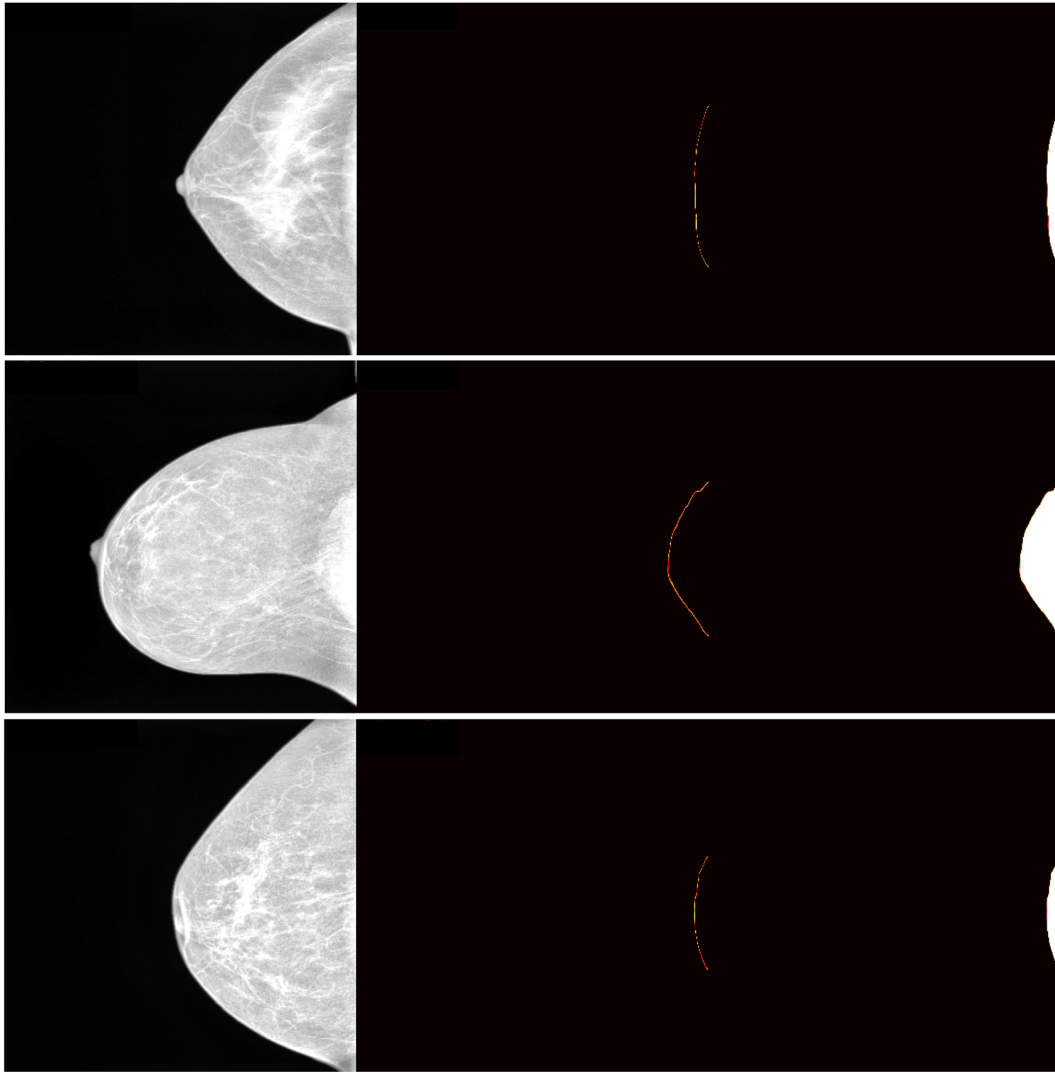


Figura 6.8: Exemplos de segmentações boas e mapas de incertezas bem contidas em relação a predição das imagens de mamografia

Com as considerações acima, a primeira medida de incerteza S_1 a ser calculada consiste na extração do contorno da estrutura segmentada em P_{mp} , que representa uma segmentação ideal, para comparar com as regiões de maior temperatura em U_{mp} . A comparação é feita mediante a diferença dos contornos retirados de P_{mp} e U_{mp} . Caso o comprimento de P_{mp} seja mais próximo de U_{mp} , indica que a segmentação feita sobre o peitoral indica uma incerteza baixa. Quanto maior o comprimento das regiões mais quentes de U_{mp} , em relação ao comprimento do contorno de segmentação de P_{mp} , indica que a segmentação do peitoral foi obtida com alta incerteza.

Para obter a estrutura que represente uma segmentação do peitoral ideal, é necessário estimar o fecho convexo em P_{mp} da estrutura segmentada de maior largura e extrair os pixels que representam a borda do fecho convexo. Assim, é possível adquirir o formato de segmentação ideal do peitoral $C(P_{mp})$ em 1D para estimar o comprimento deste formato conforme a figura 6.9.

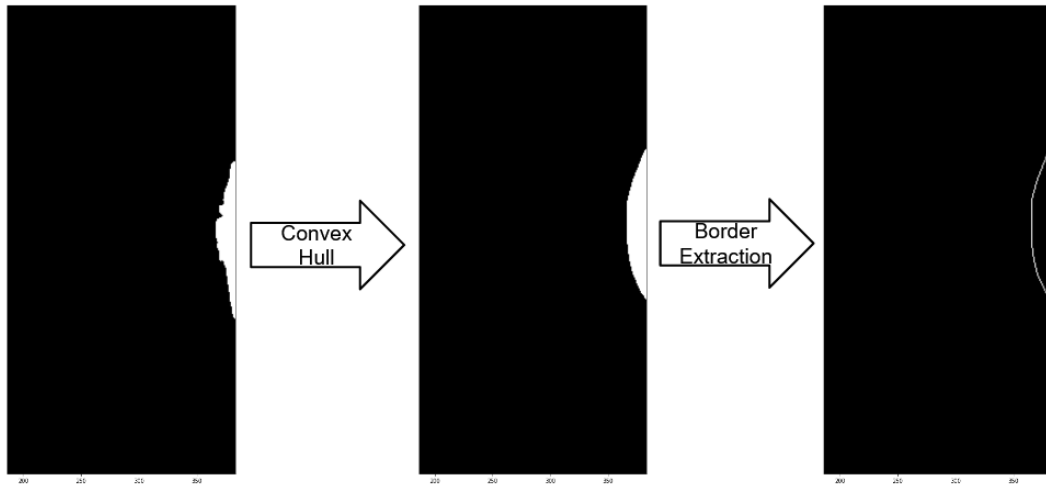


Figura 6.9: Ilustração para obtenção do comprimento de uma segmentação ideal baseado em P_{mp}

A seguir, para obter a estrutura que representa as regiões mais quentes de U_{mp} , realiza-se o procedimento morfológico de *fechamento* (*closing*) em U_{mp} , que consiste em duas etapas, primeiro, a segmentação de U_{mp} será submetida ao processo de *Dilatação*, dilatando os pixels para preencher possíveis pixels escuros, e por segundo, aplicar a operação de *Erosão* para obter a estrutura que representa as regiões mais quentes de U_{mp} . Feito isto, uma última operação de *Thinning* é necessária para compactar as regiões mais intensas em um formato 1D, que consiste em extrair de U_{mp} um eixo medial que representará os pixels mais quentes em 1D que permitirão calcular o comprimento $C(U_{mp})$ conforme ilustrado na figura 6.10.

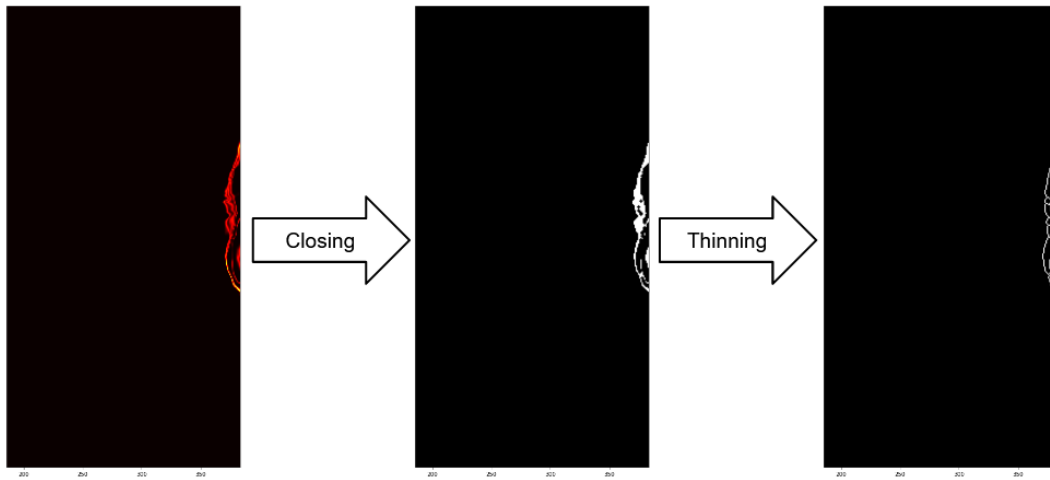


Figura 6.10: Ilustração para obtenção do comprimento das áreas mais quentes em U_{mp}

Logo, para calcular a pontuação de incerteza com base nos critérios acima, pode-se usar uma função de peso gaussiana para restringir o valor de S_1 entre 0 e 1:

$$S_1 = 1 - \exp \left(-0.5 * \frac{||C(P_{mp})| - |C(U_{mp})||^2}{\max(0.3 * |C(P_{mp})|, 10)} \right) \quad (6-6)$$

Mapas de incerteza U_{mp} com características caóticas tendem a aproximar S_1 a 1, indicando que a predição foi feita de maneira duvidosa, por U_{mp} não se assemelhar a um peitoral ideal de P_{mp} através da diferença de comprimento entre $C(P_{mp})$ e $C(U_{mp})$. Quanto maior esta diferença, mais S_1 se aproximará de 1. Caso contrário, S_1 se aproxima de 0, indicando que a segmentação teve menos incerteza, pelo fato do $C(U_{mp})$ estar menos distante do $C(P_{mp})$. A figura 6.11 demonstra um caso onde S_1 tende à 1, indicando incerteza alta por este critério por conta de $C(U_{mp})$ ser consideravelmente distante à $C(P_{mp})$.

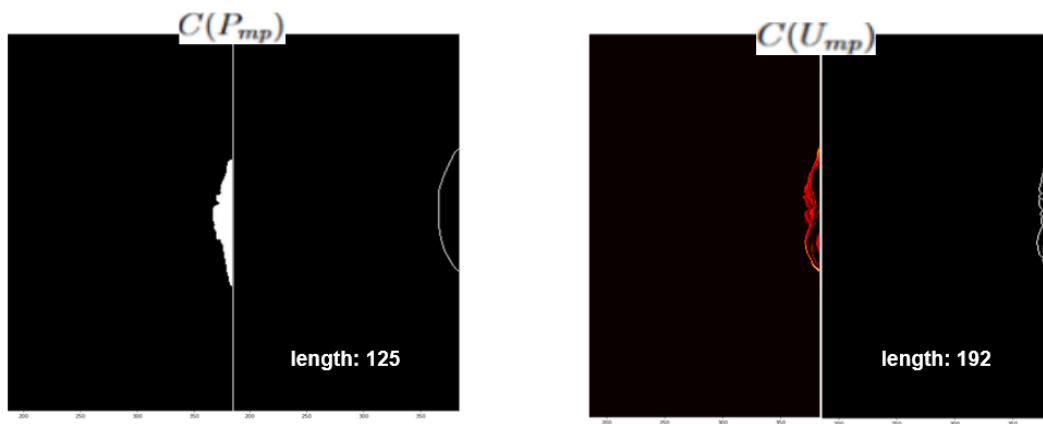


Figura 6.11: Exemplo onde S_1 tende à 1, pelo fato do comprimento $C(U_{mp})$ se distanciar de $C(P_{mp})$.

A segunda medida de incerteza S_2 consiste em realizar uma análise sobre a topologia de P_{mp} , relacionando o número de Euler e componentes conexas na predição do peitoral. Segmentações de peitorais com buracos ou mais de uma componente conexa indicam uma segmentação ruim, visto que uma segmentação ideal consiste de uma estrutura regular e de apenas um único peitoral. Não pode existir mais de um peitoral na predição. Logo, considere a equação abaixo para obtenção da medida, baseada na função 2-5:

$$N_e(P_{mp}) = C - H \quad (6-7)$$

Sendo C a quantidade de componentes conexas e H a quantidade de buracos em P_{mp} . Caso a diferença de $N_e(P_{mp})$ resulte em um valor inteiro diferente de 1 ou a quantidade de componentes for diferente de 1, o score S_2 receberá o valor 1, indicando que o modelo não fez uma segmentação boa. Caso contrário, S_2 é ajustado para 0. A figura 6.12 ilustra um cenário onde a imagem à esquerda é a imagem a ser segmentada pelo modelo, e à direita a máscara de segmentação da mesma, na qual a medida de incerteza S_2 seria exatamente 1, verificando que existem dois componentes conexos ao invés de um.

A terceira medida de incerteza S_3 consiste em calcular a solidez baseada na área da segmentação de $A(P_{mp})$ com a área do fecho convexo $A(F_c(P_{mp}))$ da mesma. Adaptando a equação 2-3, obtém-se:

$$S_p(P_{mp}, F_c) = \frac{A(P_{mp})}{A(F_c)} \quad (6-8)$$

Logo, S_3 assume valor próximo a 0 indicando incerteza baixa, caso a solidez S_p dê valores aproximados a 1 e vice-versa. A solidez é útil neste caso para avaliar se o peitoral possui irregularidades na sua segmentação, como predição ruidosa.

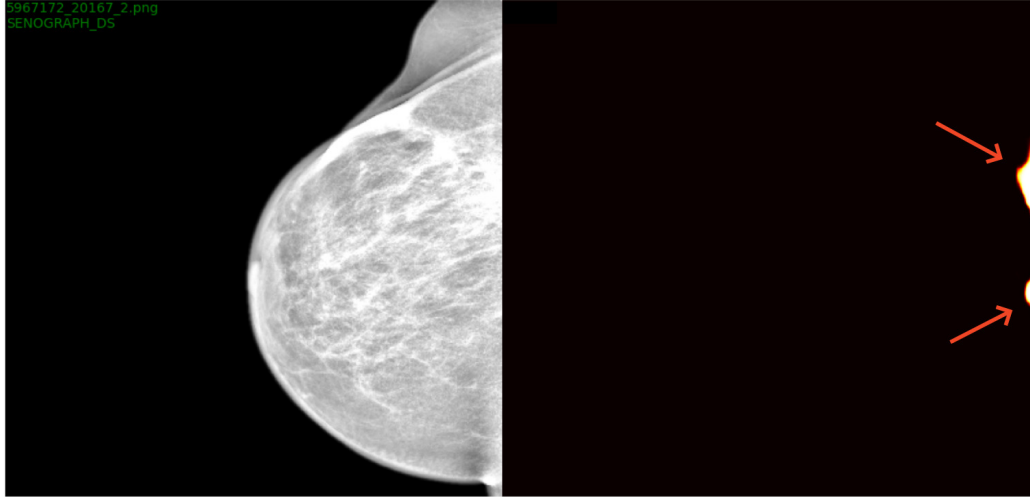


Figura 6.12: Exemplo onde a análise topológica aborda duas segmentações de peitoral

A medida tende a penalizar a incerteza, aproximando S_3 a 1, caso a segmentação se diferencie de um formato convexo ou tenha irregularidades. S_3 irá tender à 0, caso contrário. A figura 6.13 exemplifica um caso onde a área S_p é calculada em relação a $A(P_{mp})$ e $A(F_c)$, representado pelo contorno rosa, onde a medida irá tender à 1 por conta dos vãos existentes dentro do fecho convexo, indicando solidez baixa. Para obtenção do score S_3 , foi utilizada a seguinte equação que normaliza a pontuação em valores de 0 a 1:

$$S_3 = 1 - \exp\left(-0.5 * \frac{(|1 - S_p|)^2}{0.1^2}\right) \quad (6-9)$$

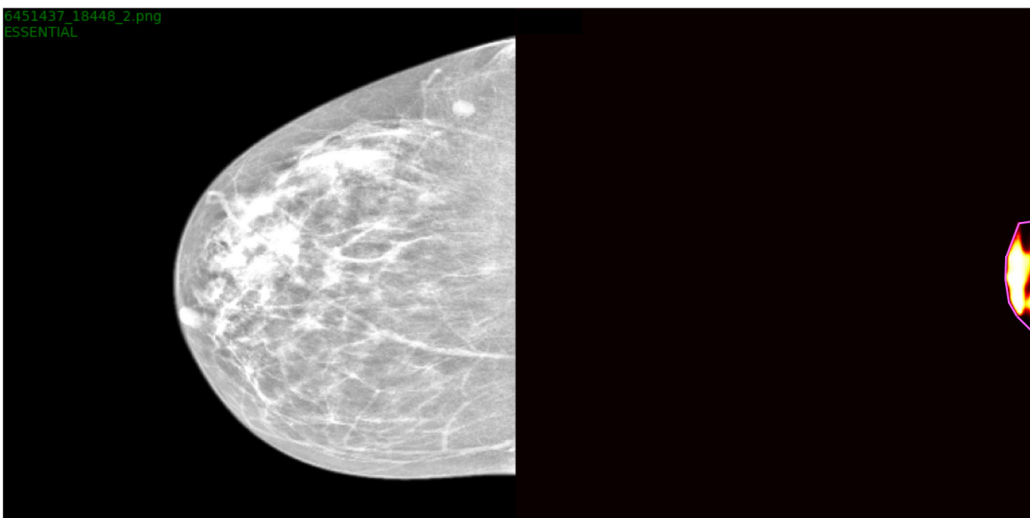


Figura 6.13: Exemplo do processo de cálculo da medida de incerteza baseada na solidez.

A quarta medida de incerteza S_4 se baseia na Distância de Hausdorff entre os pontos correspondentes do fecho convexo F_c de P_{mp} e da segmentação de P_{mp} . Aplicando a equação 2-2 para estimar S_4 obtém-se:

$$D_h(P_x, P_y) = \max\{D(x, y), D(y, x)\} \quad (6-10)$$

De maneira análoga, considera-se P_x como o conjunto de pontos de P_{mp} e P_y o conjunto de pontos do F_c retirados de P_{mp} . A ideia consiste em fazer uma comparação em pares entre P_x e P_y para obter a maior distância correspondente entre os pontos. Desta maneira, seria possível detectar alguns casos de contornos em que a segmentação P_{mp} é constituída de vales e a similaridade a partir da distância de P_{mp} com um peitoral ideal.

Portanto, S_4 se aproxima de 1, indicando incerteza alta, caso D_h resulte numa distância longa. Caso contrário, S_4 irá tender a 0, onde D_h é mínima. Na figura 6.14 é possível observar uma distância considerável entre os pares de pontos que representam a maior distância entre P_x e P_y , e como D_h é suficientemente grande. Este exemplo majoraria a medida incerteza S_4 , pois, os pares de pontos que representam a maior distância entre o peitoral ideal (fecho convexo) e a segmentação de P_{mp} é relevante, evidenciando uma segmentação bem acidentada e incompleta. À vista disso, a equação para obter S_4 , consiste de:

$$S_4 = 1 - \exp\left(-0.5 * \frac{(|0| - |D_h(P_x, P_y)|)^2}{(P_{mp} * 0.012)^2}\right) \quad (6-11)$$

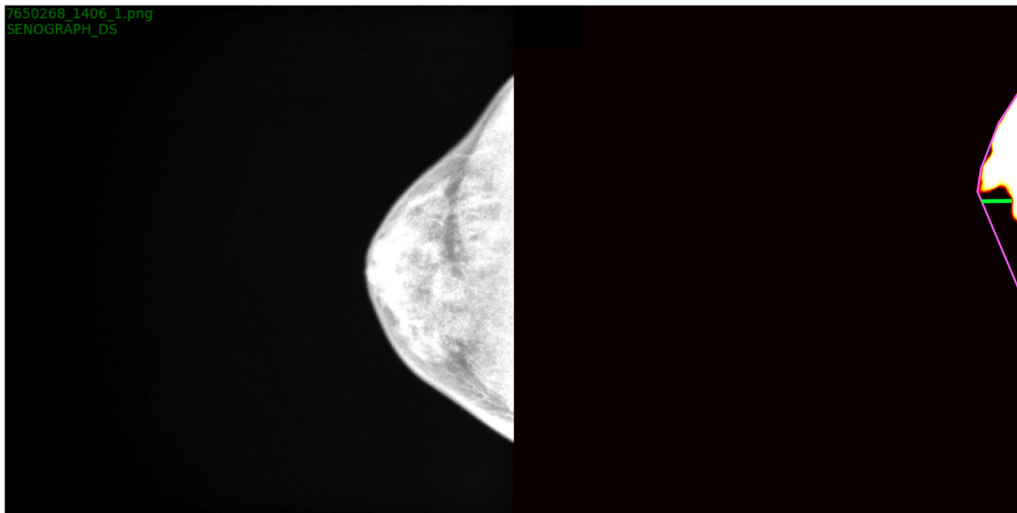


Figura 6.14: Distância de Hausdorff representada pela reta em verde, indicando a maior distância entre os pontos correspondentes em P_x e P_y

A quinta medida de incerteza S_5 se baseia na avaliação da excentricidade de P_{mp} . A excentricidade pode ser calculada pela equação 2-4, adaptada desta

maneira:

$$E_{cc}(P_{mp}) = \frac{D(E_b)}{D(E_a)} \quad (6-12)$$

Logo, $D(E_a)$ consiste na distância do eixo maior, como se fosse a maior reta possível dentro de uma elipse que contem P_{mp} nos seus limites, já $D(E_b)$ consiste na distância do eixo menor, a maior reta possível que seja perpendicular à $D(E_a)$. Obtendo $D(E_a)$ e $D(E_b)$, a razão entre os dois retorna a excentricidade de P_{mp} , na qual, quanto maior seu valor, menor será S_5 e vice versa. O critério acima se demonstra importante, pois nenhuma das medidas citadas acima tenta detectar a incerteza de peitorais minúsculos, de até 15 pixels de tamanho, por exemplo. Por fim, caso a excentricidade de P_{mp} seja alta, a incerteza S_5 recebe um valor próximo a 0 e vice-versa. Portanto, a equação que representa S_5 é definida como:

$$S_5 = \exp\left(-0.5 * \frac{(|0| - |E_{cc}(P_{mp})|)^2}{0.3^2}\right) \quad (6-13)$$

Um exemplo onde a excentricidade é baixa, aumentando a medida de incerteza S_5 , é observada na figura 6.15, onde a segmentação P_{mp} deduziria uma razão entre os eixos próximos a 1, indicando incerteza alta pelo fato do peitoral segmentado ser muito pequeno.

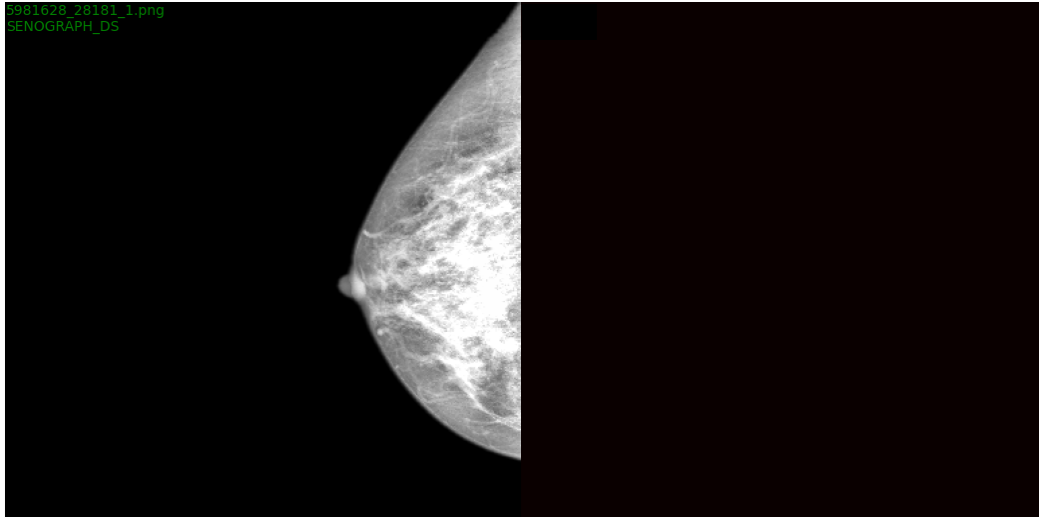


Figura 6.15: Caso onde a segmentação P_{mp} possui uma excentricidade baixa.

Todos as medidas de incerteza computadas acima são normalizadas para valores entre 0 e 1 através de uma função gaussiana que normaliza os scores de incerteza calculados acima.

No entanto, para atribuir um score de incerteza do peitoral para cada predição feita pelo modelo, é necessário uma medida global para o peitoral, cujo valor pode ser obtido via soma das medidas de incerteza citadas acima:

$$S_{peitoral}(Y) = w_1 * S_1 + w_2 * (w_3 * S_2 + w_4 * S_3 + w_5 * S_4 + w_6 * S_5) \quad (6-14)$$

Onde Y consiste na imagem feita a predição, w_n como parâmetro de peso da medida, S_n cada medida de incerteza e o $S_{peitoral}$ a pontuação de incerteza do peitoral normalizada entre 0 e 1. Os pesos w_1 e w_2 são iguais a 0.5, e os pesos $\{w_3, w_4, \dots, w_6\}$ são 0.25, mantendo equilíbrio entre os pesos, sem penalizar ou diminuir tolerância entre as medidas S_n .

Um caso de contorno que pode ser observado, consiste em estimar o score de incerteza $S_{peitoral}(Y)$ para segmentações de peitoral muito pequenas ou nulas. Para estes casos, onde a predição do peitoral é inexistente ou é composta de poucos pixels, $S_{peitoral}(Y)$ é forçado a assumir o valor -1, pois como não há informação de predição, as medidas de incerteza teriam dificuldades em serem calculadas por conta da falta de uma informação vital, principalmente, as medidas que dependem exclusivamente de P_{mp} . Logo, do ponto de vista da segmentação, o peitoral é uma estrutura que pode estar presente na imagem ou não, tornando a quantificação da incerteza para esta estrutura mamária bem peculiar para estes casos.

6.4.2

Score de incerteza sobre a papila

A estratégia proposta para estimar uma pontuação de incerteza para predições da papila, consiste em calcular três medidas. Mas, antes, conceitua-se em relação a papila, o mapa de predição como P_p e o mapa de incerteza U_p .

1. Comparação da distância entre o mapa de segmentação e o mapa de incerteza da papila.
2. Análise topológica via número de Euler e componentes conexas sobre P_p
3. Solidez de P_p .

Uma segmentação de papila boa P_p , consiste em um formato próximo a de um círculo, com poucas deformidades quanto sua forma, sem buracos e como uma única componente conexa. O mapa U_p de baixa incerteza consiste em ser similar a P_p , ou seja, as regiões mais quentes da incerteza de U_p se situarem exatamente próximas nas bordas de segmentação de P_p e temperaturas altas no contorno de U_p em menor tamanho possível. A figura 6.16 demonstra casos de segmentação e respectivos mapas de incerteza considerados bons. A

representação das imagens possui a mesma ordem como foi definida em relação à figura 6.8.

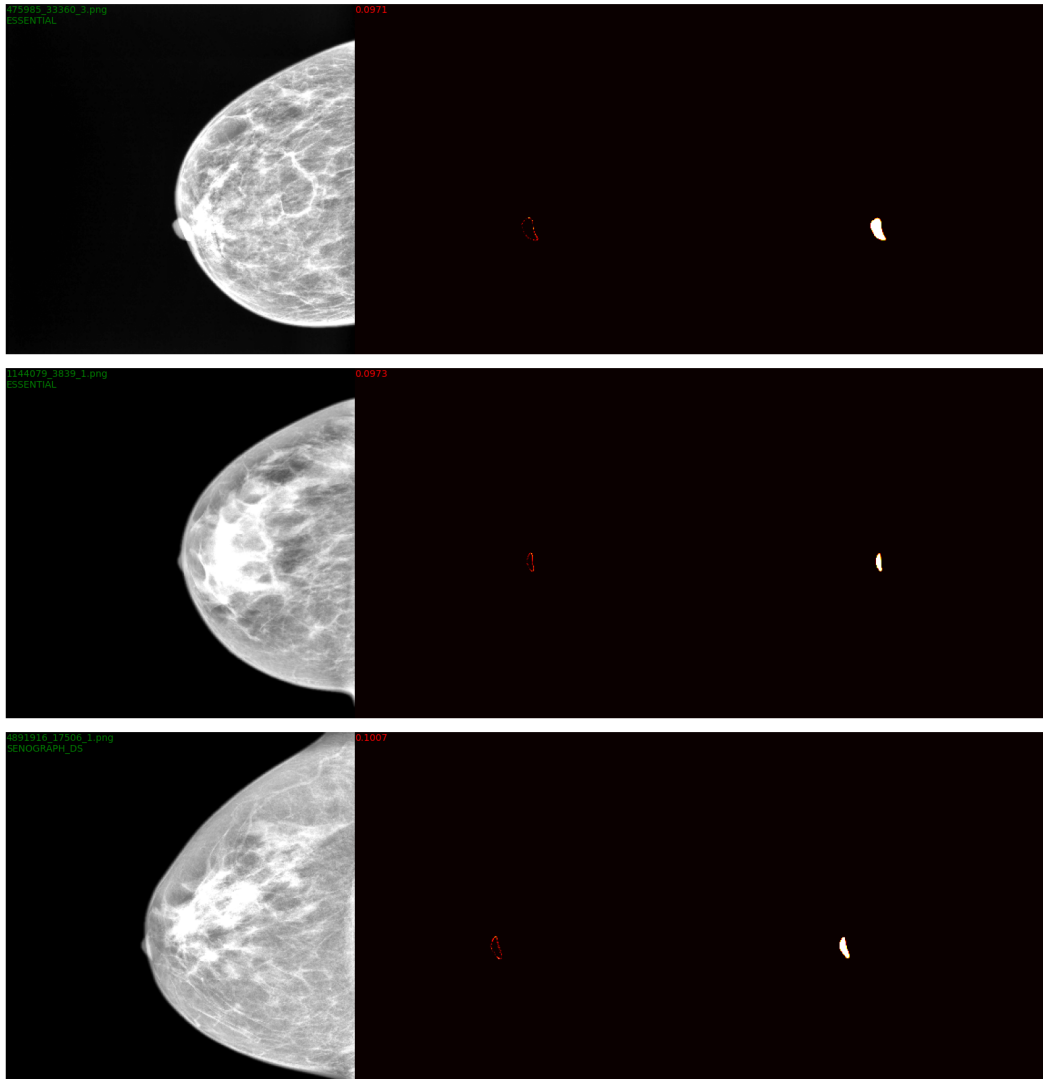


Figura 6.16: Exemplos de segmentações boas e mapas de incertezas bem contidas em relação às suas respectivas imagens

Para computar a pontuação de incerteza da papila, o desvio das boas características em P_p e U_p nos retorna indicativos de incerteza que podem ser calculados a partir de análises topológicas, solidez e distância entre P_p e U_p .

A proposta consiste em computar três medidas normalizadas que possibilitem a construção de uma pontuação de incerteza S_{papila} para as papilas preditas pelo modelo.

A primeira medida de incerteza S_1 baseia-se em extrair de U_p a contagem dos pixels que estão distantes de P_p . Considera-se uma boa segmentação da papila casos onde U_p está bem próximo à P_p na mesma região. Inicialmente, o procedimento de *fechamento* é aplicado em U_p para manter somente as regiões mais quentes. A seguir, P_p se sobrepõe à U_p , e o procedimento de *dilatação*

é utilizado em P_p para inflar os pixels e suprimir os pixels vizinhos de U_p pertos da segmentação P_p . Com isto, é possível calcular a área $A(U_p)$ via contagem dos pixels restantes de U_p que se encontram distantes da região de segmentação de P_p . Logo, quanto maior a área calculada, maior S_1 será e vice-versa. Esta medida é importante como uma tentativa de avaliar o quão distante uma segmentação em P_p se encontra do mapa de incerteza U_p . A estratégia pode ser elucidada a partir da figura 6.17, onde é demonstrado um exemplo na qual S_1 tenderá a 1 pelo fato dos pixels mais quentes em U_p , na figura à direita, situarem-se distantes da região segmentada em P_p destacada na imagem do meio.

$$S_1 = 1 - \exp\left(-0.5 * \frac{||0| - |A(U_p)||^2}{T^2}\right) \quad (6-15)$$

Onde T é a tolerância da área do mapa de incerteza U_p , cujo o valor é 5.

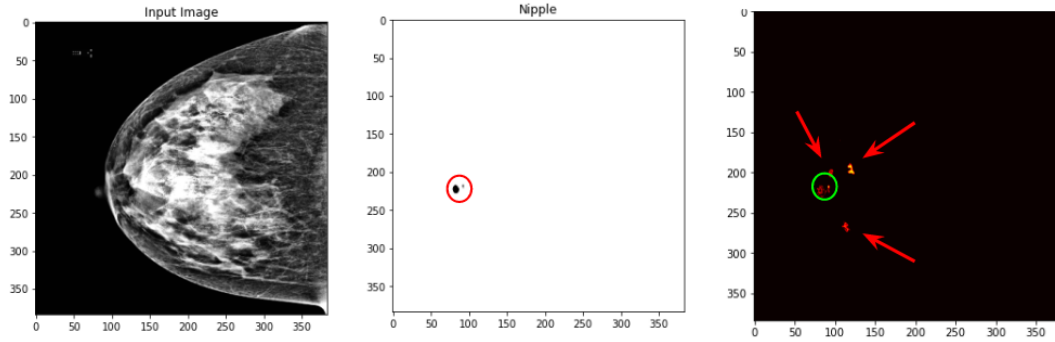


Figura 6.17: Exemplo onde S_1 tende a 1, por conta das regiões mais quentes obtidas em U_p se localizarem distantes da região de segmentação em P_p

A segunda medida de incerteza S_2 consiste em contar a quantidade de buracos e componentes conexas a partir da análise topológica do número de Euler em P_p . Pois, uma das características da predição da papila consiste em ser uma única componente conexa. Logo, é feita uma análise sobre quantos buracos e componentes conexas existem em P_p . Caso 2-5 diferir de 1, S_2 será 1, indicando alta incerteza de predição, caso a fórmula resultar exatamente em 1, S_2 será 0, designando que a partir da topologia da papila, a predição consiste de uma única papila e sem buracos, atendendo um dos critérios de segmentação boa da papila. A figura 6.18 demonstra um exemplo onde a medida de incerteza S_2 é 1 por conta da segmentação possuir duas componentes conexas ao invés de uma, sendo a imagem à esquerda, o exame a ser segmentado pelo modelo, e à direita, a segmentação da papila.

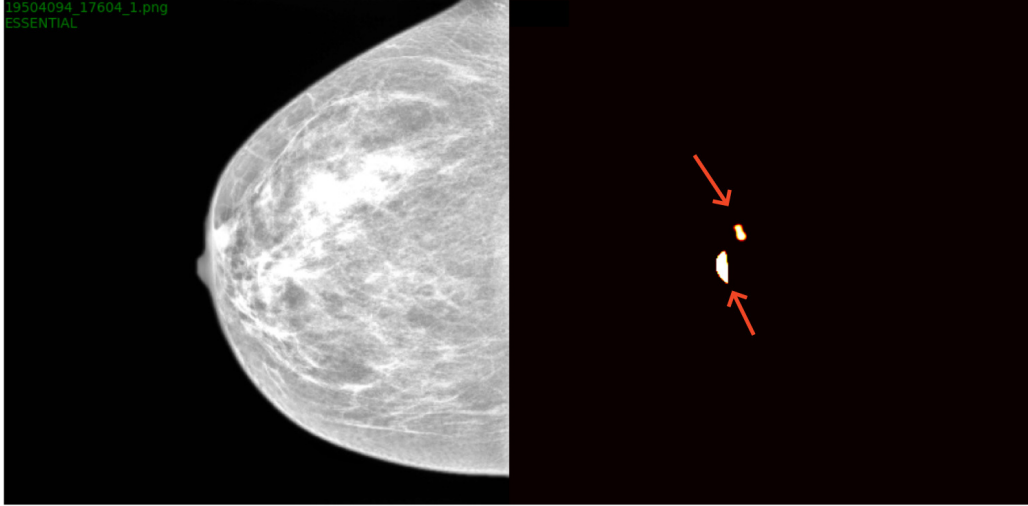


Figura 6.18: Exemplo onde observa-se duas componentes conexas em P_p .

A terceira medida de incerteza S_3 baseia-se a partir da equação 2-3. A pontuação é calculada a partir da razão entre área da predição $A(P_p)$ e a área do fecho convexo $A(F_c)$ que contorna P_p , conforme a equação:

$$S_{ppl}(P_p, F_c) = \frac{A(P_p)}{A(F_c)} \quad (6-16)$$

A medida tende a aumentar caso P_p se diferir de uma estrutura convexa em sua predição, pois um dos critérios de segmentação boa sobre a papila consiste de uma região convexa livre de lacunas internas ou várias componentes conexas. Logo, S_3 pode ser formulado de tal maneira:

$$S_3 = 1 - \exp\left(-0.5 * \frac{||1 - |S_{ppl}(P_p, F_c)||^2}{0.5^2}\right) \quad (6-17)$$

Para um melhor entendimento, a figura 6.19 ilustra o fecho convexo de P_p , aproximado pelo contorno rosa e no seu interior, uma lacuna grande entre as duas segmentações de papila, resultando em uma solidez significativamente baixa, aumentando a incerteza S_3 .

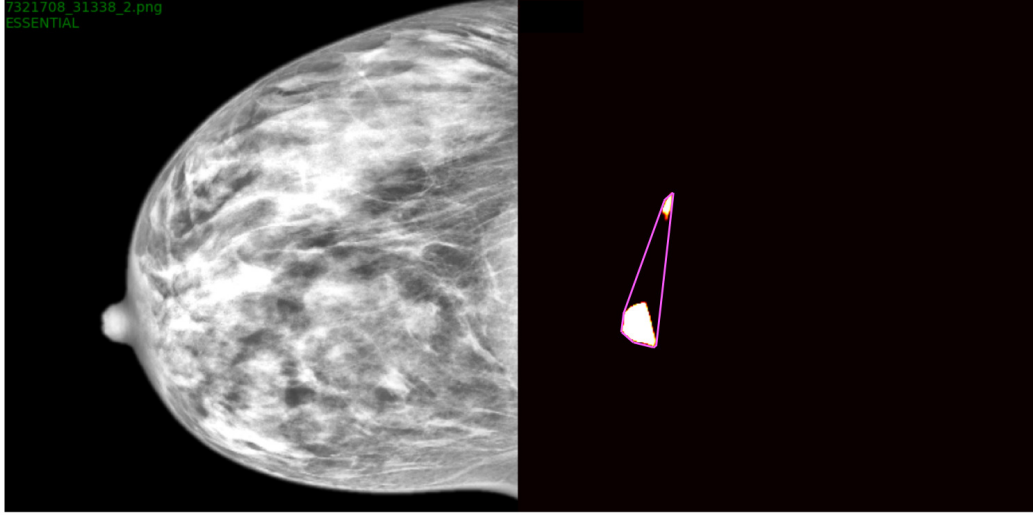


Figura 6.19: Um possível caso onde a solidez calculada é baixa, aumentando a medida de incerteza S_3 .

Assim como a normalização foi feita para cada medida de incerteza computada para o score do peitoral. A normalização usando função gaussiana é aplicada também para cada medida de incerteza relacionada à papila com valores entre 0 a 1, exceto para medidas de análise topológica.

Da mesma maneira, para que as predições feitas pelo modelo possam ser pontuadas e identificar as predições mais duvidosas, um score de incerteza global da papila, reunindo todas as três medidas acima, é designado via média sobre as medidas de incerteza da papila normalizadas entre 0 a 1.

$$S_{papila}(Y) = \frac{1}{n} \sum_{n=1}^n S_n \quad (6-18)$$

Onde Y consiste na imagem feita a predição, n o número de medidas, S_n n -ésima medida de incerteza e S_{papila} a pontuação de incerteza do papila que assume valores de 0 a 1.

Do ponto de vista de segmentação, casos onde de fato não existe papila ou modelo faz uma segmentação muito pequena, abaixo de um tamanho permitido, S_{papila} é calculado mediante a seguinte equação:

$$S_{papila}(Y) = 0.5 + 0.5 * (1 - \exp\left(-0.5 * \frac{||0| - |P(U_p)||^2}{T^2}\right)) \quad (6-19)$$

Na qual $P(U_p)$ é a quantidade de pixels cuja entropia difere de 0 em U_p após procedimento de *fechamento*.

A equação 6-19 garante que este caso seja revisado pelos anotadores em algum momento durante os ciclos de active learning, apesar de ser rara a não aparição da papila em exames de mamografia.

Agora que é possível estimar pontuações de incertezas pertinentes a segmentações de estruturas mamárias como peitoral e papila, torna-se factível escolher as imagens com predições duvidosas feitas pelo modelo de segmentação, encaminhar aos anotadores para etiquetagem, adicioná-las ao dataset da rede, retreinar o modelo e assim repetir o ciclo até que uma condição seja alcançada. Com esta técnica adaptada, espera-se que o modelo se aperfeiçoe, priorizando a acurácia de alguma estrutura mamária a partir da pontuação de incerteza escolhida, reduzindo a quantidade de dados a serem anotados.

6.5 Resultados

Nesta seção, elucidam-se os experimentos de active learning realizados para obter os resultados da abordagem.

Para demonstrar os resultados obtidos a partir de active learning, será esclarecido o setup dos experimentos que envolveram o uso da abordagem.

6.5.1 Configuração dos experimentos

O conjunto de dados não anotados (pool) somam exatamente 4863 imagens de mamografia. Para dar início ao experimento, foram subtraídas da pool, 300 imagens previamente anotadas, restando 4563 dados na pool, para criar o dataset inicial com propósito de treinar e habilitar o modelo de segmentação semântica de fazer as inferências sobre os dados restantes da pool. Para gerar as métricas de avaliação do modelo por cada ciclo da abordagem, foi gerado um dataset de validação com 170 imagens aleatórias à parte da pool. Este conjunto de validação permanece imutável e é mantido para todos os ciclos e experimentos realizados por esta seção para que não haja viés sobre os resultados de cada rodada de active learning. Portanto, as métricas obtidas por experimento são mediante o conjunto de validação. Sobre a quantidade de dados que é subtraída da pool para ser adicionada ao dataset a cada iteração da abordagem, é regrado via algum critério definido sobre os scores de incerteza computados mediante a predição dos dados da pool, mantendo-se fixo em 150 imagens. Por exemplo, selecionar da pool as 150 imagens que tiveram maior incerteza de papila. Em relação à quantidade de imagens geradas via TTA para aquisição dos mapas de incerteza por cada imagem da pool, foi definida para 5, utilizando a técnica de ajuste de gama na imagem cujos valores podem assumir randomicamente entre $[0.6, 1.4]$ entre cada amostra de TTA.

Para poder comparar os resultados com um modelo ideal, foi treinado um modelo com todos os dados da pool anotados e as métricas de avaliação da

rede foram computadas mediante o dataset de validação. Assim será possível ter um limite superior para comparar métricas de avaliação a cada iteração, de um modelo cujos dados adicionados no dataset são randômicos e um modelo que utiliza o critério de seleção mediante os scores de incerteza definidos no experimento para alimentar o dataset.

Os resultados são obtidos pela técnica proposta na figura 6.1 e cada subseção a seguir apresentará o experimento com o critério utilizado para seleção dos dados informativos que possivelmente o IoU da estrutura cresça, com base no critério de escolha dos dados pelo score de incerteza da estrutura em particular. Ou seja, caso o critério se baseie na incerteza da papila, espera-se que a cada ciclo da abordagem, o IoU da papila aumente ou se mantenha estável em relação ao limite superior e do modelo treinado com dados randômicos a cada rodada.

A seguir, os experimentos foram batizados para melhor identificação.

Para resumir, foram definidos para os experimentos AL-Random, AL-Papila, AL-Peitoral e AL-Pec-Npl o tamanho inicial do dataset em 300 imagens e a cada ciclo da abordagem, a partir de um critério de seleção, 150 imagens mais informativas da pool foram adicionadas ao dataset.

6.5.2

Experimento AL-Random

Este experimento consiste em obter os resultados de IoU para cada ciclo de active learning e comparar com os resultados de outros experimentos.

O critério de seleção dos dados a serem subtraídos da pool e adicionados ao dataset para cada iteração do experimento é puramente randômico.

As figuras 6.20, 6.21 e 6.22 ilustram os IoUs obtidos.

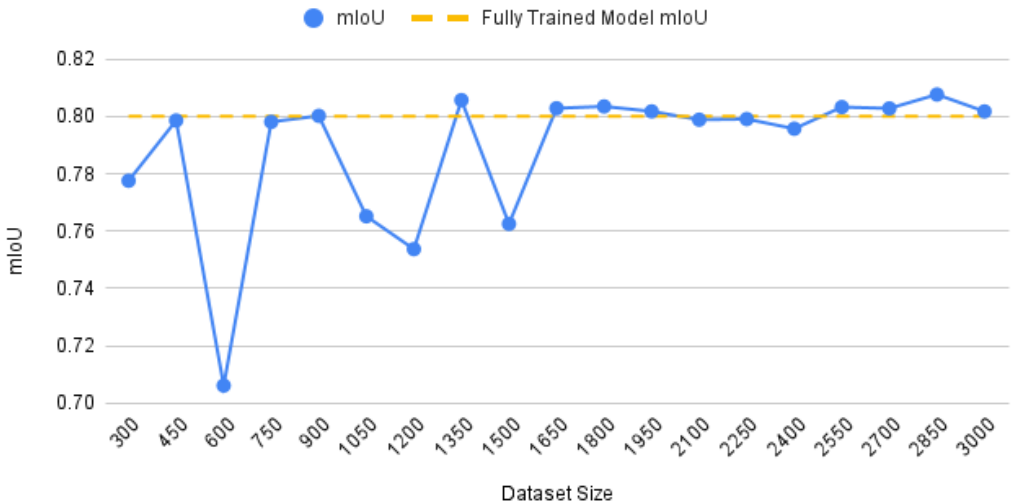


Figura 6.20: O IoU médio de todas as estruturas mamárias por ciclo.

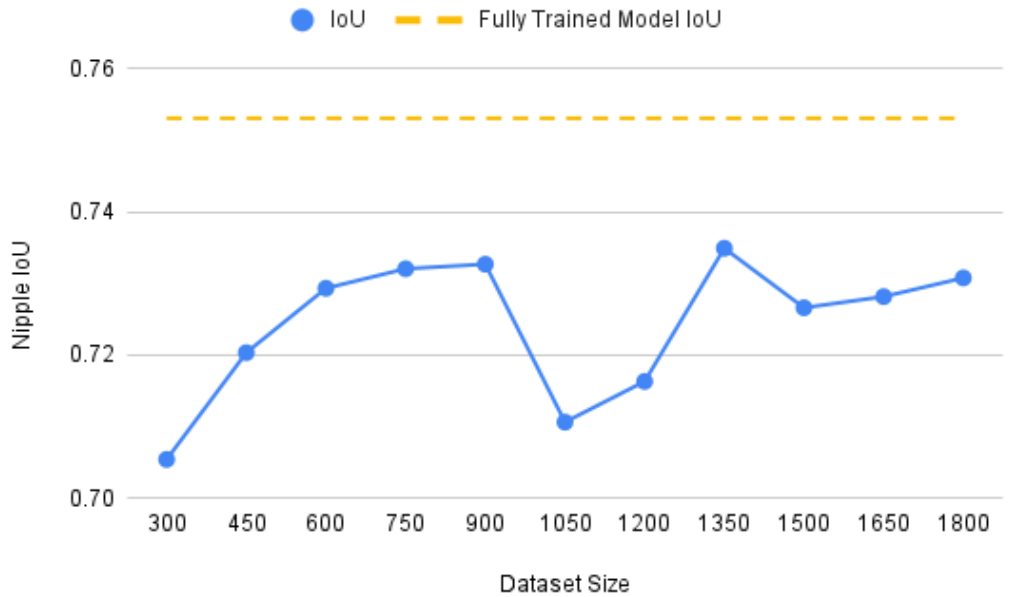


Figura 6.21: O IoU exclusivo da estrutura mamária papila por ciclo.

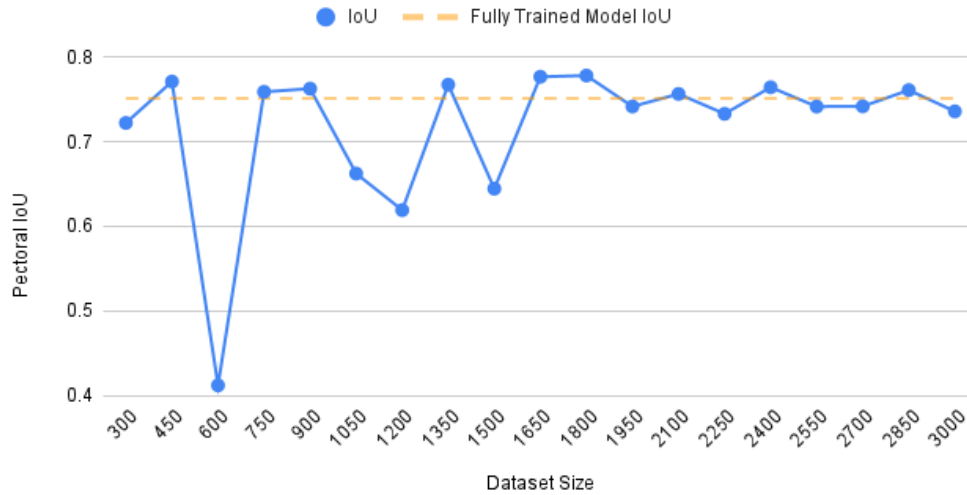


Figura 6.22: O IoU exclusivo da estrutura mamária músculo peitoral por ciclo.

6.5.3

Experimento AL-Papila

O critério de seleção dos dados a serem subtraídos da pool e adicionados ao dataset para cada iteração deste experimento, consiste em selecionar as 150 imagens de mamografia que o modelo teve mais dúvida com base no score de incerteza da papila S_{papila} . Ao ordenar pela pontuação escolhida acima de modo decrescente, garante-se a obtenção das 150 imagens.

As figuras 6.23 e 6.24 demonstram os resultados obtidos neste experimento.



Figura 6.23: O IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via S_{papila} .

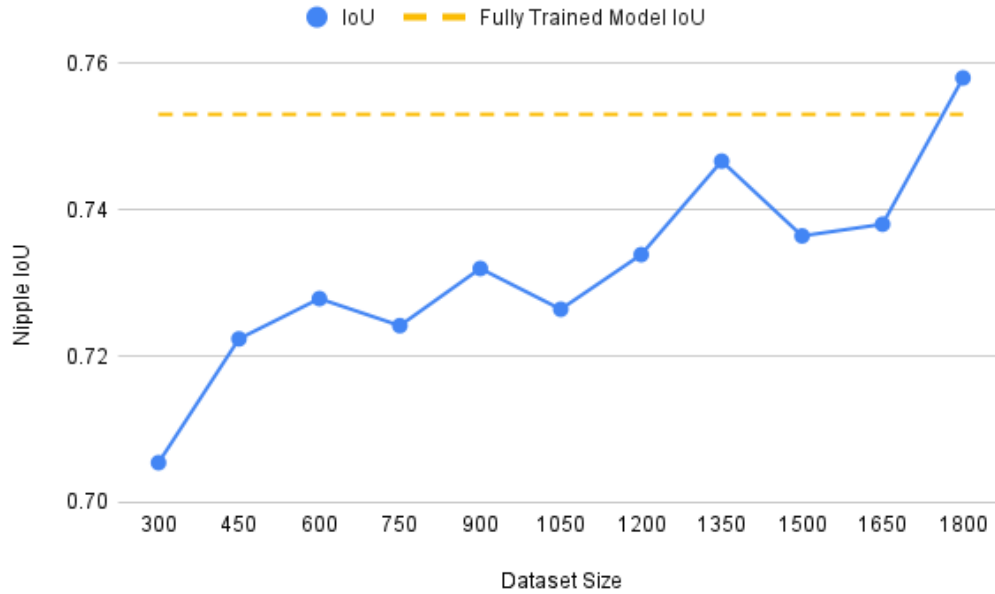


Figura 6.24: O IoU exclusivo da estrutura mamária papila por ciclo, selecionando as imagens mais incertas via S_{papila} .

6.5.4 Experimento AL-Peitoral

Para este experimento, o critério consiste em selecionar as 150 imagens de mamografia em que o modelo teve mais dúvida com base no score de incerteza do músculo peitoral $S_{peitoral}$. Ao computar todas as pontuações de incerteza da pool, os dados são novamente ordenados de modo decrescente e selecionados os 150 mais incertos. As figuras 6.25 e 6.26 ilustram os resultados.

6.5.5 Experimento AL-Pec-Npl

Neste experimento, o critério consiste em selecionar as 150 imagens de mamografia que o modelo teve mais dúvida com base no score de incerteza do músculo peitoral $S_{peitoral}$ e papila S_{papila} . Para isso, parcelam-se as 150 imagens selecionadas em 75 imagens com maior $S_{peitoral}$ e outras 75 imagens de maior S_{papila} . Ao computar todas as pontuações de incerteza da pool, os dados são ordenados de modo decrescente duas vezes, no primeiro momento em relação a $S_{peitoral}$ e outra sobre S_{papila} .

As figuras 6.27 a 6.29 ilustram os resultados.

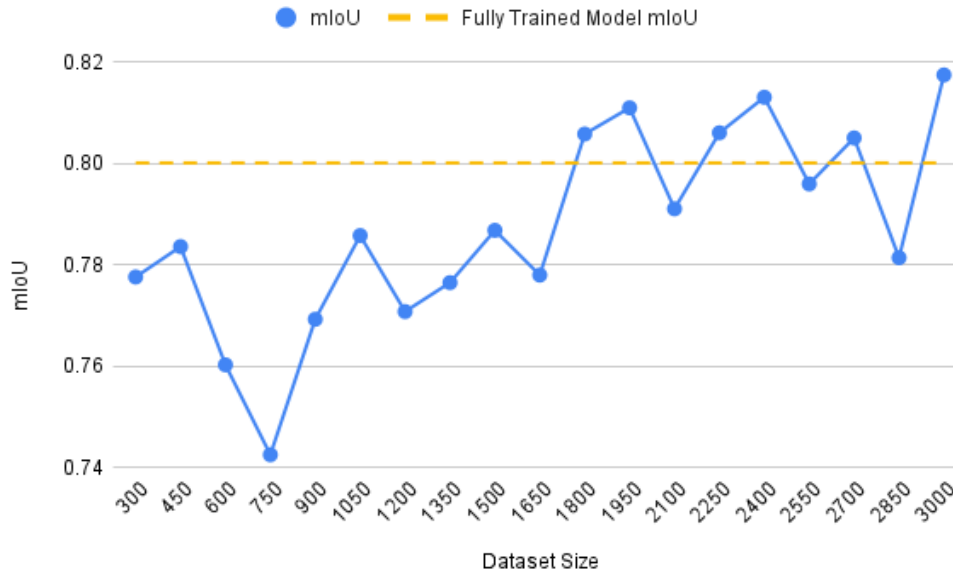


Figura 6.25: IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$.

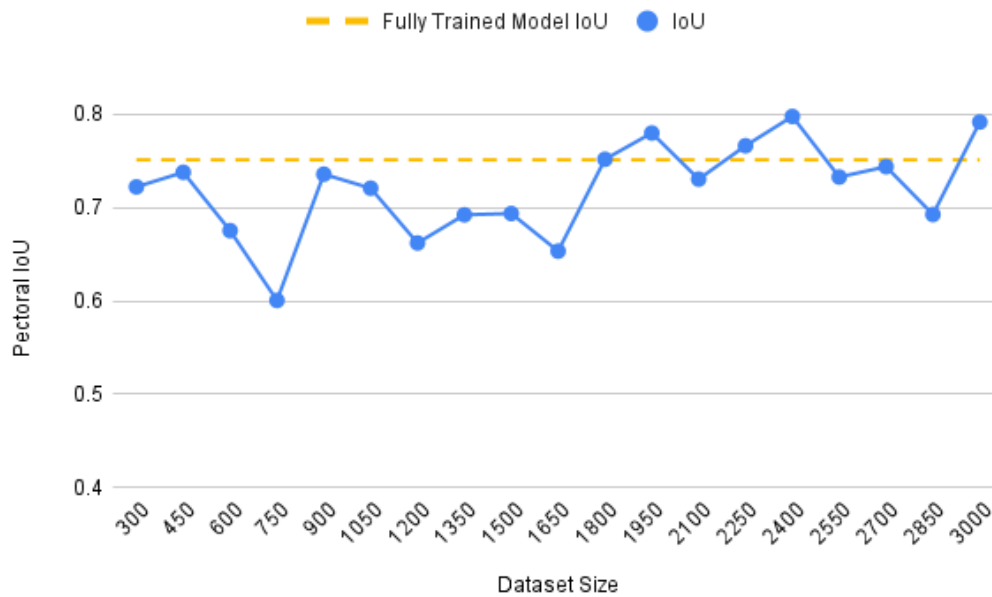


Figura 6.26: O IoU exclusivo da estrutura mamária peitoral por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$.

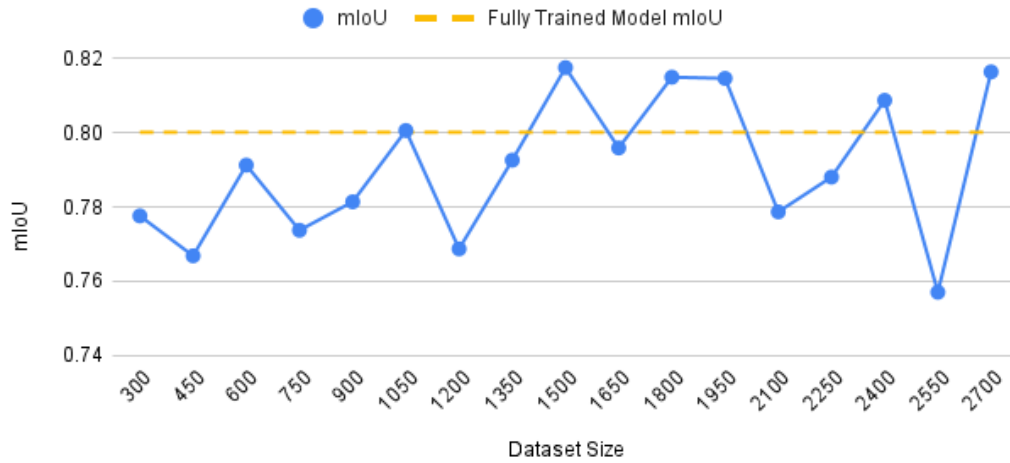


Figura 6.27: IoU médio de todas as estruturas mamárias por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} .

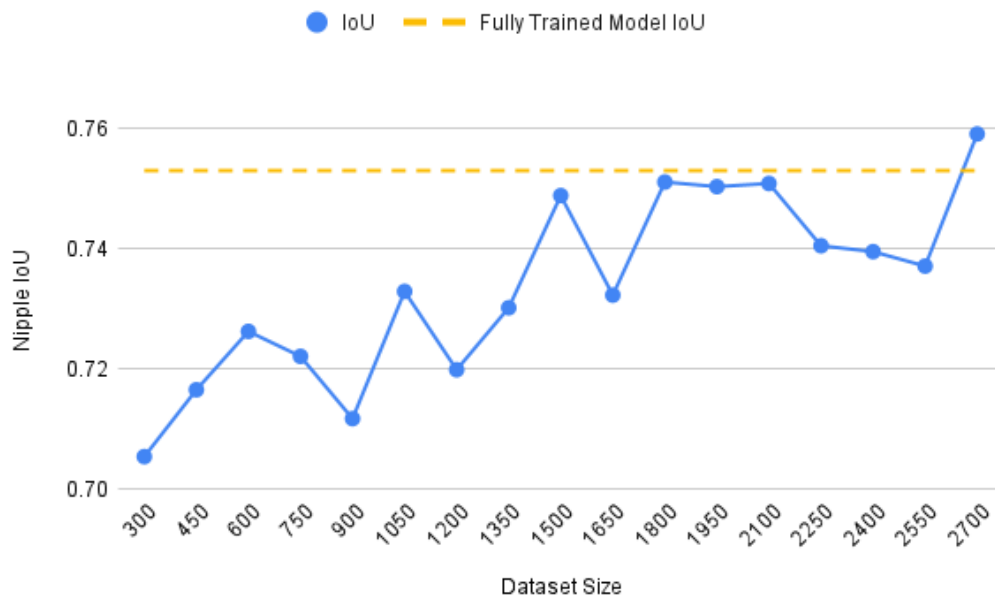


Figura 6.28: IoU exclusivo da estrutura mamária papila por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} .

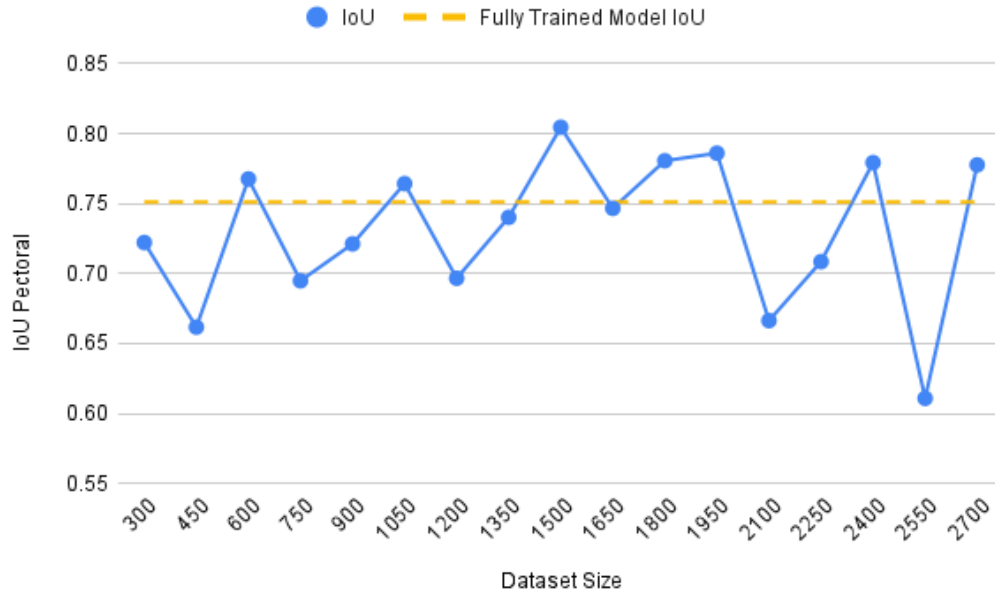


Figura 6.29: IoU exclusivo da estrutura mamária músculo peitoral por ciclo, selecionando as imagens mais incertas via $S_{peitoral}$ e S_{papila} .

6.6

Combinação dos resultados

As figuras de 6.30 a 6.33 mostram as comparações das IoUs da papila e músculo peitoral entre os experimentos anteriores.

Most Uncertain 150 Images by Nipple Score

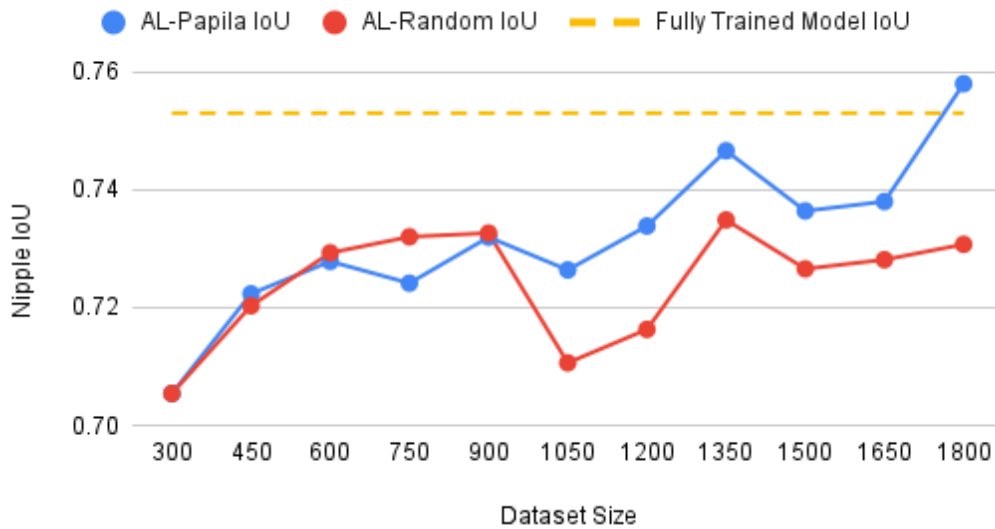


Figura 6.30: Comparação do IoU da estrutura papila entre os experimentos AL-Papila e AL-Random.

Most Uncertain 150 Images by Pectoral Score

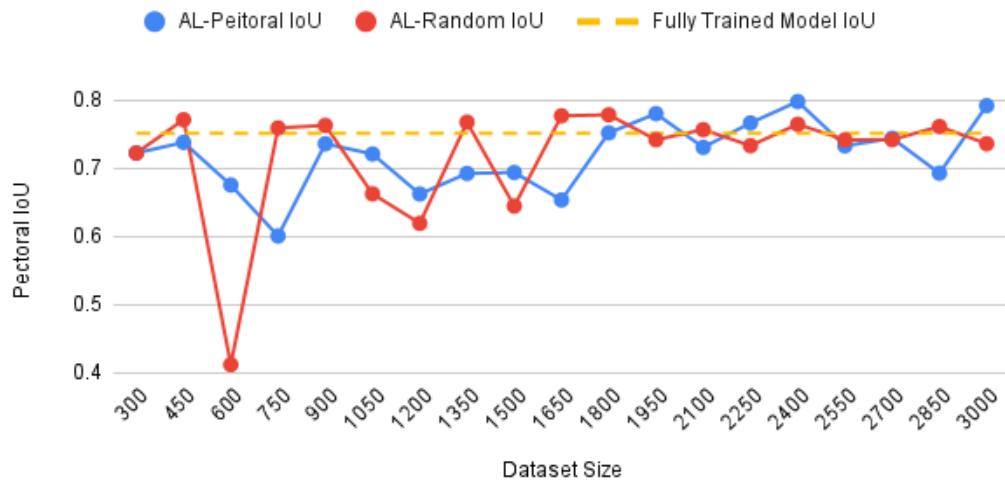


Figura 6.31: Comparação do IoU da estrutura músculo peitoral entre os experimentos AL-Pectoral e AL-Random.

Most Uncertain 150 images by Both Scores

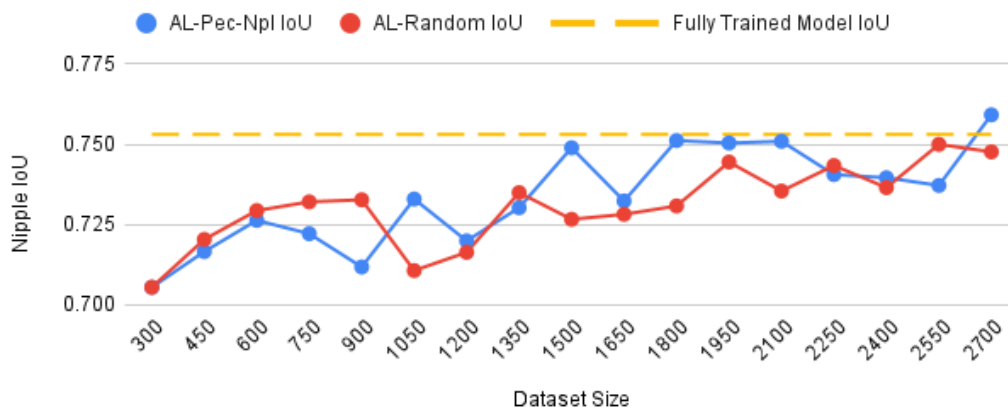


Figura 6.32: Comparação da IoU da estrutura papila entre os experimentos AL-Pec-Npl e AL-Random.

Most Uncertain 150 images by Both Scores

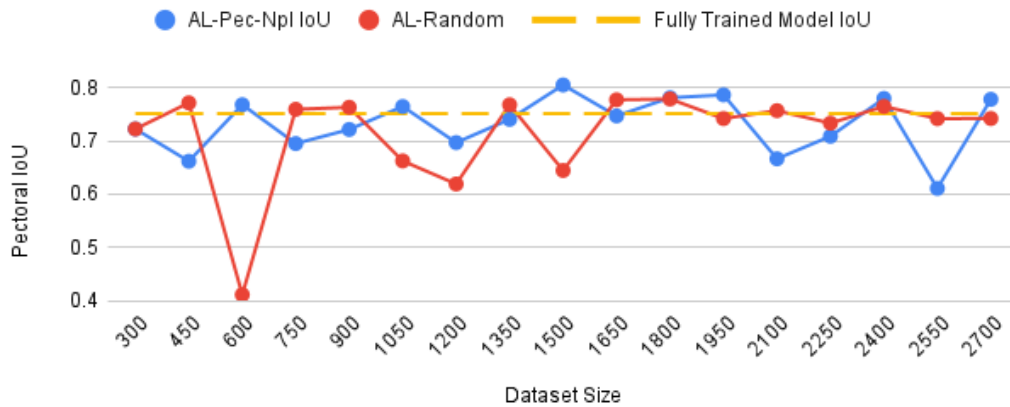


Figura 6.33: Comparação da IoU da estrutura músculo peitoral entre os experimentos AL-Pec-Npl e AL-Random.

6.7

Análise dos resultados

Observando a figura 6.30, interrompe-se o experimento AL-Papila com a condição de parada quando a acurácia de segmentação IoU ultrapassa o limite superior demarcado na figura, na qual esta margem é representada pelo modelo completamente treinado, ou seja, o modelo treinado com todas imagens da pool anotadas e adicionadas ao dataset.

Entende-se que a curva azul do IoU representada na figura 6.30, demonstra um contraste suficiente para indicar que o critério escolhido consegue filtrar as imagens mais informativas em relação à papila, verificando que a curva atinge além do limite superior com IoU da papila em 0,7580.

Em relação à figura 6.31, o critério de parada definido para interromper o experimento foi atingir a quantidade de 3000 imagens anotadas no dataset. Avaliando a comparação entre ambas as curvas, não há contraste suficiente entre os IoUs dos experimentos, para afirmar que o score de incerteza do peitoral estaria selecionando as imagens mais informativas para o treinamento do modelo a cada etapa. Observa-se que as curvas estão muito próximas e, em certos momentos, o IoU do peitoral de AL-Random se sobressai e se mantém mais estável que o IoU de AL-Peitoral após 1650 imagens no dataset.

Uma observação em relação ao IoU de AL-Peitoral na figura 6.31 consiste nas amplitudes detectadas após 1800 imagens no dataset. À medida atinge marcas bem próximas a 0.8, mantém uma certa consistência nas margens do limite superior, apesar de variar com maior amplitude que o IoU de AL-Random.

Sobre a figura 6.32, entende-se que mesmo selecionando uma quantidade reduzida de imagens mais informativas sobre a papila, o IoU desta estrutura consegue alcançar o limite superior nas margens de 0.75 e a partir de 1350 imagens no dataset, e é observado um distanciamento do IoU da papila em relação ao experimento AL-Random.

Na figura 6.33, o IoU do peitoral consegue manter uma certa consistência em relação ao IoU do experimento AL-Random até 1950 imagens no dataset. A partir deste ponto, o IoU do peitoral de AL-Pec-Npl se revela instável, não conseguindo manter a consistência como a curva do experimento adversário. Observa-se, portanto, que o critério definido para este experimento não é muito indicado, pelo fato de apenas acrescentar valor apenas no IoU da papila.

Além disso, o mIoU das estruturas mamárias no experimento AL-Pec-Npl na figura 6.27 demonstra instabilidades de grande amplitude a partir 1950 imagens no dataset, oscilando aproximadamente entre 0.76 e 0.82, se comparadas à figura 6.20. Isso indica que o critério definido para o experimento AL-Pec-Npl está afetando o IoU de outras estruturas mamárias.

Um dos experimentos teste feitos para tentar melhorar e explorar o critério de seleção do peitoral, evidencia uma diminuição em relação ao IoU da estrutura por ciclo de active learning. O critério utilizado consistia selecionar 75 imagens de maior $S_{peitoral}$ e 75 imagens em que a segmentação do peitoral feita pelo modelo era nula ou $S_{peitoral}$ estava entre 0.4 e 0.7. A figura 6.34 mostra como o IoU decresce ao adicionar imagens no dataset caso o critério configurado não consiga separar as imagens mais informativas.

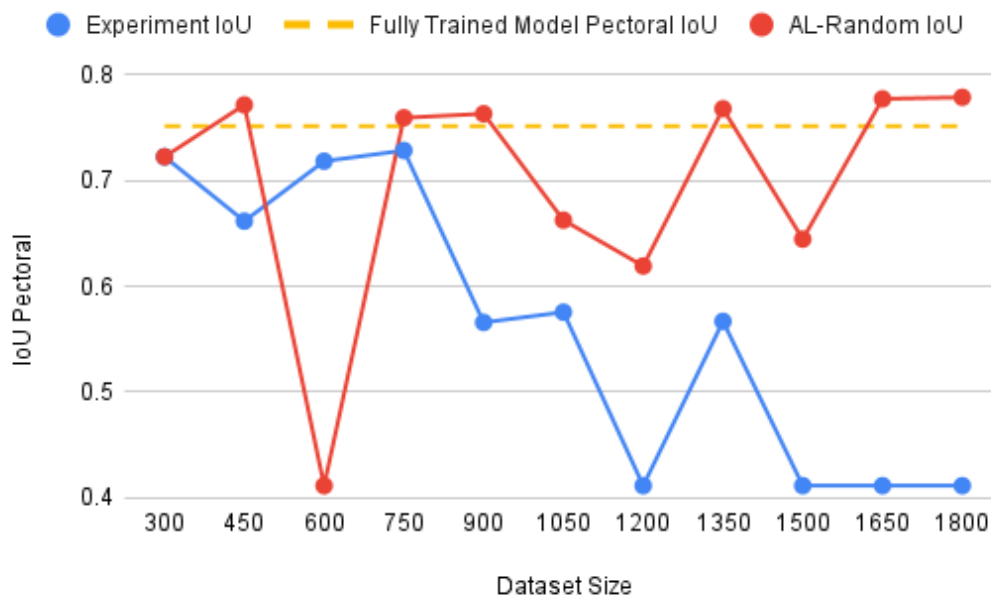


Figura 6.34: IoU da estrutura peitoral ao selecionar imagens não informativas.

Entende-se que o processo de active learning aplicado para músculo peitoral, neste contexto, possui peculiaridades nas quais, nos exames, a aparição dessa estrutura não é garantida. Logo, estimar medidas de incertezas e critérios de seleção para esta estrutura anatômica não é trivial. Pelo fato de existir a probabilidade do modelo não encontrar o peitoral, fazer segmentações nulas ou minúsculas e informações utilizadas para estimar $S_{peitoral}$ podem não ser suficientes para filtrar imagens mais informativas desta estrutura anatômica. A partir do experimento teste feito, conforme a figura 6.34, verifica-se que ao adicionar imagens que não possuíam músculo peitoral visualmente confirmado ou imagens com segmentações nulas/muito pequenas filtradas, segundo o critério estabelecido ao dataset, estas imagens estariam tornando o acerto do $S_{peitoral}$ a cada rodada de active learning dificultoso, além dessas imagens estarem diminuindo a acurácia de segmentação IoU para esta estrutura mamária específica.

Em relação à papila, mediante os resultados neste contexto, active learning consegue ser aplicado com resultados bons, conseguindo filtrar as imagens mais pertinentes de serem anotadas priorizando esta estrutura em particular.

Este trabalho propôs duas abordagens, mecanismos de anotação assistida por modelos de segmentação e active learning, que tornam eficientes os processos de anotação e construção de datasets de segmentação semântica que envolvem imagens de mamografia computadorizada.

A primeira abordagem consiste em treinar um modelo rudimentar de segmentação de dobras de pele, para imitar os delineamentos feitos pelos anotadores peritos sobre o artefato mencionado. Entende-se que a técnica consegue melhorar a eficiência do processo de anotação, pelo fato de consumir menos tempo para delinear as dobras de peles em exames de mamografia.

Além disso, os experimentos para obtenção dos resultados de tempo decorrido para anotar dobras de pele foram estendidos com métricas adicionais para avaliar o uso da abordagem para outras estruturas mamárias como papila, músculo peitoral e tecido fibroglandular. Concluiu-se que houve diminuição de custos de anotação de um modo geral para anotar os exames em relação ao tempo de anotação. Isso indica que as segmentações sugeridas pelo modelo, em parte, auxilia o anotador a completar as anotações sugeridas dependendo da estrutura mamária como peitoral e tecido fibroglandular. No entanto, a técnica adicionou tempo de anotação nas papilas, pelo fato da segmentação feita pelo modelo não ser satisfatória e recorrer à total correção da estrutura por parte dos anotadores. O fato do tempo de processamento do modelo estar acrescentando tempo para completar as anotações nos exames de mamografia indica que o mecanismo não atua bem como melhoria para estruturas como papila, não compensando o tempo gasto para o modelo processar a imagem e sugerir a segmentação nesses casos. Nas demais estruturas, houve redução nos tempos de anotações e para anotações de exames de mamografia, em relação à interatividade, foi atenuada de modo geral a quantidade de interações durante as anotações com recurso de anotações assistidas por modelos. Porém, observa-se que a partir dos resultados dos experimentos estendidos, a quantidade de interações de mouse e teclado não demonstram muita relevância como uma medida atrelada à eficiência em anotações. Apesar da medida servir de métrica de avaliação em contexto de segmentação semântica, a mesma não se aplica ao contexto da melhoria proposta em anotações assistidas por modelos de segmentação. Necessita-se de um aprofundamento melhor sobre métricas de interatividade no contexto desta pesquisa. Evidencia-se que, para este cenário, a principal métrica para medir eficiência em anotações consiste em tempo de

anotação.

Active learning neste trabalho atua como uma melhoria, direcionando os exames de mamografia mais importantes de serem anotadas, com a finalidade de construir datasets mais compactos e seletos para uma estrutura mamária específica. Desta forma, são necessários menos exames a serem anotados priorizando uma estrutura mamária. O benefício acima pode ser conquistado a partir dos scores de incertezas calculadas para papila e peitoral, que servem como indicação das imagens a serem anotadas que o modelo teve maior dificuldade em fazer as inferências para estas estruturas anatômicas. Com base nos resultados dos experimentos, o score de incerteza da papila demonstrou bons contrastes ao selecionar as imagens por ciclo de maior pontuação, superando a acurácia de segmentação IoU da papila de um modelo completamente treinado com menos imagens no dataset. Acerca do score de incerteza do músculo peitoral, não houve contraste suficiente entre os resultados para afirmar que a métrica estaria filtrando as imagens mais informativas de serem anotadas. Devido a peculiaridades do músculo peitoral para vista CC, onde a estrutura pode não estar presente com uma certa frequência, ou o fato do modelo não encontrar o peitoral no exame, torna-se não trivial o uso de active learning para este caso em específico. Observa-se mediante os resultados do experimento teste, conforme a figura 6.34 apresentada no final do capítulo 6, ao acrescentar ao dataset uma parcela de imagens CC que não possuíam visualmente o músculo peitoral ou a segmentação era nula, a cada etapa de active learning, é evidenciado que a cada ciclo da abordagem, o IoU do músculo peitoral é bastante afetado, diminuindo à cada iteração do processo. Isso levanta a hipótese de que o aprendizado do modelo estaria sendo afetado com o acréscimo de imagens desta natureza ao dataset. Logo, conclui-se que active learning, aplicado neste contexto da mamografia, consegue filtrar as imagens mais importantes de serem anotadas mediante o score de incerteza da papila, priorizando o IoU desta estrutura. Porém o mesmo não se pode afirmar em relação ao músculo peitoral, pois os experimentos não evidenciam contrastes suficientes para afirmar que o uso do score de incerteza do peitoral estaria filtrando as imagens mais pertinentes de serem anotadas, visando uma melhoria no IoU para esta estrutura.

Esta dissertação abre portas para algumas frentes de trabalhos futuros. Por exemplo, recomenda-se estender os scores de incertezas de active learning para calcular pontuações para outras estruturas mamárias como tecido fibroglandular. Além de um estudo mais aprofundado para melhoria do score de incerteza do peitoral.

Algumas melhorias pertinentes aos experimentos conduzidos nos capí-

tulos 4 e 5, como melhoria nos mecanismos de ativação da cronometragem, criação das etiquetas e uma quantidade maior de participantes, poderiam enriquecer ainda mais este trabalho. Além de explorar mais métricas para avaliação de eficiência em processos de anotação. Em adição, entende-se que ao tornar mais eficiente o processo de anotação, não se sabe muito sobre efeitos colaterais pertinentes à qualidade de anotação nos exames. A melhoria do mecanismo pode não garantir um aumento na eficácia do processo, ou seja, o processo se torna mais eficiente, porém o mesmo pode não ocorrer em realizar anotações visando a qualidade das anotações. Um estudo mais profundo para avaliar esta questão poderia ser uma frente a se investigar.

Por fim, uma outra frente seria a junção de active learning com a técnica de anotações assistidas, para tentar explorar quais ganhos e aportes poderiam ser adquiridos tentando encontrar uma relação com a quantidade de imagens reduzida a ser anotada por active learning com o tempo de anotação e interação consumida para anotar exames de mamografia mais informativos por ciclo.

Referências bibliográficas

ADHIKARY, K.; KUMA, A. Feature extraction and classification technique in neural network. **International Journal of Computer Applications**, Citeseer, v. 35, n. 3, p. 29–35, 2011.

BUDD, S.; ROBINSON, E. C.; KAINZ, B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. **Medical Image Analysis**, Elsevier, p. 102062, 2021.

ESTEVA, A. et al. Deep learning-enabled medical computer vision. **NPJ digital medicine**, Nature Publishing Group, v. 4, n. 1, p. 1–9, 2021.

GAL, Y.; ISLAM, R.; GHAHRAMANI, Z. Deep bayesian active learning with image data. In: PMLR. **International Conference on Machine Learning**. [S.l.], 2017. p. 1183–1192.

GARCIA-GARCIA, A. et al. A review on deep learning techniques applied to semantic segmentation. **arXiv preprint arXiv:1704.06857**, 2017.

GAWLIKOWSKI, J. et al. A survey of uncertainty in deep neural networks. **arXiv preprint arXiv:2107.03342**, 2021.

GRAHAM, R. L.; YAO, F. F. Finding the convex hull of a simple polygon. **Journal of Algorithms**, Elsevier, v. 4, n. 4, p. 324–331, 1983.

JANG, H.-J.; CHO, K.-O. Applications of deep learning for the analysis of medical data. **Archives of pharmacal research**, Springer, v. 42, n. 6, p. 492–504, 2019.

JESINGER, R. A. Breast anatomy for the interventionalist. **Techniques in vascular and interventional radiology**, Elsevier, v. 17, n. 1, p. 3–9, 2014.

MASOOD, S. et al. A survey on medical image segmentation. **Current Medical Imaging**, Bentham Science Publishers, v. 11, n. 1, p. 3–14, 2015.

MO, Y. et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning. **Neurocomputing**, Elsevier, v. 493, p. 626–646, 2022.

MONARCH, R. M. **Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI**. [S.l.]: Simon and Schuster, 2021.

PAGARE, R.; SHINDE, A. A study on image annotation techniques. **International Journal of Computer Applications**, Citeseer, v. 37, n. 6, p. 42–45, 2012.

PHILBRICK, K. A. et al. Ril-contour: a medical imaging dataset annotation tool for and with deep learning. **Journal of digital imaging**, Springer, v. 32, n. 4, p. 571–581, 2019.

RAPSON, C. J. et al. Reducing the pain: A novel tool for efficient ground-truth labelling in images. In: IEEE. **2018 international conference on image and vision computing New Zealand (IVCNZ)**. [S.l.], 2018. p. 1–9.

- REN, P. et al. A survey of deep active learning. **arXiv preprint arXiv:2009.00236**, 2020.
- RITI, Y. F. et al. Feature extraction for lesion margin characteristic classification from ct scan lungs image. In: IEEE. **2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)**. [S.l.], 2016. p. 54–58.
- ROY, A. G. et al. Bayesian quicknat: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. **NeuroImage**, Elsevier, v. 195, p. 11–22, 2019.
- SAGER, C.; JANIESCH, C.; ZSCHECH, P. A survey of image labelling for computer vision applications. **Journal of Business Analytics**, Taylor & Francis, p. 1–20, 2021.
- SAKINIS, T. et al. Interactive segmentation of medical images through fully convolutional neural networks. **arXiv preprint arXiv:1903.08205**, 2019.
- SAMBATURU, B. et al. Efficient and generic interactive segmentation framework to correct mispredictions during clinical evaluation of medical images. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S.l.], 2021. p. 625–635.
- SCANDALEA, M. L. D. et al. Deep active learning for axon-myelin segmentation on histology data. **arXiv preprint arXiv:1907.05143**, 2019.
- SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of big data**, SpringerOpen, v. 6, n. 1, p. 1–48, 2019.
- SINGNOO, J.; FINLAYSON, G. D. Understanding the gamma adjustment of images. In: SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY. **Color and Imaging Conference**. [S.l.], 2010. v. 2010, n. 1, p. 134–139.
- SUN, J. et al. Interactive medical image segmentation via point-based interaction and sequential patch learning. **arXiv preprint arXiv:1804.10481**, 2018.
- TAGHANAKI, S. A. et al. Deep semantic segmentation of natural and medical images: a review. **Artificial Intelligence Review**, Springer, v. 54, n. 1, p. 137–178, 2021.
- TAHA, A. A.; HANBURY, A. An efficient algorithm for calculating the exact hausdorff distance. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 37, n. 11, p. 2153–2163, 2015.
- TAHERDOOST, H. Sampling methods in research methodology; how to choose a sampling technique for research. **How to Choose a Sampling Technique for Research (April 10, 2016)**, 2016.

TAJBAKHS, N. et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. **Medical Image Analysis**, Elsevier, v. 63, p. 101693, 2020.

TAJBAKHS, N. et al. Guest editorial annotation-efficient deep learning: The holy grail of medical imaging. **IEEE Transactions on Medical Imaging**, IEEE, v. 40, n. 10, p. 2526–2533, 2021.

VOSSKÜHLER, A. et al. Ogama (open gaze and mouse analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. **Behavior research methods**, Springer, v. 40, n. 4, p. 1150–1162, 2008.

WANG, G. et al. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. **Neurocomputing**, Elsevier, v. 338, p. 34–45, 2019.

WANG, G. et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. **IEEE transactions on medical imaging**, IEEE, v. 37, n. 7, p. 1562–1573, 2018.

XIONG, J. **Active Learning for Semantic Segmentation**. 2019.

YANG, L. et al. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: SPRINGER. **International conference on medical image computing and computer-assisted intervention**. [S.l.], 2017. p. 399–407.

YEGHIAZARYAN, V.; VOICULESCU, I. An overview of current evaluation methods used in medical image segmentation. **Department of Computer Science, University of Oxford**, 2015.