**Guião da Defesa**

**Slide 1 -** Bom dia a todos, hoje defenderei a minha dissertação, cujo título é “Characterization of Retinal Fluids in Optical Coherence Tomography Images”, orientado pela professora Tânia Melo e co-orientado pela professora Ana Maria Mendonça.

**Slide 2 -** Nesta apresentação irei falar do problema que motiva esta pesquisa e dos objetivos definidos com base neste problema. Explicarei a revisão da literatura levada a cabo nos temas a abordar. Posteriormente, falarei dos materiais e métodos utilizados nas experiências e apresentarei os resultados obtidos nas mesmas, discutindo-os. Terminarei com as conclusões em cada tema, as limitações do estudo desenvolvido e quais as direções futuras de trabalho.

**Slide 3 –** Doenças como a degenerescência macular associada à idade, edema macular diabético, e a oclusão da veia central da retina são algumas das principais causas de perda de visão, estando associadas ao aparecimento de fluído na retina. Estes fluídos podem ser visualizados a partir de tomografias de coerência ótica, ou OCT. As características destes fluídos, como o seu tipo e volume, são assim importantes biomarcadores associados à progressão destas doenças. Os fluídos podem ser classificados de acordo com a sua localização na retina em fluído intra-retiniano (IRF), sub-retiniano (SRF), ou descolamento do epitélio pigmentar (PED). No entanto, a caracterização destes fluídos é uma tarefa laboriosa, que requer a segmentação semântica manual de todas as imagens de um volume de OCT, estando assim condicionado pela experiência médica e pela qualidade e resolução do OCT.

**Slide 4 –** Assim, definiram-se como objetivos gerais desta dissertaçãoa segmentação de fluído em scans de OCT, com o intuito de estimar o seu volume e aumentar a resolução dos scans de OCT, gerando os scans intermédios, aumentando a confiança nos volumes obtidos. Para isto, desenvolvi modelos de deep learning de segmentação automática dos fluídos retinianos e utilizei o melhor modelo desenvolvido para estimar o volume dos fluído nos OCT avaliados. Posteriormente, implementei um modelo generativo com o intuito de gerar imagens intermédias entre duas slices conhecidas e investiguei o impacto desta geração no volume de fluído estimado.

**Slide 5 –** Começou-se então com a pesquisa bibliográfica na segmentação de fluído, seguindo a metodologia de uma revisão sistemática. Pesquisaram-se os artigos com a query aqui vista, nas bases de dados abaixo. Excluíram-se e incluíram-se os artigos seguindo os critérios exibidos à direita, restando apenas 20 artigos.

**Slide 6 –** Agruparam-se os artigos de acordo com a modalidade de segmentação de fluído, em binário ou multi-classe, e de acordo com a arquitetura das redes utilizadas. Concluiu-se então que a maioria dos artigos utiliza apenas redes convolucionais neuronais (CNN), baseados na U-Net. Estes artigos inovam maioritariamente nos módulos utilizados na rede. Os artigos que usam Transformers, associam-nos às CNNs, por vezes substituindo o encoder e noutras vezes a secção do bottleneck. A delimitação da retina é também uma prática comum, utilizada para limitar a informação da retina. Os limites da retina podem ser detetados com vários métodos, e a sua introdução na rede pode ser feita a partir de cropping, colocando os valores fora da retina como zero, ou fornecendo esta informação num canal adicional.

**Slide 7/8 –** Na geração de slices intermédias, as abordagens consideradas podem ser agrupadas em quatro grupos, de acordo com a abordagem desenvolvida ou a sua funcionalidade. No primeiro grupo, encontram-se as abordagens de geração de slices intermédias que, a partir das slices vizinhas, geram uma intermédia. No segundo grupo, além das gerar slices intermédias, aumenta-se a resolução intra-plano, nos planos de pior resolução, e, consequentemente, aumenta a resolução inter-plano. Por fim, existe o grupo das CNNs 3D, que são muito computacionalmente dispendiosas, e o grupo das aplicações para vídeo, no qual o princípio de funcionamento é o mesmo, mudando a grandeza física considerada: as diferentes imagens são separadas por tempo, ao invés de distância, como nas imagens médicas.

**Slide 9 –** Para treinar modelos de segmentação e generativos são necessárias grandes quantidades de dados. Para a segmentação, estes têm que ser anotados com os fluídos a segmentar, de preferência com variadas qualidades de imagem e com fluídos com diferentes características. Para os modelos generativos, é necessário que a distância entre B-scans consecutivos seja constante, dentro de um volume. Para isso, escolheu-se o dataset RETOUCH que contém uma grande quantidade de dados obtidos com diferentes dispositivos. Testaram-se ainda os melhores modelos de segmentação num dataset privado do Hospital de São João. Este dataset é pequeno, mas apresenta grande diversidade.

**Slide 10 –** Para a segmentação de fluídos, utilizou-se a baseline U-Net, pela sua presença de destaque na segmentação de fluídos retinais e, em geral, na imagem médica. A loss de treino foi inspirada num artigo, previamente implementado, que tem duas componentes principais: coeficiente de Dice calculado para o foreground, para promover uma segmentação de fluído adequada, e cross-entropy, para que ocorra uma classificação adequada dos pixéis.

**Slide 11 –** Patches com várias formas foram experimentadas. Inicialmente extraíram-se aleatoriamente patches pequenas da região de interesse, ROI, e depois extraíram-se patches maiores, de cima para baixo em cada B-scan, de modo que toda a área da imagem ficasse em pelo menos uma patch.

**Slide 12 –** Por fim, extraíram-se ainda patches verticais de cada B-scan, após se redimensionarem todos para uma dimensão comum. Extraíram-se diferentes números de patches de cada imagem: começou-se por extrair 4 patches disjuntas, mas extraíram-se também 7 e 13 patches. Nestes últimos casos, as patches eram extraídas da esquerda para a direita em intervalos constantes, até que não fosse possível extrair mais patches.

**Slide 13 –** Para a geração de slices intermédias, recorreu-se a uma simples GAN, inspirada num dos trabalhos mencionados da literatura, de vídeo frame interpolation. Numa GAN, existem duas redes distintas: o gerador, que neste caso é muito semelhante a uma U-Net, e o discriminador, que se trata de uma série de convoluções que terminam na probabilidade de a imagem ser real. O gerador, que tenta criar as imagens intermédias recebendo as duas slices adjacentes, tenta também enganar o discriminador. Este último tenta distinguir se as imagens são verdadeiras ou falsas. À medida que o gerador aprende a gerar imagens mais semelhantes às reais, o discriminador torna-se mais criterioso. Melhoram assim, ambas as redes. Testou-se ainda a geração de imagens utilizando uma U-net, inspirado na literatura.

**Slide 14 –** Durante o treino,o gerador recebe a patches da slice anterior e posterior à que se pretende gerar. Neste caso, utilizaram-se patches de 64 por 64, extraindo-se o máximo de patches possíveis da rede. O gerador é treinado com uma loss composta por 4 componentes: adversarial, que diz respeito à capacidade de enganar o discriminador, multi-scale structural similarity index measure loss, que procura preservar as estruturas das imagens reais nas escalas geradas, mean absolute error, que compara as imagens ao nível dos pixéis, e a gradient difference loss, que compara as diferenças entre um píxel e os pixéis que o rodeiam numa imagem real às mesmas diferenças numa imagem falsa.

**Slide 15 –** O cálculo de volume de fluído é feito a partir da estimativa da área de um fluído em cada B-scan, multiplicando por metade da distância à slice anterior e metade da distância à slice posterior. Por fim, soma-se o volume calculado em todos os B-scans de um volume OCT. A loss do discriminador é apenas a binary cross entropy.

**Slide 16 –** Para as experiências, utilizou-se uma five fold validation, aplicada aos volumes de OCT. Na segmentação, apenas os volumes com fluído segmentado foram incluídos. Na geração, todos os 112 volumes foram incluídos. No entanto, a fold reservada para comparação entre experiências nas experiências de segmentação, foi também reservada nas experiências de geração.

**Slide 17 –** Assim, na experiência 1 realizou-se segmentação de fluídos, utilizando apenas uma U-Net, variando-se a forma do input. Na experiência 2, a segmentação foi realizada utilizando uma U-Net para cada fluído, variando-se a loss. A experiência 3 diz respeito à geração de imagens utilizando uma GAN, enquanto na experiência 4 utilizou-se uma U-Net. As experiências 5 e 6 dizem respeito, respetivamente, à estimativa do volume de fluído nos volumes de OCT reservados, e à estimativa de volumes de fluído em volumes gerados e super-resolvidos utilizando a GAN.

**Slide 18 –** Passando agora para os resultados das experiências, na primeira experiência as patches revelaram-se demasiado pequenas para capturar as transições entre a retina e o background, levando a classificações erradas e segmentações fora da retina. Na segunda experiência, os resultados foram melhores, mas as grandes patches não permitiram grande detalhe na segmentação. Ainda, a retina ficava regularmente dividida em patches diferentes, dificultando a aprendizagem do modelo.

**Slide 19 –** Os melhores resultados foram obtidos nas patches verticais, onde o modelo aprendeu corretamente a anatomia da retina e as patches eram pequenas o suficiente para permitir uma detalhada segmentação dos fluídos mais pequeno. Testou-se então com 4, 7 e 13 patches extraídas e verificou-se que à medida que o número de patches por B-scan aumenta, a velocidade de progressão da rede aumenta também, mas com maior custo computacional. As melhores performances foram obtidas nas redes treinadas com 4 e 7 patches.

**Slide 20 –** Experimentaram-se ainda várias rotações aleatórias como transformações de input, verificando-se uma grande influência desta segmentação na performance. Sem rotação, a performance era pior, enquanto com 5 graus de rotação o modelo obtinha piores Dice scores, mas mantinha uma maior coerência anatómica e realizava menos sobre segmentações.

**Slide 21 –** A segmentação com modelos dedicados a cada fluído obteve piores performances que as redes multi-classe, devido à falta de competição entre classes e a falta de conhecimento das estruturas anatómicas, motivada pela classificação dos pixéis em diferentes fluídos. Quando se mudou a loss, a performance piorou ainda mais, com constantes sobre segmentações de pequenas quantidades de fluído. Estes erros não são tão penalizados sem a componente Dice.

**Slide 22 –** Comparando a performance com um artigo com uma implementação semelhante da literatura, mas que usou três redes combinadas a partir de majority voting. Apesar desta ter custos computacionais mais elevado, as redes implementadas ao longo da dissertação obtiveram performances comparativas em IRF. Em SRF, a nossa implementação teve uma melhor performance por ser treinada em patches maiores, mas em PED, isto tornou-se uma desvantagem. Em geral, o treino por patches torna-se muito mais vantajoso que o treino com a imagem completa.

**Slide 23 –** A performance no dataset do São João foi influenciada por diversos fatores como a qualidade das imagens, cuja definição da coróide e níveis de ruído são muito diferentes daqueles vistos em treino, ter sido segmentada apenas por um indivíduo, que resulta, assim, em menores quantidades de fluído segmentadas, e diferentes critérios de segmentação de PED.

**Slide 24 –** Assim, ainda que a segmentação dos fluídos tenha sido boa, existiu sempre sobre segmentação na coróide e de PED em geral, causados pelas diferenças mencionadas. Para mitigar a segmentação da coróide, utilizou-se uma rede de segmentação da retina, eliminado todos os fluídos fora da mesma. Isto melhorou quer o coeficiente de Dice, quer a coerência anatómica.

**Slide 25 –** Passando agora para as experiências de geração de imagens com a GAN, obtivemos B-scans visualmente realistas. No entanto, a distância entre B-scans no volume, que é uma característica associada ao volume de OCT, interfere fortemente na qualidade do scan gerado. Isto é, quando a distância entre slices é maior, as transições visuais são também maiores e, assim, mais difíceis de prever. Por este motivo, os scans de Spectralis, que têm uma distância inter-slice maior, têm pior qualidade. A geração de IRF é a parte mais difícil da tarefa de geração, uma vez que este fluído é por norma pequeno e rodeado por finos tecidos da retina, cuja posição é difícil de prever. Assim, os fluídos de maiores dimensões são mais fáceis de gerar. Percebemos ainda que, por a loss ter uma componentemuito pesada de comparação ao nível dos pixéis, que a geração se focava demasiado na previsão do ruído speckle do que no conteúdo mais difícil de prever da retina (IRF).

**Slide 26 –** A geração com a U-Net levou a resultados muito menos satisfatórios uma vez que, para reduzir a loss utilizada, mean absolute error, a rede produzia imagens muito desfocadas, devido à quantidade de ruído nesta técnica de imagem. Para contrariar isto, uma loss perceptual deveria ser implementada.

**Slide 27 –** Quanto ao cálculo de volume de fluído nos volumes OCT originais, os valores foram semelhantes aos calculados com as máscaras verdadeiras. As diferenças que se vê neste gráfico dizem respeito, assim, às limitações do modelo de segmentação. Os erros mais significativos ocorrem no IRF, onde a segmentação automática de fluído é pior e os volumes são menores.

**Slide 28 –** Já nos volumes falsos e super-resolvidos, a performance é muito semelhante à estimada com o modelo de segmentação, sendo que a maioria das diferenças aparece, então, no fluído IRF, onde a geração é pior, e nos volumes OCT obtidos com o dispositivo Spectralis, onde a distância entre slices é maior. Isto prova que o maior limitador da estimativa de fluído é então o modelo de segmentação.

**Slide 29 –** Passando por fim para as conclusões: na segmentação, provou-se a influência do tamanho das patches e das transformações aleatórias na performance do modelo e obtiveram-se resultados semelhantes aos da literatura, mesmo com menor dispêndio computacional. Na geração de imagens, viu-se que a distância entre slices é muito importante, que as losses ao nível dos pixéis são facilmente influenciadas pelo ruído speckle, e que as regiões mais pequenas são mais difíceis de gerar. Por fim, a estimação de fluído está naturalmente dependente das performances dos modelos de segmentação e de geração, mas é mais influenciado pelo modelo de segmentação. A estimativa do fluído é menos precisa em IRF onde, devido às suas características anatómicas, é mais difícil de gerar e segmentar.

**Slide 30 –** Todo este progresso não foi feito sem limitações, onde verificamos que na segmentação a U-Net não é capaz de segmentar com precisão regiões grandes e pequenas ao mesmo tempo e os modelos não generalizaram muito bem em fontes externas. Quanto à geração, os modelos prendem-se demasiado à representação de ruído, em vez de estruturas relevantes. Para mitigar estes problemas, foi sugerida a utilização de redes de segmentação mais complexas, o condicionamento do input com o objetivo de fornecer mais informação à rede e treinar o modelo em dados de fontes diversas. Para a geração, sugere-se a implementação de losses percetuais fortes, como a LPIPS, e a introdução de blocos de atenção.

**Slide 31 –** Chego assim ao fim da minha apresentação e, qualquer questão que tenham, por favor façam. Obrigado pela atenção e presença de todos.